ASTES

Advances in Science, Technology & Engineering Systems Journal

VOLUME 5-ISSUE 6|NOV-DEC 2020

www.astesj.com
ISSN: 2415-6698

EDITORIAL BOARD
Editor-in-Chief

Prof. Passerini Kazmerski
University of Chicago, USA

Editorial Board Members

Prof. Rehan Ullah Khan
Qassim University, Saudi Arabia

Prof. María Jesús Espinosa
Universidad Tecnológica Metropolitana, Mexico

Dr. Hongbo Du
Prairie View A&M University, USA

Dr. Nguyen Tung Linh
Electric Power University, Vietnam

Tariq Kamal
University of Nottingham, UK

Sakarya University, Turkey

Dr. Mohmaed Abdel Fattah Ashabrawy
Prince Sattam bin Abdulaziz University, Saudi Arabia

Mohamed Mohamed Abdel-Daim
Suez Canal University, Egypt

Dr. Omeje Maxwell
Covenant University, Nigeria

Prof. Majida Ali Abed Meshari
Tikrit University Campus, Iraq

Dr. Heba Afify
MTI university, Cairo, Egypt

Regional Editors

Dr. Hung-Wei Wu
Kun Shan University, Taiwan

Dr. Maryam Asghari
Shahid Ashrafi Esfahani, Iran

Dr. Shakir Ali
Aligarh Muslim University, India

Dr. Ahmet Kayabasi
Karamanoglu Mehmetbey University, Turkey

Dr. Ebubekir Altuntas
Gaziosmanpasa University, Turkey

Dr. Sabry Ali Abdallah El-Naggar
Tanta University, Egypt

Aamir Nawaz
Gomal University, Pakistan

Dr. Gomathi Periasamy
Mekelle University, Ethiopia

Dr. Walid Wafik Mohamed Badawy
National Organization for Drug Control and Research, Egypt

Dr. Abhishek Shukla
R.D. Engineering College, India

Abdullah El-Bayoumi
Cairo University, Egypt

Ayham Hassan Abazid
Jordan university of science and technology, Jordan

Editorial

Advances in Science, Technology and Engineering Systems Journal (ASTESJ) is an online-only journal dedicated to publishing significant advances covering all aspects of technology relevant to the physical science and engineering communities. The journal regularly publishes articles covering specific topics of interest.

Current Issue features key papers related to multidisciplinary domains involving complex system stemming from numerous disciplines; this is exactly how this journal differs from other interdisciplinary and multidisciplinary engineering journals. This issue contains 208 accepted papers in Information System and Electrical Engineering domains.

Editor-in-chief
Prof. Passerini Kazmersk

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Volume 5 Issue 6 **November-December 2020**

CONTENTS

Qualitative Properties of a Cell Proliferating Model with Multi-phase Transition and Age Structure 01
Youssef El Alaoui, Larbi Alaoui

A Study on Intelligent Dialogue Agent for Older Adults' Preventive Care – Towards Development of a Comprehensive Preventive Care System – 09
Sho Hirose, Daisuke Kitakoshi, Akihiro Yamashita, Kentarou Suzuki, Masato Suzuki

Supervised Learning Techniques for Stress Detection in Car Drivers 22
Pamela Zontone, Antonio Affanni, Riccardo Bernardini, Leonida Del Linz, Alessandro Piras, Roberto Rinaldo

Analysis of Green Building Effect on Micro grid Based on Potential Energy Savings and BIM 30
Ihsan Mizher Baht, Petre Marian Nicolae, Ileana Diana, Nameer Baht

Effective Segmented Face Recognition (SFR) for IoT 36
Fei Gao, Jiangjiang Liu

Business Intelligence for Generating Comprehensive Report in Electronic Completion and Handover 45
Fadhillah Moulita Andiani, Faizal Abid, Hendri, Abba Suganda Girsang

Development of Group-Based Differentiated Learning (GBDL) Models 52
Toto Ruhimat, Deni Darmawan

The COVID-19 Effect in Mexican SMEs 63
Abel García-Villagrán, Patricia Cano-Olivos, José Luis Martínez-Flores, Diana Sánchez-Partida

Automated Extraction of Heavyweight and Lightweight Models of Urban Features from LiDAR Point Clouds by Specialized Web-Software 72
Sergiy Kostrikov, Rostyslav Pudlo, Dmytro Bubnov, Vladimir Vasiliev, Yury Fedyay

Using TOST in Teaching Operating Systems and Concurrent Programming Concepts 96
Tzanko Golemanov, Emilia Golemanova

Technology Adoption in Education: A Systematic Literature Review 108
Kayode Emmanuel Oyetade, Tranos Zuva, Anneke Harmse

Understanding the usage, Modifications, Limitations and Criticisms of Technology Acceptance Model (TAM) 113
William Ratjeana Malatji, Rene Van Eck, Tranos Zuva

The Impact Assessment of the Errors in Determining the Mass and Zero Lift-Drag Coefficient on the Aircraft's Performance Data 118
Klyagin Viktor Anatolievich, Laushin Dmitry Andreevich

Infrared Uplink Implementation for Software Defined Visible Light Communication Systems 127
Oswaldo René Banda-Sayco

Experimental Investigation of RC Footings Resting on Sand Strengthened with Concrete Jacketing 133
Mohamed Attia Fouda, Mahmoud Elkateb, Tamer Elkateb, Ayman Khalil

Electro-tactile Stimulation for Augmenting Finger Motoric Learning 143
Daniel Sutopo Pamungkas, Arjon Turnip

Aviation MRO: Impact of Physical Environment Factors on Job Performance in Aircraft Maintenance Organization 148
Kamal Jaiswal, Serdar Dalkilic, Evangelos Papageorgiou, Balgopal Singh

Inferring Topics within Social Networking Big Data, Towards an Alternative for Socio-Political Measurement 155
Khalid Ait Hadi, Rafik, Abdellatif El Abderrahmani

Research of the Effect of Rotation and Low-Frequency Vibration on the Robotic Assembly Process 160
Mikhail Vladimirovich Vartanov, Trung Ta Tran, Van Dung Nguyen

A Novel Approach to Augment the Opto-Electronic Properties of Stannic Oxide (SnO_2) Thin Films by Vanadium Doping 169
A. Victor Babu, S. Murugan, D.C. Bernice Victoria

Applications of the Heuristic Optimization Approach for Determining a Maximum Flow Problem Based on the Graphs' Theory 175
Simona Kirilova Filipova-Petrakieva

A Toolkit for the Automatic Analysis of Human Behavior in HCI Applications in the Wild 185
Andrea Generosi, Silvia Ceccacci, Samuele Faggiano, Luca Giraldi, Maura Mengoni

Real-Time Identification and Classification of Driving Maneuvers using Smartphone 193
Munaf Salim Najim Al-Din

The Effect of E-Service Quality on Customer Satisfaction and Loyalty (Case Study at E-Marketplace XYZ in Indonesia) 206
Hanny Juwitasary, Christian Christian, Edi Purnomo Putra, Hilman Baskara, Mohammad Wildan Firdaus

Trace-Driven Simulation of LoRaWAN Air Channel Propagation in an Urban Scenario 211
Eugen Harinda, Hadi Larijani, Ryan M. Gibson

Multi-Directional Light Sensing Using A Rotating Sensor 221
Hoang Anh Dung, Nguyen Manh Cuong, Nguyen Phan Kien

Vehicle Rollover Detection in Tripped and Untripped Rollovers using Recurrent Neural Networks 228
Kailerk Treetipsounthorn, Thanisorn Sriudomporn, Gridsada Phanomchoeng, Christian Dengler, Setha Panngum, Sunhapos Chantranuwathana, Ali Zemouche

Hybridization of Improved Binary Bat Algorithm for Optimizing Targeted Offers Problem in Direct Marketing Campaigns 239
Moulay Youssef Smaili, Hanaa Hachimi

IT GRC Smart Adviser: Process Driven Architecture Applying an Integrated Framework 247
Meriyem Chergui, Aziza Chakir

Multi Operated Virtual Power Plant in Smart Grid 256
Yevhen Fediv, Olha Sivakova, Mykhailo Korchak

A Fuzzy Controller Based SAPF for Power Quality Enhancement of Distribution System Integrated with Wind Energy Source 261
Vikas Kumar Sharma, Lata Gidwani

Pre-University Students' Learning Styles and Attitude towards Mathematics Achievements 269
Nurhilyana Anuar, Nurashikin Abdullah, Sharifah Norasikin Syed Hod

A Comprehensive Review of Traditional Video Processing 274
Helen Kottarathil Joy, Manjunath Ramachandra Kounte

Overmind: A Collaborative Decentralized Machine Learning Framework 280
Puttakul Sakul-Ung, Amornvit Vatcharaphrueksadee, Pitiporn Ruchanawet, Kanin Kearpimy, Hathairat Ketmaneechairat, Maleerat Maliyaem

The Role of Promotion in Mobile Wallet Adoption – A Research in Vietnam 290
Ha Hoang, Tan Trinh Le

Complex Order PI$^{?+j?}$ D$^{?+j?}$ Design for Surface Roughness Control in Machining CNT Al-Mg Hybrid Composites 299
Ravi Sekhar, Tejinder Paul Singh, Pritesh Shah

Dense Deep Neural Network Architecture for Keystroke Dynamics Authentication in Mobile Phone 307
Lubna Abdelkareim Gabralla

Emotion Recognition on FER-2013 Face Images Using Fine-Tuned VGG-16 315
Gede Putra Kusuma, Jonathan, Andreas Pangestu Lim

Multi-Model Security and Social Media Analytics of the Digital Twin 323
Jim Scheibmeir, Yashwant Malaiya

The Contribution of Wind Energy Capacity on Generation Systems Adequacy Reliability using Differential Evolution Optimization Algorithm 331
Athraa Ali Kadhem, Noor Izzri Abdul Wahab, Ahmed Abdalla

Enhanced Power Utilization for Grid Resource Providers 341
Tariq Alwada'n, Thair Khdour, Abdulsalam Alarabeyyat, Ali Rodan

DC-DC Buck Converter Driver with Variable Off-Time Peak Current Mode Control 347
Osvaldo Gasparri, Paolo Del Croce, Andrea Baschirotto

Blueprint Model: An Agile-Oriented Methodology for Tackling Global Software Development Challenges 353
Andre Figliuolo da Cruz, Cristiano Pereira Godoy, Lanier Menezes dos Santos, Lucas Frota Marinho, Marco Santarelle Jardim, Elisangela Paiva da Silva, C?cero Augusto Pahins, Paulo Fonseca, Felipe Taliar Giuntini

Vietnamese Text Classification with TextRank and Jaccard Similarity Coefficient 363
Hao Tuan Huynh, Nghia Duong-Trung, Dinh Quoc Truong, Hiep Xuan Huynh

Intrusion Detection and Classification using Decision Tree Based Key Feature Selection Classifiers 370
Manas Kumar Nanda, Manas Ranjan Patra

Comparison of Gaze Points Among Viewing Conditions (Resolution, Display Size, Viewer Position) During Video Viewing 391
Miho Shinohara, Yusuke Nosaka, Reiko Koyama, Riko Nakagawa, Takuya Sarugaku, Mitsuho Yamada

Method for Improving the Quality of the Product Obtained by Abrasive Treatment with Impregnated Tools 398
Viktor Butenko, Liana Gusakova, Dmitry Durov, Boris Safoklov, Oleg Dolgov

Level of Resilience and Family Functionality in Adolescents of two Educational Institutions of a Vulnerable Area in Lima Province 403
Rosa Perez-Siguas, Hernan Matta-Solis, Eduardo Matta-Solis, Melissa Yauri-Machaca, Anika Remuzgo-Artezano, Lourdes Matta-Zamudio

To the Question of Multi-Criteria Optimization of Aircraft Components in Order to Optimize its Life Cyclee 408
Sergey Alekseevich Serebryansky, Alexander Vladimirovich Barabanov

The Probe Mark Discoloration on Bond Pad and Wafer Storage 416
Wen-Fei Hsieh, Henry Lin, Vincent Chen, Irene Ou, & Yung-Song Lou

A Simulation Based Proactive Approach for Smart Capacity Estimation in the Context of Dynamic Positions and Events 423
Naeem Ahmed Haq Nawaz, Hamid Raza Malik, Ahmed Jaber Alshaor, Kamran Abid

Effective Learning of Tax Regulations using Different Chatbot Techniques 439
Rafael Mellado-Silva, Antonio Faúndez-Ugalde, María Blanco-Lobos

Electronic Warfare Methods Combatting UAVs 447
Miroslav Kratky, Vaclav Minarik, Michal Sustr, Jan Ivan

Fault-Tolerant Control of Permanent Magnet Synchronous Motor Drive under Open-Phase Fault 455
Amr Saleh, Nada Sayed, Ghada Ahmed Abdel, Mona Nagieb Eskander

Extending the Classifier Algorithms in Machine Learning to Improve the Performance in Spoken Language Understanding Systems Under Deficient Training Data 464
Sheetal Jagdale, Milind Shah

Reliability Improvement of Radial Distribution System by Reconfiguration 472
Srividhya. P, Mounika. K, Kirithikaa. S, Narayanan. K, Gulshan Sharma, Girish Ganesan. R, Tomonobu Senjyu

American Sign Language Recognition Based on MobileNetV2 481
Kin Yun Lum, Yeh Huann Goh, Yi Bin Lee

An Adaptive Nonlinear Sensorless Controller of Doubly Fed Induction Generator Driven By Wind Turbine 489
Radouane Ourhdir, Mohammed Rachidi

An Economic Theory Perspective for the Fight Against Poverty in the Peruvian Andes 497
Robert Antonio Romero-Flores

Microcontroller Based Data Acquisition and System Identification of a DC Servo Motor Using ARX, ARMAX, OE, and BJ Models 507
Mokhlis Salah-eddine, Said Sadki, Bahloul Bensassi

Facebook Translation Service (FTS) Usage among Jordanians during COVID-19 Lockdown 514
Zakaryia Almahasees, Helene Jaccomard

Accelerating Decision-Making in Transport Emergency with Artificial Intelligence 520
Alexander Raikov

Determinism of Replicated Distributed Systems–A Timing Analysis of the Data Passing Process 531
Adriano A. Santos, António Ferreira da Silva, António P. Magalhães, Mário de Sousa

Variation Between DDC and SCAMSMA for Clustering of Wireless MultipathWaves in Indoor and Semi-Urban Channel Scenarios 538
Jojo Blanza, Lawrence Materum

Interface for Visualization of Wireless Propagation Multipath Clustering Outcomes 544
Jojo Blanza, Lawrence Materum

The Effect of Different Starches in the Environmental and Mechanical Properties of Starch Blended Bioplastics 550
Adriana C. Neves, Tew Ming, Marta Mroczkowska, David Culliton

Proposal of a New Descriptive-Correlational Model of Population Lifestyle Analysis and Disease Diagnosis 555
Selene Tamayo Castro, Kristian Aldapa Salcido, Linda García Rodríguez

sharpniZer: A C# Static Code Analysis Tool for Mission Critical Systems 561
Arooba Shahoor, Rida Shaukat, Sumaira Sultan Minhas, Hina Awan, Kashif Saghar

Experimental Study on Mechanical Behavior of Polypropylene-based Blends with Talc Fillers 571
Pham Thi Hong Nga, Van-Thuc Nguyen

Ontology-based Data Management Tool for Studying Radon Concentration 577
Felix Fernandez-Pena, Alex Maigua-Quinteros, Pilar Urrutia-Urrutia, Diana Coello-Fiallos

Strategic Plan for the Achievement of the Competitiveness of Small Companies with Respect to Large Ones 584
Alan Guadalupe Ochoa Navarro, Juan De Dios Cota Apodaca, Dario Fuentes Guevara

"Traffic Congestion Triangle" Based on More than One-Month Real Traffic Big Data Analysis in India 588
Tsutomu Tsuboi

Effective Application of Information System for Purchase Process Optimization 594
Pearl Keitemoge, Daniel Tetteh Narh

Fast and Efficient Maximum Power Point Tracking Controller for Photovoltaic Modules 606
Khalid Chennoufi, Mohamed Ferfra

Inventory Management Practices during COVID 19 Pandemic to Maintain Liquidity Increasing Customer Service level in an Industrial Products Company in Mexico 613
Ignacio Alvarez-Placencia, Diana Sánchez-Partida, Patricia Cano-Olivos, José-Luis Martínez-Flores

Chatbot Developments in The Business World 627
Azani Cempaka Sari, Natashia Virnilia, Jasmine Tanti Susanto, Kent Anderson Phiedono, Thea Kevin Hartono

Burnout Among Primary School Teachers in the Wazzane Region in Morocco: Prevalence and Risk Factors 636
Abdeslam Amri, Zakaria Abidli, Mounir Bouzaaboul, Rabea Ziri, Ahmed Omar Touhami Ahami

Performance Analysis and Enhancement of Spline Adaptive Filtering based on Adaptive Step-size Variable Leaky Least Mean Square Algorithm 642
Sethakarn Prongnuch, Suchada Sitjongsataporn

Automatic Stochastic Dithering Techniques on GPU: Image Quality and Processing Time Improved 652
Giorgia Franchini, Roberto Cavicchioli, Jia Cheng Hu

Improved System Based on ANFIS for Determining the Degree of Polymerization 664
Marcel Nicola, Marian Du??, Maria-Cristina Ni?u, Ancu?a-Mihaela Aciu, Claudiu-Ionel Nicola

Network Modeling with ANP to Determine the Appropriate Area for the Development of Dry Port in Thailand 676
Jenjira Sukmanee, Ramil Kesvarakul, Nattawut Janthong

Social Influence Factor of e-Tourism Application Case Study University Student 684
Kristianus Oktriono, Surjandy Surjandy, Meyliana Meyliana, Michele Carolina, Stephanie Stephanie

Hand-Based Biometric Recognition Technique – Survey 689
Katerina Prihodova, Miloslav Hub

Automatic License Plate Detection and Recognition for Jordanian Vehicles 699
Khalil Mustafa Ahmad Yousef, Bassam Jamil Mohd, Yusra Abd-Al-Haleem Al-Khalaileh, Ahlam Hani Al-Hmeadat, Bushra Ibrahim El-Ziq

Adaptive Identification Method of Vehicle Model for Autonomous Driving Robust to Environmental Disturbances 710
Yohei Yamauchi, Mitsuyuki Saito

Cognitive Cybernetics in the Foresight of Globalitarianism 718
Zdenko Balaž, Krystian Wawrzynek

Ultra Wide Band-based Control of Emulated Autonomous Vehicles for Collision Avoidance in a Four-Way Intersection 724
Jashandeep Bhuller, Paolo Dela Peña, Vladimir Christian Ocampo II, Julio Simeon, Lawrence Materum

Minimizing Collisions of Self-Driving Cars by a Control System Using Predetermined Two-Dimensional Grid Localization 731
Jashandeep Bhuller, Paolo Dela Peña, Vladimir Christian Ocampo II, Julio Simeon, Lawrence Materum

Comparison of Support Vector Machine-Based Equalizer and Code-Aided Expectation Maximization on Fiber Optic Nonlinearity Compensation Using a Proposed BER Normalized by Power and Distance Index 738
Mark Renier M. Bailon, Lawrence Materum

Analysis of Long-term Equilibrium Relationship Between KRW, RMB, JPY Exchange Rates and International Financial Market Variables: Comparative Analysis of KRW, RMB, JPY 744
Moon-Kyum Kim, Woong Ryeol Kim, Moon-Kyum Kim

Optimization of Multi-user Face Identification Systems in Big Data Environments 762
Majdouline Meddad, Chouaib Moujahdi, Mounia Mikram, Mohammed Rziza

Surge Protection Device for Ex Application 768
Teik Hua Kuan, Kuew Wai Chew, Kein Huat Chua

Optimal Irrigation Strategy using Economic Model Predictive Control 781
Luisella Balbis

Dependency Head Annotation for Myanmar Dependency Treebank 788
Hnin Thu Zar Aye, Win Pa Pa

A Model-Driven Approach for Reconfigurable Systems Development 801
Ismail Ktata, Naoufel Kharroubi

Algorithm Design for Accurate Steps Counting Based on Smartphone Sensors for Indoor Applications 811
Hani Muhsen, Odeh Al-Amaydeh, Rakan Al-Hamlan

Polarity Switch within Social Networks 817
Sara Abas, Malika Addou, Zineb Rachik

Roadmap for Industrial Engineering Education Sustainability Withdrawn
Hani Shafeek

Towards a Smart Campus for Qassim University: An Investigation of Indoor Navigation System 831
Mohammed Hadwan, Rehan Uallah Khan, Khalil Ibrahim Mohammad Abuzanouneh

Effect of Heat Retention Time and Pouring Temperature on Graphite Shape and Mechanical Properties of Gray Cast Iron 838
Hong-Nga Thi Pham

Updated Analysis of Business Continuity Issues Underlying the Certification of Invoicing Software, Considering a Pandemic Scenario 845
Nelson Russo, Leonilde Reis

An Enhanced Conceptual Security Model for Autonomous Vehicles 853
Abdulla Obaid Al Zaabi, Chan Yeob Yeun, Ernesto Damiani, Gaemyoung

Low Power Fast Settling Switched Capacitor PTAT Current Reference Circuit for Low Frequency Applications 865
Muhammed Mansoor C. B., Hanumantha Rao G., Rekha S.

Blockchain Application in Higher Education Diploma Management and Results Analysis 871
Fernando Richter Vidal, Feliz Gouveia, Christophe Soares

Evaluating the Effectiveness of Query-Document Clustering Using the QDSM Measure 883
Claudio Gutierrez-Soto, Marco Palomino, Arturo Curiel, Hector Riquelme Cerda, Fernando Bejar Rain

Method of Analysis and Classification of Acoustic Emission Signals to Identify Pre-Seismic Anomalies 894
Marapulets Yury, Senkevich Yury, Lukovenkova Olga, Solodchuk Alexandra

A Fast Adaptive Time-delay-estimation Sliding Mode Controller for Robot Manipulators 904
Dang Xuan Ba

Investigation of the Effect of FBG Profiles, Temperature and Transmission Distance for Environmental Sensing & Monitoring 912
Muhammad Arif Riza, Yun Ii Go

Barriers and Supports in Engineering Career Development: An Exploration of First-Year Students 920
Rosmery Ramos-Sandoval, Jano Ramos-Diaz

Prediction of Vessel Dynamic Model Parameters using Computational Fluid Dynamics Simulation 926
Nu'man Amri Maliky, Nanda Pratama Putra, Mochamad Teguh Subarkah, Syarif Hidayat

Design and Implementation of DFT Technique to Verify LBIST at RTL Level 937
Nagaraj Vannal, Saroja V Siddamal

Empathy in Nursing Students that do the Non-Medical Internship in Three Universities in Lima, 2019 944
Liseth Acuña-Medina, Yumira Arias-Quispe, Yackeline Espeza-Veláquez, Brian Meneses-Claudio, Hernan Matta-Solis, Eduardo Matta-Solis

Investigating the Optical Behavior of Single/Multi-Dimensional Photonic Crystal Structures for Photovoltaic Applications 951
Gehad Ali Alsayed, Zahraa Ismail, Sameh O. Abdellatif

Standalone Operation of Modified Seven-Level Packed U-Cell Inverter for Solar Photovoltaic System 959
Kishan Bhushan Sahay, Pankaj Kumar Singh, Rakesh Maurya

Hand Gesture Classification using Inaudible Sound with Ensemble Method 967
Jinwon Cheon, Sunwoong Choi

Docker-C2A : Cost-Aware Autoscaler of Docker Containers for Microservicesbased Applications 972
Mohamed Hedi Fourati, Soumaya Marzouk, Mohamed Jmaiel, Tom Guerout

NPC five level inverter using SVPWM for Grid-Connected Hybrid Wind-Photovoltaic Generation System 981
Elamri Oumaymah, Oukassi Abdellah, Bouhali Omar, El Bahir Lhoussain

Mobile Money Wallet Attack Resistance using ID-based Signcryption Cryptosystem with Equality Test 988
Seth Alornyo, Mustapha Adamu Mohammed, Francis Botchey, Collinson Colin M. Agbesi

Advanced Design of Current-mode Pass-band Filter using Ant Colony Optimization Technique 995
Kritele Loubna, Benhala Bachir, Zorkani Izeddine

Feature Gate Computational Top-Down Model for Target Detection 1001
Aarthi Ramachandran, Amudha Joseph, Shunmuga Velayutham

Design of a Remote Real-time Groundwater Level and Water Quality Monitoring System for the Philippine Groundwater Management Plan Project 1007
Carlos M. Oppus, Maria Aileen Leah G. Guzman, Maria Leonora C. Guico, Jose Claro N. Monje, Mark Glenn F. Retirado, John Chris T. Kwong, Genevieve C. Ngo, Annael J. Domingo

Determinants of Technological and Innovation Performance of the Nepalese Cellular Telecommunications Industry from the Customers' Perspective 1013
Rewan Kumar Dahal, Ganesh Bhattarai, Dipendra Karki

An Analysts' Skills: Bespoke Software vs Packaged Software at Small Software Vendors 1021
Issam Jebreen

Ozone, Nitrogen Dioxide, and PM2.5 Measurement at Three Urban Parks in Manila, Philippines using Portable Sensors 1027
Maria Cecilia Galvez, Daniel Paulo Tipan, Angelo Ashtin Valera, Edgar Vallar, Alma Nacua

Evaluation of the Living Condition in Fishermen Settlement in the Coastal Area of Marisa City, Gorontalo Province 1033
Irwan Wunarlan, Sugiono Soetomo, Iwan Rudiarto

Design of a Mobile Application for the School Enrollment Process in Order to Prevent Covid-19 1042
Alexi Delgado, Enrique Lee Huamaní, Alfredo Chiara-Sotomayor, Florencio Roman-Casahuamán

Estimating LAI of Rice Using NDVI Derived from MODIS Surface Reflectance 1047
Rushikesh Kulkarni, Kiyoshi Honda

Developing a Modular Material-Based Independent Training Model for Primary School Teacher Training 1054
Asep Herry Hernawan, Mustari Bosra

Projection of Wireless Multipath Clusters Using Multi-Dimensional Visualization Techniques 1064
Aaron Don M. Africa, Emmanuel T. Trinidad, Lawrence Materum

Review of Different Methods for Optimal Placement of Phasor Measurement Unit on the Power System Network 1071
Ademola Abdulkareem, Divine Ogbe, Tobiloba Somefun

Human Emotion Recognition Based on EEG Signal Using Fast Fourier Transform and K-Nearest Neighbor 1082
Anton Yudhana, Akbar Muslim, Dewi Eko Wati, Intan Puspitasari, Ahmad Azhari, Murein Miksa Mardhia

5G, Vehicle to Everything Communication: Opportunities, Constraints and Future Directions 1089
Boughanja Manale, Tomader Mazri

Development of Secondary Processing Data Methods under Single Point Thunderstorm Activity Monitoring 1096
Anatoly Panyukov, Alexander

Evaluation of the Physico-Chemical Properties of Soil and Apple Leaves (Malus Domestica) in Beni Mellal-Khenifra Region, Morocco 1103
Berrid Nabyl, Lougraimzi Hanane, El-Khabbazi Houda, Abidli Zakaria, Hamidi Otman, Keltoum Rahali, El Mahjoub Aouane

Multi-Criteria Decision Analysis Coupled with GIS and Remote Sensing Techniques for Delineating Suitable Artificial Aquifer Recharge Sites in Tafilalet Plain (Morocco) 1109
Aicha Ousrhire, Hassane Oulidi Jarar, Abdessamad Ghafiri

Support Vector Machine Integrated with SASS-SM Approach for Parkinson's Disease Classification Withdrawn
Kadhem Al-Daffaie, Hadi Ratham Al-Ghayab

Experimental Analysis of Thin Layer Drying of Ginger Rhizome in Convective Environment 1132
Gbasouzor Austin Ikechukwu, Sabuj Mallik, Jude Ejikeme Ebem Njoku, Joshua Depriver

A Novel Approach of Smart Logistics for the Health-Care Sector Using Genetic Algorithm 1143
Marouane EL Midaoui, Mohammed Qbadou, Khalifa Mansouri

The Design of a Hybrid Model-Based Journal Recommendation System 1153
Adewale Opeoluwa Ogunde, Mba Obasi Odim, Oluwabunmi Omobolanle Olaniyan, Theresa Omolayo Ojewumi, Abosede Oyenike Oguntunde, Michael Adebisi Fayemiwo, Toluwase Ayobami Olowookere, Temitope Hannah Bolanle

An Efficient Performance of OFDM-Shaped Symbol for 5G Green Communication Compared to FBMC 1163
Mohamed Yasin Ibrahim Afifi, El-Sayed Soliman Ahmed Said, Abd El-Hady Abd El-Azim Ammar

PV Integrated Recursive Least Mean Square Estimation Based Shunt Active Power Filter 1171
Ragam Rajagopal, K. Palanisamy, S. Paramasivam

Defeating Anti-Debugging Techniques for Malware Analysis Using a Debugger 1178
Jong-Wouk Kim, Jiwon Bang, Mi-Jung Choi

Development of a Technology and Digital Transformation Adoption Framework of the Postal Industry in Southern Africa: From Critical Literature Review to a Theoretical Framework 1190
Kgabo Mokgohloa, Grace Kanakana-Katumba, Rendani Maladzhi

Evolution of Teaching Approaches for Science, Engineering and Technology within an Online Environment: A Review 1207
Rendani Wilson Maladzhi, Grace Mukondeleli Kanakana-Katumba

Development and Performance Analysis of HRPL Using 6LoWPAN CC2538 Module for IoT Ecosystem 1217
Nin Hayati Mohd Yusoff, Nurul Azma Zakaria

Priority-based Scheduling Algorithm for NOMA-integrated V2X 1225
Ala Din Trabelsi, Hend Marouane, Faouzi Zarai

Creativity in Prototypes Design and Sustainability – The case of Social Organizations 1237
Clara Silveira, Leonilde Reis, Vitor Santos, Henrique S. Mamede

Computer Vision for Industrial Robot in Planar Bin Picking Application 1244
Le Duc Hanh, Huynh Buu Tu

Crystallinity and Hardness Enhancement of Polypropylene using Atmospheric Pressure Plasma Discharge Treatment 1250
Oscar Xosocotla, Horacio Martinez, Bernardo Campillo

LEACH Based Protocols: A Survey 1258
Nour Najeeb Abdalkareem Qubbaj, Anas Abu Taleb, Walid Salameh

Towards a Documents Processing Tool using Traceability Information Retrieval and Content Recognition Through Machine Learning in a Big Data Context 1267
Othmane Rahmaoui, Kamal Souali, Mohammed Ouzzif

Time Granularity-based Privacy Protection for Cloud Metering Systems 1278
Hesham Aly El Zouka, Mustafa Mohamed Hosni

Analysis of the Operational Impact of ETC Lanes on Toll Station 1286
Alimam Mohammed Karim, Alimam Mohammed Abdellah, Seghiouer Hamid

Empirical Probability Distributions with Unknown Number of Components 1293
Marcin Kuropatwi´nski, Leonard Sikorski

Basic Study of 3-D Non-Invasive Measurement of Temperature Distribution using Ultrasound Images during HIFU Heating 1306
Ryosuke Sakakibara, Yasuhiro Shindo, Kazuo Kato, Pak Kon Choi, Akira Takeuchi

Wideband Active Switch for Electronic Warfare System Applications 1312
Mahadev Sarkar, Gaurav Anand, Sivakumar Ramadoss

Evaluation of Simple Space Interpolation Methods for the Depth of Precipitation: Application for Boyacá, Colombia 1322
Pedro Mauricio Acosta Castellanos, Alejandra Castro Ortegon, Hugo Fernando Guerrero Sierra

SH-CNN: Shearlet Convolutional Neural Network for Gender Classification 1328
Chaymae Ziani, Abdelalim Sadiq

Quality of Nursing Care in Hospitalized Patients of the Carlos Lanfranco La Hoz Hospital, 2019 1335
Amancio Izquierdo-Príncipe, Jaqueline Garcia-Núñez, Brian Meneses-Claudio, Hernan Solis-Matta, Lourdes Matta-Zamudio

Quality of Life in Patients with Type 2 Diabetes of the Central Hospital of the Peruvian Air Force, 2019 1340
Jared Zavala-Izaguirre, Fanny Mego-Llanos, Sarita Cornejo-Quispitongo, Brian Meneses-Claudio, Hernan Solis-Matta, Lourdes Matta-Zamudio

Social skills and Resilience in Adolescents of Secondary Education of the Kumamoto I 3092 Educational Institution, of the Puente Piedra District – Lima 2019 1345
Evelyn Roncal-Cespedes, Gloria Castillo-Laban, Brian Meneses-Claudio, Hernan Matta-Solis, Lourdes Matta-Zamudio, Eduardo Matta-Solis

Social Skills and Resilience in Adolescents of an Educational Institution in North Lima, 2019 1350
Lili Sifuentes-Gomez, Doris Vega-Davila, Betty Flores-Paz, Brian Meneses-Claudio, Hernan Matta-Solis, Eduardo Matta-Solis

Improved Design and Recommendations for Street Lighting in Gitega City 1356
Ntawuhorakomeye Noel, Ndayiragije Leonidas, Belov Mikhail Petrovich

Investigating Students' Computational Thinking through STEM Robot-based Learning Activities 1366
Sasithorn Chookaew, Suppachai Howimanporn, Santi Hutamarn

Effect of Starch Oxidation Degree on the Properties of Hydrogels from Dialdehyde Starch and Polyvinyl Alcohol 1372
Jahel Desire Carrera, Daniela Alejandra Viteri Narváez, Marco Leon, José Francisco Alvarez-Barreto

Image Tag Recommendation based on Ranked Categorical Nearest Neighbors and Weighted Tag Features 1381
Anupama D. Dondekar, Balwant A. Sonkamble

Analysis of Physical and Mechanical Properties of Galvanic-Plasma Wear-Resistant Coatings 1387
Mahmood Shaker Albdeiri, Sergey Sergeev, Vladimir V. Krasilnikov

A prediction of Cutting Force, System Vibration, and Productivity in Five-Axis Milling Process of the Spiral Bevel Gear 1394
Nguyen Van Thien, Dung Hoang Tien, Nhu-Tung Nguyen, Nguyen Van Que, Do Duc Trung, Pham Thi Thieu Thoa

A Study on the Tool Wear in Milling Process of the Gleason Spiral Bevel Gear 1402
Hoang Xuan Thinh, Pham Van Dong, Tran Ve Quoc

Recommendation System for SmartMart-A Virtual Supermarket 1408
Poonam Ghuli, Manoj Kartik R, Mohammed Amaan, Mridul Mohta, N Kruthik Bhushan, Poonam Ghuli, Shobha G

Solution of the Semiconductor-Device Equations by the Numerov Process 1414
Nicol`o Speciale, Rossella Brunetti, Massimo Rudan

Handling Priority Data in Smart Transportation System by using Support Vector Machine Algorithm 1422
Sara Ftaimi, Tomader Mazri

Fabrication and Optimization of High Frequency ZnO Transducers for Both Longitudinal and Shear Emission: Application of Viscosity Measurement using Ultrasound 1428
Hatem Dahmani, Ibrahim Zaaroura, Abbas Salhab, Pierre Campistron, Julien Carlier, Malika Toubal, Souad Harmand, Vincent Thomy, Marc Neyens, Bertrand Nongaillard

Parameter Estimation for Industrial Robot Manipulators Using an Improved Particle Swarm Optimization Algorithm with Gaussian Mutation and Archived Elite Learning 1436
Abubakar Umar, Zhanqun Shi, Lin Zheng, Alhadi Khlil, Zulfiqar Ibrahim Bibi Farouk

Comparative Study Between Three Methods for Optimizing the Power Produced from Photovoltaic Generator 1458
El hadji Mbaye Ndiaye, Mactar Faye, Alphousseyni Ndiaye

An Investigation of the Effect of Optimal Plane Spacing Between Electrode Planes for the EIT Industrial Applications 1466
Yew Lek Chong, Renee Ka Yin Chin

Social Engineering to Establish Digital Culture in Higher Education 1474
Tedi Priatna, Dian Sa'adillah Maylawati, Hamdan Sugilar, Muhammad Ali Ramdhani

Simulation Based Energy Consumption Optimization for Buildings by Using Various Energy Saving Methods 1480
Omar Khaled Barakat, Ahmed El-Biomey Mansour, Mahmoud Mohamed Abd Elrazik, Ashraf Aboshosha, Amir Yassin Hassan

Impact of Changing Microstructural Compositions of Lime Based Mortar on Flexibility: Case Study of Sustainable Lime-Cement Composites) 1488
Sule Adeniyi Olaniyan

Comparative Analysis Spline Methods in Digital Processing of Signals 1499
Hakimjon Zaynidinov, Sayfiddin Bakhromov, Bunyod Azimov, Sarvar Makhmudjanov

Impact of Salinity on Stabilized Leachate Treatment from Ozonation Process 1511
Iva Yenis Septiariva, I Wayan Koko Suryawan, Novi Kartika Sari, Ariyanti Sarwono

Implementation of a Journalist Business Intelligence in Social Media Monitoring System 1517
Abba Suganda Girsang, Sani Muhamad Isa, Natasya, Megga Eunike Cristilia Ginzel

Prophet Architecture in Normalized Meter Energy Consumption Prediction on Building 1529
Fernando Lioexander, Abba Suganda Girsang

A Software-Defined Network Approach for The Best Hospital Localization Against Coronavirus (COVID-19) 1537
Bilal Babayigit, Eda Nur Hascokadar

NemoSuite: Web-based Network Motif Analytic Suite 1545
Wooyoung Kim, Yi-Hsin Hsu, Zican Li, Preston Mar, Yangxiao Wang

Application of Deep Belief Network in Forest Type Identification using Hyperspectral Data 1554
Xianxian Luo, Songya Xu, Hong Yan

A Computational Modelling and Algorithmic Design Approach of Digital Watermarking in Deep Neural Networks 1560
Revanna Sidamma Kavitha, Uppara Eranna, Mahendra Nanjappa Giriprasad

ESP2: Embedded Smart Parking Prototype 1569
Tarek Frikha, Hedi Choura, Najmeddine Abdennour, Oussama Ghorbel, Mohamed Abid

A Framework for Adoption and Diffusion of Mobile Applications in Africa 1577
Chinedu Wilfred Okonkwo, Magda Huisman, Estelle Taylor

Features Preference using Conjoint Analysis Method for E-marketplace Social Care System 1593
Angelina Ervina Jeanette Egeten, Harjanto Prabowo, Ford Lumban Gaol, Meyliana

High-Performance Computing: A Cost Effective and Energy Efficient Approach 1598
Safae Bourhnane, Mohamed Riduan Abid, Khalid Zine-Dine, Najib Elkamoun, Driss Benhaddou

Time-to-Event Analysis for Recovery from Coronavirus Disease (COVID-19): A Case Study on Wuhan and Elsewhere in China from Jan 1 to Feb 11, 2020 1609
Murtada Khalafallah Elbashir, Saleh N. Almuayqil

Elasticity Based Med-Cloud Recommendation System for Diabetic Prediction in Cloud Computing Environment 1618
Karamath Ateeq, Manas Ranjan Pradhan, Beenu Mago

On Design of IoT-based Power Quality Oriented Grids for Industrial Sector 1634
Nesma N. Gomaa, Khaled Y. Youssef, Mohamed Abouelatta

FPGA-Based Homogeneous and Heterogeneous Digital Quantum Coprocessors 1643
Valeriy Hlukhov

Type 2 Diabetes Risk and Physical Activity in outpatients treated in Health Centers in a District of North Lima, 2020 1651
Deisy Chipana-Collahua, Mariluz Chipana-Collahua, Rosa Villegas-Ortiz, Brian Meneses-Claudio, Hernan Matta-Solis

A Novel Way to Design ADS-B using UML and TLA+ with Security as a Focus 1657
Pranay Bhardwaj, Carla Purdy, Nawar Obeidat

Real-time Gradient-Aware Indigenous AQI Estimation IoT Platform 1666
Hasan Tariq, Abderrazak Abdaoui, Farid Touati, Mohammad Abdullah Al Hitmi, Damiano Crescini, Adel Ben Mnaouer

Japanese Abstractive Text Summarization using BERT 1674
Yuuki Iwasaki, Akihiro Yamashita, Yoko Konno, Katsushi Matsubayashi

Sentiment Analysis in English Texts 1683
Arwa Alshamsi, Reem Bayari, Said Salloum

Analysis and Evaluation of Competitiveness in Medical Tourism Industry in Taiwan 1690
Yen-Hung Chen, Tin-Chang Chang

Education Value Chain Model for Examination, Grading, and Evaluation Process in Higher Education based on Blockchain Technology 1698
Meyliana, Yakob Utama Chandra, Cadelina Cassandra, Surjandy, Erick Fernando, Henry Antonius Eka Widjaja, Harjanto Prabowo

Analysis of Learning Difficulties in Object Oriented Programming in Systems Engineering Students at UNTELS 1704
Teodoro Diaz-Leyva, Omar Chamorro-Atalaya

Efficient and Scalable Ant Colony Optimization based WSN Routing Protocol for IoT 1710
Afsah Sharmin, Farhat Anwar, S M A Motakabber

Design of Power Efficient Routing Protocol for Smart Livestock Farm Applications 1719
Shahenda S. Abou Emira, Khaled Y. Youssef, Mohamed Abouelatta

New Properties of Crimes in Virtual Environments 1727
Roman Dremliuga, Natalia Prisekina, Andrei Yakovenko

A Machine Vision Approach for Underwater Remote Operated Vehicle to Detect Drowning Humans 1734
Yaswanthkumar S K, Keerthana M, Vishnu Prasath M S

An Evaluation of Some Machine Learning Algorithms for the Detection of Android Applications Malware 1741
Olorunshola Oluwaseyi Ezekiel, Oluyomi Ayanfeoluwa Oluwasola, Irhebhude Martins

Touristic's Destination Brand Image: Proposition of a Measurement Scale for Rabat City (Morocco) 1750
Abdellatif Elouali, Smail Hafidi Alaoui, Noura Ettahir, Abderrazzak Khohmimidi, Nadia Motii, Keltoum Rahali, Mustapha Kouzer

The Designing of Institute's Educational Mascots for Brand Identity 1759
Nop Kongdee, Suparada Prapawong, Manissaward Jintapitak

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 01-08 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Qualitative Properties of a Cell Proliferating Model with Multi-phase Transition and Age Structure

Youssef El Alaoui*,[1], Larbi Alaoui[2]

[1]*Faculty of Sciences, University Mohammed V - Agdal, Rabat, 10010, Morocco*

[2]*International University of Rabat, Sala Al-Jadida, 11100, Morocco*

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 18 October, 2020
Online: 08 November, 2020

Keywords:
Semigroup
Dual semigroup
Asynchronous exponential growth

ABSTRACT

In this paper we study a cell division cycle modeled by a system of partial differential equations with an age structure. This model translates the many regulatory mechanisms within the cell cycle where it introduces the notion of phases. The individual cell can be either in I phases where the transition between theses phases are ordered and unidirectional. The model is related to the suns and stars caluculus via the dual semigroups of operators that are considered as solution of an abstract integral equation equivalent to a Volterra type equation of the form $w(t) = \Phi(w_t)$. We will determine the core operator Φ and prove that the semigroup solution of the model possesses the asynchronous exponential property. The model permits different types of controls where the provided framework allows better control on the model parameters and yields the characterization of the intrinsic rate of natural increase through properties of the core operator Φ. Finally, we demonstrate that the asymptotic behavior of the model is governed by the simple dominant eigenvalue and its associated eigenvector, that leads to the dispersion of the cell structure through the future generations.

1 Introduction

This paper is considered as an extension of the work originally presented in 2019 International Conference of Computer Science and Renewable Energies ICCSRE [1], that represents more refined work where we investigated the long time behavior of a cell cycle model using a well founded mathematical theory.
The cell cycle is an ordered set of molecular events that a cell undergoes until it divides into two daughter cells. It has been recognized that the cell cycle can be divided into two major phase: The interphase and the mitosis phase denoted by M. The interphase is commonly composed of three phases : G_1, S, and G_2. It is the longest part of the cell cycle and is devoted to duplication of the cell's component. During the S phase, the chromosomes are duplicated. It is known that throughout the M phase, the cell split, producing two daughter cells [2]. In each phase the progression in the division cycle is reglemented by a serial of molecular regulation events which are ordered, directional and punctuated by checkpoints. These latter make use of some protein dephosphorylation and degradation where they are under control of the molecular circadian clocks. In fact, several animal experiments showed that the role of these clocks are essential to exert a rhythmic regulating control of each step [3], [4]. These steps are intracellular cascades which ensures the smooth conduct of the proliferation. The extracellular factors such growth factors play an important role in the controling strategy which ensures the environmental requirements are adapted to the proliferation of the cell. Thus, the regulation of the cell cycle is a crucial process to the survival of the cell. It is known that there exists two key classes of regulatory molecules: cyclins (regulatory subunits) and cyclin dependent kinases (catalytic subunits) denoted by CDKs where through phosphorylation activates or inactivates target proteins to permit entry in the next phase of the cell cycle.

Through the cell cycle, there exists different restriction points. The first one is called start checkpoint. It regulates the progression in the G1 phase. The transition from G_1 to S is regulated by CDK4/cyclin D and CDK6/cyclin D where CDK is expressed in cells and cyclin D is synthesized due to the growth factors. Various cyclins are produced at each phase: S, G_2 and M, resulting in the periodic formation of distinct combinations of cyclin-CDK complexes that trigger the cell cycle events [2, 5].

These various mechanisms of regulation are showed to be inefficient in cancer cells. In addition, the disruption of circadian clocks is showed to enhance the tumour growth [6]–[8]. In fact, the circadian rhythms play an important role in the cell cycle control where they contribute to the cell cycle regulation by the synchronisation of the

*Corresponding Author: Youssef EL ALAOUI, is.youssefelalaoui@gmail.com

control system. However, these dynamics have still to be understood and investigated. Further, various speculation tried to explain this phenomenon which give rise to the concept of synchronisation, that is, to the speculation that less synchronised phases are within cell population, the faster is proliferation. It is known that circadian proteins control CDKs and theses latter control phase transitions. In the following, we want to investigate the question if circadian clocks control the proliferation of the cell population.

The paper is concerned with a system of partial differential equations (PDE) describing the case of a multi transition phases cycle [9] in which cells can transit from one phase to the next one. The cell cycle is assumed to be divided into I different phases where the individual cell can be distinguished from one another according to their age a. These cells can growth, die and transit to the next phase where the rates describing the associated physiological process depend only on the cell's age. Then, the model read as follows (see figure 1)

$$\begin{aligned} &\frac{\partial}{\partial t}p_i(t,a)+\frac{\partial}{\partial a}p_i(t,a)=-\mu_i(a)p_i(t,a)-\mathcal{K}_{i\to i+1}(a)p_i(t,a)\\ &p_i(t,0)=\int_0^\infty \mathcal{K}_{i-1\to i}(a)p_{i-1}(t,a)da \quad \text{for } 2\le i\le I \\ &p_1(t,0)=2\int_0^\infty \mathcal{K}_{I\to 1}(a)p_I(t,a)da \end{aligned} \tag{1}$$

Here $p_i(t,a)$ is the density of the population of cells at the ith phase of age a at the instant t, that is, $\int_{a_1}^{a_2} p_i(t,a)da$ is the number of cells at the ith phase with an age in the interval (a_1,a_2) at the instant t. The transition from the state i to the state j is given by the rate $\mathcal{K}_{i\to j}(a)$, that is $\mathcal{K}_{i\to j}(a)dt$ is the probability that a cell having an age a at instant t in the phase i transit in the time interval $(t,t+dt)$ to the phase j, in which they start with age 0, the last phase when $i=I$ refer to mitosis as it is expressed by the boundary conditions in (1). At the ith state, the cells go to apoptosis at the rate $\mu_i(a)$, that is $\mu_i(a)dt$ is the probability that a cell having age a at instant t die in the time interval $(t,t+dt)$ in the phase i. Since the effective separation of the mother cell into two daughter cells, that is, since the cell birth, the individual cell can quantitated by a continuous variable corresponding to cell age.

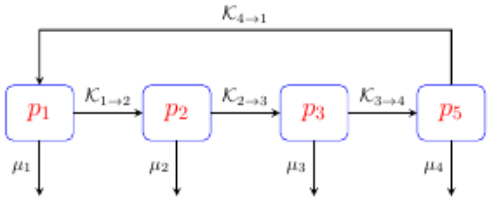


Figure 1: Flow chart of the model describing the evolution of cells along each phase of the cycle and the transition to the next phase between the cell densities. Here I=4, this corresponds to G1-S-G2-M.

It is known that cells transit through the cell cycle with variability in intermitotic times which correspond to different biological processes such the proteic (G_1 phase) or DNA (S phase) or microtuble (M phase) synthesis . As contrary to the model given in [9], the progression speed is assumed to be constant, that is one does not take into account the external growth factors. Indeed, the progression speed can be enhanced by the external growth factors in the early of the G_1 phase.

The authors in [9]–[16] considered a partial differential equations modeling a proliferating cell population with taking into consideration their external controls which can be hormonal, circadian, pharmacological . . . that leads to investigate the consequences of a periodic control on these models. Their results assert that the effect of circadian control on tumour proliferation is indirect.

Here, the death rates are supposed only age dependent. However, the case time-periodic death rates studied in [10, 11], using the Floquet theory, had shown a higher first eigenvalue and thus enhanced proliferation of the population.

For the one phase cell transition model [16, 17], it has been proved that if we influence the cell cycle with some control of a periodic function, we obtain that cells with cell cycle duration slightly lower than the control function period are selectively more advantageous, in addition, they hypothetize that the effect of disruption of circadian rhythms on tumour growth enhancement is indirect. The model for I phases was first introduced in [9], where the authors studied the long time behaviour of this model with the effect of a circadian control using an entropy method. Their results show that it imposes a circadian rythm to the cell cycle.

Mathematical cell cycle models have been treated since the 1960s by authors such as Fredrickson, Bell, Sinko and Streifer. These models were traditionally formulated as partial differential equations for the age and/or size distribution of the cell population. Diekmann et al. [18] used semigroup methods to give conditions for when there is, and also when there is not, asynchronous exponential growth of the population in a constant environment. More recently, Diekmann and coworkers have argued in favour of using (abstract) integral equations for the birth rate instead of PDE, also, and in particular, for the nonlinear models resulting from interaction via the environment.

The purpose of this paper is to create a mathematical framework for the analysis of this type of models by using the theoretically well established semigroup method based on the suns and stars calulcus using the perturbed dual semigroups of operators on the spaces of the form $\mathcal{E}:=L^1((-h,0),\mathcal{F})$. These semigroups are solutions of the integral equation [19]–[21] of the type :

$$p(t)=P_0(t)f+\int_0^t P_0^{\odot *}(t-\tau)\Phi(p(\tau))d\tau, \quad \text{for } t\ge 0 \tag{2}$$

where $h>0$, $\mathcal{F}$ is a Banach space and $\Phi:\mathcal{E}\to\mathcal{Z}$ and $\mathcal{Z}$ is some bigger Banach space. It has been proved that this method is well suited for applications, in particular in the field of cell biology models [22, 23].

In the following, the main result is to provide necessary and sufficient conditions on the parameters building the cell division cycle model (1) in order to yield the property of asynchronous exponential growth.The asynchronous exponential property is explained mathematically by the fact that a simple dominant eigenvalue govern the spectrum of the infinitesimal generator of the semigroup solution of the model. In the context of the semigroup theory, this operator will be the associated infinitesimal generator. We can say that the asynchronous exponential growth occurs when the following estimate $\|e^{-\lambda_0 t}p(t,\cdot)-\mathcal{P}p_0\|$ decreases to 0 exponentially as t tends to infinity, where the vector population is defined by $p(t,a)=(p_i(t,a))_{1\le i\le I}^{tr}$ and $p_0(a)$ is the initial given distribution of the population. This latter

means that when the cell population is multiplied by an exponential factor, know an asymptotic stabilization around a unidimensional projection. It is noted that this projection is only dependent on the initial structure of the concerned population.
We will use the theory developed in the works [19]–[21] to exhibit the link between the solutions of the model and the perturbed dual semigroups. First, we determine the operator Φ. Secondly, we derive different results, only based on the properties of the operator Φ ranging from existence to qualitative properties and particulary give a characterization of the Malthusian parameter as a spectral radius of some bounded operator.
Throughout this paper, the two parameters $\mathcal{K}_{i\to j}(a)$ and $\mu_i(a)$ are supposed to satisfy the assumption

$\left(\mathbf{H}_{\mathcal{K},\mu}\right)$: (1) The rates $\mathcal{K}_{i\to j}, \mu_i$ defined from $(0,\bar{a})$ to $\mathbb{R}$ are continuous.
(2) $\mathcal{K}, \mu \in L^\infty((0,\bar{a}),\mathbb{R})$ such that there exists $0 < \bar{\mathcal{K}} < \infty$ where $0 < \mathcal{K}_{i\to j}(a) \le \bar{\mathcal{K}}$ and$\mathcal{K} \not\equiv 0$ in addition there exists $\bar{\mu} > 0$ such that $0 \le \mu_i \le \bar{\mu}$

2 Well posedness

In this section, the mathematical analysis of the multi phases proliferating cell cycle model will be provided using the theory developed for the perturbed dual semigroups [24]–[27] associated to the equation of Volterra type :

$$p(t) = \phi(p_t) \tag{3}$$

where the Hale notation $p_t(s) := p(t+s)$ is used (see [28]–[35]), and the operator $\phi : \mathcal{E} := L^1((-h,0),\mathcal{F}) \to \mathcal{F}$. Our construction will permit to link the solutions of the model (1) to the solutions of the integral equation

$$p(t) = P_0(t)f + j^{-1}\left(\int_0^t P_0^{\odot *}(t-s)\Phi(p(s))ds\right) \tag{4}$$

where P_0 satisfies the properties of a semigroup of bounded linear operators on $\mathcal{E}$ with $\mathcal{E}^\odot \subset \mathcal{E}^*(\subset$ stands as the linear subspace) on which P_0^* yields the strong continuity, that is,

$$\mathcal{E}^\odot = \{f^* \in \mathcal{E}^* : \lim_{t\downarrow 0}\|P^*(t)f^* - f^*\| = 0\}$$

It can be shown that this subspace is closed in the norm topology of $\mathcal{E}^*$ and weak* dense subspace on $\mathcal{E}^*$ [24]. In addition, the semigroup on $\mathcal{E}^\odot$ is denoted by $P_0^\odot := P_0^*|_{\mathcal{E}^\odot}$. It is obvious that this restriction build a semigroup that is strongly continuous on the subspace $\mathcal{E}^\odot$. Since we consider an element of $\mathcal{E}$ as a bounded linear functional on $\mathcal{E}^\odot$. We can embed $\mathcal{E}$ into $\mathcal{E}^{\odot *}$ via the map q of $\mathcal{E}$ into $\mathcal{E}^{\odot *}$ defined by :

$$< f^\odot, qf >=< f, f^\odot >, \qquad \text{for } f \in \mathcal{E}, f^\odot \in \mathcal{E}^\odot$$

It is obvious that $q(\mathcal{E}) \subset \mathcal{E}^{\odot\odot}$. When $q(\mathcal{E}) = \mathcal{E}^{\odot\odot}$ then $\mathcal{E}$ is said to be $\odot$-reflexive with respect to $P(t)$.

Go back to the study of the model (1). The state space considered is $\mathcal{E} := L^1([0,\tilde{a}],\mathbb{R}^I)$. The parameter $\tilde{a} > 0$ is the maximal age that an individual cell can attain. This space is endowed with norm

$$\|f\|_{\mathcal{E}} = \int_0^{\tilde{a}} \|f(a)\|_{\mathbb{R}^I} da, \qquad f \in \mathcal{E}$$

Next, we introduce the operator $\chi(a)$ that yields the equation

$$\begin{aligned} \chi'(a) &= -\chi(a)\mathcal{M}(a) \\ \chi(0) &= Id(\text{ stands for identity}) \end{aligned} \tag{5}$$

with

$$\mathcal{M}(a) = \mathrm{Dg}\,(m_{i,i}(a))_i \qquad 1 \le i \le I \tag{6}$$

where Dg stands for the diagonal matrix and the entries of the matrix $\mathcal{M}$ hold

$$m_{i,j}(a) := \begin{cases} \mu_i(a) + \psi_{i\to i+1}(a) & \text{if } i = j \\ 0 & \text{else} \end{cases} \tag{7}$$

The aim of the introduction of this transformation is to demonstrate that $\{P(t)\}_{t\ge 0}$ the semigroup solution of the model (1) is equivalent to $(P_\phi(t))_{t\ge 0}$ the semigroup solution of (4). Thus, via this machinery and under the properties of the operator ϕ, we will establish the asynchronous exponential property only by handling with the equivalent semigroup and go back to the original one by using the cited transformation.
Before going to give the main results, we need some settings. First, let set $p(t,a) := (p_i(t,a))_{1\le i\le I}^{tr}$. So via this vector, our initial model (1) can be reformulated as an vector type Mckendrick Von Foerster model

$$\frac{\partial}{\partial t}p(t,a) + \frac{\partial}{\partial a}p(t,a) = -\mathcal{M}(a)p(t,a) \tag{8}$$

with the boundary condition at age 0

$$p(t,0) = \int_0^{\tilde{a}} \mathcal{B}(a)p(t,a)da \tag{9}$$

The matrix $\mathcal{M}$ is given above by (6) and (7) and the matrix $\mathcal{B}$ is given by

$$B(a) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 2K_{I\to 1} \\ K_{1\to 2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & K_{2\to 3} & 0 & \cdots & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & \cdots & 0 & K_{I-1\to I} & 0 \end{pmatrix}$$

After we gave a compact form to the original model (1). Now, we consider the transformation

$$m(t,.) := \Gamma p(t,.) \tag{10}$$

where $(\Gamma g)(a) := \chi(a)g(a)$ with $a \in (0,\bar{a})$ and the operator χ satisfies the first order differential equation (5). Via this latter, we arrive at the equivalent system where it is more easily to handle than the original one (8). The model is formulated as given

$$\begin{aligned} \frac{\partial}{\partial t}m(t,a) + \frac{\partial}{\partial a}m(t,a) &= 0 \\ m(t,0) &= \int_0^{\tilde{a}} \theta(a)m(t,a)da \end{aligned} \tag{11}$$

The kernel θ is given by

$$\theta(a) = \mathcal{B}(a)\chi^{-1}(a), \quad \text{for } a \in (0,\tilde{a}) \tag{12}$$

The diagonal matrix χ^{-1} is given by

$$\chi^{-1}(a) = \mathrm{Dg}\left(\tilde{z}_{i,i}(a)\right)$$
$$\tilde{z}_{i,j} := \begin{cases} e^{-\int_0^a m_{i,i}(\tau)d\tau} & \text{if } i = j \\ 0 & \text{else} \end{cases} \tag{13}$$

Definition 2.1 *The strongly continuous semigroup $\{P(t)\}$ is said to be a translation semigroup on $L^1((-r,0),\mathcal{F})$ with $r > 0$ and $\mathcal{F}$ is a Banach space if the following $(P(t)f)(x) = f(x+t)$ is satisfied when $x + t < 0$.*

Diekmann et al. have demonstrated in [20] that the space $\mathcal{E}$ is $\odot$-reflexive with respect to P_0, where the operator P_0 is defined as the translation semigroup to the right and 0 on $\mathbb{R}^+$. Its generator $\mathcal{A}_0$ is given by

$$\mathcal{D}(\mathcal{A}_0) = \{f \in \mathcal{E} : f \in AC \text{ and } f(0) = 0\}$$
$$\mathcal{A}_0 f = -f'$$

where AC stands for the absolutely continuous space. We have the following identification (see [19] and [20])

$$L^1([0,\bar{a}],\mathbb{R}^I)^{\odot} \to C_0((-\tilde{a},0],\mathbb{R}^I)$$
$$L^1([0,\bar{a}],\mathbb{R}^I)^{\odot *} \to \mathrm{NBV}([0,\tilde{a});\mathbb{R}^I)$$

The semigroup $P_0^{\odot *}$ is the translation to the right and 0 on $\mathbb{R}^+$. Its generator $\mathcal{A}_0^{\odot *}$ is given by

$$\mathcal{D}(\mathcal{A}_0^{\odot *}) = \{f^{\odot *} \in \mathcal{E}^{\odot *} : f^{\odot *}(a) = \int_0^a g(s)ds \text{ and } g \in \mathcal{E}^{\odot *}\}$$
$$\mathcal{A}_0^{\odot *} f^{\odot *} = -f^{\odot *} = g$$

Using integration and the natural embedding q from $\mathcal{E}$ into $\mathcal{E}^{\odot *}$ and taking into account the boundary condition, we arrive at

$$m(t) = P_0(t)g + q^{-1}\int_0^t P_0^{\odot *}(t-s)\Phi(m(s))ds \tag{14}$$

where the core operator Φ defined from $\mathcal{E}$ to $\mathcal{E}^{\odot *}$ is given as a perturbation with finite dimensional range

$$\Phi(g) = \phi_1(g)\mathcal{H}_1 + \phi_2(g)\mathcal{H}_2 + \ldots + \phi_I(g)\mathcal{H}_I \tag{15}$$

The Heaviside function $\mathcal{H}_i$ is given by $\mathcal{H}_i(a) = e_i$ if $a \in (0,\tilde{a})$ else $\mathcal{H}_i(a) = 0$ where $\{e_1, e_2, \ldots, e_I\}$ is the canonical basis of $\mathbb{R}^I$. The canonical injection q is defined by $(qx)(a) = \int_0^a x(s)ds,\ s \in [0,\tilde{a})$ where the inverse is clearly differentiation.
The operator given in (11) is defined from $\mathcal{E}$ to $\mathbb{R}^I$ and it is given by the integral formula

$$\Phi g = \int_0^{\tilde{a}} \theta(s)g(s)ds \tag{16}$$

the kernel θ is given by (12). In the following, we adopt the following Hale notation

$$m_t(a) := m(t-a), \quad \text{for all } a \in [0,\tilde{a}]$$

with

$$m_0(a) = g(a) \qquad a \in [0,\tilde{a}] \tag{17}$$

In fact, the solutions of the formulation given by the integral equation (14) is proven to be equivalent to the solutions of the renewal equation [20]

$$\begin{aligned} m(t) &= \Phi(m_t), \quad \text{for } t > 0 \\ m_0(a) &= g(a), \quad a \in [0,\tilde{a}] \end{aligned} \tag{18}$$

where the latter generates a translation semigroup [28] given by

$$\mathcal{T}_\Phi g(a) = \begin{cases} g(a-t) & \text{if } a - t > 0 \\ \Phi(\mathcal{T}_\Phi(t-a)g) & \text{if } a - t \le 0 \end{cases}$$

The following result gives us the existence and uniqueness of the perturbed semigroup solution of the abstract integral equation (14)

Theorem 2.2 *Since the assumption $(\mathbf{H}_{\mathcal{K},\mu})$ holds. We obtain that $\phi \in \mathcal{L}(\mathcal{E},\mathcal{F})$. Then for the initial data $g \in \mathcal{E}$, the equation (14) has unique solution $m(t)$ on $[0,T)$ for some positive real number T where the operator $\mathcal{A}_\phi$ characterized as the following*

$$\begin{aligned} \mathcal{D}(\mathcal{A}_\phi) &= \{\varphi \in W^{1,1}((0,\bar{a}),\mathbb{R}^I), \varphi(0) = \phi\varphi\} \\ \mathcal{A}_\phi\varphi &= -\varphi', \qquad \varphi \in \mathcal{D}(\mathcal{A}_\phi) \end{aligned} \tag{19}$$

is the inifinitesimal generator of the semigroup P_ϕ on $\mathcal{E}$.

Proof. The hypothesis given by $(\mathbf{H}_{\mathcal{K},\mu})$ allows to us to obtain the first assertion about the boundedness of ϕ. Then, by using regular arguments on contraction mapping theory, we obtain the existence of the perturbed semigroup $(P_\phi(t))_{t\ge 0}$ solution of (14) on $\mathcal{E}$. □

3 Compactness and irreducibility

It is known that the eventual compactness result is often used to determine the asymptotic behavior of semigroups. In fact, it implies that the growth bound of the semigroup denoted by $\omega(T(t))$ equals the spectral bound denoted by $s(A)$, where

$$\omega(T) = \inf\{w : \|T(t)\| \le Me^{wt}, M \ge 0\}$$

and

$$s(A) = \begin{cases} \sup\{\Re e\lambda : \lambda \in \sigma(A)\} & \text{if } \sigma(A) \ne \emptyset \\ -\infty & \text{if } \sigma(A) = \emptyset \end{cases}$$

Definition 3.1 *We say that the semigroup $\{P(t)\}_{t\ge 0}$ is eventually compact if for some $t_0 \ge 0$ the operators $P(t)$ are compact for all $t \ge t_0$.*

We will give a criterion to obtain the compactness of the semigroup.

Lemma 3.2 *Since $\Phi \in \mathcal{L}(\mathcal{X},\mathbb{R}^I)$ is with finite Range. The semigroup $\{P(t)\}_{t\ge 0}$ given by (14) is eventually compact.*

Proof. The core operator Φ is spanned by the linearly independent set $\{\mathcal{H}_i\}_{1\le i\le I}$ in $\mathcal{X}^{\odot *}$, that is, Φ is with Range$(\Phi) < \infty$, where the core operator Φ takes the form given by (15). We know that there exists $\Phi_i^* \in L^\infty([-\tilde{a},0],\mathbb{R}^I)$ such that :

$$\Phi_i(\zeta) = \int_0^{\tilde{a}} < \zeta(a), \Phi_i^*(-a) > da, \qquad \forall \zeta \in \mathcal{X}$$

We conclude that :

$$\Phi(\zeta) = \sum_{i=1}^{n} \int_0^{\tilde{a}} < \zeta(a), \Phi_i^*(-a) > \mathcal{H}_i da, \qquad \forall \zeta \in \mathcal{X}$$

Let $t > 2\tilde{a}$, we have $\forall \tau, \tau' \in [0, \tilde{a}],\ \tau' < \tau$ for all f in the unit ball of $\mathcal{X}$. Then,

$$\begin{aligned}
&\|[P(t)\zeta](\tau) - [P(t)\zeta](\tau')\|_{\mathbb{R}^n} \\
&= \|q^{-1} \int_0^t P_0^{\odot *}(t-s) \left\{ (\Phi(P(s)\zeta)ds)(\tau) - (\Phi(P(s)\zeta)ds)(\tau') \right\} \|_{\mathbb{R}^n} \\
&= \|q^{-1} \left(- \int_{t-\max\{t,\tau\}}^{t} \Phi(P(s)\zeta)ds + \int_{t-\max\{t,\tau'\}}^{t} \Phi(P(s)\zeta)ds \right) \|_{\mathbb{R}^n} \\
&= \|q^{-1} \left(\int_{t-\max\{t,\tau'\}}^{t-\max\{t,\tau\}} \Phi(P(s)\zeta)ds \right) \|_{\mathbb{R}^n} \\
&= \|q^{-1} \left(\sum_{i=1}^{n} \int_0^{\tilde{a}} \int_{t-\max\{t,\tau'\}}^{t-\max\{t,\tau\}} < P(s)\zeta(a), \Phi_i^*(-a) > \mathcal{H}_i ds\, da \right) \|_{\mathbb{R}^n} \\
&= \|q^{-1} \left(\sum_{i=1}^{n} \int_0^{\tilde{a}} \int_{t-\max\{t,a+\tau'\}}^{t-\max\{t,a+\tau\}} < \Phi(P(s)\zeta), \Phi_i^*(-a) > \mathcal{H}_i ds\, da \right) \|_{\mathbb{R}^n}
\end{aligned}$$

Let $\alpha_1 = t - \tau$ and $\alpha_2 = t - \tau'$

$$\begin{aligned}
&\|[P(t)\zeta](\tau) - [P(t)\zeta](\tau')\|_{\mathbb{R}^n} \\
&= \|q^{-1} \left(\sum_{i=1}^{n} \int_{\alpha_1}^{\tilde{a}+\alpha_1} \int_0^a < \Phi(P(s)\zeta), \Phi_i^*(\alpha_1 - a) > \mathcal{H}_i ds\, da \right. \\
&\left. - \sum_{i=1}^{n} \int_{\alpha_2}^{\tilde{a}+\alpha_2} \int_0^a < \Phi(P(s)\zeta), \Phi_i^*(\alpha_2 - a) > \mathcal{H}_i ds\, da \right) \|_{\mathbb{R}^n} \\
&= \|q^{-1} \left(\sum_{i,j=1}^{n} \int_{\alpha_1}^{\tilde{a}+\alpha_1} \int_0^a < \Phi_j(P(s)\zeta)\mathcal{H}_j, \Phi_i^*(\alpha_1 - a) > \mathcal{H}_i ds\, da \right. \\
&\left. - \sum_{i,j=1}^{n} \int_{\alpha_2}^{\tilde{a}+\alpha_2} \int_0^a < \Phi_j(P(s)\zeta)\mathcal{H}_j, \Phi_i^*(\alpha_2 - a) > \mathcal{H}_i ds\, da \right) \|_{\mathbb{R}^n} \\
&= \|q^{-1} \left(\sum_{i,j=1}^{n} \int_0^a \left\{ \int_{\alpha_2}^{\tilde{a}+\alpha_2} < \Phi_j(P(s)\zeta)\mathcal{H}_j, \Phi_i^*(\alpha_1 - a) - \Phi_i^*(\alpha_2 - a) > \right. \right. \\
&\mathcal{H}_i da + \int_{\alpha_1}^{\alpha_2} < \Phi_j(P(s)\zeta)\mathcal{H}_j, \Phi_i^*(\alpha_2 - a) > \mathcal{H}_i da \\
&\left. \left. + \int_{\tilde{a}+\alpha_1}^{\tilde{a}+\alpha_2} < \Phi_j(P(s)\zeta)\mathcal{H}_j, \Phi_i^*(\alpha_2 - a) > \mathcal{H}_i da \right\} ds \right) \|_{\mathbb{R}^n} \\
&\leq \sup_{1 \leq i,j \leq n} \|\mathcal{H}_i\| \|\mathcal{H}_j\| \sum_{i,j=1}^{n} \left(\sup_{0 \leq s \leq 2\tilde{a}} |\Phi_j(P(s))\zeta)| \times \right. \\
&\left\{ \int_{\alpha_2}^{\alpha_2+\tilde{a}} |\Phi_j^*(\alpha_1 - a) - \Phi_j^*(\alpha_2 - a)| da + \int_{\alpha_1}^{\alpha_2} |\Phi_i^*(\alpha_2 - a)| da \right. \\
&\left. \left. + \int_{\tilde{a}+\alpha_1}^{\tilde{a}+\alpha_2} |\Phi_i^*(\alpha_2 - a)| da \right\} \right)
\end{aligned}$$

Hence, we obtain that $\{P(t)\zeta\}_{t\geq 0}$ is equicontinuous since $|\alpha_2 - \alpha_1|$ tends to zero uniformly for ζ in the unit ball of $C([0, \tilde{a}], \mathbb{R}^n)$. The rest of the proof is due to applying a famous result of Arzela-Ascoli, we obtain then that $\{P(t)\}_{t\geq 0}$ is eventually compact. □

Proposition 3.3 *Let assume that* $(\mathbf{H}_{\mathcal{K},\mu})$ *is satisfied. Then the strongly continuous semigroup P and* $\mathcal{T}_\Phi$ *are eventually compact.*

Proof. The core operator Φ given by the formula (15) is a perturbation of finite rank. Then by using the following result [35, Proposition 4], we obtain the eventual compactness of the semigroup solution $\{T_\Phi(t)\}_{t\geq 0}$. The perturbed semigroup $\{W(t)\}_{t\geq 0}$ is eventually compact by applying the result in Lemma3.2. □
In the follwing we define the family $(\tilde{\Psi}_\lambda)_{\lambda\in\mathbb{C}}$ of operators such that their entries are defined by

$$\tilde{\Psi}_{\lambda,i,j} := < \nabla_{\lambda,j}, \nabla_i^* >, \quad \lambda \in \rho(\mathcal{A}_0) \tag{20}$$

with

$$\nabla_{\lambda,i} = q^{-1} R(\lambda, \mathcal{A}_0^{\odot *}) \nabla_i^{\odot *}, \quad \lambda \in \rho(\mathcal{A}_0)$$
$$\{\nabla_i^{\odot *}\}_{1\leq i\leq I} \text{ is a linearly independent set in } \mathcal{X}^{\odot *}$$

If we adopt this formulation to the integral equation given by (14) such that

$$\begin{aligned}
\nabla_i^{\odot *} &= \mathcal{H}_i \\
\nabla_{\lambda,j} &= \left(q^{-1} R(\lambda, \mathcal{A}_0^{\odot *}) \mathcal{H}_j \right)(s) = e^{\lambda s} e_j \\
\nabla_i^* &= \Theta_{i,j}
\end{aligned}$$

where Θ is given by

$$\begin{pmatrix}
0 & 0 & \cdots & 0 & 2\psi_{I\to 1}\bar{z}_{I,I} \\
\psi_{1\to 2}\bar{z}_{1,1} & 0 & \cdots & \cdots & 0 \\
0 & \psi_{2\to 3}\bar{z}_{2,2} & \ddots & & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & \psi_{I-1\to I}\bar{z}_{I-1,I} & 0
\end{pmatrix}$$

Therefore we obtain

$$\tilde{\Psi}_\lambda x = \Phi(e^{\lambda \cdot} \otimes x) \quad \text{for } x \in \mathbb{R}^I$$

where the relation $\otimes$ is defined by

$$(e^{\lambda \cdot} \otimes x)(s) := e^{\lambda s} x \quad \text{for } s \in \mathbb{R}.$$

Next we will give a characterization of the spectrum of the infinitesimal generator $\mathcal{A}_\Phi$ associated to the perturbed semigroup W_Φ by using the family of operators defined above. This characterization permits to show that the spectrum is a point spectrum and there exists a characteristic equation, the roots of which are the eigenvalue of the infinitesimal generator of the associated perturbed semigroup. Furthemore, we will give a criterion which assure the simplicity of these eigenvalues.

Proposition 3.4 *λ belongs to $\sigma(\mathcal{A}_\Phi)$ if and only if λ belongs to $\sigma_p(\mathcal{A}_\Phi)$ if and only if*

$$\prod_{i=1}^{I} \int_0^{\tilde{a}} \exp\left(- \int_0^x \lambda + m_i(\tau) d\tau \right) \mathcal{K}_{i\to i+1}(x) dx = \frac{1}{2} \tag{21}$$

Proof. By transforming the integral equation (14) using the Laplace transformation, we obtain the following resolvent equation

$$(I - q^{-1} R(\lambda, \mathcal{A}_0^{\odot *})\Phi) R(\lambda, \mathcal{A}) = R(\lambda, \mathcal{A}_0), \text{ for } \lambda \in \rho(\mathcal{A}_0)$$

Then it is clear that λ belongs to the spectrum of $\mathcal{A}_\Phi$ if and only if the operator $I - q^{-1} R(\lambda, \mathcal{A}_0^{\odot *})\Phi$ is not invertible. By straightforward computation it is turn out that it is satisfied if and only if $\det(\mathbb{I} - \tilde{\Psi}_\lambda) = 0$ that is equivalent to the assertion (21) since the term $q^{-1} R(\lambda, \mathcal{A}_0^{\odot *})\Phi$ is bounded and $\sigma(\mathcal{A}_0) = \emptyset$. By the eventual compactness of the semigroup the spectrum of $\mathcal{A}_\Phi$ is ponctual. □

Proposition 3.5 *The spectral bound of $\mathcal{A}_\Phi$ equal to the growth bound of W_Φ, that is,*

$$s(\mathcal{A}_\Phi) = \omega(W_\Phi)$$

such that there exists a real dominant eigenvalue $\tilde{\lambda}$ of $\mathcal{A}_\Phi$ solution of

$$\tilde{\lambda} = s(\mathcal{A}_\Phi)$$
$$\prod_{i=1}^{I} \int_0^{\tilde{a}} \exp\left(-\int_0^x \lambda + m_i(\tau)d\tau\right)\mathcal{K}_{i\to i+1}(x)dx = \frac{1}{2} \qquad (22)$$

Proof. Since our semigroup solution of (14) is eventually compact, the result given in [20, Theorem 2.7] provides to us that $s(\mathcal{A}_\phi) = \omega(W_\phi)$. The map given by the spectral radius of the operator $\tilde{\Psi}_\lambda$ is continuous. In addition, the operator $\tilde{\Psi}$ decreases and it satisfies

$$r(\tilde{\Psi}_\lambda) \text{ tends to } +\infty \quad \text{as } \lambda \text{ tends } -\infty$$
$$r(\tilde{\Psi}_\lambda) \text{ tends to } 0 \quad \text{as } \lambda \text{ tends } +\infty$$

Then there exists a unique real root $\tilde{\lambda}$ that holds (22) where $\tilde{\lambda}$ is an eigenvalue of $\mathcal{A}_\phi$. Using the fact that the spectral radius of $\tilde{\Psi}$ is monotone, we have the result

$$\tilde{\lambda} = \sup\{\Re e\lambda : \lambda \in \sigma(\mathcal{A}_\phi)\} = s(\mathcal{A}_\phi)$$

Finally, since the semigroup is eventually compact, the real eigenvalue $\tilde{\lambda}$ solution of (22) is dominant.□

We can say from the above result that less control on the transition and mortality rates, the higher the growth rate, that is the proliferation is faster within cells with less synchronised phases.

Proposition 3.6 *The operator $\tilde{\Psi}_\lambda$ is irreducible such that $\tilde{\lambda}$ is a pole of order 1.*

Proof. The irreducibility is a strong positivity property for the semigroup that provides to us a convenient criterion about the simplicity of the poles. Let $g, g^* \in \mathbb{R}^I$ we have

$$< \tilde{\Psi}_\lambda g, g^* >= \int_0^{\tilde{a}} \exp(-\lambda a) \sum_{i=1}^{I-1} \mathcal{K}_{i\to i+1} g_{i+1}^* \bar{z}_{i,i} g_i + 2\mathcal{K}_{I\to 1} g_1^* \bar{z}_{I,I} g_I da > 0$$

□

Next, we investigate the behavior in the long term of the perturbed semigroup solution of (14), where we show that the property of asynchronous exponential property is fufilled

Theorem 3.7 *The solution semigroup $\{P_\phi(t)\}_{t\geq 0}$ of* (14) *holds*

$$\|\exp(-\lambda_0 t)P_\phi(t) - Q\| \leq \beta e^{-\gamma t} \text{ for } \gamma > 0, \beta \geq 1, \forall t \geq 0$$

such that Q given by

$$Qg = \alpha(g)(e^{\lambda_0 \cdot} \otimes \varphi_{\lambda_0})$$

and

$$\alpha(g) = \frac{< \psi_{\lambda_0}^*, \phi(\theta \mapsto \int_0^\theta \exp(\lambda_0(\theta - s))g(s)ds >}{< \psi_{\lambda_0}^*, \phi(\theta \mapsto \theta \exp(-\lambda_0\theta) \otimes \psi_{\lambda_0} >}, \quad \psi \in \mathcal{E} \qquad (23)$$

where $\psi_{\lambda_0} > 0$ ($\psi_{\lambda_0}^* > 0$) is the eigenvector of $\tilde{\Psi}_{\lambda_0}$ (respectively of $\tilde{\Psi}_{\lambda_0}^*$) such that $< \psi_{\lambda_0}^*, \psi_{\lambda_0} >= 1$.

Proof. Our state space $\mathcal{X}$ can be decomposed as the following

$$\mathcal{X} = N_{\tilde{\lambda}} \oplus R_{\tilde{\lambda}}$$

where

$$\begin{cases} N_{\tilde{\lambda}} = \ker(\mathcal{A}_\Phi - \tilde{\lambda}I) = e^{\tilde{\lambda}\cdot} \otimes \ker(I - \tilde{\Psi}_{\tilde{\lambda}}) \\ \quad = \{\alpha e^{\tilde{\lambda}\cdot} \otimes \psi_{\tilde{\lambda}} : \alpha \in \mathbb{R}\} \\ R_{\tilde{\lambda}} = \text{Range}(\mathcal{A}_\Phi - \tilde{\lambda}\mathbb{I}) \end{cases}$$

where $\varphi_{\tilde{\lambda}}$ is the positive eigenvector of $\tilde{\Psi}_{\tilde{\lambda}}$ associated to the eigenvalue $\tilde{\lambda}$ solution of (22). The projection of g on the subspaces $N_{\tilde{\lambda}}$ and $R_{\tilde{\lambda}}$ gives us

$$g = \alpha(g)\exp(-\tilde{\lambda}.) \otimes \psi_{\tilde{\lambda}} + \rho$$

Then our semigroup solution of (14) can be rewritten as

$$P_\Phi(t)g = \alpha(g)\exp(-\tilde{\lambda}(t - .))\psi_{\tilde{\lambda}} + (P_\Phi(t)|_R)(\rho)$$

Using the fact that $\omega(P_\Phi(t)|_R) < \tilde{\lambda}$. Then we obtain that

$$P_\Phi(t)g = \alpha(g)\exp(-\tilde{\lambda}(t - .))\psi_{\tilde{\lambda}} + o(\exp(\tilde{\lambda}t))$$

Then we obtain

$$\|\exp(-\tilde{\lambda}t)P_\Phi(t)g - Qg\|_{\mathcal{E}} \leq \beta\exp(-\delta t)\|g\|_{\mathcal{E}} \quad \text{for } \gamma > 0, \beta \geq 1, g \in \mathcal{E}$$

where $Qg = \alpha(g)\exp(-\tilde{\lambda}.)\psi_{\tilde{\lambda}}$. Then

$$\alpha(g) = \frac{< \psi_{\tilde{\lambda}}^*, \phi(\theta \mapsto \int_0^\theta \exp(\tilde{\lambda}(\theta - s))g(s)ds >}{< \psi_{\tilde{\lambda}}^*, \phi(\theta \mapsto \theta\exp(-\tilde{\lambda}\theta) \otimes \psi_{\tilde{\lambda}} >}, \quad g \in \mathcal{E}$$

such that $\psi_{\tilde{\lambda}}^*$ is a positive eigenvector of $\tilde{\Psi}_{\tilde{\lambda}}^*$ that yields the condition $< \psi_{\tilde{\lambda}}^*, \psi_{\tilde{\lambda}} >= 1$. □

Proposition 3.8 *The semigroup solution of* (1) *possesses the asynchronous exponential property.*

Proof. Using the relation $(\Gamma f)(s) = \chi(s)f(s), \forall s \in (0, \tilde{a})$ we obtain the similar semigroup

$$P(t)f = \Gamma^{-1}(P_\Phi(t)\Gamma f), \quad \forall f \in \mathcal{E}$$

Then by considering the formula $\tilde{Q}f = \Gamma^{-1}(Q(\Gamma f))$ we obtain the estimate

$$\begin{aligned} &\|\exp(-\tilde{\lambda}t)P(t)f - \tilde{Q}f\|_{\mathcal{E}} = \|\Gamma(\exp(-\tilde{\lambda}t)P(t)f - Qf)\|_{\mathcal{E}} \\ &\leq \|\exp(-\tilde{\lambda}t)(P_\Phi(t)\Gamma f) - (\alpha(\Gamma f)\exp(-\tilde{\lambda}.)\psi_{\tilde{\lambda}})\|_{\mathcal{E}} \\ &\leq \beta\exp(-\gamma t)\|f\|_{\mathcal{E}} \end{aligned}$$

We can conclude that the solution of (1) yields the asynchronous exponential property. □

As generations follow one another, the synchronisation between phases are less expressed within the evolving cell population, that is we observe a variability in the structure when a cell transition from the ith phase to the $(i+1)^{th}$ phase, and becomes fully desynchronised. In the following a threshold phenomenon appear and it can be formulated as the spectral radius of some integral operator

Proposition 3.9 *Let $r(\tilde{\Psi}_\lambda)$ be the spectral radius of $\tilde{\Psi}_\lambda$ defined by* (20)*. Then the following holds :*

1. *If $\int_0^{\tilde{a}} \exp\left(-\int_0^x m_i(\tau)d\tau\right)\mathcal{K}_{i\to i+1}(x)dx < \frac{1}{2}$ the trivial equilibrium state is asymptotically stable.*
2. *If $\int_0^{\tilde{a}} \exp\left(-\int_0^x m_i(\tau)d\tau\right)\mathcal{K}_{i\to i+1}(x)dx > \frac{1}{2}$ the trivial equilibrium state is unstable.*

Proof. We have shown that $\tilde{\lambda} = s(\mathcal{A}_\phi)$. Due to the monotonicity of the operator $\tilde{\Psi}_\lambda$ we have that the spectral bound $s(\mathcal{A}_\phi) < 0$ if

$$r(\tilde{\Psi}_0) = \int_0^{\tilde{a}} \exp\left(-\int_0^x m_i(\tau)d\tau\right)\mathcal{K}_{i\to i+1}(x)dx < \frac{1}{2}$$

which corresponds to the trivial equilibrium state. Then, since from

$$\|P(t)f\|_\mathcal{E} \leq e^{\tilde{\lambda}t}\|f\|_\mathcal{E}$$

we have $\lim_{t\to\infty} P(t)f = 0$ and it is unstable when

$$\int_0^{\tilde{a}} \exp\left(-\int_0^x m_i(\tau)d\tau\right)\mathcal{K}_{i\to i+1}(x)dx > 1/2 \quad \square$$

4 Conclusion

The study of the cell population dynamics by taking into account the proliferation process and generations overlapping take an important part in the mathematical modeling and analysis. Various analytical and probabilist methods were performed to this aim. In particular, the semigroup method, that demonstrate it is theoretically well established combining between different theories such as operator theory, functional analysis, spectral theory ...

In this paper, the dual semigroup theory provide a useful framework to examine these processes. In particular, the context of the control of the cell cycle where many complex regulatory processes are involved where the structuration by age permits to us to track the individual cell position through the multiple generations. We proved the asynchronous exponential property that consists as an important trait within the cell populations. In fact, the synchronisation or desynchronisation is essential in tumour therapy in which drugs can synchronize cell phases and others can destroy cells in the same phases.

Here, we investigated the role of the circadian clock in the control of the proliferation within the cell population. Mathematically, we can conclude that the disruption of the circadian clock, in particular the desynchronisation between the transition from a phase to another one has a major effect on the proliferation of cells. In addition, we demonstrate that as time evolves, the cells are more desynchronised than the younger generations. In this context, natural questions arise about the differences in synchronisation between healthy and cancerous cells. This question is essential in order to understand the cell cycle and to use such knowledge to make optimisation methods for cancer therapy.

In this paper, we consider only age dependent parameter, in perspective we want to examine the effect of time dependent and different periodic controls on the cell proliferation and to compare theses controls between them.

References

[1] Y. El Alaoui, L. Alaoui, "Mathematical analysis of an age structured multi phases cell cycle model with checkpoints." in 2019 IEEE International Conference of Computer Science and Renewable Energies (ICCSRE-IEEE), 2019, doi:10.1109/iccsre.2019.8807639 .

[2] H. Lodish, A. Berk, C.A. Kaiser, M. Krieger, M.P. Scott, A. Bretscher, H. Ploegh, P.T. Matsudaira, Molecular cell biology, W.H. Freeman, New York, 2007.

[3] S. M. Reppert, D. R. Weaver, Coordination of circadian timing in mammals, Nature, **418**(6901):935-41,2002, doi: 10.1038/nature00965.

[4] G. Filipski, P.F. Innominato, M.W. Wu, X. M. Li, S. Iacobelli, L.J. Xian, F. Levi, Effect of light and food schedules on liver and tumour molecular clocks in mice, Journal of the National Cancer Institute, **97**, 507–517, 2005, doi:10.1093/jnci/dji083.

[5] D. Morgan, The cell cycle: Principles of control, Primers in Biology series, Oxford University Press, 2006.

[6] E. Filipski, V.M. King, X.M. Li, T.G. Granda, M. Mormont, X.H. Lin, B. Claustrat, M.H. Hastings, F. Lévi, Host circadian clock as a control point in tumor progression, Journal of the National Cancer Institute **94**(9): 690-7, 2002, doi: 10.1093/jnci/94.9.690.

[7] E. Filipski, F. Delaunay, V.M. King, M. -W. Wu, B. Claustrat, A. Gréchez-Cassiau, C. Guettier, M.H. Hastings, F. Lévi, Effects of chronic jet lag on tumor progression in mice, Cancer Research **64**, 7879-7885, 2004, doi:10.1158/0008-5472.CAN-04-0674.

[8] E. Filipski, P.F. Innominato, M.W. Wu, X.M. Li, S. Iacobelli, L.J. Xian, F. Lévi, Effects of light and food schedules on liver and tumor molecular clocks in mice, Journal of the National Cancer Institute **97** (7), 507–517, 2005, doi:10.1093/jnci/dji083.

[9] J. Clairambault, B. Laroche, S. Michler and B. Perthame, A mathematical model of the cell cycle and its control, Research Report, **4892**, INRIA. 2003, URL :https://hal.inria.fr/inria-00071690.

[10] J. Clairambault, P. Michel, B. Perthame, Circadian rhythm and tumour growth, Comptes Rendus de l'Académie des Sciences, Série I, Mathématique, Elsevier, **342**, 17–22, 2006, URL: https://hal.archives-ouvertes.fr/hal-00113511.

[11] J. Clairambault, P. Michel, B. Perthame, A mathematical model of the cell cycle and its circadian control, in: I.A. Deutsch, L. Brusch, H. Byrne, G. de Vries, H.P. Herzel, Mathematical Modeling of Biological Systems, **I**, Birkhäuser Boston, 247–259, 2007, doi:10.1007/978-0-8176-4558-8_21

[12] J. Clairambault, S. Gaubert, B. Perthame, An inequality for the Perron and Floquet eigenvalues of monotone differential systems and age structured equations, Comptes Rendus Mathematique, **345**(10):549–554, 2007, doi: 10.1016/j.crma.2007.10.001.

[13] F. Bekkal Brikci, J. Clairambault, B. Perthame, Analysis of a molecular structured population model with possible polynomial growth for the cell division cycle, Mathematical and Computer Modelling, **47**(7–8):699–713, 2008, doi:10.1016/j.mcm.2007.06.008.

[14] F. Bekkal Brikci, J. Clairambault, B. Ribba, B. Perthame, An age-and-cyclin-structured cell population model for healthy and tumoral tissues,Journal of Mathematical Biology, **57**(1):91–110, 2008, doi:10.1007/s00285-007-0147-x.

[15] M. Doumic, Analysis of a population model structured by the cells molecular content, Mathematical Modelling of Natural Phenomena,**2**(3):121–152,2007, doi:10.1051/mmnp:2007006.

[16] J. Clairambault, S. Gaubert and T. Lepoutre, Comparison of Perron and Floquet eigenvalues in age-structured cell division models, Mathematical Modelling of Natural Phenomena, **4**:183-209,2009, doi: 10.1051/mmnp/20094308.

[17] J. Clairambault, S. Gaubert and T. Lepoutre, Circadian rhythm and cell population growth, Mathematical and Computer Modelling, **53**(7-8):1558-1567, 2011, doi:10.1016/j.mcm.2010.05.034.

[18] O. Diekmann, H.J.A.M. Heijmans, and H.R. Thieme, On the stability of the cell size, Journal of Mathematical Biology **19**: 227-248, 1984, doi: 10.1007/BF00277748.

[19] O. Diekmann, S.A. Van Gils, S.M. Verduyn Lunel, AND H.-O. Walther, Delay Equations, Functional, Complex and Nonlinear Analysis, Springer-Verlag, New York, 1995.

[20] O. Diekmann , P. Getto and M. Gyllenberg, Stability and bifurcation analysis of volterra functional equations in the light of suns and stars, SIAM Journal on Mathematical Analysis, **39**(4): 1023–1069, 2007, doi: 10.1137/060659211.

[21] O. Diekmann, M. Gyllenberg, Equations with infinite delay: blending the abstract and the concrete, Journal of Differential Equations, **252**(2):819–851, 2012, doi: 10.1016/j.jde.2011.09.038.

[22] A. M. de Roos, O. Diekmann, P. Getto, M.A. Kirkilionis, Numerical equilibrium analysis for structured consumer resource models, Bulletin of Mathematical Biology **72**(2):259–297, 2010, doi: 10.1007/s11538-009-9445-3.

[23] O. Diekmann, M. Gyllenberg, J.A.J. Metz, S. Nakaoka and A.M. de Roos, Daphnia revisited: local stability and bifurcation theory for physiologically structured population models explained by way of an example, Journal of Mathematical Biology, **61**(2):277–318, 2010, doi: 10.1007/s00285-009-0299-y.

[24] Ph. Clement, O. Diekmann, M. Gyllenberg, H. J. A. M. Heijmans, and H. R. Thieme, Perturbation theory for dual semigroups. I. The sun-reflexive case, Mathematische Annalen, **277**:709–725, 1987, doi:10.1007/BF01457866.

[25] Ph. Clement, O. Diekmann, M. Gyllenberg, H. J. A. M. Heijmans, and H. R. Thieme, Perturbation theory for dual semigroups. II. Time-dependent perturbations in the sun-reflexive case, Proceedings of the Royal Society of Edinburgh Section A: Mathematics, **109**: 145–172, 1988, doi: 10.1017/S0308210500026731

[26] Ph. Clement, O. Diekmann, M. Gyllenberg, H. J. A. M. Heijmans, and H. R. Thieme, Perturbation theory for dual semigroups. III. Nonlinear Lipschitz continuous perturbations in the sun-reflexive case, In Proceedings of Volterra Integrodifferential Equations in Banach Spaces and Applications 1987, 67–89, 1989.

[27] Ph. Clement, O. Diekmann, M. Gyllenberg, H. J. A. M. Heijmans, and H. R. Thieme, Perturbation theory for dual semigroups. IV. The intertwining formula and the canonical pairing, in Trends in Semigroup Theory and Applications Trieste-Italy, **116**:95–116, 1989.

[28] Y. El Alaoui and L. Alaoui, Asymptotic Behavior in a cell proliferation Model with Unequal division and Random Transition using translation semigroup,Indian Journal of Sciences and Technology, **10**(28):1-8, 2017, doi: 10.17485/ijst/2017/v10i28/101042.

[29] L. Alaoui and Y. El Alaoui, AEG property of a cell cycle model with quiescence in the light of translation semigroups, International Journal of Mathematical Analysis, **9**(51): 2513-2528, 2015, doi:10.12988/ijma.2015.59221.

[30] L. Alaoui, Nonlinear homogeneous retarded differential equations and population dynamics via translation semigroups, Semigroup Forum, **63**:330–356, 2001, doi: 10.1007/s002330010083.

[31] L. Alaoui, Age-dependent population dynamics and translation semigroups, Semigroup Forum, **57**: 186-207, 1998, doi: 10.1007/PL00005973.

[32] L. Alaoui, A cell cycle model and translation semigroups, Semigroup Forum, **54**, 1997, doi: 10.1007/BF02676597.

[33] L. Alaoui, Generators of translation semigroups and asymptotic behavior of the Sharpe-Lotka model, Differential Integral Equations, **9**:343-362, 1996.

[34] L. Alaoui, Population dynamics and translation semigroups, dissertation, University of Tübingen, 1995.

[35] L. Alaoui and O. Arino, Compactness and spectral properties for positive translation semigroups associated with models of population dynamics, Differential Integral Equation, **6**:459-480, 1993.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 09-21 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

A Study on Intelligent Dialogue Agent for Older Adults' Preventive Care - Towards Development of a Comprehensive Preventive Care System -

Sho Hirose[*,1], Daisuke Kitakoshi[1], Akihiro Yamashita[1], Kentarou Suzuki[2], Masato Suzuki[1]

[1]*National Institute of Technology, Tokyo College, Hachioji, Tokyo, 193-0997, Japan*

[2]*Kyorin University, Mitaka, Tokyo, 181-8612, Japan*

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 18 October, 2020
Online: 08 November, 2020

Keywords:
Intelligent Dialogue Agent
Comprehensive Preventive Care System
Reinforcement Learning

ABSTRACT

Preventive care approaches have attracted much attention in Japan, which is one of the world's most super-aged societies. These approaches aim to decrease the number of people who require nursing care or other human support. Our research group has developed several kinds of preventive care systems, including a fall prevention system, a cognitive training system, and an Intelligent Dialogue Agent (IDA). In this paper, we focus on familiarizing older adults with the IDA – a conversational agent that encourages older adults to engage it in natural conversations, while monitoring their health in the process – and we introduce a Speech Content Coordinating Function (SCCF) to the IDA to further improve older adult's familiarity with and interest in the agent. We also have a plan to develop a Comprehensive Preventive Care System (CPCS) which can encourage proactive conversation and provide effective monitoring of older adults as well as habitual cognitive training. To evaluate the basic characteristics and impressions of the CPCS, a prototype version that consists of the IDA and the Mechanism for Cognitive Training (MCT) is developed. The results of the experiments indicated that the SCCF can adequately adjust speech content depending on the user's circumstances. We also confirmed that the prototype CPCS achieves synergistic effects (e.g., more detailed monitoring of older adult's health condition, increased frequency of cognitive training) when its components are used together.

1 Introduction

This paper is an extension of work originally presented at the 2019 International Conference on Technologies and Applications of Artificial Intelligence [1].

Japan is currently known as the most advanced super-aged society. In 2018, Japanese people aged 65 and older accounted for 28.1% of the total population [2]. As the aging demographic increases, the number of people requiring long-term care also increases, as does the number of people who need nursing care or other human support, and that number in Japan has reached 6.41 million, as certified by the Long-Term Care Insurance Act [3]. Preventive care approaches, which aim to decrease the number of people who require nursing care or other human support by reducing the decline in individuals' mental and/or physical functions, have attracted a lot of attention. These approaches also aim to control increasing medical costs and the associated burden because the number of older adults is increasing while the working population is expected to decline due to declining total fertility rates. In recent years, many researchers in various fields have focused on the prevention not only of dementia but also of fall, one of the causes of the need for nursing care, and have shown a certain degree of effectiveness in preventive care [4]– [12]. However, some older adults experience physical and/or psychological distress due to preventive care activities, and therefore have trouble engaging in these activities on a regular basis.

At the same time, in the past decade, various voice-interaction-based devices and softwares (Dialogue Agents) such as smart speakers (e.g., Google Home and Amazon Echo) and voice assistants (e.g., Siri and Cortana) [13] have been developed and employed to provide useful information and a variety of services in response to users' prompts. In addition, there has been much research into the use of Dialogue Agents for older adults with dementia and mild cognitive impairment [14]– [17]. However, it should be noted that these devices and the Dialogue Agents generally do not talk to users

[*]Corresponding Author: Sho Hirose, Email: s20613@tokyo.kosen-ac.jp

without a specific request or prompt. However, we believe that introducing a function that makes it possible for Dialogue Agents to speak to users spontaneously in order to elicit a reaction will play an important role in endowing the agents with a sense of humanity and familiarity.

Against these backgrounds, previous studies examined Intelligent Dialogue Agents (IDAs) that allow natural and flexible communication with and monitoring of older adults by using a Smart Speaker, a friendly interface, to investigate older adults' impressions of the Smart Speaker and the basic characteristics of the IDA [18]. The IDA employed a Spontaneous Talking Function (STF) that allows the IDA to talk to older adults without users' requests to encourage them to actively dialogue. Our research group has shown that the STF is effective in some ways, particularly in providing increased convenience and familiarity to older adults. Our research group is also developing a cognitive training system based on a memory game on a tablet device [19] and a Fall Prevention System for regular preventive care exercise [3] to prevent dementia and fall, one of the main factors leading to a need for nursing care. Additionally, our research group is conducting a study to estimate the behavior of older adults who live alone from data gathered by supervision sensors installed in their homes [20]. In this paper, we introduce a Speech Content Coordinating Function (SCCF) into the IDA in order to further improve the user's familiarity with and interest in the IDA. The SCCF determines the content of speech by using a reinforcement learning algorithm according to the user's preferences and condition. In order to investigate the basic characteristics of the SCCF and users' impression of it, and to evaluate whether the SCCF was able to acquire appropriate policy depending on its users, several experiments were conducted. In the experiments, we developed a text-based agent in the form of a chat-bot and introduced the SCCF into it.

In addition, this paper considers the feasibility of a Comprehensive Preventive Care System (CPCS). Although we aim to develop the CPCS by integrating our preventive care systems, it is currently in the conceptual phase. Once fully developed, the CPCS will contribute to the encouragement of active conversation and the provision of not only effective monitoring of the health condition of older adults but also regular cognitive training and fall prevention exercise. As the first step to examine the adequacy and impressions of the CPCS, we conducted the experiment with a prototype version of the CPCS using both the IDA with a sensor and the Mechanism for Cognitive Training (MCT) which is at present equivalent to the cognitive training system.

At present, neither the IDA nor the prototype CPCS has been tested by older adults, because their basic characteristics and impressions have not been evaluated. Therefore, we enrolled students as participants in the experiments as the first phase to confirm the future applicability of the IDA and CPCS for older adults.

2 Preliminaries (Reinforcement Learning)

Before describing our study, we will first introduce the concept of Reinforcement Learning (RL). RL is a machine learning method [21]. An RL agent attempts to acquire an appropriate policy (a state-to-action map) based on observations and trial-and-error interactions with its environment. A policy characterizes an agent's behavior and is described using the function:

$$\pi : \mathcal{S} \times \mathcal{A} \rightarrow \boldsymbol{R}, \tag{1}$$

where $\mathcal{S}$ and $\mathcal{A}$ denote sets of states and actions, respectively. The pair $(s, a)(\forall s \in \mathcal{S}, \forall a \in \mathcal{A})$ is referred to as a rule, which states that "if an agent observes a state s, it outputs an action a". At each time step t, an agent observes state s_t, and selects action a_t using π.

RL algorithms are roughly classified into two approaches, exploration-oriented and exploitation-oriented approaches. The former approaches, which include Q-leaning, Sarsa, and so on, aim to optimize the total discounted reward typically in environments modeled by Markov Decision Process (MDP). The latter ones, such as Bucket Brigade and REINFORCE, are based on the notion of credit assignment and aim to increase learning speed and to acquire reasonable policies even in non-MDP environments although it is not ensured to acquire the optimal policy by using them. Because the environments we treat in this study seem to be non-MDP ones, the exploitation oriented RL algorithms are applied. Brief explanations about REINFORCE and Bucket Brigade, which are employed in the IDA and MCT (described later), respectively, are presented below.

2.1 REINFORCE

REINFORCE [22] is an exploitation-oriented approach based on a policy-gradient reinforcement learning algorithm. Policy-gradient reinforcement learning algorithms such as natural policy gradients [23], and stochastic gradient methods [24] learn policies directly by updating parameters θ that characterize a policy π_θ, as opposed to many conventional reinforcement learning algorithms such as Q-learning [25] or Sarsa [26]. The values of θ are related to conditional probabilities $p(a|s)$. In other words, the policy-gradient reinforcement learning algorithms aim to acquire a probability distribution that selects the optimal action to take in each state so as to maximize the sum of the rewards. The goal of these algorithms is thus to obtain the parameters θ, which characterize the probabilistic policy π_θ that maximizes the sum of rewards, updated according to the following formula (2):

$$\theta \leftarrow \theta + \eta \Delta J(\theta), \tag{2}$$

where η is the learning rate. REINFORCE updates θ by $\Delta J(\theta)$ shown below:

$$\Delta J(\theta) = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} \left(r_{m,t} - b\left(\theta\right) \right) \ln \pi_\theta(a_{m,t}|s_{m,t}), \tag{3}$$

where $s_{m,t}$, $a_{m,t}$, and $r_{m,t}$ are state, action and reward at the t-th time step in the m-th episode ($t = 1, ..., T; m = 1, ..., M$), respectively. In the above formula, an episode denotes the sequence of rules selected between rewards, and one time step is counted when the agent selects a rule. The term $b(\theta)$ reduces the variance of the estimate and serves to stabilize the policy learning process, called the baseline ($b(\theta) = \Sigma_{m=1}^{M} \Sigma_{t=1}^{T} r_{m,t}$), and π_θ denotes the agent's policy. In the present study, the probability that the IDA will select its action a given a state s is computed by the softmax function (4) with the

inverse temperature parameter β, because both the state and action spaces are discrete.

$$\pi_\theta(a|s)(= p(a|s)) = \frac{\exp(\beta\theta_{(a|s)})}{\sum_{a'}\exp(\beta\theta_{(a|s)})} \tag{4}$$

REINFORCE is used to coordinate the IDA's speech content depending on the user's circumstances in this study.

2.2 Bucket Brigade

In the Bucket Brigade algorithm [27], a policy is characterized by rule-weights $w(s_t, a_t)$, which are used as an agent's policy π. When an agent acquires reward R after it selects rule (s_t, a_t) at time step t, the weights of rule $w(s_t, a_t)$ and $w(s_{t-1}, a_{t-1})$ are reinforced as follows:

$$\begin{aligned} w(s_t, a_t) &= w(s_t, a_t) - C_{bid} \times w(s_t, a_t) - \\ &\quad C_{tax} \times w(s_t, a_t) + R, \qquad (5) \\ w(s_{t-1}, a_{t-1}) &= w(s_{t-1}, a_{t-1}) + \\ &\quad C_{bid} \times w(s_t, a_t), \qquad (6) \end{aligned}$$

where C_{bid} and C_{tax} denote a "bid" parameter specifying a degree of propagating the weight $w(s_t, a_t)$ to the weight $w(s_{t-1}, a_{t-1})$, in which the rule (s_{t-1}, a_{t-1}) plays a role of "trigger" to select the next rule (s_t, a_t), and a "tax" parameter which determines a discount rate for the weight of the rule selected at step t, respectively.

As described earlier, the Bucket Brigade is one of the algorithms categorized as an exploitation-oriented approach. The above formulae mean that this algorithm updates a rule weight at each step, even when no reward is given, and also that this algorithm is relatively easy to implement compared to other conventional RL methods. Moreover, the agent with the Bucket Brigade can acquire reasonable policies in many cases even if the agents do not have a sufficient number of trials because it actively employ rewarded past experience. This study applies the algorithm to the function for adjusting the difficulty level of cognitive training for respective users in the mechanism providing the users with cognitive training, which is described in the next section.

3 Preventive Care Systems

In this section, we present the preventive care systems that are currently under development or about to start development. We first overview the Intelligent Dialogue Agent (IDA), which our research group is currently developing, then introduce the Mechanism for Cognitive Training (MCT) as another preventive care framework, and finally provide brief outline of the Comprehensive Preventive Care System (CPCS), which will consist of the above two components and will be developed as our final goal.

3.1 Intelligent Dialogue Agent (IDA)

We have developed the Intelligent Dialogue Agent (IDA) with the aim of encouraging natural and flexible conversation with its users (older adults) and of monitoring older adult's health condition based on their behavior and dialogue history. The IDA talks spontaneously, using a variety of contents such as notification of daily routines (e.g., taking medicine, walking), trivia, news, weather and traffic information, and food information. The sensor mounted on the IDA also can monitor older adults within its detection area.

3.1.1 Architecture of the IDA

Figure 1 shows the final design of the IDA. The IDA consists primarily of an input/output (I/O) unit, a learning unit, and a database.

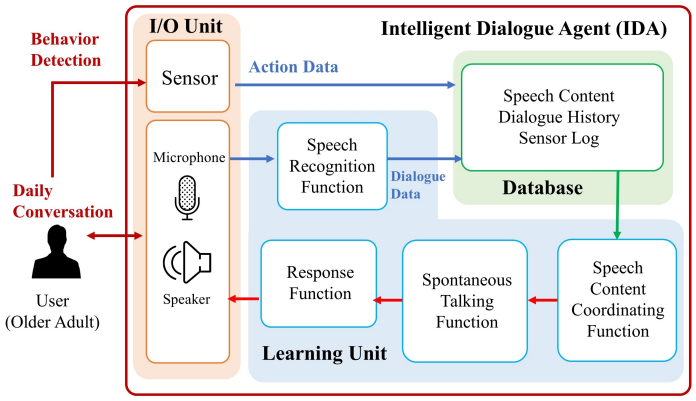


Figure 1: Final design of the IDA

(1) I/O Unit

The I/O unit is composed of a microphone, a speaker and a sensor. The user's motion is recorded by the sensor in order to track his or her characteristic behavior, such as lifestyle activities (e.g., waking and sleeping times, regular exercise). The pyroelectric sensor we adopted in this study, and which is embedded in the IDA, is both inexpensive and easy to use. Moreover, our research group has already employed a similar type of sensor and confirmed that it worked effectively to capture the motions of individuals [20]. This sensor can detect objects separate from the user, as well as even the slightest movement of people separate from the user. Although the IDA currently employs a microphone and speaker which are implemented by the included smart speaker (Google Home Mini) as its I/O unit, these components have certain limitations; for example, the smart speaker cannot acquire the user's response and dialogue history in real time. We thus plan to develop and implement an IDA that includes the highly desirable speech recognition and synthesis functions without using a smart speaker.

(2) Database

The database stores logs of the user's activity as detected by the sensors, and the history of dialogue between the IDA and the user, which is converted from speech information into text data by the Speech Recognition Function, a component of the Learning Unit. The data stored in the database are used to estimate the user's condition and to adjust the behavior of the IDA. We adopted a NoSQL cloud-hosted database known as the Firebase Realtime Database (https://firebase.google.com) for this component because it is compatible with Google Home Mini, which we employed in the IDA.

(3) Learning Unit

The learning unit has four functions: the Spontaneous Talking Function (STF), the Speech Recognition Function, the Response Function, and the Speech Content Coordinating Function (SCCF). A brief explanation of the four functions is as follows. The IDA can output its speech content without prompting by the user at the appropriate time by using the STF. The STF triggers speech to the user based on the user's daily routines and behavioral logs stored in the database. The Speech Recognition Function converts voice information obtained from the microphone, such as questions from the user or the user's responses to the IDA's speech, into textual information. The Response Function transforms the next speech content into voice information and transfers it to the speaker. In addition, the SCCF provides appropriate speech content depending on the user's preferences and condition. We expect that introducing the SCCF into the IDA will improve the convenience and familiarity of the IDA and hence the user's willingness to continue using the IDA. In order to adjust the content, timing, and frequency of speech, the following history information should be taken into account:

1. The behavioral history of the user (i.e., the frequency and number of times the user is detected) obtained by sensors on the I/O unit.
2. The history and frequency of dialogue between the user and the IDA.
3. What kinds of speech content were previously produced by the IDA and at which time slots.
4. Basic information about the user (e.g., age, gender, address, preferred topics for conversation)

3.1.2 The Speech Content Coordinating Function (SCCF)

The results of the experiments and questionnaires conducted in our previous study confirmed that the provision of appropriate information by the IDA increases the willingness of users to continue participating in dialogue with the IDA [1]. In the present study, we introduced the SCCF into the IDA and investigated its basic characteristics and performance.

In order to acquire the adequate behavior (policy) to determine natural responses depending on the particular user's preferences and/or circumstances, we apply the REINFORCE algorithm described in Section 2.1 to the SCCF as the policy learning algorithm. To conduct policy learning, we defined state, action, and reward for the IDA. The IDA's action is defined as the speech category $c \in \boldsymbol{C}$ that the agent employs to talk to the user ($\boldsymbol{C}$: A set of speech categories). There are 13 different speech categories: $\boldsymbol{C}$ = {time signal, greeting, today's weather, weekly weather forecast, today's trivia, domestic news, international news, economic news, technology news, science news, entertainment news, sports news, and each user's daily routine work}. The state of the IDA is defined as the pair of speech categories (c_1, c_2) that constituted the agent's output in the last two time steps. This state definition is based on the assumption that there is some relationship between past speech content and current content in human-to-human conversation ($c_1 \in \boldsymbol{C}, c_2 \in \boldsymbol{C}$). The user's feedback on the speech content spoken by the IDA is employed as the reward. All feedback is classified into one of three categories: positive (e.g., good, useful, beneficial), negative (e.g., bad, not useful, detrimental), or uninteresting, and the reward values obtained by the IDA are determined according to the feedback categories. Before establishing the reward setting, we asked a speech therapist at a long-term care insurance facility for older adults to comment on "how older adults come to have a good impression of speech content". She responded as follows:
[Comments of speech therapist]

"Even if older adults express negative responses to the information provided by the agent's speech, they may not necessarily have a negative evaluation of the corresponding speech (i.e., any user's reaction to the speech is evidence of being interested in the speech content)."

Based on this information from the speech therapist, we set up our SCCF to have two reward settings, one employing the speech therapist's comments and the other based on another assumption under which "the users' interest in speech directly corresponds to interest in the content of the speech." When the IDA acquires a reward, the weight parameter θ for each state and action are updated, and the probabilities for selecting actions are changed.

We substituted the softmax function (4) into the REINFORCE updating equation ((2), (3)) and the number of time steps T was set at one because we defined one episode as ending with one instance of speech. The following formula derived from (2), (3), and (4) is applied to the policy learning (parameter update) in the SCCF.

$$\theta \leftarrow \theta + \eta' \frac{1}{M} \sum_{m=1}^{M} (r_{m,t} - b(\theta))(1 - \pi_\theta(a|s)) \tag{7}$$

3.2 The Mechanism for Cognitive Training (MCT)

Our research group has developed the Mechanism for Cognitive Training (originally proposed as the cognitive training system) that encourages older adults to engage in cognitive training to help prevent dementia [19]. As shown in Figure 2, the MCT provides older adults with a brain-twister-based memory game as cognitive training, provided in such a way that they can regularly perform this memory game through repetitive interactions with a software agent on a tablet device. The software agent also serves as both an opponent who plays the game with the older adult users and a conversation partner; the goal of the latter function is to increase the familiarity of older adults with the MCT. Our system is expected to prevent dementia by preserving and improving cognitive functions such as short-term memory and encouraging and improving the user's motivations to use them. Furthermore, the MCT has a difficulty adjustment function that can set the appropriate game difficulty level according to differences between users and each individual user's current circumstances in order to create an environment that makes it easier for users to continue cognitive training for long periods of time, since they neither get bored because the game is too easy nor give up because it is too difficult. The database stores information on each user's game-playing conditions, such as memory time, response time, accuracy rate, and difficulty, and uses these data to adjust the difficulty level and provide feedback on the memory game results to the users.

The difficulty adjustment function employs the Bucket Brigade algorithm described in Section 2.2 in order to coordinate the game difficulty level. The Bucket Brigade algorithm is known to have two

characteristics: (i) advancing policy learning even when no reward is given; and (ii) additionally acquiring reasonable policies in many cases, even if the agents do not have a sufficient number of trials, by actively employing rewarded past experiences.

The software agent in the MCT is based on the concept of Human-Agent Interaction (HAI), which aims to design an appropriate framework of interactions between humans and agents. The MCT currently has a function to coordinate speech content depending on the game-playing frequency and/or time interval between bouts of gameplay so that a user can play the game (engage in the cognitive training exercise) comfortably. However, since it is intended that the software agent will be implemented on a tablet device, its functional and performance limitations may prevent it from fully achieving the desired functionality. Therefore, both functions for coordinating speech content and sensors for detecting the user's motions will be implemented in the IDA. In addition, the functions currently embedded in the software agent on the MCT will be implemented in the CPCS as functions of the IDA.

As a result of several experiments, we have confirmed that older adults were able to increase the difficulty level and the accuracy of the cognitive training by using the MCT. Furthermore, introducing the difficulty adjustment function made it possible to appropriately coordinate the training difficulty depending on the circumstances of the older adult playing the game. Although the use of MCT clearly had a positive effect on maintaining/improving the short term memory and cognitive function of older adults, we also confirmed that experiments over a longer period with a much large number of participants will be needed in order to make a detailed assessment of the MCT's effectiveness in dementia prevention.

3.3 Comprehensive Preventive Care System (CPCS)

The Comprehensive Preventive Care System (CPCS) aims to allow older adults to enjoy engaging in cognitive training while being monitored by their family and friends through active conversation. The CPCS consists of an IDA, which can carry out natural conversation with users, and an MCT (detailed in section 3.3) which provides users with a memory game on a tablet device as a kind of cognitive training (Figure 2).

Our research group is carrying out research on and development of each component with a view towards their eventual integration into the CPCS. Because the integration of the IDA and the MCT will allow the CPCS to employ the data stored in both in parallel, this system will be able to produce a synergistic effect between the cognitive training and monitoring functions. For example, the cognitive training status obtained by the MCT will be combined with the dialogue/behavior history acquired by the IDA, making possible a more detailed understanding of the user's condition (e.g., the user's degree of fatigue, whether he or she has taken a break while engaging in cognitive training, and so on). One of the final purposes of the CPCS is to notify family and friends of the circumstances of an older adult user obtained by the system, which also aims to accomplish the following three goals:

1. To maintain and improve the cognitive function of older adults;
2. To promote active conversation between older adults and their family, friends and the IDA; and
3. To create an effective environment where older adults and their families and/or community groups can be mutually involved to better monitor the older adult's health condition, and promote their care and dementia prevention.

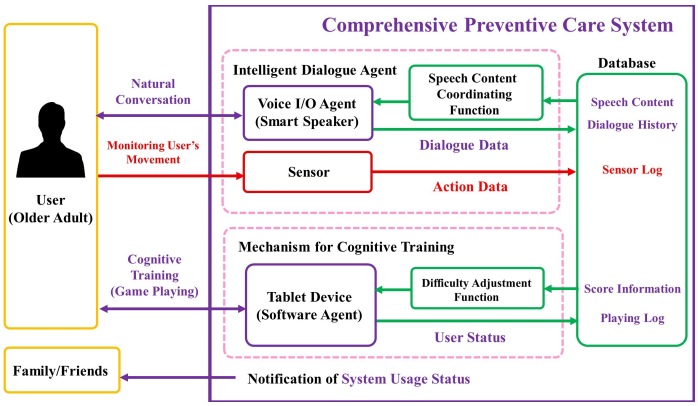


Figure 2: the Diagram of the CPCS

4 Experiments

As described in Section 1, we have not yet had an opportunity to employ older adults in the evaluation of the SCCF and the CPCS. In addition, many aged households lack a Wi-Fi environment which would be needed in order to use and evaluate the CPCS at home. For these reasons, we conducted several experiments with students to confirm that the SCCF can appropriately coordinate the speech content depending on the user's circumstances and to evaluate the basic characteristics and impressions of the CPCS. Moreover, we used the prototype version of the CPCS (described in a later section) in these experiments, because at present we have not evaluated the full version of the CPCS, but only its respective components (the IDA and the MCT).

Before conducting the experiment, a simple written questionnaire was administered to older adults in order to confirm their awareness of smartphones, the internet environment at their home, and the degree of interest/expectation in regard to spontaneous talking by the IDA (smart speaker). The questionnaire results showed that (i) older adults' degree of recognition of smart speakers was low; (ii) the majority of them did not have an internet (Wi-Fi) connection that would permit the use of a smart speaker at home; however, (iii) they had a high level of interest about conversation with the IDA (smart speaker), particularly its spontaneous talking function. Regarding item (iii), the older adults had on average 2.3 topics that they wanted the IDA to talk about, and no older adults answered that the talking function was "unnecessary" or that they had "no topics which they wanted to make the smart speaker talk about". The detailed results of the questionnaire are presented in Appendix (A).

4.1 Evaluation of the SCCF

In our first experiment, we aimed to evaluate the basic characteristics of the SCCF by employing a prototype of the IDA (henceforth referred to simply as the IDA) developed as a chat-bot using the Slack message application. The chat-bot periodically talks to provide information and learns from its users' feedback. The experiment was conducted for six days and our subjects were 10 students (average age: 18.9).

As the speech content of the chat-bot, six out of 13 categories were randomly selected before starting the experiment. The total number of states was $6 \times 6 = 36$ because the chat-bot had up to two previous categories (c_1, c_2) as pairs. From the set of selected speech categories $\boldsymbol{C'}$ ($|\boldsymbol{C'}| = 6$), the chat-bot selects speech category c according to the current policy. The initial value of the parameters θ which characterize a policy π_θ equals θ_0. Therefore, the chat-bot chooses a speech category with uniform probability at the first stage of the experimental period. Each participant responded to the speech content provided by the IDA with one of three kinds of feedback (P/N: positive/negative impression; U: uninterested). Reward values were determined by the collective feedback of users. The following two types of reward settings were utilized based on comments received from the speech therapist.

[Setting R_A]

Under this setting, in accordance with the speech therapist's assumption that "older adults would be interested in a topic if their evaluation was not U," a positive reward (+1) was assigned if the participant's feedback was P or N, and no reward was given otherwise.

[Setting R_B]

Under this setting, positive (+1) reward, negative (-1) reward, or no reward was given if the user's evaluation of a topic was P, N, or U, respectively.

The participants gave feedback for approximately 300 speech contents for each setting. To avoid any influence of the order of the participants' evaluations in settings R_A and R_B, the participants were grouped into two equally sized groups, and the order was changed in different groups. The parameters used in the policy learning and action selection are shown in Table 1.

Table 1: Parameters used in reinforcement learning

Parameter	Value
η'	0.22
M	3
θ_0	10
β	15

In addition, subjective evaluations was conducted after the end of the experimental period (three days) for each reward setting and after the end of the experiment to examine the participants' impressions of the SCCF. We also administered the questionnaire regarding the level of interest in each category at the end of the experiment because we did not want it to affect the participants' evaluation of the chat-bot's speech.

Which speech categories provided by the IDA were most preferred varied among users. The favorite speech categories and/or favored order of talking could also vary within the same user depending on the current circumstances (e.g., on holiday, at work, while eating, before bedtime, and so on). For these reasons, it was difficult to acquire a policy for providing optimal speech contents in the experiments; however, we can say that we were able to acquire a policy better expressing the user's preference by the learning in the SCCF if the speech categories preferred by the user were easy to be provided from the IDA compared to the other categories. The experiment thus confirmed the probability of the IDA generating speech for each category after learning the preferences of users while investigating the favorite categories of users via the questionnaire. By comparing the above results, we discuss the adequateness of the SCCF by evaluating whether the SCCF can learn the preference of users and how accurately the learning result matches the actual preference of users.

4.2 Preliminary Experiments with the CPCS

Several experiments were conducted to investigate the synergistic effect of using the MCT in conjunction with the IDA. As part of our preparation for conducting the experiment with older adults, we asked 10 students (the same as the participants in section 3.1) to participate in our preliminary experiments. However, after excluding one participant (participant F) due to an equipment malfunction that resulted in incomplete sensor data, a total of nine participants were included in these experiments.

A prototype version of the CPCS was employed for these experiments. The system consists primarily of the MCT and the IDA, but does not have functions that allow the two mechanisms to communicate or that allow either of the components to use data collected by the other. Using both mechanisms in parallel would contribute to the combined use of the log data obtained by each. We conducted two kinds of experiments to investigate differences in impressions of the CPCS when the IDA talked spontaneously and when it did not. This experiment was conducted in two periods, a speech period and a nonspeech period, of three days each for a total experimental duration of six days. However, to ensure that the order of the two periods did not affect the subjects' impressions of the CPCS, the order was randomized for each participant. The IDA employed in the experiment had a simple STF consisting of a smart speaker (Google Home Mini) and a pyroelectric sensor without introducing the SCCF, and provided information (speech content $c \in \boldsymbol{C''} = \boldsymbol{C} \backslash$ {each user's daily routine work}) selected randomly every 15 min between 9:00 and 23:00 during the speech period. Additionally, we asked participants to listen to the IDA's speech at least eight times a day. Sensors detected the participants' behavior throughout the experimental period (sampling interval: 1 s; detection range: 115° within a 2-m view range). The difficulty adjustment function was included in the MCT. We asked the participants to play a 10-min memory game every day near the location where the sensor was installed.

Written questionnaires were also administered after the first half of the experiment and at the end of the experiment to inquire about the number of times participants heard the IDA speaking, whether the difficulty level of the memory game was appropriate, and their

Table 2: The actual speech probability of the chat-bot's speech at the end of the experiment for each participant.

Reward setting	R_A						R_B					
Participant / ranking	1st place	2nd place	3rd place	4th place	5th place	6th place	1st place	2nd place	3rd place	4th place	5th place	6th place
A	**0.680**	0.040	0.050	0.060	0.080	0.090	**0.300**	**0.280**	**0.180**	0.090	0.040	0.110
B	**0.505**	**0.257**	0.040	0.099	0.059	0.030	**0.310**	**0.380**	0.140	0.030	0.070	0.070
C	0.030	0.000	0.050	**0.416**	0.040	**0.455**	**0.300**	**0.180**	**0.210**	0.150	0.120	0.040
D	**0.198**	0.139	**0.208**	**0.238**	0.079	0.129	**0.416**	0.069	**0.238**	0.079	0.079	0.109
E	0.108	**0.206**	**0.235**	0.157	0.118	**0.167**	**0.168**	**0.218**	0.129	0.139	0.139	**0.208**
F	0.109	0.059	**0.327**	0.119	**0.218**	**0.168**	**0.168**	0.149	**0.208**	0.099	**0.267**	0.099
G	**0.300**	**0.260**	0.100	**0.180**	0.080	0.080	0.139	**0.287**	0.119	**0.257**	0.119	0.069
H	0.000	0.000	0.000	0.020	**0.980**	0.000	0.149	0.129	**0.228**	**0.277**	0.099	0.119
I	**0.190**	**0.170**	0.120	0.100	**0.250**	**0.170**	**0.304**	0.127	0.147	**0.196**	0.157	0.069
J	**0.240**	**0.256**	0.112	0.096	0.088	**0.208**	0.160	**0.260**	0.080	0.070	**0.200**	**0.230**
Average	**0.236**	0.139	0.124	0.148	**0.199**	0.150	**0.241**	**0.208**	**0.168**	0.139	0.129	0.112

willingness to continue to use the CPCS.

Similarly to the case of IDA, how efficiently the CPCS prototype can monitor a user's behavior would be influenced by its installation site and the daily habits of the user. However, synergistic effects are also expected through the concomitant use of the IDA and the MCT, and could have a positive impact on the frequency with which users engage in cognitive training. To clarify these issues, we evaluated and discussed the impressions of users, the effectiveness of the system at monitoring users, and the relationships between dialogue with the IDA and the frequency of cognitive training by examining data such as the sensor log, the result of the questionnaire/interview with the participants, and the training frequency in parallel.

5 Results and Discussion

This section describes our experimental results and discusses our investigation of the basic characteristics and performance of the SCCF as well as the effectiveness of the CPCS prototype presented in Section 4.

5.1 Performance Evaluation of the SCCF

Let us first discuss the performance and adequacy of the SCCF in providing appropriate information according to the user's preferences and condition. Table 2 lists the probability of the chat-bot talking about each category over the last 100 times in the three-day experimental period. This table shows the results for all participants. In this table, "ranking" corresponds to the participants' interest levels for each speech category obtained by the questionnaire administered after the experiment. This table revealed that the probability of talking acquired by the learning tends to be high when the rank of the interest level is high, especially under the reward setting R_B. The bold digits in the table indicate areas in which probability increased from its initial value (16.66...%). Table 3 represents the results of the questionnaire concerning the participants' interest levels for each category on a five-point scale (the other questionnaire results are shown in Appendix (B)). Digits in boldface indicate areas with an interest level greater than 3. Table 2 shows that, for reward setting R_B, the probabilities of the top three speech categories tend to be larger than the initial probability value while the probabilities of the bottom three categories tend to be smaller. Additionally, the results presented in Tables 2 and 3 show that, under reward setting R_B, there is a positive correlation between the probabilities of speech from the chat-bot at the end of the experiment and interest in the actual content of the speech under reward setting R_B. In other words, the probability that the chat-bot provides the participants with topics of interest increased throughout the experiment, while the probability that it provided topics for which many participants' feedback was U (no interest) decreased. Therefore, we believe that the chat-bot with reward setting R_B was able to provide information that reflected the users' thoughts and states.

On the other hand, the chat-bot with reward setting R_A failed to acquire the appropriate policy. An important feature in the chat-bot with setting R_A was that the probability of a particular category being provided was too large (up to 98%) when targeting a specific participant compared to the results obtained with setting R_B (up to 41%). Critically, reward setting R_A does not include negative rewards; regardless of whether the user's feedback is positive or negative, a positive reward is given. Thus, it is much more likely that the chat-bot with the reward setting R_A will be given a positive reward than the chat-bot with reward setting R_B. However, it seems that the speech probability was changed too drastically for the chat-bot with reward setting R_A under the current parameter settings.

Table 3: Degree of interest in participants' speech categories (on a 5-point scale).

Participant / ranking	1st place	2nd place	3rd place	4th place	5th place	6th place
A	**5**	**4**	**4**	3	2	2
B	**5**	**4**	3	3	3	2
C	**4**	**4**	**4**	3	1	2
D	**5**	**5**	**4**	1	2	1
E	**5**	3	3	**4**	3	2
F	**5**	**5**	**4**	3	2	1
G	**4**	**4**	**4**	**4**	3	2
H	**4**	**4**	3	2	1	1
I	**5**	**4**	**4**	2	1	1
J	**4**	**4**	**5**	**4**	3	3
Average	**4.6**	**4.1**	**3.8**	2.9	2.1	1.7

5.2 Basic Characteristics of the CPCS

This section discusses whether the prototype version of the CPCS can monitor and/or understand a user's circumstances in detail using the log data collected by the IDA and MCT, and evaluates users' impressions of the system as well. The sensor data of participant F were incomplete due to an equipment malfunction and are therefore not included in these results.

5.2.1 *Differences in the Participants' Amounts of Activity*

Let us consider the amount of activity detected by the pyroelectric sensor on the CPCS. A participant's amount of activity Act^x is defined as the number of times the sensor detects his/her motion for x min. The maximum and minimum values of Act^x equal $60x$ and 0, respectively. Since the sensor may also detect objects other than the user, the value of Act^x must be corrected depending on the experimental circumstances. Table 4 shows the maximum amount of activity per 30 min for each participant over the course of the experiment. As shown in the table, Participant B had the greatest amount of activity (number in red) while Participant A had the lowest amount (blue). Figure 3 also shows sketches of the rooms usually used by Participants A and B, and the location and orientation of the sensor in those rooms.

Installation of the sensor in a location where it can detect the participant's motion appropriately is essential in order for behavior to be estimated accurately. Since Participant B answered in the interview after the experimental period that she often reads for long periods of time at her desk, we believe that this caused her amount of activity to be greater than that of the other participants because she was active within the detection range of the sensor. These results suggest that environmental factors, such as the location and direction of the sensor, may influence the recorded amount of activity (some noise and error in sensing may also influence the results).

Therefore, in the case that the CPCS in its current form would be employed in practice by older adults, it is necessary to consider in advance where the sensors should be installed to effectively track the older adults' amount of activity. At the same time, if the CPCS could be installed at the appropriate location, we can expect to properly understand their behavior.

Table 4: Maximum amount of activity per 30 min for each participant.

Participant	Act^{30}	Participant	Act^{30}
A	1210	G	1301
B	1800	H	1353
C	1312	I	1711
D	1386	J	1503
E	1672		

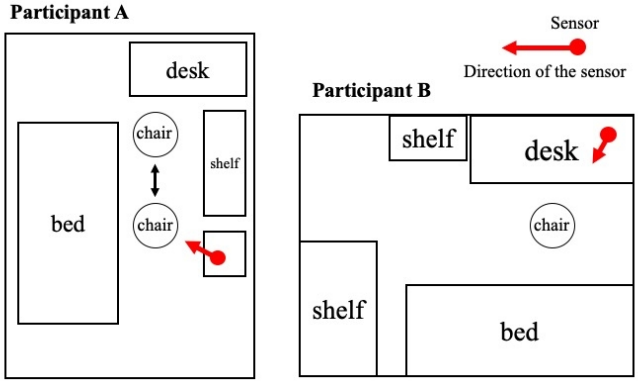


Figure 3: Diagram of the rooms where the experiment was conducted for Participants A and B

5.2.2 *Sensor Logs and Time Spent on the Memory Game*

Figure 4 depicts the average number of cognitive training memory games played every 30 min on weekdays (bar graph) and activity patterns (i.e., transition in the amount of activity) recorded by the sensor (line graph). The results shown in this figure are the average values for seven participants who showed similar behavioral trends. Participant I was excluded because his academic year and life patterns were different from those of the others. Here, interval (b) in Figure 4 is characterized by low activity because the participants were at school and not near the CPCS installation site. In addition, as can be seen at interval (c), there was an overall tendency for the participants to play the memory game more often at night, increasing their activity level at the same time. In contrast, interval (a) shows a larger amount of activity compared to interval (b), but almost no gameplay history except for Participant E. This result indicates that the participants were preparing to go to school during the corresponding period.

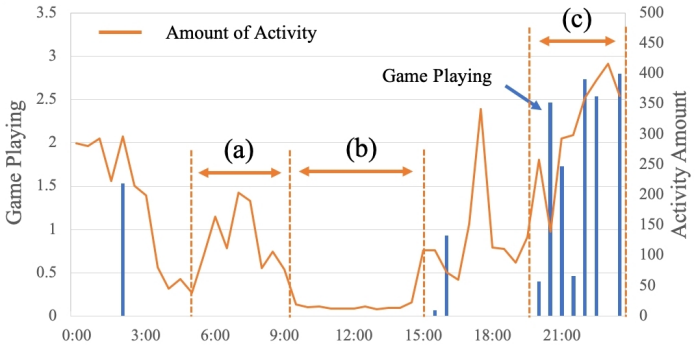


Figure 4: Average amount of game playing and activity patterns on weekdays.

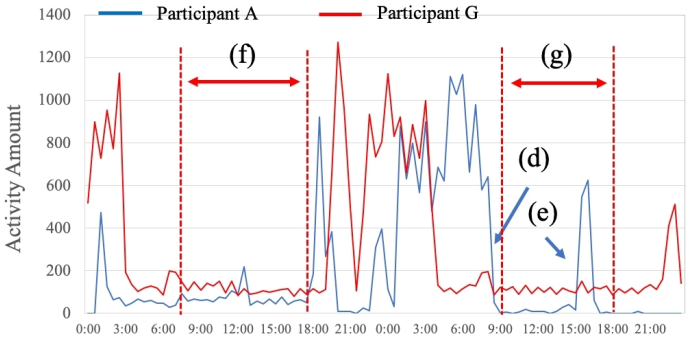


Figure 5: Weekend activity patterns for Participants A and G

At the same time, we can also see significant differences between the participants' behavior on weekday vs. weekends in terms of their life rhythms, such as sleeping hours and daily habits. Figure 5 shows the weekend activity patterns for Participants A and G, which were particularly distinctive. These patterns allow us to estimate the participants' awakening time and bedtime. For example, Participant A's activity decreased early Sunday morning ((d) in Figure 5) and increased around 15:00 ((e) in Figure 5). In the interview after the experiment, Participant A confirmed that this activity pattern corresponded to his sleep pattern. Similarly, the

other participants' sleep patterns could also be estimated from their activity logs. These results demonstrate the usefulness of the users' stored behavioral log data.

On the other hand, it was not possible to fully estimate Participant G's weekend sleep pattern. Figure 5 suggests that Participant G's bedtime was 3:00 on Saturday, and her waking time was around 19:00. However, in the interview after the experiment, we learned that she was awake but not in the room with the sensor at certain times of the day, such as from 7:30 to 17:00 (interval (f) in Figure 5) on Saturday and from 8:00 to around 18:00 (interval (g) in Figure 5) on Sunday. The fact that Participant G spent less time in the room where the CPCS was installed may have contributed to her lower amount of activity.

These results demonstrate that the use of sensor logs and gameplay histories obtained by the CPCS allows us to estimate not only daily habits such as waking and sleeping times but also whether users go out for long periods of time.

5.2.3 *Using Sensor Logs to Capture Users' Lifestyle Habits*

This section focuses on the sensor logs obtained in the IDA, a component of the CPCS, while playing the memory game on the MCT, and before and after playing. Figure 6 shows the amount of activity recorded every one minute for about 20 min while Participant B was playing the memory game. The memory game has 13 levels of difficulty and records three degrees of fatigue: 0, Bad; 1, Normal; 2, Good. Figure 7 shows the transitions in difficulty level and degrees of fatigue during the time that Participant B played the memory game. Note that all the intervals in Figure 7 correspond to interval (h) in Figure 6. The red circles on the red line plot in Figure 7 indicate the times at which the memory game was started. In Participant B's interview after the experimental period, she noted that she played the memory game at her desk (near the location where the CPCS was installed). During and after the gameplay (interval (h) and the second half of (h) in Figure 6), the amount of activity increased compared to the period of time before play, while after 22:12 (interval (i) in Figure 7), the participant temporarily stopped playing the memory game. Figure 6 confirms that the amount of activity remains high. We can speculate that the cause of this pattern was that Participant B was feeling fatigued from the memory game and may have stopped playing the game to do some other task or take a break.

Figure 8 depicts the amount of activity recorded every one min for about 20 min while Participant E was playing the memory game. In contrast to the results obtained for Participant B, Participant E's activity tended to decrease during the game-playing time ((j) in Figure 8) compared to his activity before and after playing. The interviews revealed that Participant E tended to play memory games on the bed. We inferred that this position was responsible for the low activity levels recorded by the system, since the location of the sensor would have made it difficult to detect someone on a bed. The interviews of the other participants indicated that the sensor logs adequately expressed their behaviors. It was clear from these results that the data obtained from the IDA and the MCT can be used to extrapolate detailed information about the user's circumstances.

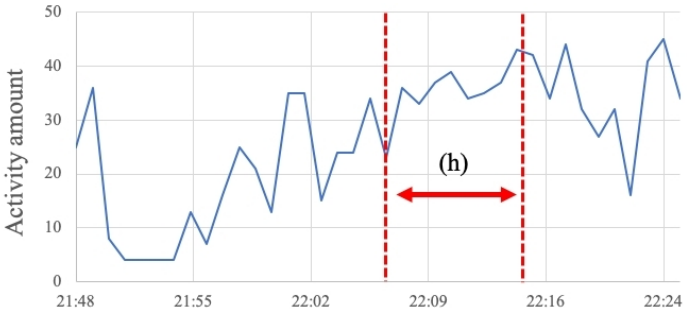


Figure 6: Activity before, during, and after Participant B's game playing.

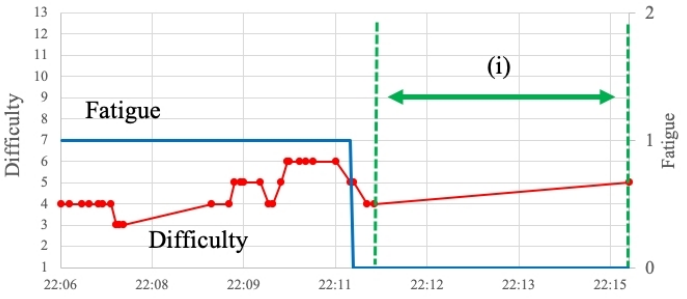


Figure 7: Changes in difficulty and fatigue during the memory game (Participant B).

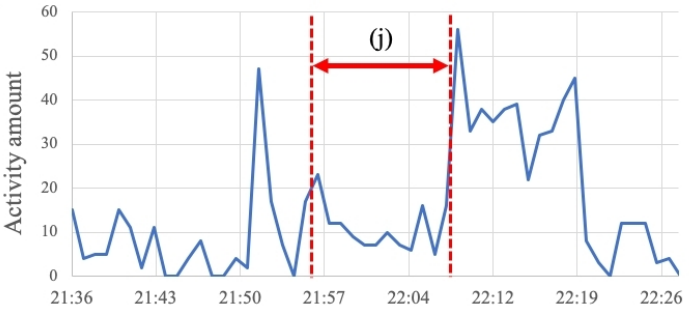


Figure 8: Activity patterns for Participant E.

5.2.4 *Questionnaire Investigation of the CPCS*

Next we will discuss the participants' impressions of the CPCS and the synergistic effects of combining the IDA and the MCT based on the results of subjective evaluations of the nine participants (college students) conducted after the experimental period. Among the nine participants, five stated that they felt there was a need to use the CPCS; however, there was only one participant who wanted to use both the IDA and the MCT as individual components of the CPCS. Thus, although in some cases the student participants expressed interest in the components of the CPCS, it can be said that, on the whole, they did not have a positive impression of the prototype version of the CPCS. One reason for this may be that students attend classes on a daily basis and have frequent opportunities to talk with the people around them. Further research needs to be conducted to investigate differences in the trends across age groups.

Table 5 shows the average number of times the memory game provided by the MCT was played per day by each participant with

Table 5: Differences in the number of times the memory game was played with and without speech from the IDA.

Participant	A	B	C	D	E	G	H	I	J	Average
Not talking	14.7	5.0	23.0	21.0	0.0	37.0	21.0	31.0	30.0	20.3
Talking	**20.7**	**10.0**	**26.7**	**31.0**	**15.7**	**41.0**	**29.0**	**40.3**	**41.3**	**28.4**
Difference (talking - not talking)	6.0	5.0	3.7	10.0	15.7	4.0	8.0	9.3	11.3	8.1

and without speech from the IDA. When the IDA spoke, the number of times each participant played was greater than that when the IDA did not speak (average: 8.1 times). We consider that the above result was caused by the synergy of the combination of the IDA and the MCT, because the dialogue between the IDA and the participant promoted their habitual game playing. The other questionnaire results can be seen in Appendix (C).

6 Conclusion

In the present study, we introduced the Speech Content Coordinating Function (SCCF) into the Intelligent Dialogue Agent (IDA), and developed a prototype version of the Comprehensive Preventive Care System (CPCS) consisting of the IDA and the Mechanism for Cognitive Training (MCT). The SCCF was developed as a text-based agent in the form of a chat-bot that employed reinforcement learning based on the policy gradient method. This system was tested in an experiment with student participants. The experimental results confirmed that the SCCF acquires appropriate policies based on the user's preferences and conditions. Furthermore, the results of experiments using the prototype version of the CPCS confirmed that taking advantage of data obtained from both the IDA and the MCT can lead to synergistic effects to monitor users in greater detail, better understand their circumstances, and increase the frequency of cognitive training.

In future research, we will further improve the convenience and familiarity of the SCCF by determining not only the content of speech but also the appropriate frequency and timing of speech based on the user's circumstances at any given time. At the same time, because this improvement may cause an exponential increase in the number of states and actions compared to the learning environment of the experiments conducted in the present study, we plan to improve the current learning algorithm or apply more powerful ones, such as deep reinforcement learning algorithms. The adaptability and familiarity of the improved SCCF will need to be evaluated through more practical experiments with older adults in a setting that more closely replicates real-life circumstances.

We also expect that further synergistic effects will occur by allowing the components of the prototype CPCS to communicate and share their data. It will also be necessary investigate the effects of the CPCS on older adults and their impressions of this system by fully examining the different trends in people of different age groups. We also plan to integrate a fall prevention system, which is currently being developed in parallel by our research group, into the CPCS; however, since this system involves physical exercise and a certain risk of injury, we will proceed with its integration in stages, starting with older adults who are at little risk of injury. Similarly, we will continue to work on the introduction of other preventive care systems and expect to see synergies similar to those obtained in the present study.

Acknowledgment The authors would like to thank the staff of the Fujisawa Care Center, and the members of the Hachioji Senior Citizens' Club and Cafe Kajiyashiki for their cooperation.

References

[1] D. Kitakoshi, S. Hirose, A. Yamashita, M. Suzuki, K. Suzuki, "Development of an Intelligent Dialogue Agent for Older Adults: Evaluation of Functions to Control Spontaneous Talk and Coordinate Speech Content," in 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), 1–6, IEEE, 2019, doi:10.1109/TAAI48200.2019.8959870.

[2] S. Bureau, The Statistical Handbook of Japan 2019, Ministry of Internal Affairs and Communications, 2019.

[3] D. Kitakoshi, T. Okano, M. Suzuki, "An empirical study on evaluating basic characteristics and adaptability to users of a preventive care system with learning communication robots," Soft computing, **21**(2), 331–351, 2017, doi:10.1007/s00500-015-1631-7.

[4] T. Suzuki, H. Kim, H. Yoshida, T. Ishizaki, "Randomized controlled trial of exercise intervention for the prevention of falls in community-dwelling elderly Japanese women," Journal of bone and mineral metabolism, **22**(6), 602–611, 2004, doi:10.1007/s00774-004-0530-2.

[5] A. Sixsmith, N. Johnson, "A smart sensor to detect the falls of the elderly," IEEE Pervasive computing, **3**(2), 42–47, 2004, doi:10.1109/MPRV.2004.1316817.

[6] N. K. Suryadevara, S. C. Mukhopadhyay, "Wireless sensor network based home monitoring system for wellness determination of elderly," IEEE sensors journal, **12**(6), 1965–1972, 2012, doi:10.1109/JSEN.2011.2182341.

[7] W.-H. Wang, P.-C. Chung, Y.-L. Hsu, M.-C. Pai, C.-W. Lin, "Inertial-sensor-based balance analysis system for patients with Alzheimer's disease," in 2013 Conference on Technologies and Applications of Artificial Intelligence, 128–133, IEEE, 2013, doi:10.1109/TAAI.2013.36.

[8] R. Naqvi, D. Liberman, J. Rosenberg, J. Alston, S. Straus, "Preventing cognitive decline in healthy older adults," CMAJ, **185**(10), 881–885, 2013, doi:10.1503/cmaj.121448.

[9] Y. S. Delahoz, M. A. Labrador, "Survey on fall detection and fall prevention using wearable and external sensors," Sensors, **14**(10), 19806–19842, 2014, doi:10.3390/s141019806.

[10] T. Watanabe, K. Kamata, S. A. Hasan, S. Shibusawa, M. Kamada, T. Yonekura, M. Yamada, "Design of an antagonistic exercise support system using a depth image sensor," in Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare, 162–169, 2016.

[11] A. Chelli, M. Patzold, "A Machine Learning Approach for Fall Detection and Daily Living Activity Recognition," IEEE Access, **7**, 38670–38687, 2019, doi:10.1109/ACCESS.2019.2906693.

[12] D. Yacchirema, J. DePuga, C. Palau, M. Esteve, "Fall detection system for elderly people using IoT and ensemble machine learning algorithm," Personal and Ubiquitous Computing, **23**(5-6), 801–817, 2019, doi:10.1007/s00779-018-01196-8.

[13] G. López, L. Quesada, L. A. Guerrero, "Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces," in International Conference on Applied Human Factors and Ergonomics, 241–250, Springer, 2017, doi:10.1007/978-3-319-60366-7_23.

[14] M. K. Wolters, F. Kelly, J. Kilgour, "Designing a spoken dialogue interface to an intelligent cognitive assistant for people with dementia," Health Informatics Journal, **22**(4), 854–866, 2016, doi:10.1177/1460458215593329.

[15] S. Nakatani, S. Saiki, M. Nakamura, K. Yasuda, "Generating Personalized Virtual Agent in Speech Dialogue System for People with Dementia," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 10917 LNCS, 326–337, 2018, doi:10.1007/978-3-319-91397-1_27.

[16] A. Russo, G. D'Onofrio, A. Gangemi, F. Giuliani, M. Mongiovi, F. Ricciardi, F. Greco, F. Cavallo, P. Dario, D. Sancarlo, V. Presutti, A. Greco, "Dialogue Systems and Conversational Agents for Patients with Dementia: The Human-Robot Interaction," Rejuvenation Research, **22**(2), 109–120, 2019, doi:10.1089/rej.2018.2075.

[17] F. Tang, I. Uchendu, F. Wang, H. H. Dodge, J. Zhou, "Scalable diagnostic screening of mild cognitive impairment using AI dialogue agent," Scientific Reports, **10**(1), 2020, doi:10.1038/s41598-020-61994-0.

[18] S. Yamada, D. Kitakoshi, A. Yamashita, K. Suzuki, M. Suzuki, "Development of an Intelligent Dialogue Agent with Smart Devices for Older Adults: A Preliminary Study," in 2018 Conference on Technologies and Applications of Artificial Intelligence (TAAI2018), 50–53, IEEE, 2018, doi: 10.1109/TAAI.2018.00020.

[19] D. Kitakoshi, R. Hanada, K. Iwata, M. Suzuki, "Cognitive training system for dementia prevention using memory game based on the concept of human-agent interaction," Journal of Advanced Computational Intelligence and Intelligent Informatics, **19**(6), 727–737, 2015, doi:10.20965/jaciii.2015.p0727.

[20] H. Eguchi, S. Yoshikawa, A. Yamashita, D. Kitakoshi, Y. Ho, Y. Hujimoto, T. Yamaguchi, M. Kanae, T. Sato, "Analysis of pyroelectric sensor data installed in thehomes of elderly individuals," in Proc. the First International Symposium on AI for ASEAN Development, 2018.

[21] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, MIT press Cambridge, 1998.

[22] R. J. Williams, "Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning," Machine Learning, **8**(3-4), 229–256, 1992, doi:10.1007/BF00992696.

[23] S. M. Kakade, "A natural policy gradient," in Advances in neural information processing systems, 1531–1538, 2002.

[24] H. Kimura, K. Miyazaki, S. Kobayashi, "Reinforcement learning in POMDPs with function approximation," in International Conference on Machine Learning, volume 97, 152–160, 1997.

[25] C. J. Watkins, P. Dayan, "Q-learning," Machine learning, **8**(3-4), 279–292, 1992, doi:10.1007/BF00992698.

[26] G. A. Rummery, M. Niranjan, On-line Q-learning using connectionist systems, volume 37, University of Cambridge, Department of Engineering Cambridge, UK, 1994.

[27] R. Nakano, "Efficient learning of behavioral rules," in From Animals to Animats 6: SAB2000 Proceedings Supplement of the Sixth International Conference on Simulation of Adaptive Behavior, p178-184. Paris, France, 2000.

Appendix

We administered three questionnaires concerning (A) older adults' impression of the smart speaker and the IDA, (B) the evaluation of the SCCF, and (C) the preliminary experiments with the CPCS, before, during, or after the experiments described in Section 4. Several major results were already given in Section 4, and other results related to the discussion in Section 4 are shown below. Tables A1 - A4, B1 - B3, and C1 - C8 list the questions and typical answers for each.

(A) A questionnaire administered to investigate older adults' impression of the smart speakers and the IDA

Affiliation of the participants

A preventive care circle for older adults and a senior citizens' club

Number of participants

28 (9 from a preventive care circle, and 19 from the senior citizens' club)

Gender and age

male: 12; female: 16; average age: 76.6 (male: 76.4; female: 77.4)

Table A1: Items in the questionnaire administered to older adults.

Item	Question
Awareness of smart speakers.	(1) Are you familiar with smart speakers? (2) What kind of speech content would you like the IDA to spontaneously talk about? (multiple choice allowed) (3) Would you want to use a smart speaker if one was in your home?
Impression of the spontaneous talking function	(4) If a smart speaker talked to you spontaneously, what would be your reaction? Could you select all items applicable to you? (5) If a smart speaker could talk to you spontaneously, what would you like to talk about?
Information about the environment	(6) Do you have a Wi-Fi environment at home?
Information about the participant	(7) Could you tell me your daily routine? (8) Could you tell me your hobbies?

Table A2: Answers to question (1) in Table A1 (Answerer: 9 older adults in preventive care circle).

Answer	Percentage of respondents (%)
I know about them.	0.0
I have heard the name.	11.1
I don't know about them.	77.8
No answer	11.1

Table A3: The number of contents the older adults selected for question (2) in Table A1.

Index	Number of responses
Maximum	6.0
Minimum	1.0
Average	3.4

Table A4: Answers to question (6) in Table A1.

Answer	Percentage of respondents (%)
I have Wi-Fi.	35.7
I don't have Wi-Fi.	46.4
I don't know.	7.1
No answer	10.7

(B) A questionnaire administered to evaluate the SCCF

Affiliation of the participants

National Institute of Technology, Tokyo College

Number of participants

10

Gender and age

male: 8; female: 2; average age: 18.9

Table B1: Items in the questionnaire administered after the experiment to evaluate the SCCF.

Item	Question
Effect of the learning	(1) Comparing the first and third of these 3 days, did you feel the speech contents from the chat-bot changed? (2) The IDA's setting to learn your preferences for its speech content was different between the first 3 days and the last 3 days. In which period do you think the IDA provided more interesting speech contents, the first 3 days or the last 3 days? (3) Could you tell me the reason for your answer to the above question?
Impression of the chat-bot's speech content	(4) Which of the following best describes your impression of the IDA's speech content on this particular day?
Participants' interest in the chat-bot's speech content	(5) Could you rank the topics provided by the chat-bot in terms of how much they interested you? (6) Could you rate each topic provided by the chat-bot during the experiment on a 5-point scale?
Participants' interest in the topics in daily life	(7) Could you rank the topics provided by the chat-bot in order according to how much they usually interest you? (8) Could you rate the topics provided by the chat-bot on a 5-point scale based on how much they interest you in your daily life?

Table B2: Older adults' impression of the chat-bot at each reward setting (answers to questions (1) and (2) in Table B1, 3-point scale, with 3 being best).

Item / Reward setting	R_A	R_B
Changes in speech content from chat-bot	2.3	1.9
Impression of the chat-bot's speech content	2.4	2.3

Table B3: Which reward settings R_A or R_B was preferred by older adults (answers to question (6) in Table B1).

Reward setting	Number of choosers
R_A	3
R_B	6
Neither	1

(C) A questionnaire administered regarding the preliminary experiments with the CPCS

Affiliation of the participants

National Institute of Technology, Tokyo College

Number of participants

10 (As for the questionnaires during and after the preliminary experiment, the number of participant equals 9.)

Gender and age

male: 8; female: 2; average age: 18.9

The parts of Tables C7 and C8 expressed as red-colored text are the questions and the answers discussed in Section 5.2.4.

Table C1: Items in the questionnaire administered before the preliminary experiment

Item	Question
Information about the environment	(1) Please place a check mark beside the following items if they were located at the system installation site.
Information about the participant	(2) Could you tell me your daily routine? (3) Could you tell me the approximate time when you go home? (4) Could you tell me your schedule after going home on mm/dd? (5) Could you tell me how many people you live with?
Awareness of smart speakers	(6) Do you know about smart speakers? (7) Do you have a smart speaker? (8) Would you want to use a smart speaker if you had one at home? (If you already have one, do you want to keep using it?)

Table C2: Equipment in the participant's home. Answers to question (1) in Table C1.

Equipment	The number of people with the equipment.
Display that can be connected with HDMI	9
HDMI cable	9
Wired/Wireless keyboard.	8
Wired mouse.	7

Table C3: Awareness of smart speakers. The number of respondents by choice for each question ((6), (7), (8) in Table C1).

Question	Answer	The number of respondents
(6)	I know about them.	5
	I only know the name.	5
	I don't know about them.	0
(7)	Yes.	2
	No.	8
(8)	I want to use one.	6
	I'm not sure.	2
	I don't want to use one.	0
	I have one and I want to keep using it.	2
	I have one, but I don't want to keep using it.	0

Table C4: Items in the questionnaire administered during the preliminary experiment

Item	Question
Frequency of listening to the IDA's speech	(1) Could you tell me the average number of times per day you listened to the speech from the IDA?
Impression of the spontaneous talking function	(2) Which of the following best describes your impression of the spontaneous talk from the IDA? (3) What is your impression of the IDA's spontaneous talk? (4) Could you tell me the reason you chose that item as the answer to the above question? (5) Could you tell me the contents that you want the IDA to talk about? (if any)
Willingness to keep using the CPCS	(6) Do you want to keep using the cognitive training mechanism and the IDA?
Impression of the CPCS	(7) Which of the following best describes your impression of the prototype of the CPCS (the cognitive training mechanism and the IDA)? (8) Could you tell me the reason for your answer?

Table C5: The number of times each older adult listened to the IDA's speech during the experimental period (answers to question (1) in Table C4).

Index	Value
Maximum	16.0
Minimum	1.0
Average	7.0

Table C6: The number of respondents by choice for each question ((2), (3), (6), (7) in Table C4)

Index	Answer	The number of respondents
(2)	I'd like the IDA to talk more frequently.	3
	The frequency of speech from the IDA is high.	2
	I'd like the IDA to increase its amount of speech.	3
	I'd like the IDA to decrease its amount of speech.	1
	The IDA provides useful information.	3
	The IDA provides uninteresting information.	4
(3)	Very good. It would be useful.	0
	A bit good.	3
	No opinion.	5
	Usually not good.	1
	Very bad. Unnecessary.	0
	Yes, both of them very much.	0
(6)	Yes, both of them a little.	1
	Yes, only the cognitive training mechanism very much.	2
	Yes, only the cognitive training mechanism a little.	1
	Yes, only the IDA very much.	0
	Yes, only the IDA a little.	4
	I don't want to use either of one of them very much.	1
	I don't want to use either one of them at all.	0
(7)	I thought it was interesting and a necessary part of the system.	4
	I thought it was interesting, but not really necessary for the system.	0
	I didn't feel the necessity of the IDA.	3
	I didn't feel the necessity of the cognitive training mechanism.	1
	I thought it was not interesting at present, but the system would be needed in the future.	1
	I thought it was not interesting and I do not feel that it is necessary.	0

Table C7: Items in the questionnaire conducted after the preliminary experiment

Item	Question
Impression of the training difficulty adjustment function	(1) The cognitive training mechanism has a function to coordinate training difficulty. How did you feel the difficulty of training (memory game) was coordinated by the function? (2) How would you like to the training difficulty to changed?
Required functions in the cognitive training mechanism	(3) Could you tell me your opinion about the functions you want us to add and/or your requests for improvement?
Willingness to keep using the CPCS	(4) Which of the following systems would you want to use on a regular basis?
Impression of the spontaneous talking function	(5) Could you tell me your impression of the IDA having the spontaneous talking function compared to that without the function? (6) Could you tell me the reason for your answer to the above question?
Required functions in the CPCS	(7) Could you tell me your opinion about the functions you want us to add and/or your requests for improvement in the CPCS?

Table C8: The number of respondents by choice for each question ((1), (2), (4), (5) in Table C7).

Index	Answer	The number of respondents
(1)	It was easy throughout the game.	0
	It was appropriate.	8
	It was difficult throughout the game.	0
	It was sometimes easy, sometimes difficult.	1
(2)	I want to be able to change the difficulty by myself.	3
	It is better that the cognitive training mechanism changes the difficulty.	3
	Both are OK.	0
	It depends on the situation.	3
(4)	The IDA and the cognitive training mechanism.	1
	The IDA.	3
	The cognitive training mechanism.	5
	I don't want to use either of them.	0
(5)	I thought it was interesting and felt that it was a necessary part of the system.	3
	I thought it was interesting, but didn't feel that it was necessary.	2
	I didn't feel the necessity of the IDA.	2
	I thought it was not interesting at present, but the function will be needed in the future.	2
	I thought it was not interesting and also not necessary.	0

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 22-29 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Supervised Learning Techniques for Stress Detection in Car Drivers

Pamela Zontone[1], Antonio Affanni[1], Riccardo Bernardini[1], Leonida Del Linz[2], Alessandro Piras[1], Roberto Rinaldo[*1]

[1]*DPIA, Polytechnic Department of Engineering and Architecture, University of Udine, Udine, 33100, Italy*

[2]*VI-grade Srl, Via G. Galilei 42, Tavagnacco, Udine, 33010, Italy*

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 21 October, 2020
Online: 08 November, 2020

Keywords:
Stress Detection in Car Drivers
Electrodermal Activity
Heart Rate Variability
Motion Artifact
Supervised Learning Algorithm
Machine Learning
Deep Learning

ABSTRACT

In this paper we propose the application of supervised learning techniques to recognize stress situations in drivers by analyzing their Skin Potential Response (SPR) and the Electrocardiogram (ECG). A sensing device is used to acquire the SPR from both hands of the drivers, and the ECG from their chest. We also consider a motion artifact removal algorithm that allows the generation of a single cleaned SPR signal, starting from the two SPR signals, which could be characterized by artifacts due to vibrations or movements of the hands on the wheel. From both the cleaned SPR and the ECG signals we compute some statistical features that are used as input to six Machine Learning Algorithms for stress or non-stress episodes classification. The SPR and ECG signals are also used as input to Deep Learning Algorithms, thus allowing us to compare the performance of the different classifiers. The experiments have been carried out in a firm specialized in developing professional car driving simulators. In particular, a dynamic driving simulator has been used, with subjects driving along a straight road affected by some unanticipated stress-evoking events, located at different positions. We obtain an accuracy of 88.13% in stress recognition using a Long Short-Term Memory (LSTM) network.

1 Introduction

Lately[1], the assessment of stress in car drivers is getting more attention, because a highly stressed driver could be less focused, more likely to drive in a risky way [1], and/or to get involved into accidents [2]. The emotional state, corresponding to a condition of tension and pressure in an individual, causes reactions of the sympathetic nervous system. These are defined as stress in literature [3].

Different approaches have been proposed to detect stress in individuals, using physical or physiological parameters [4]. Machine Learning (ML) algorithms are often employed [5]– [8]. In [9], a Support Vector Machine (SVM) classifier is used, with Skin Conductance Response (SCR), Electroencephalogram (EEG), and Heart Rate (HR) as inputs. In [10], various signals are utilized, such as Blood Volume Pulse (BVP), Skin Conductance (SC), Skin Temperature (ST), Electromyography (EMG), and Respiration (RESP). Deep Learning (DL) algorithms have also been successfully applied for stress recognition. Physiological signals are used, such as Electrocardiogram (ECG) [11] and Electrodermal Activity (EDA) [12], or facial expressions [13], or a collection of physiological and motion data, including ECG, EDA, EMG, RESP, Temperature (TEMP), and Acceleration (ACC) [14]. The system described in [15] uses the time differences between consecutive R-waves of QRS complexes, also called RR intervals, for mental stress identification in firefighters. In several works, a comparison between ML and DL methods for stress recognition has been carried out, also showing that Convolutional Neural Networks (CNNs) can outperform ML conventional methods [16, 17].

There are various stress analysis approaches when we consider car driving scenarios as well. Many use physical characteristics, e.g., facial or eye images [18, 19]. In others, physiological characteristics are considered [20–22]. Among these, EDA can be used jointly with different signals [23, 24]. Authors in [25] study the relation between the vehicle's acceleration and the driver's heart rate in elders. In [26], EEG patterns, belonging to different driving conditions, are classified using an SVM, a Neural Network (NN), and a Random Forest (RF). Some DL algorithms have also been successfully applied for stress recognition in the driving context as well, and are particularly relevant to this work. Authors in [27] present an end-to-end architecture, composed of convolutional layers, that uses ECG signals for stress detection. Their experiments are designed for identification of mental fatigue and for discrimina-

[*]Corresponding Author: Roberto Rinaldo, roberto.rinaldo@uniud.it
[1]This paper is an extension of the work originally presented in [43].

tion between driving and resting conditions. A multimodal system based on CNNs and Long Short-Term Memory (LSTM) networks, using ECG, vehicle data, and contextual data has been proposed in [28]. In [29], CNNs are considered to detect a driver's braking intention, in different driving conditions, using EEG data.

In this work we present a system that allows the recognition of the emotional and stress conditions of subjects while driving. The main features of our system, which differentiate our work from others in the literature, are the relative simplicity of the sensors used, the possibility of being used nearly in real time on 15 s signal blocks, and the ability to discriminate stressful events of short duration during driving. The experiments are organized in a company that designs professional driving simulators. We use a simulator that reproduces a road track in a 3D environment. Along the track we place different tasks to go through, which are expected to cause a reaction in the subjects. Two different signals are considered, the Skin Potential Response (SPR) and the ECG. We have proven that these signals are suitable for the analysis of the emotional condition of a driver and that can be acquired with an acceptably intrusive setup, with no intervention on the vehicle [30].

We analyze the SPR signal instead of the widely utilized SCR, because it can be logged in a more straightforward way, with no electric current in the skin. In addition, the electrode impedance and skin impedance changes have a lower impact on this signal [31]. In detail, our objective is to compare the performance of several ML algorithms in recognizing the stress responses of the drivers that undergo various stress situations. To do this, some features are extracted from both the SPR and ECG signals. In addition, we consider deep learning techniques for driver stress recognition. These require the raw recorded signals as input, contrary to what happens in traditional machine learning techniques, where a feature extraction process is needed. Specifically, we use a CNN and an LSTM network. Compared to other similar studies cited above [24, 27], we apply machine and deep learning techniques for detecting stress during short tasks in a controlled and simulated environment.

The paper is structured as follows. In the next section, a description of the fundamental blocks of the proposed system is provided. We introduce the sensing device utilized for both SPR and ECG recordings, the Motion Artifact (MA) removal algorithm developed to reduce motion artifacts, and the learning algorithms implemented for classification. Section 3 reviews the experimental setup. Section 4 presents the results of our comparative study, using different classification algorithms, on drivers' stress recognition. Finally, in Section 5 we draw some conclusions.

2 Description of the system

The block diagram of the developed system is shown in Figure 1. The inputs are the two SPR signals, recorded from the hands of a test subject ($s_1(t)$ and $s_2(t)$ in the figure), and the ECG signal recorded from the chest of the subject. The signals are acquired synchronously through a sensing device that uses a textile garment with electrodes, in addition to other electrodes placed under the gloves on both hands.

The two SPR signals can exhibit some artifacts that can be generated when the subject is driving, e.g., due to the hands movements or vibrations. To overcome this problem, we designed a MA removal algorithm that, taking into account the local energy of the two SPR signals, picks the smoother one, with the aim to remove spurious peaks and asymmetric signal spikes, possibly related to MA. As a matter of fact, the SPR component due to the sympathetic nervous system activity in the two hands is highly correlated. In the end, a learning classification algorithm is used to recognize the presence of stress in certain time intervals. In detail, every single time interval is classified into two possibile classes, which are the stress class (or "1") and the non-stress class (or "0"). In case of ML algorithms, this is carried out by selecting some statistical features from the cleaned SPR signal ($s(t)$ in Figure 1) and from the ECG signal. These features have been picked out, among others, since they well represent the stress signal peculiarities. In case of DL algorithms, instead, the raw SPR and ECG recordings are directly processed. In the next subsections, we will describe in detail each one of these fundamentals blocks.

2.1 The sensing device

The sensor used for data acquisition is the VI-BioTelemetry system (https://www.vi-grade.com/en/solutions/vi-biotelemetry/). This commercial sensor, shown in Figure 2, has been designed in our group and engineered by the VI-grade company. Its detailed description can be found in several recent scientific papers (e.g., [31, 32]).

Summarizing its features and specifications, the sensor is battery operated and transmits data to a laptop via WiFi connection. It acquires two SPR channels and three ECG channels. The SPR signals are acquired from the hands of the tested subject (through Ag/AgCl wet electrodes positioned on the palm and the back of each hand) and the ECG data are acquired using a commercial vest having textile electrodes in contact with the chest of the subject (using an adhesive conductive gel). The analog front end conditions the SPR and ECG signals with proper bandwidth and gain. Analog signals are converted using a 12-bit A/D and an on board DSP which acquires the signals at 1 kSa/s rate. The A/D converter is characterized by an Integral Nonlinearity (INL) lower than 1 LSB (Least Significant Bit) and an Effective number of bits (ENOB) equal to 11.1. The DSP sends the acquired signals to the WiFi module which is responsible to send all the time-aligned data to a server in the local area network.

Regarding the ECG channels, they are basically designed as band-pass differential amplifiers with maximum input range ± 5 mV (i.e., gain 370) and bandwidth in the [0.03, 160] Hz range. The input impedance of each channel is 100 MΩ, in order to reduce the load error to less than 1%, since the skin impedance is lower than 1 MΩ. The resolution of the ECG signals acquired on the chest skin results to be in the order of 2.5 μV and the accuracy of the ECG readout has been characterized in the past [32] to be as low as 0.1%. As for the ECG signal, we first detect the R-peak locations using the Pan-Tompkins algorithm [33].

The instantaneous RR signal is obtained by converting the RR interval time series with cubic spline interpolation at 100 Sa/s. Similarly to what is done in [34], where RR values and DL are used for Congestive Heart Failure (CHF) detection, interpolation is needed to synchronize signal segments, in our case with the SPR signal blocks. Note that this can be important to preserve correct time and

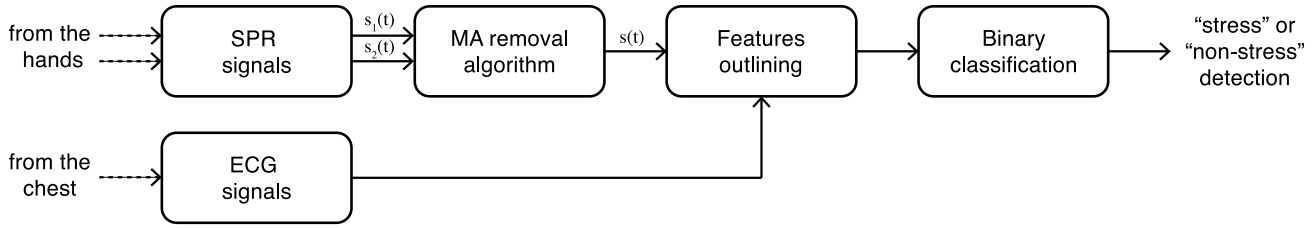


Figure 1: Block diagram of the presented system.

frequency information (see [35] and references therein), which the DL classifiers can possibly extrapolate from the data.

Regarding the SPR channels, the sensors acquire the electrodermal signals on both hands as differential voltages between the palm and the back of each hand. The SPR signals are supposed to be in the ±10 mV range. They are amplified with a gain equal to 160. The acquisition bandwidth is in the [0.08, 40] Hz range; the input impedance of each channel is 100 MΩ. The resolution of the SPR data acquired on the hands results in the order of 5 μV and the accuracy of the SPR readout has been characterized in the past [31] to be in the order of 0.2%. The SPR signals are subsampled at 100 Sa/s before further processing.

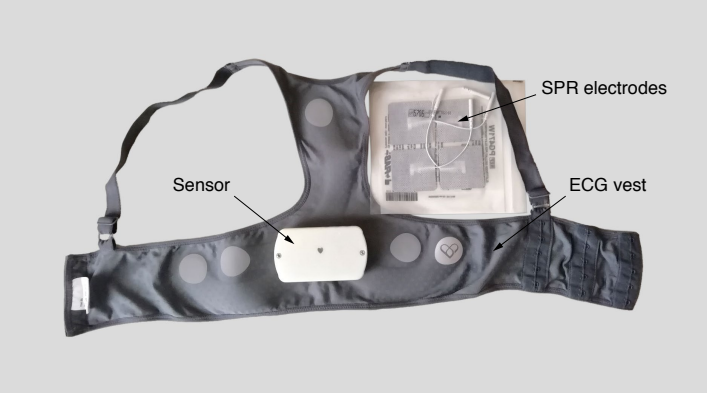


Figure 2: The VI-BioTelemetry sensor used in the experiments. The system is composed by an acquisition box, a vest with textile electrodes for ECG acquisition on the chest and Ag/AgCl electrodes for SPR acquisition on the hands.

2.2 Motion artifact removal block

As introduced above, some artifacts due to the movements and actions of the subject's hands while driving can become visible on the SPR signals, and could alter their shape. To take this issue into account, we propose an MA algorithm (see also [36, 37]) that is built on two assumptions: the first being that motion artifact enhances the local energy of the signal. The second being that the motion artifacts rarely appear simultaneously in the SPR signal of both hands. We observed this behaviour in our testing most of the time, and the proposed solution appears to be a good compromise to lessen the MA. Considering the RMS values σ_1 and σ_2 of the SPR input signals s_1 and s_2 on a $L = 100$ samples wide moving window, corresponding to 1 s at a sample rate of 100 Sa/s, we establish a smoothing threshold function $f(x)$ in order to obtain $f(x) \approx 0$ when $0 \leq x \ll 1$ and $f(x) \approx 1$ when $1 \ll x \leq 2$, namely,

$$f(x) = \frac{1}{1 + e^{-2(x-1)}} \tag{1}$$

This function is employed in a correction term β, that enables us to reduce motion artifacts. At the nth sample, this correction term is calculated as:

$$\beta(n) = \begin{cases} f\left(\frac{\sigma_1(n)}{\sigma_2(n)}\right) & \text{if} \quad \sigma_2(n) \neq 0 \\ 1 & \text{if} \quad \sigma_2(n) = 0 \end{cases} \tag{2}$$

We can deduce that $\beta \to 0$ when $\sigma_1 < \sigma_2$ and $\beta \to 1$ when $\sigma_1 > \sigma_2$. The correction term β, as in (2), is then used to determine the clean single output $s(n)$ of the procedure as

$$s(n) = \beta(n) \cdot s_2(n) + \left[1 - \beta(n)\right] \cdot s_1(n) \tag{3}$$

In the end, the output in (3) imitates the signal characterized by the lower energy in the 1 s examined interval.

2.3 Classification algorithms

In the proposed system, we utilize a classification technique to detect the presence (or absence) of a driver's stress condition in a given time interval during a car driving simulation. For this classification task, we utilize and compare six different supervised machine learning algorithms: an SVM classifier, a Naïve Bayes classifier, a k-Nearest Neighbors (k-NN) classifier, a Decision Tree (DT), an RF classifier, and an Artificial Neural Network (ANN). All these algorithms have been widely used in the literature for classification purposes. We also use two deep learning algorithms, in particular a 1D CNN and an LSTM network.

Our goal is to classify all of the time intervals in two possible classes: the stress class (or "1") and the non-stress class (or "0"). For this, we use the processed and cleaned SPR signal, and the ECG signal, with the latter one processed to remove the ectopic beats similarly to what is proposed in [35]. In addition, a further normalization preprocessing step is applied to both SPR and RR signals, because they can differ considerably among the individuals. This step consists of the standardization of each signal, based on the mean and variance of the preceding 5 minutes of the current signal value. We decided to process a 15 s long interval at a time, and we take a new interval every 5 s. In this way each interval covers the preceding one by 10 s. We do this to keep a reasonably low processing delay. The overall number of features extracted from these time intervals is nine: five from the SPR signal and four from the ECG signal. The five statistical features considered for the SPR signal

are: the interval variance, the energy, the mean absolute value, the mean absolute derivative and the max absolute derivative. The four statistical features from the ECG signal are: the mean value of RR intervals, the standard deviation of RR intervals (SDNN), the root mean square of subsequent RR interval differences (RMSSD) and the mean value of the Heart Rate [38]. We gather the features of the two signals together to construct the feature vectors. Each feature is then normalized in the range [0, 1]. All of the ML classifiers have been set up using Matlab (2017.a), except for the ANN that has been set up using Python. Hyperparameters are optimized automatically by the Matlab routine functions used for classification. In addition, for the SVM classifier, a Radial Basis Function (RBF) kernel has been used. The Bayesian optimization has also been used during the training procedure. For the ANN, we implement a four-layer perceptron trained with the back-propagation algorithm. We utilize Keras and the optimization package called "hyperopt" [39], allowing the use of the Bayesian optimization in this case as well. The optimization process also allows us to select the number of nodes in the two hidden layers (ranging from 16 to 128). The drop-out percentage has also been computed by the optimization package. Regarding the cross validation methodology, the 10-fold cross validation is used for all of the ML algorithms. For the ANN classifier we set the batch-size equal to 128.

The DL algorithms are also set up using Python. In detail, the architecture of the CNN used for classification has been inspired by the one proposed in [28], and its parameters are briefly reported in Table 1. The CNN contains two identical sequences of four operations: a convolutional layer, an Exponential Linear Unit (ELU), a batch normalisation layer and a max pooling layer. These are followed by a fully-connected layer (of 128 nodes), and the output layer where the softmax activation function is used. As anticipated, the inputs of the CNN are no longer the feature vectors, but the unprocessed samples of both the SPR and RR signals recorded from the subject at the same moment, in time intervals 15 s long. For each time interval, the values of the SPR and the RR signals must be arranged side by side. Since each 15 s long time interval is made of N = 1500 samples, with a sample rate of 100 Sa/s for both the SPR and RR signals, we obtain an input matrix with a size of 1500×2, to be sent as input to the network. We proceed picking a new interval every 5 seconds and applying the same procedure. The labels corresponding to each interval are the same already derived for the ML algorithms. Other details about how the CNN model is defined are the following: we use the Adam optimizer, the learning rate has been set equal to 0.001, the categorical cross-entropy has been chosen as loss function, the number of epochs has been set equal to 20 and the batch-size to 64 (https://keras.io).

Another DL algorithm, i.e., an LSTM network, has been used in the experiments. This is a particular type of Recurrent Neural Network (RNN) that overcomes the issue of the short-term memory of the RNNs (also known as the vanishing gradient problem). The LSTM internal gates keep the relevant information in a long sequence of data, throwing away the non-relevant data, and this can help us recognizing stress with a long dynamic temporal behavior. The architecture of the implemented LSTM network is reported in Table 2. The same settings introduced before for the CNN (the learning rate, the number of epochs, etc.) have been used for the LSTM as well, with batch-size 128. Regarding this network, we also consider a variant of it which uses the hyperopt optimization package (it will be denoted as LSTM2 in Section 4). In this case the number of nodes (ranging from 16 to 128) and the dropout value (varying between 0 and 1), for each layer, have been computed by the optimization package.

Table 1: Architecture of the implemented CNN

CNN	
Convolutional layer	Filters = 20, Kernel size = 10, Stride = 2
Exponential Linear Units (ELU)	Alpha = 0.1
Batch norm. + dropout (0.15)	
Max pooling	Pool size = 2, Stride = 2
Convolutional layer	Filters = 20, Kernel size = 10, Stride = 2
Exponential Linear Units (ELU)	Alpha = 0.1
Batch norm. + dropout (0.15)	
Max pooling	Pool size = 2, Stride = 2
Flatten	
Dense + dropout (0.5)	Nodes = 128, Activation = ReLU
Dense	Nodes = 2, Activation = Softmax

We employ the "leave-one-person-out" technique for all of the learning algorithms, such that the training of each classifier is performed using the data of all individuals, except one, on which the classifiers will be tested. This process is carried out for each individual, and in the end the overall system performance is determined by computing the average of the results originated from all of the individuals.

Table 2: Architecture of the implemented LSTM network

LSTM	
LSTM layer	Nodes = 128, Activation = ReLU
Dropout	0.4
LSTM layer	Nodes = 128, Activation = ReLU
Dropout	0.4
Dense	Nodes = 2, Activation = Softmax

3 Experimental setup

The experiments, as introduced above, are carried out in a firm that develops professional driving simulators, both in hardware and software. For our tests, we use a dynamic driving simulator that allows the reproduction of the same movements of a real car. In fact, it provides motion feedback to the drivers through a nine degrees-of-freedom moving platform. We enrolled 16 healthy individuals, 22-47 years old, some from the University of Udine and some from the University of Padua. The subjects were asked to drive on a 28 km long straight highway track, with four stress-inducing obstacles demanding some concentration to get over. These four obstacles are shown in Figure 3. From left to right and from top to bottom, they are: Sponsor block (from right to left), Tire labyrinth, Double lane change (right to left) and Sponsor block (from left to right).

All subjects prepare themselves before entering the simulator. They wear the sensing devices both on chest and hands. They gave consent having their physiological signals recorded, and the test were performed according to the principles of the Declaration of Helsinki.

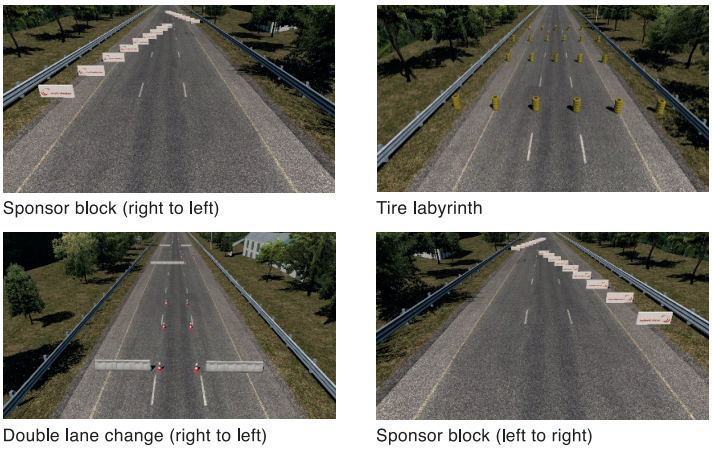


Figure 3: The four stress-inducing events located along the road.

Figure 4 displays the cleaned SPR signal, after the MA removal block, and the instantaneous HR signal of a subject (derived from the RR signal) during the simulation. The start of an event is signaled with a vertical line with a circle marker, and the end of an event with a vertical line with a square marker. We also assume an event starts when the obstacle enters the field of view of the subject (and this happens approaching a distance of 800 meters from the obstacle), and ends 30 seconds after surpassing the obstacle. Passing each obstacle takes approximately 65 seconds. The peculiar stress peaks shape in the SPR signal during the stress event can be seen evidently in the figure, just as the distinctive rise in Heart Rate.

4 Experimental Results

The classifiers, which allow the recognition of stress in each 15 s long interval, are built using the cleaned SPR signal and the ECG signal of each subject. For the traditional ML algorithms the features of the signals are extracted from each interval. For the DL algorithms there are no features to be extracted, and the raw signals themselves are considered for each interval. In all cases, the intervals overlap each other by 10 s as already specified.

We assume that all the intervals happening in a stretch of the road where obstacles are present, to have the value "1", with stress, and that all the intervals happening in a stretch of the road without obstacles to have the value "0", with no stress. As already mentioned, we employ the "leave-one-person-out" method in the classification process. So, for the training, we examine 15 subjects at a time out of the 16 in total. In this way, we obtain a total number of intervals ranging from 1176 to 1209, for both stress and non-stress classes. The total number of intervals for training is not fixed because each individual overcomes the obstacles with slightly different timing and speed. During the test we obtain approximately 187 intervals to analyze for each individual (this is an average value taking all of the subjects into consideration), where the majority of them, about two thirds, belong to the non-stress class.

According to the assumed ground-truth, the True Positives (TP), the False Negatives (FN), the True Negatives (TN) and the False Positives (FP) can be computed, as well as the performance indicators like Accuracy, Sensitivity and Specificity:

$$\text{Accuracy } (\%) = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \cdot 100 \tag{4}$$

$$\text{Sensitivity } (\%) = \frac{\text{TP}}{\text{TP+FN}} \cdot 100 \tag{5}$$

$$\text{Specificity } (\%) = \frac{\text{TN}}{\text{FP+TN}} \cdot 100 \tag{6}$$

Due to the fact that our test data are unbalanced (in fact we have less stress intervals than non-stress intervals), two additional indicators are also important, the Balanced Accuracy (BA) and the Geometric Mean (GM), defined as

$$\text{BA } (\%) = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP+FN}} + \frac{\text{TN}}{\text{FP+TN}}\right) \cdot 100 \tag{7}$$

$$\text{GM } (\%) = \sqrt{\frac{\text{TP}}{\text{TP+FN}} \cdot \frac{\text{TN}}{\text{FP+TN}}} \cdot 100 \tag{8}$$

Table 3 presents the overall performance of our system (MEAN ± STD), calculated by averaging the results of the 16 subjects, for all of the considered ML algorithms. The accuracy and balanced accuracy values among the classifiers are close, even if the ANN classifier is slightly better than the other ones in terms of accuracy, and the RF in terms of balanced accuracy. Moreover, the sensitivity of the RF is higher than the others, as it is the specificity of the Naïve Bayes classifier. The GM values are similar for all of the ML algorithms, with the RF classifier still performing marginally better than the others.

As introduced before, as a first attempt to use a deep learning algorithm to recognize stress and non-stress episodes in our scenario, we set up a 1D CNN and an LSTM. We compute the overall performance of the CNN and the LSTM using the test set, and we obtain the values reported in Table 3. So, comparing the results of the DL algorithms with the ones obtained applying the aforementioned ML algorithms, we can notice that the CNN architecture performs similarly to the other ML architectures. In particular, the accuracy of the ANN and the specificity of the Naïve Bayes classifier exceed the values corresponding to the CNN classifier, whereas the CNN sensitivity is better than that obtained with the other ML classifiers. However, when looking at the LSTM, we can see that it works well in detecting the stress episodes, outperforming all the other algorithms in terms of accuracy, sensitivity, balance accuracy and geometric mean. The specificity of the LSTM, instead, is lower than the Naïve Bayes one. The variant LSTM2 performs even better, and all of the performance indicators are higher than the other DL and ML algorithms.

The overall performance of the system looks favourable. It should be noted that some of our hypotheses are critical and can affect the performance indicators. The first hypothesis we make is that all the time intervals happening during the overcoming of an obstacle by a driver should originate stress, even if it is possible that the driver is not constantly stressed during the manoeuvre. The second one is that a driver is not stressed when not overcoming an obstacle. This might not be true at all times, since a driver can be in a condition of stress during those time intervals for causes unknown to us.

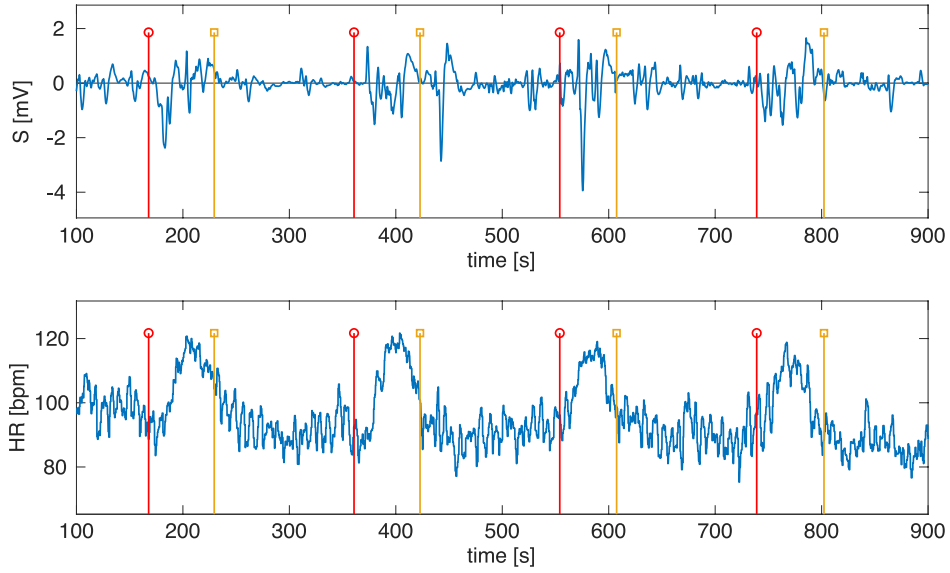


Figure 4: SPR and HR signals (at the top and the bottom, respectively) for a subject during the driving simulation. Each event onset is evidenced with a red vertical line with a circle marker, and each event offset with a yellow vertical line with a square marker.

Table 3: Performance of all the supervised learning algorithms

	Accuracy (%)	Sensitivity (%)	Specificity (%)	BA (%)	GM (%)
SVM	72.86 ± 8.77	86.69 ± 7.79	66.46 ± 13.02	76.57 ± 6.91	75.40 ± 7.85
Naïve Bayes	74.04 ± 6.53	82.83 ± 8.04	70.27 ± 10.85	76.55 ± 4.65	75.81 ±5.01
k-NN	72.07 ± 6.01	86.60 ± 8.30	65.99 ± 8.77	76.30 ± 4.96	75.26 ± 5.23
DT	70.76 ± 8.88	85.16 ± 6.61	64.80 ± 13.13	74.98 ± 5.98	73.69 ± 7.20
RF	74.25 ± 6.17	87.05 ± 7.60	68.83 ± 8.83	77.94 ± 5.03	77.11 ± 5.35
ANN	74.45 ± 8.65	86.19 ± 8.13	69.00 ± 13.38	77.59 ± 6.50	76.56 ± 7.30
CNN	73.97 ± 6.23	87.99 ± 6.40	67.76 ± 8.30	77.87 ± 4.82	76.99 ± 5.27
LSTM	82.56 ± 3.04	88.61 ± 7.55	69.79 ± 17.61	79.20 ± 5.96	77.46 ± 9.23
LSTM2	88.02 ± 0.97	93.60 ± 1.92	76.26 ± 4.22	84.93 ± 1.55	84.43 ± 1.78

Another possible critical aspect of our hypotheses is that part of the signal responses may be due to the increased physical effort during the events. However, in some previous works, e.g., [40], we observed that the intensities of the signals during designated stress events were significant, and could not be confused with the levels recorded during non-stress intervals, even if a comparable physical activity was required for driving. In [41], we also evaluated a car driver's stress condition in a simulated autonomous driving scenario. In that experiment, we submitted questionnaires to the drivers and evaluated other parameters such as the User Experience (UX) measure and the NASA Task Load Index. Questionnaires supported our findings derived by analyzing the SPR and ECG signals (see also [42]).

To account for these considerations, and enhance the stress recognition ability, we apply a re-label step to the output of the classifiers (see Figure 5). To do so, we assume stress is happening if a minimum of four consecutive intervals are marked as positive (thus covering a time span of 30 s). In the end, the blocks of four or more consecutive positive intervals are kept, whereas blocks of up to three consecutive positive intervals are instead re-labelled as negative.

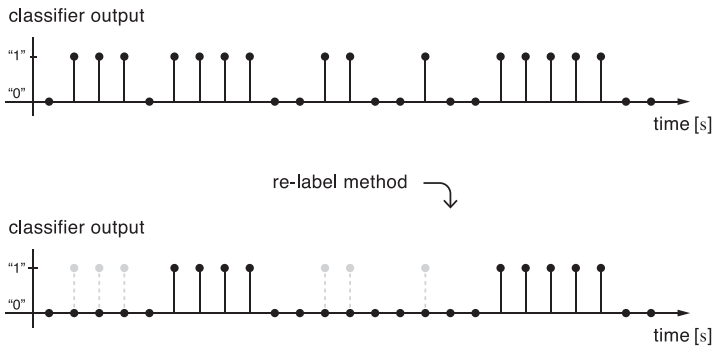


Figure 5: The re-label method applied to the output of each classifier.

In Table 4 the performance indicators applying the re-label method are presented. The values are slightly higher than the ones in Table 3. On the other hand, sensitivity is a little smaller for all of the classifiers, since we noticed that a number of stress intervals, which were correctly detected before, are now discarded by the re-label step. The performance indicators of the LSTM2 are clearly higher, when compared to the other algorithms, in this scenario as well.

Table 4: Performance of all the supervised learning techniques when the re-label method is applied to the output of each classifier

	Accuracy (%)	Sensitivity (%)	Specificity (%)	Balanced Accuracy (%)	Geometric Mean (%)
SVM	77.61 ± 8.14	85.11 ± 8.82	74.05 ± 11.67	79.58 ± 6.89	79.00 ± 7.30
Naïve Bayes	77.39 ± 6.69	82.83 ± 8.04	70.27 ± 10.85	76.55 ± 4.65	75.81 ±5.01
k-NN	75.86 ± 5.84	83.85 ± 11.04	72.61 ± 8.95	78.23 ± 5.47	77.56 ±5.63
DT	76.17 ± 7.87	82.42 ± 8.09	73.69 ± 11.41	78.06 ± 5.85	77.50 ±6.16
RF	77.15 ± 6.39	86.16 ± 7.42	73.32 ± 8.89	79.74 ± 5.35	79.22 ±5.59
ANN	78.17 ± 9.13	84.73 ± 9.72	75.16 ± 14.23	79.94 ± 6.88	79.15 ± 7.50
CNN	79.47 ± 6.45	85.31 ± 8.33	77.09 ± 8.48	81.20 ± 5.79	80.87 ± 5.90
LSTM	83.21 ± 3.27	88.23 ± 7.78	72.62 ± 17.39	80.42 ± 5.94	78.91 ± 9.00
LSTM2	88.13 ± 0.87	93.30 ± 1.92	77.20 ± 4.05	85.25 ± 1.42	84.83 ± 1.60

5 Conclusions

We presented a scheme to recognize individuals' stress while driving in a car driving simulator. The proposed system uses relatively simple sensors, with an acceptably intrusive setup. Due to the short signal block analysis, it can be used nearly in real time and can discriminate stressful events of short duration. In our system, we record two SPR signals for each individual, one for each hand, and the ECG signal. We then apply an MA removal algorithm to remove the artifacts that could appear in the SPR signals during the drive, so that the output is a single cleaned SPR signal. Some features are extracted from both the cleaned SPR signal and the ECG, and sent as input to an ML algorithm. The raw signals are instead directly used as input to DL algorithms. In particular, we evaluate the performance of an SVM classifier, a Naïve Bayes classifier, a *k*-NN classifier, a DT, an RF classifier and an ANN. The performance of all the ML classifiers are similar one to each other, however, the ANN classifier performs better than the others in terms of accuracy. The RF classifier outperforms the others in terms of sensitivity, balanced accuracy and geometric mean, while the Naïve Bayes classifier in terms of specificity. We also report the results obtained with a 1D CNN and an LSTM. The CNN outperforms the other ML classifiers in terms of sensitivity. The LSTM, when some parameters are selected by the optimization procedure, outperforms all the other learning algorithms. A re-labelling procedure is also considered, and the performance indicators of all of the algorithms are compared again in this case. We obtain the highest accuracy, equal to 88.13%, using the optimized LSTM.

Some critical aspects of the proposed experiment are due to the assumption that stress is present only during the arranged events, and that part of the signal responses may be due to the increased physical effort while driving through the events. We have observed that these issues still allow us to interpret the results with a good level of confidence. We conclude that the overall performance of our system is encouraging, showing that by using the SPR and ECG signals we can identify the stress in drivers, deriving from some demanding driving activities, with acceptable accuracy.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment We are grateful to VI-grade S.r.l., Tavagnacco (UD), Italy, and in particular to Fabio Formaggia, Diego Minen, Michela Minen, and Carlo Savorgnan, for their assistance in data acquisition with their professional driving simulator.

References

[1] M. Navon, O. Taubman-Ben-Ari, "Driven by emotions: The association between emotion regulation, forgivingness, and driving styles," Transportation Research Part F: Traffic Psychology and Behaviour, **65**, 1–9, 2019, doi:10.1016/j.trf.2019.07.005.

[2] C. D. Katsis, N. Katertsidis, G. Ganiatsas, D. I. Fotiadis, "Toward emotion recognition in car-racing drivers: A biosignal processing approach," IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, **38**(3), 502–512, 2008, doi:10.1109/TSMCA.2008.918624.

[3] W. Boucsein, Electrodermal Activity, Springer, 2012.

[4] S. Greene, H. Thapliyal, A. Caban-Holt, "A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health," IEEE Consumer Electronics Magazine, **5**(4), 44–56, 2016, doi:10.1109/MCE.2016.2590178.

[5] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, J.-H. Chen, "EEG-based emotion recognition in music listening," IEEE Transactions on Biomedical Engineering, **57**(7), 1798–1806, 2010, doi:10.1109/TBME.2010.2048568.

[6] J. Zhang, W. Wen, F. Huang, G. Liu, "Recognition of Real-Scene Stress in Examination with Heart Rate Features," in 2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, 2017, doi:10.1109/IHMSC.2017.13.

[7] H. Becker, J. Fleureau, P. Guillotel, F. Wendling, I. Merlet, L. Albera, "Emotion recognition based on high-resolution EEG recordings and reconstructed brain sources," IEEE Transactions on Affective Computing, **11**(2), 244–257, 2020, doi:10.1109/TAFFC.2017.2768030.

[8] Z. Qin, M. Li, L. Huang, Y. Zhao, "Stress level evaluation using BP neural network based on time-frequency analysis of HRV," in 2017 IEEE International Conference on Mechatronics and Automation (ICMA), Takamatsu, 2017, doi:10.1109/ICMA.2017.8016090.

[9] M. Moghimi, R. Stone, P. Rotshtein, "Affective recognition in dynamic and interactive virtual environments," IEEE Transactions on Affective Computing, **11**(1), 45–62, 2020, doi:10.1109/TAFFC.2017.2764896.

[10] C. Maaoui, A. Pruski, F. Abdat, "Emotion Recognition for Human-Machine Communication," in 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, Nice, 2008, doi:10.1109/IROS.2008.4650870.

[11] G. Giannakakis, E. Trivizakis, M. Tsiknakis, K. Marias, "A novel multi-kernel 1D convolutional neural network for stress recognition from ECG," in 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), Cambridge, United Kingdom, 2019, doi:10.1109/ACIIW.2019.8925020.

[12] K. T. Chui, M. D. Lytras, and R. W. Liu, "A generic design of driver drowsiness and stress recognition using MOGA optimized deep MKL-SVM," Sensors, **20**(5):1474. 2020, doi:10.3390/s20051474.

[13] L. Lombardi, F. Marcolin, Psychological Stress Detection by 2D and 3D Facial Image Processing, Progresses in Artificial Intelligence and Neural Systems, Springer, 2020.

[14] M. T. Uddin, S. Canavan, "Synthesizing Physiological and Motion Data for Stress and Meditation Detection," in 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), Cambridge, United Kingdom, 2019, doi:10.1109/ACIIW.2019.8925245.

[15] A. Oskooei, S. Mai Chau, J. Weiss, A. Sridhar, M. Rodríguez Martínez, B. Michel, "DeStress: deep learning for unsupervised identification of mental stress in firefighters from heart-rate variability (HRV) data," arXiv:1911.13213v1, 2019.

[16] B. Hwang, J. You, T. Vaessen, I. Myin-Germeys, C. Park, B.-T. Zhang, "Deep ECGNet: An optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals," Telemedicine and e-Health, **24**(10), 753–772, doi:10.1089/tmj.2017.0250.

[17] F. Al Machot, A. Elmachot, M. Ali, E. Al Machot, K. Kyamakya, "A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors," Sensors, **19**(7):1659, 2019, doi:10.3390/s19071659.

[18] T. Kundu, C. Saravanan, "Advancements and recent trends in emotion recognition using facial image analysis and machine learning models," in 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Mysuru, 2017, doi:10.1109/ICEECCOT.2017.8284512.

[19] Y. Ishii, T. Ogitsu, H. Takemura, H. Mizoguchi, "Real-time eyelid open/closed state recognition based on HLAC towards driver drowsiness detection," in 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO), Bali, 2014, doi:10.1109/ROBIO.2014.7090707.

[20] D. S. Lee, T. W. Chung, B. G. Lee, "Stress events detection of driver by wearable glove system," IEEE Sensors Journal, **17**(1), 194–204, 2017, doi:10.1109/JSEN.2016.2625323.

[21] Y. J. Yu, Z. Yang, B.-S. Oh, Y. K. Yeo, Q. Liu, G.-B. Huang, Z. Lin, "Investigation on driver stress utilizing ECG signals with on-board navigation systems in use," in 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV), Phuket, 2016, doi:10.1109/ICARCV.2016.7838780.

[22] D. Tomoi, W. Wen, H. Yamakawa, S. Hamasaki, K. Takakusaki, Q. An, Y. Tamura, A. Yamashita, H. Asama, "Estimation of stress during car race with factor analysis," in 2015 International Symposium on Micro-NanoMechatronics and Human Science (MHS), Nagoya, 2015, doi:10.1109/MHS.2015.7438258.

[23] G. Rigas, Y. Goletsis, D. I. Fotiadis, "Real-time driver's stress event detection," IEEE Transactions on Intelligent Transportation Systems, **13**(1), 221–234, 2012, doi:10.1109/TITS.2011.2168215.

[24] J. A. Healey, R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," IEEE Transactions on Intelligent Transportation Systems, **6**(2), 156–166, 2005, doi:10.1109/TITS.2005.848368.

[25] D.-W. Koh, S.-G. Lee, "An evaluation method of safe driving for senior adults using ECG signals," Sensors, **19**(12):2828, 2019, doi:10.3390/s19122828.

[26] Z. Halim, M. Rehan, "On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning," Information Fusion, **53**, 66–79, 2020, doi:10.1016/j.inffus.2019.06.006.

[27] H.-M. Cho, H. Park, S.-Y. Dong, I. Youn, "Ambulatory and laboratory stress detection based on raw electrocardiogram signals using a convolutional neural network," Sensors, **19**(20):4408, 2019, doi:10.3390/s19204408.

[28] M. N. Rastgoo, B. Nakisa, F. Maire, A. Rakotonirainy, V. Chandran, "Automatic driver stress level classification using multimodal deep learning," Expert Systems with Applications, **138**:112793, 2019, doi:10.1016/j.eswa.2019.07.010.

[29] L. G. Hernández, O. M. Mozos, J. M. Ferrández and J. M. Antelis, "EEG-based detection of braking intention under different car driving conditions," Frontiers in Neuroinformatics, **12**:29, 2018, doi:10.3389/fninf.2018.00029.

[30] P. Zontone, A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, F. Formaggia, D. Minen, M. Minen, C. Savorgnan, "Car driver's sympathetic reaction detection through electrodermal activity and electrocardiogram," IEEE Transactions on Biomedical Engineering, 1–12, 2020, doi:10.1109/tbme.2020.2987168.

[31] A. Affanni, G. Chiorboli, "Design and characterization of a real-time, wearable, endosomatic electrodermal system," Measurement, **75**, 111–121, 2015, doi:10.1016/j.measurement.2015.07.047.

[32] A. Affanni, "Wearable instrument to measure simultaneously cardiac and electrodermal activities," in 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Benevento, Italy, 2016, doi:10.1109/MeMeA.2016.7533749.

[33] J. Pan, W. J. Tompkins, "A real-time QRS detection algorithm," IEEE Transactions on Biomedical Engineering, **BME-32**(3), 230–236, 1985, doi:10.1109/TBME.1985.325532.

[34] W. Chen, G. Liu, S. Su, Q. Jiang, and H. Nguyen, "A CHF detection method based on deep learning with RR intervals," 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, 2017, doi:10.1109/EMBC.2017.8037578.

[35] M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-Aho, P. A. Karjalainen, "Kubios HRV – Heart rate variability analysis software," Computer Methods and Programs in Biomedicine, **113**(1), 210–220, 2014, doi:10.1016/j.cmpb.2013.07.024.

[36] A. Affanni, A. Piras, R. Rinaldo, P. Zontone, "Dual channel Electrodermal activity sensor for motion artifact removal in car drivers' stress detection," in 2019 IEEE Sensors Applications Symposium (SAS), Sophia Antipolis, France, 2019, doi:10.1109/SAS.2019.8706023.

[37] P. Zontone, A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, "Low-complexity classification algorithm to identify drivers' stress using electrodermal activity (EDA) measurements," Biomedical Engineering and Computational Intelligence, **20**, 25–33, 2020, doi:10.1007/978-3-030-21726-6_3.

[38] M. Malik, "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use," European Heart Journal, **17**(3), 354–381, 1996, doi:10.1161/01.CIR.93.5.1043.

[39] J. Bergstra, D. Yamins, D. D. Cox, "Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms," in 2013 12th Python in Science Conference (SCIPY), Austin, Texas, 2013, doi:10.1088/1749-4699/8/1/014008.

[40] A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, P. Zontone, "Driver's stress detection using skin potential response signals," Measurement, **122**, 264–274, 2018, doi:10.1016/j.measurement.2018.03.040.

[41] P. Zontone, A. Affanni, R. Bernardini, L. Del Linz, A. Piras, R. Rinaldo, "Stress evaluation in simulated autonomous and manual driving through the analysis of skin potential response and electrocardiogram signals," Sensors, **20**(9):2494, 2020, doi:10.3390/s20092494.

[42] A. Affanni, P. Zontone, R. Fenici, D. Brisinda, D. Bacchin, L. Gamberini, P. Pluchino, M. Bruschetta, C. Savorgnan, F. Formaggia et al., "Assisted / autonomous vs. human driving assessment on the DiM driving simulator using objective / subjective characterization," in 2019 10th International Munich Chassis, Munich, Germany, 2019, Pfeffer P. (eds), Springer Vieweg, Wiesbaden, 307–321, 2020, doi:10.1007/978-3-658-26435-2_23.

[43] P. Zontone, A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, "Stress Detection Through Electrodermal Activity (EDA) and Electrocardiogram (ECG) Analysis in Car Drivers," in 2019 European Signal Processing Conference (EUSIPCO), A Coruña, Spain, 2019, doi:10.23919/EUSIPCO.2019.8902631.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 30-35 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Analysis of Green Building Effect on Micro grid Based on Potential Energy Savings and BIM

Ihsan Mizher Baht[*,1], Petre Marian Nicolae[1], Ileana Diana Nicolae[2], Nameer Baht[3]

[1]*Department of Electrical engineering, Craiova University, Craiova, 200396, Romania*

[2]*Department of Computer Science and IT, Craiova University, Craiova, 200396, Romania*

[3]*Department of Computer Science, Universities west of Timisoara, Timisoara, 300088, Romania*

ARTICLE INFO

Article history:
Received: 28 August, 2020
Accepted: 08 October, 2020
Online: 08 November, 2020

Keywords:
Energy Management Systems
Micro Grids
Green Buildings
Optimization Software

ABSTRACT

In this paper we are demonstrating the importance of building and achieving a renewable electrical power system that works in synchronization with the grid. The average of consumption, generation and the negative environmental effect of generating power were analyzed. This paper is divided into three parts, one was analyzed with Green Building Studio software and the other with Building Information software. Using BIM software, the levels of energy consumption were analyzed for a building in Baghdad – Iraq and in Craiova _ Romania, alongside pollution levels produced, and the possibility of energy generation utilizing PV cells and wind, the output power that can be achieved in respect to temperature and irradiance levels. Using Green Building software, energy consumption system was analyzed in the level of green building.

1. Introduction

The concept of power and energy management refers to the use of highly effective and reliable blueprints of managing the available power and energy in such a way that the outcomes are optimized while also preventing wastage. Despite the fact that there are many factors and issues that cause power and energy loss, the absence of an effective management blueprint is one of the main contributors [1]. When an effective blueprint is absence, it then follows that the framework of power loss is heighted. Following is an evaluation of the distinctive factors and issues affecting the management of power and energy. It is vitally important to accentuate the fact that there are numerous issues and factors that impact the management of power and energy. While some of the factors are preventable in the short-term, others can only be addressed in the long-term. Unless these factors are identified in the first place, the blueprint for management of power and energy cannot be enhanced. Over the years, many problems have characterized power and energy systems in different parts of the world because such systems were not aligned towards an effective management blueprint [2]. There is thus the need to identify the distinctive issues and factors that play a direct role in the framework of power and energy management.

[*]Corresponding Author: Ihsan Baht, Contact No. +9647705808038 Email: imb121378@yahoo.com.

2. Factors Effecting Power and Energy Management

As far as the factors affecting the management of power and energy are concerned, one of the most important issues of emphasis pertains to location. Depending on the location, a given blueprint for management of power and energy can either be effective or ineffective [3]. In line with this attribute, there is always the need to ensure that the process of management of power and energy is strongly aligned towards the specific needs and priorities of a given location. Such an approach is essential towards the prevention of potential problems such as power and energy loss. The factors affecting the framework of management of power and energy also encompass the source of power. Different sources of power are characterized by distinctive attributes and issues. It then follows that the specific source must be determined in order to develop a framework of management of power and energy that is effective in the short-term as well as in the long-term [4]. A notable way in which the source of power impacts the framework of management of power and energy is that it determines whether a given management strategy can help in the reduction of energy wastage or not. This goes a long way towards determining the quality of outcomes in the long-term as well as in the short-term.

Competence of personnel is a major issue that determines the approach used within the context of management of power and

energy. For instance, the competence of staff members in a power grid determines the manner in which they approach decision making processes concerning power management [5]. Additionally, it is also notable that the competence of personnel goes a long way towards determining whether they understand the technical elements of management of power and energy or not.

As far as the factors influencing the blueprint of management of power and energy are concerned, another notable aspect of emphasis pertains to the availability of resources. In some instances, the scarcity of resources can compromise the extent to which the management of power and energy is aligned towards the intended goals. This is because there are extensive resource requirements at any stage of managing energy and power. It is also vitally important to highlight the fact that the availability of resources impacts the manner in which planning is actualized into the blueprint for management of power and energy [6]. In view of these stipulations, it is hence evident that the availability of resources is always a notable aspect of emphasis that determines the manner in which the blueprint for management of power and energy is actualized. The engineering vision of analyzing and constructing green buildings is a foundation for energy management such as to guarantee the achieving of great economic benefits [7, 8].

3. Building Performance Analysis (BPA) and Optimization Energy (OE) in Green Building Studio (GBS)

The Green building studio (GBS) which powers the energy analysis through (cloud-based service), the GBS with Insight 360 Revit Autodesk is able to test daily and weekly then make review, comparing with the previous results, there are two types of testing in green building studio (GBS):

Service side: energy results from GBS compare against the baseline. Analytics: written the energy results as simulation and compared with the baseline. The green building studio analyze the performance of the building as total annual energy consumption (electrical, fuel, CO2), total annual cost and potential energy savings.

The GBS analyzes automatically 37 separate energy simulations that can run in the cloud through the weather data station (in Baghdad use for analyzing the weather station GBS_06M12_12_002295).

The performance factors of analysis are:

- Light efficiency (from 0.7w/sf-1.9w/sf)
- Plug load efficiency (from 0.6w/sf-2.6w/sf)
- Building orientation (0 degree, 135 degree….etc.)
- Operations schedule (24/7, 12/7, 12/5 ….etc.).
- Daylighting occupancy and non-control.
- Infiltration load.
- HVAC.
- Other (civil and mechanical parameters).

3.1. Green Building Analysis Based On Bulding Orientation

The first effective factor before creating the green building, the Building orientations, selects the optimum degree to the building construction to provide the best energy use intensity, total annual cost and total annual energy. Table 1 represents the orientation of a green building in Baghdad city.

Table 1: The Total Annual Energy Cost (Building Orientations)

Degree	EUI MJ/M^2 /YE.	Total annual cost $		
		Electrical	Fuel	Energy
0	1030	7233	120	7354
135	1054	7201	130	7332
180	1026	6752	136	6889
225	985	6401	134	6535
270	1024	6976	127	7104
315	1016	7057	121	7178
45	1012	6982	123	7106
90	1060	7346	127	7473

From the Table 1 at 225 degree building orientation, the energy use intensity (EUI) 985 MJ/m^2 /year, total annual cost electric 6401 and the energy cost for one year 6535. The solar irradiance, temperature, flow air, infiltrations etc. at 225 degree is the optimum point for construction of green building to save the cost of energy and fuel.

Table 2 represents the annual energy consumptions at a green building with the different degree orientation at the same location. At 225 degree the total annual energy consumption in kWh 46.590(minimum consumptions), and at 90-degree orientation 53.465(maximum consumptions)

Table 2: The Total Annual Energy Consumptions (Building Orientations)

Degree	Total annual energy	
	Electrical kWh	Fuel MJ
0	52.649	87.929
135	52.416	95.127
180	49.146	99.485
225	46.590	97.616
270	50.777	93.077
315	51.362	88.650
45	50.822	89.573
90	53.465	92.914

From the results (Table 1, Table 2) the minimum energy consumption, minimum EUI, and minimum energy fuel cost is at the 225 degree. Green Building Studio (GBS), Insight 360 is able to analyze the weather effect on the building to provide the best benefit from the environment and the effects of climate, wind …etc.

The parameters analysis extensions a scope of possibilities to find very active building as a green building, for example the climate can determine the type and capacity of the lighting installation depending on the degree of the building orientations.

Figure 1 and Figure 2 Represent the total annual energy consumption MJ and total annual electrical consumption. The Y-axis total energy (MJ), X-axis time (year) for example in Jan the red bar shows the space heat ,yellow bar area light, blue bar space cooling.

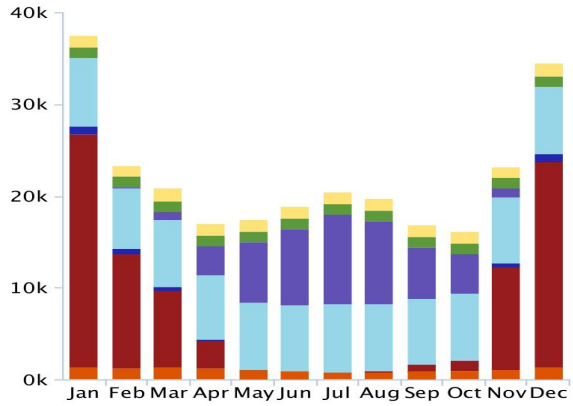


Figure 1: The Total Annual Energy Consumption (Mj)

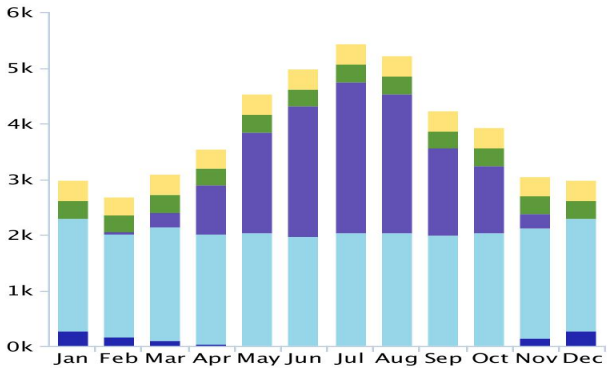


Figure 2: The Total Annual Electrical Consumption (Kwh)

3.2. Green Building Analysis Based On Daylighting

The second factor effect on the energy consumption in green building is daylighting analysis, the daylighting analysis is divided into four parts:

- Daylighting and occupancy control daylighting & occupancy control.
- Daylighting and occupancy control daylighting control.
- Daylighting and occupancy control none.
- Daylighting and occupancy control & occupancy control.

Table 3 presents the amount of total annual cost and total energy consumption for the four parts of daylights control regarding to GBS weather station (cloud-based service).

Table 3: The Amount of Total Annual Cost and Total Energy Consumption

		Total Annual Cost $			Total Annual Consumptions Energy	
Day Light	Eui Mj/M²/Y	Ele.	Fue.	Ener.	Elec. Kwh	Fuel Mj
1	1016	7.020	123	7.143	51.04	89.58
2	1021	7.105	122	7.227	51.710	88.89
3	1030	7.233	120	7.354	52.649	87.92
4	1023	7128	121	7250	51.883	88.74

From the Table 3 the green building studio analysis improves the maximum energy potential savings in Daylighting and occupancy control daylighting & occupancy control.

Figure. 3 and Figure. 4 represent the total energy (MJ), electrical consumption kwh for Daylighting and occupancy control daylighting & occupancy control for the one-year period.

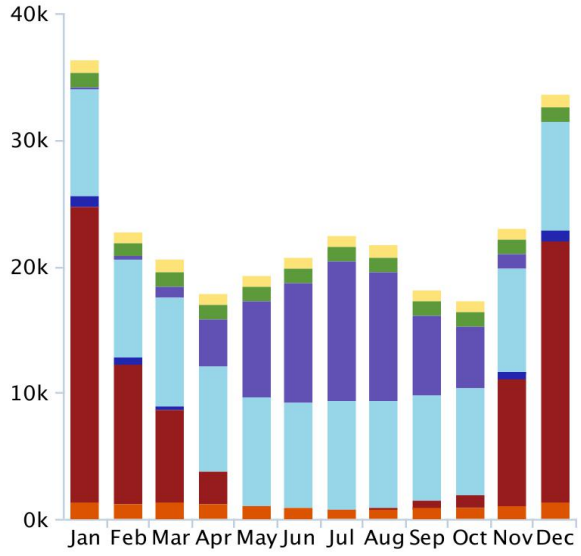


Figure. 3: The Total Energy (Mj)

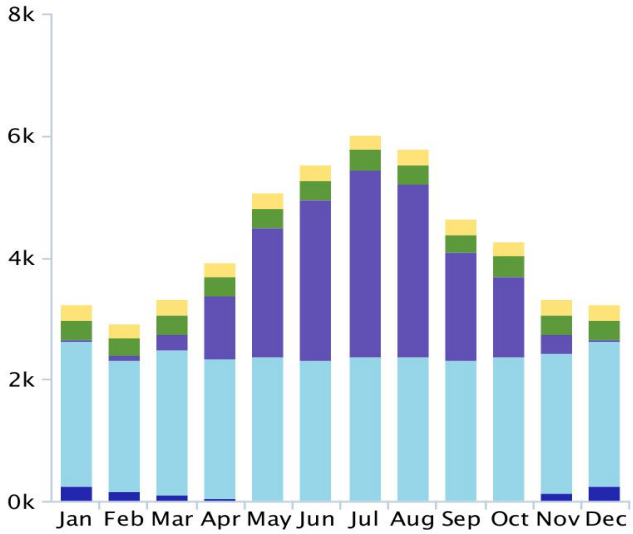


Figure 4: Electrical Consumption (Kwh)

3.3. Green Building Analysis Based On potantial energy (cost)savings

For the design of a green building and an existing building analysis, the reduce of the energy used is one of the main targets of the GBS, potential energy savings (PES) analyze all buildings energy performance (more than 37 separate energy simulations).

For the photovoltaics analysis (in Baghdad, weather station GBS_06M12_12_002295) at 225-degree orientation building, installed on the roof of the building, the panel type single crystalline -13.8 efficiency, installed panel cost 8$ per watt, applied electric cost 0.14$, the maximum payback period 50 years, installed panel cost 162.266$, installed panel area 147m^2, the annual energy production 36 kWh.

Table 4 presents the potential cost saving and annual energy with payback per surface per year, the annual energy produced depends on two factors, the panel area and solar exposure %, for example at the annual energy produce of 984(kWh), the panel area 4 m^2, solar exposure 69.9%.

Table 4: Potential Cost Saving and Annual Energy with Payback Per Surface Per Year

	Potential cost savings $		
Annual energy kWh	Per year/m^2	Per year	payback per surface per years
984	36.04	133	23.9
2914	36.04	408	23.9
364	36.04	51	23.9
447	36.04	63	23.9
16.703	32.23	2.338	26.1
3.251	32.14	455	26.2
3.207	31.80	499	26.3
3.563	31.8	499	26.3
2.767	31.80	387	26.3

3.4. Implications of Power and energy management

In view of the various factors and issues affecting the process of management of power and energy, there is also the need to identify the distinctive implications of management of power and energy [9]. These implications re essential in that they influence the planning process especially when it comes to decision making as well as allocation of resources [10]. Additionally, clarity of the implications helps in determining the best possible strategy to apply within the confines of management of power and energy.

The specific implications of management of power and energy also encompass sustainability. This refers to the use of energy resources in such a way that the needs of the future generations are considered even while scarce resources are being utilized by current generations [11]. Despite the fact that there are many factors and issues that contribute towards the sustainable use of power, the framework of management is always important. Additionally, the fact that the process enhances the extent to which energy and power is is optimized is also indicative of the manner in which the blueprint for management of power and energy enhances suitability.

3.5. Evaluation of CO2 emissions when only fuel is used for power generation

One of the most important criteria for creating a system that depends on improving the environmental levels is the effect of CO2 gas that is a by-product of electrical energy production.

Figure 5. Depicts the effect of energy production. Fuel consumption in Baghdad – Iraq is up to 6 tons [12] [13].

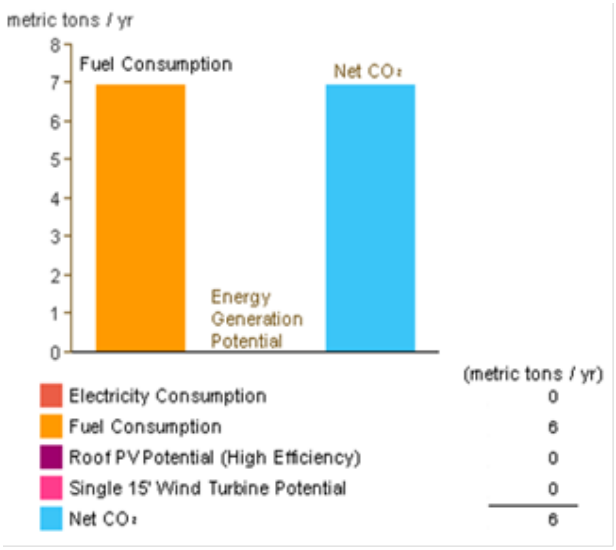


Figure 5: Energy and CO2 Consumptions (Baghdad)

Figure 6 depicts that fuel consumption is 4 tons in the city of Craiova – Romania. The generation system above was analyzed using BIM software for the same area, utilization and population. But the average effect in Craiova is less than Baghdad [12] [13].

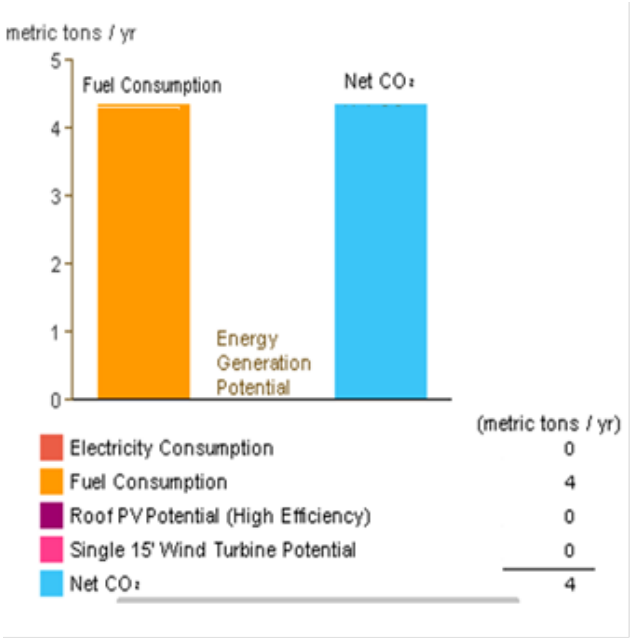


Figure 6: Energy and CO2 Consumptions (Craiova).

The table 5 below states the expected average electrical power generation if renewable energy (PV) is utilized, where these data were analyzed and average temperature and irradiance

were calculated for both cities of Baghdad and Craiova for several types of solar panels and also for wind turbine. The average production in Baghdad is higher than that in Craiova because of the nature of weather and intervals of sun energy availability and higher levels of irradiance [12] [13].

Table 5: Estimations energy produce by Pv cells And Wind

Energy Estimation	**BGW**(kwh/yer)	**CRA.** (kwh/yer)
PV cells (low efficiency)	12,377	9,009
PV cells (medium efficiency)	24,755	18,017
PV cells (high efficiency)	36,132	27,026
wind turbine Single 15’	1,956	691

4. Analysis by using BIM software

BIM software functions through analyzing of buildings and stating the effects of civil and mechanical works on the electrical power consumption. BIM software includes several criteria, in total 26 factors which improve grid performance and decrease consumption. Figure 7 illustrates the average annual energy consumption of the buildings taking into consideration building utilization, for example as a school, hospital or residence.

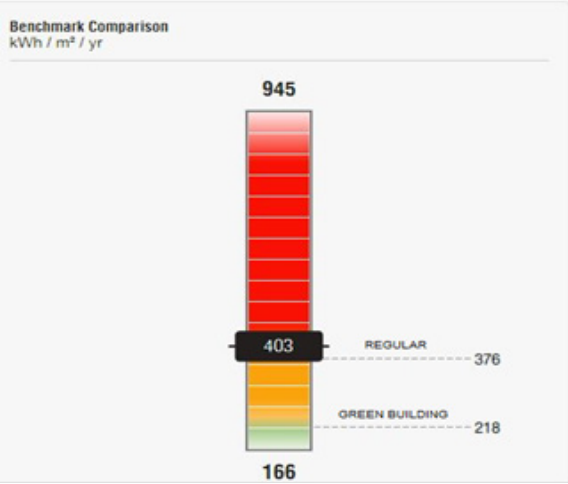


Figure 7: Energy Consumption

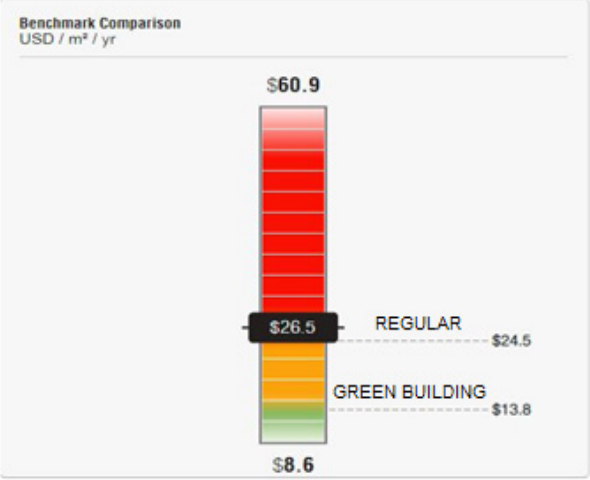


Figure 8: Costs Analysis

We can observe that average energy consumption reaches 376 kW, while when using BIM software analysis system consumption may be 218 kW/m^3 at the same efficiency. Estimated cost changes also from $24.5 to $13.6 (Figure. 8).

Figure 9 represents a comparison in analysis of a building in Baghdad where average consumption is highest through normal analysis, but when Green Building software is utilized as represented in yellow color which reaches a consumption of 310kwh/m^3, and green color when analyzed using BIM software with an annual consumption of up to 120kwh/m^3 [12], [13].

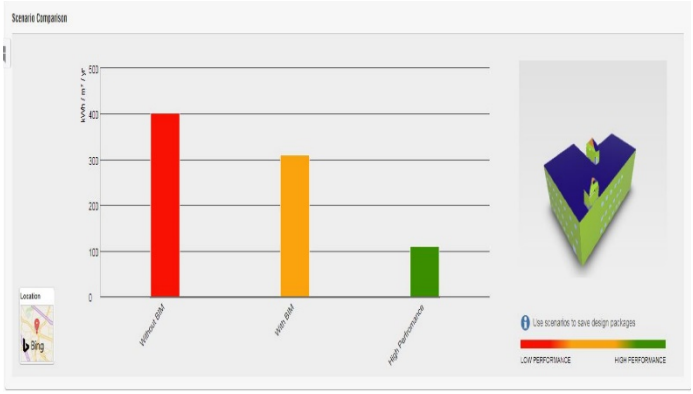


Figure 9: Analysis Performance

5. Conclusions

In this paper, the work was divided into three parts, first part analyzes a building and the effect of weather using Insight software. In the second part an analysis is done via Green Building, where a total cost was calculated to build a solar energy generation system to reduce electrical power consumption, and utilization of building angle which serves to better usage of weather. Third part is an analysis of the systems above. And depicting the economic value that BIM software provides to building efficiency, decrement of consumption, and decrement of costs

References

[1] A. Gulagi , Dmitrii Bogdanov & Christian Breyer, "A Cost Optimized Fully Sustainable Power System for Southeast Asia and the Pacific Rim," Energies, MDPI, Open Access Journal, **10**(5), 1-25, April 2017, doi:10.3390/en10050583. 2017. 10.583 .

[2] T. Dragicevic, J. Guerrero, J.Vasquez, and D. Skrlec, “Supervisory control of an adaptive-droop regulated dc microgrid with battery management capability,” IEEE Trans. on Power Electron., **29**(2), 695–706, 2014. doi: 10.1109/TPEL.2013.2257857.

[3] H. Yousefi, M. Ghodusinejad, " Feasibility Study of a Hybrid Energy System for Emergency Off-grid Operating Conditions. Majlesi Journal Of Electrical Engineering", **11**(3), 7-14 ,2017.

[4] A. Wentland,"Imagining and enacting the future of the German energy transition: electric vehicles as grid infrastructure. Innovation" The European Journal Of Social Sciences, **29**(3), 285-302, 2016, doi: 10.1080/13511610.2016.1159946 .

[5] E. Ganesan., "Modeling, control, and power management for a grid-integrated photo voltaic, fuel cell, and wind hybrid system", Turkish Journal Of Electrical Engineering & Computer Sciences, **24**(6), 4804-4823, 2016 , doi: 10.3906/elk-1404-409.2016.

[6] N. Çetinkaya, "Improving of renewable energy support policy and a performance analysis of a grid-connected 1 mwp pv power plant in Konya Selcuk University" , Journal Of Engineering, Science & Technology / Selçuk Üniversitesi Mühendislik, Bilim Ve Teknoloji Dergisi, **5**(3), 251-261, 2017,doi:10.1273/tused.10159.2016

[7] T. Dragicevic, J. Guerrero, J.Vasquez, and D. Skrlec, “Supervisory control of an adaptive-droop regulated dc microgrid with battery management capability,” IEEE Trans. on Power Electron., **29(**2), 695–706 , 2, doi: 10.1109/TPEL.2013.2257857,2014.

[8] F. Zhang, T. Zhang, R. Wang, "Optimizing Hierarchical Power Distribution of Multiple Local Energy Network Systems in GridConnected Mode”, Polish Journal of EnvironmentalStudies, **26**(5), 1981-2000, doi:org/10.15244/pjoes/69440.2017.

[9] A. Al Ameri , Aouchenni Ounissa , Cristian Nichita & Aouzellag Djamal, "Power Loss Analysis for Wind Power Grid Integration Based on Weibull Distribution," Energies, MDPI, Open Access Journal, **10**(4), 1-16, April, 2017, doi:org/10.3390/en10040463.2017.

[10] A. Lösch, & C. Schneider,"Transforming power/knowledge apparatuses: the smart grid in the German energy transition. Innovation". The European Journal Of Social Sciences, **29**(3), 262-284, 2016, doi:org/10.1080/13511610.2016.1154783.2016.

[11] M. Serraji, El Amine D. O., & Boumhidi J., "Multi swarm optimization based adaptive fuzzy multi agent system for microgrid multi-objective energy management",International Journal of Knowledge-based and Intelligent Engineering Systems, **20**(4), 229-243, 2016, doi: 10.3233/KES-160350.2016.

[12] I. M. Baht, P.M. Nicolae, M. Ş. Nicolae, and Ileana Diana Nicolae, Analysis of Load Demand and Economic Assessment Based on Weather Forecasting in Green Buildings, Annals of the University of Craiova, Electrical Engineering series, **43**(1), 147-152, 2019 .

[13] I.M. Baht, P.M. Nicolae and M. Ş. Nicolae, "Impact of Weather Forecasts and Green Building on Micro Grid Energy Management System", Proceedings of SIELMEN 2019, Chişinău, Oct , POCU380/6/13/123990. 2019.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 36-44 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Effective Segmented Face Recognition (SFR) for IoT

Fei Gao, Jiangjiang Liu*

Department of Computer Science, Lamar University, Beaumont,77710, USA

ARTICLE INFO

Article history:
Received: 01 September, 2020
Accepted: 07 October, 2020
Online: 08 November, 2020

Keywords:
Fully Convolutional Networks (FCN)
Convolution Neural Network (CNN)
Classification
Face Recognition

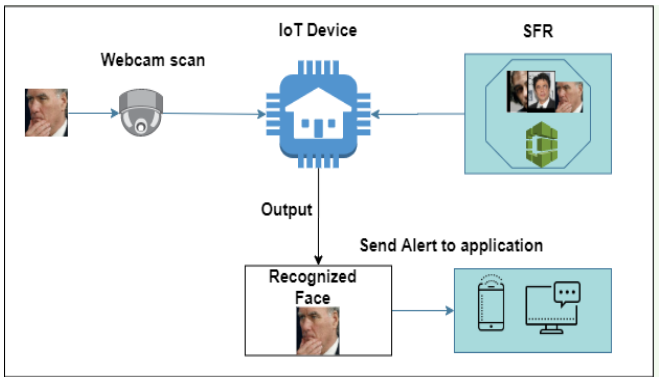
ABSTRACT

Face recognition technology becoming pervasive in the fields of computer vision, image processing, and pattern recognition. However, face recognition accuracy rates will decrease if training is done on disguised images with covered objects on a face area. This paper aims to propose a state-of-the-art face recognition methodology which could be applied in Internet of Things (IoT) devices as an input source; then face segmentation and training process will be executed in the cloud via internet; the recognition result will be sent to the connected applications which determines a safety check for personal or public security. This paper focuses on implementation of face segmentation and training process for IoT. Face extraction from the background and disguised part is applied by Fully Convolutional Networks (FCN), and then deep convolution neural network is employed for face training and testing process. This algorithm has been experimented on a challengeable face dataset. The proposed face recognition system is applied to IoT services which have multiple applications, such as, personal home security and public library space management. Compared to recognition without face segmentation, the results of proposed methodology indicate a better accuracy regarding recognition rate.

1. Introduction

IoT is the combination of hardware environment, software services and Internet. It is composed of sensors, servers, hardware equipment, software equipment, and Internet connections [1]. With the fast improvement of sensors and devices, IoT could be applied to multiple fields such as wearables, smart home, and city. Face recognition is one of the popular personal security topics for IoT system, which could be widely used in personal home and public places security fields. Also, it has been used as an identification task in different areas, especially in corresponding with computer-based security and safety systems in homes, criminal identification, and smart phone devices' face identification. For instance, cameras were used for image capturing and IoT processing devices, and Raspberry Pis, were used for comparing captured images with server database to provide directions over the GSM module and alert mobile phones [2]. Similar IoT systems with face recognition components were also proposed for security purposes in various applications, such as libraries and banks [3,4]. In this paper, we propose an effective segmented face recognition for IoT (SFR-IoT) shown in Figure 1.

Face recognition means recognize a specific person with a 2-D or 3-D face image or video base on available face data sets. If a camera is utilized to capture a suspect's face image, that image can then be used to identify the suspect by using biometrics. The application of biometrics technology such as face recognition could reduce crime rates. Instead, the lack of effective skills in capturing biometric data from face images will potentially increase the opportunity for criminal activity [1].



*Corresponding Author: Jiangjiang Liu, Email: jliu@lamar.edu

Figure 1: Segmented Face Recognition Architecture (SFR-IoT)

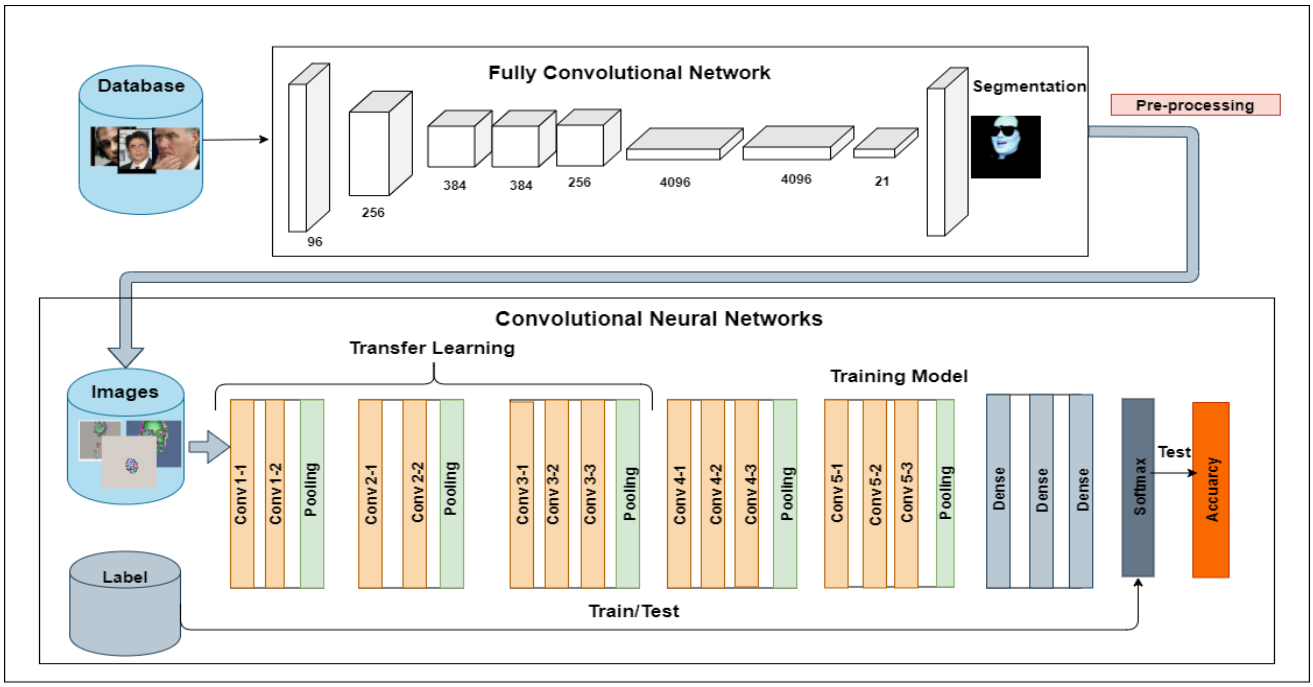


Figure 2: FCN-8Ss-VGG-16 System Architecture

Recently, a few researchers have published work related to crime detection. One study introduced a framework that first detects facial key points and then uses them to perform face recognition [2]. They mentioned that a larger number of images and camouflage images available in a data set can improve the training of the learning network and avoid the need to perform transfer learning. Results show that their framework outperforms the most advanced methods in critical point detection and facial camouflage classification.

For past investigation techniques, there are a few popular algorithms in regards to face cognition which include the geometric features method: principal component analysis (PCA) [3], linear discriminant analysis (LDA) [4], hidden Markov strategy (HMM) [5] and regular LBP features [6]. However, there are evident disadvantages in terms of utilizing these techniques. For instance, the accuracy rate of PCA will diminish significantly with lightness and a state of image change. Also, the strategy for inadequate data requires strict arrangement of information pictures, which is not applicable for basic application.

Due to the important factors of the accessibility of public data sets, as well as the improvement of Graphics Processing Unit (GPU) computation, neural networks has achieved a large renaissance, resulting in a significant increase in accuracy rates [7]. This leads to a major improvement in image recognition and finally in face recognition. For instance, Convolutional Neural Network (CNN) [8,9] is a popular model in neural networks. It is a state-of-the-art technique that has replaced the traditional algorithms on face recognition and consequently taken the network by storm, fundamentally enhancing the cutting edge in numerous applications. Several studies have applied CNN for face recognition purposes, such as Deep Face [10], DeepID3 [11] and Face Net [12]. They have achieved around 97%, 99.6%, and 99.5% success rates, respectively.

Recent studies that combined face segmentation and face recognition technology have provided an important concept for our research. For example, one study extracted face-like regions by using the extracted color information of an image in HSV color space, and the RHT algorithm is then applied to find face region [13]. Samantha proposed a technology of segmentation skin color through YcbCr as well, then applied artificial neural network classified face and non-face classes [14]. However, these researches did not integrate deep learning study for face segmentation and recognition, so this paper would like to propose a new methodology that improve the deficiency.

Even though current automated face recognition systems can recognize individual faces in controlled environments with a 99% accuracy, the accuracy drops to below 60% when images are in unconstrained environments [15].

This is caused by rich facial expressions and changing gestures; face movements with lights; angle; and distance [2].

In addition, disguised faces with glasses, scarves, and accessories will become challenging elements in face recognition tasks. Hence, in order to explore a more effective method to solve this problem, this study proposes a new system called FCN-8s-VGG-16, which is used for the segmentation technology face recognition before classification. The system contains two procedures, as shown in Figure 2. The process of this research is to build a training set of segmentation of the face area with FCN as an input of Convolution Neural Network (CNN), and then train the CNN network from VGG-Face with transfer learning. Lastly, the classification function SoftMax is applied for face probability distribution. The first step is to segment the original image for extracting face parts (eyes, nose, mouth, and skin) from the background, which applies a pre-trained FCN-8s model [16]. Secondly, the output images are fed into the network applying transfer learning [10] based on VGG-16 [8] with our fine-tuned

technology on face recognition task. Finally, the Softmax function calculates the probability distribution of the test image and generates the accuracy rate while comparing it to the label box.

2. Related Work

2.1. General Segmentation and Face Segmentation

Segmentation refers to the process of separating an image or frame into groups of pixels which are identical with respect to some standard. A few papers have applied segmentation for their research topic.

In [17], the author proposed a food calorie estimation model, which was applied on a smartphone to predicate calories of the food in images taken by a phone camera. Firstly, they estimated approximate positions of dishes by edge detection. Secondly, k-means based clustering was applied on the color pixels, thereby extracting a bounding box of a food area. Next, they used GrabCut [18] to obtain an accurate food area within the estimated bounding box. Then, they employed CNN as a calorie estimation task. They prepared 120-calorie annotated food photos as the data for the experiments. Finally, 60 test images with real food calories are estimated with the relative average error of 21.3%, regarding food calorie prediction achieved, and it is higher than the Japanese government definition.

In [19], the author applied different segmentation techniques in the aim of reorganization of the left ventricle in a VEF image. He compared two segmentation approaches: region-based segmentation and edge-based segmentation. The first detects homogeneities and the common features by simultaneously applying the chan_vese model and the thresholding techniques. In the second approach, he applied a certain algorithm such as Sobel, Canny, and Perwitt to obtain the interim between two related regions. Results showed that region-based segmentation using the chan_vese model in conjunction with thresholding gives a better performance.

Since Convolution Neural Network (CNN) has been widely researched by the scholarly field, various models and systems with different purposes are dramatically increased. Fully Convolution Network (FCN) [20] is one of most popular image segmentation technology that is widely used in studies. FCN extracts pixels from CNN layers of mixed scales, enlarges them to the size of the original image, and then applies a convolution layer to classify all this information.

There are research works pertaining to FCN for segmentation purposes that have achieved excellent results. One study presented a One-Shot Video Object Segmentation (OSVOS) based on the FCN network architecture, with transfer learned on ImageNet then fine-tuned on one training sample [21]. The performance was tested on DAVIS database, which consists of 50 full-HD video sequences and YouTube objects. The result shows that OSVOS is fast and improves the state-of-the-art technology by a significant margin (79.8% vs 68.0%). Based on these outstanding results, this paper chooses FCN as our face extraction segmentation tool.

In [22], the author proposed a method to achieve the segmentation of skin, hair, and background by applying FCN-8s and fully-connected CRF. Next, matting algorithm was facilitated in their experiment in order to receive clear hair and skin alpha masks. Finally, they demonstrated that state-of-the-art performance on LFW Parts dataset [23] holds more accurately than other approaches. The defect of the FCN algorithm influences our decision to apply Transfer Learning in our experiment, which we will talk about it in Chapter 3.

2.2. Face Recognition

Another study has shown that choosing an appropriate model in CNN architecture is vital for face recognition. This study evaluated two popular CNN architectures, Alex-Net [9] and VGG-16 [8], on face recognition. They accomplished their task by transferring learning idea to the networks trained for various classifying purposes. One study [24] used Alex-Net model to train the CASIA-WebFace [25] database, and the research shows that VGG performs better than Alex-Net. This paper inspired us to employ VGG-16 [8] as a face recognition model for our own experiment.

In [26], the author conducted face recognition by applying a CNN for feature extraction. First, they randomly selected patches from the STL-10 [27] database and used them to train a linear decoder and obtain a 400 × 192 learned weights matrix. The 64 × 64 RGB images are used to train the identifier through a convolutional layer, and then training features are extracted using these learned weights. Results show the proposed method reached a high rate of excellent sorting, ranging from about 80% to 100%.

There is another study [28] that has proposed a modified Convolutional Neural Network (CNN) architecture with the addition of two batch normalization operations on two of the layers. CNN architecture was utilized to separate particular face features, and Softmax classifier was utilized to recognize faces in the completely associated layer of CNN. The training and test process were tested on a Georgia Tech face database. After pre-processing, the researchers changed the sizes of input to 16×16×1, 16×16×3, 32×32×1, 32×32×3, 64×64×1, and 64×64×3 in order to receive the best result. Finally, 64x64x3 surpasses other sized images regarding the lowest Top-1 error rate. Our paper applied a similar strategy with this research, through altering the size of input in order to obtain the best result in our experiment.

However, to our knowledge, segmentation technology is not used in the face recognition field so far, except in our initial work [29]. This paper is an extension of work originally presented in the 2019 18th IEEE International Conference on Machine Learning and Applications. The accuracy rate will increase due to the interference decrease from unrelated face pixels, such as glasses. In order to decrease the useless effect pixels of disguised face parts, this study would like to propose an FCN-8s-VGG-16 system with segmented technology before classification. It only keeps face sections, and the other parts will be taken off. For the purpose of examination both before and after the change, the face section is sent to the classification model and accuracy will be tested for comparison.

3. The Proposed Approach

3.1. Methodology (CNN)

3.1.1. VGG-16

The core idea of convolutional networks is classifying target data samples to the different distances between classes. The

network structure includes three convolutions (conv1, conv2, conv3), two pooling layers (Pool1, Pool2), and a fully connected layer [8].

3.1.2. Convolution Layer

The input data is trained by employing a set of trainable neurons. The output of each of the feature maps corresponds to an image filter of the same size as the input of the convolution layer. The primary function of the convolution layer is to extract features from an image. Every convolution layer is trained on a feature map of the past layer, sequentially, and then it usually adds a bias parameter in order to increase accuracy by activating the function to translate results from linear to non-linear function. After that, the feature maps are fed into the next convolutional layer as input data.

3.1.3. Pooling Layer

Pooling layer decreases the dimensionality of each activation map but keeps the most vital data information. The input images are separated into a set of non-covering square shapes. Each field is downsampled by a non-linear function. For example, average or maximum is used the most. This layer accomplishes a better speculation and robust result for system.

3.1.4. ReLU Layer

Rectified linear units (ReLU) is a non-linear operation. It is an important function because it ends up with 0 if the input is less than 0. However, if the input is greater than 0, the output will be an original number. Research studies show that ReLU results can implement faster training for a huge network.

3.1.5. Fully Connected Layer

The output from convolution layer, pooling layer, and ReLU layer is high- level features of input data. The purpose of applying the Fully Connected Layer (FCL) is to sum these features for classifying the input data into different classes based on the probability of each class of each feature map. Next, FCL supports the features to a classifier, which is Softmax function. This function will conclude the probabilities of every target class over all possible target instances. Afterwards, the calculated probabilities will decide the target class for the given inputs.

3.1.6. Why VGG

There are a few CNN architectures, such as LeNet, AlexNet, Reset, and so on, which have been widely used since CNN was invented. Recently, the most state-of-the-art model [10] outperforms in the localization and recognition tracks, respectively, the ImageNet Challenge 2014 among them.

This VGG model is described by its simplicity, applying only 3×3 filters stacked over each other in increasing depth. Reducing parameter numbers and model size is employed by max pooling, two fully-connected layers, each with 4,096 nodes followed by a Softmax classifier.

3.1.7. Transfer Learning

Transfer learning is an area of artificial intelligence, which centers around the capacity for a machine learning calculation to enhance learning limits on a target data set through past exposure to an alternate way. The measured quality of a CNN implies that we can easily apply the weights from a pre-trained model and only re-prepare most elevated layers. In particular, we re-prepare all linear layers in the model and replace parameters of the highest layers in VGG-16 [8]. To get rid of the lack of geometric invariance of these approaches, fine-tuning [30] with an external data set can be utilized. The primary distinction between picture arrangement and picture recovery is the measure of information and its fluctuation. In classification, it is important to use huge data sets with high variability for different categories. However, in image recognition, the geometric invariance of an image is less essential for a training model. As the purpose of image recognition is to recognize the instance of example, less variability data will be needed.

In this paper, to use CNN functionally, we fine-tune the pre-trained CNN [10] a face data set for image recognition. Fine-tuning usually focuses on the higher layers while fixing the lower layers of a CNN. We use FCN-8s-VGG [16] architecture, which has been pre-fine-tuned for segmentation on PASCAL [31]. As the CNN model [16] is pre-trained on face dataset and performs an excellent result, we don't consider it for Fine-tuning while we are doing the same object segmentation purpose.

Table 1: The Architecture of VGG-Face Refined

Layer	Layer Type	Filters	Input	Training
1	Conv.	64	3x3x3	False
2	Conv.	64	3x3x64	False
3	Max-Pool	-	2x2	-
4	Conv.	128	3x3x64	False
5	Conv.	128	3x3x128	False
6	Max-Pool	-	2x2	-
7	Conv.	256	3x3x128	False
8	Conv.	256	3x3x256	False
9	Conv.	256	3x3x256	False
10	Max-Pool	-	2x2	-
11	Conv.	512	3x3x256	True
12	Conv.	512	3x3x512	True
13	Conv.	512	3x3x512	True
14	Max-Pool	-	-	-
15	Conv.	512	3x3x512	True
16	Conv.	512	3x3x512	True
17	Conv.	512	3x3x512	True
18	Max-Pool	-	2x2	-
19	FC	-	4096	True
20	FC	-	4096	True
21	FC	-	85	True
22	Softmax	-	-	-

3.1.8. VGG-FACE

VGG-Face [10] is a model that was pre-trained based on the VGG-16 structure and set up on a face data set acquired by the Visual Geometry Group that is comprised of more than 2.5 million pictures and 2,622 different labels. Detailed information in terms of VGG-16 model was shown in Table 1, including layer types, number of filters, and input image size.

The architecture of VGG-face contains a total of 22 layers: 13 convolutional layers for image and filter calculation, 5 max pooling layers for keeping feature information, 3 fully connected layers for summing up all previous features, and later a Softmax layer for classification [10]. The normal image resolution of the input layer is 224x224. Table 1 shows the detailed architecture of the VGG-face, including filter numbers and input size at each layer applied in our research. In this experiment, we have fixed the previous 10 layers and changed 21 fully connected layers from 1,000 to 85, since we have 85 classes in the experiment, which means 85 labels will be generated. Additionally, we fixed the parameters of the lower convolutional layers from 1-9 but trained the higher convolutional layers from 11-17 and the FCN layers from 19-21. Applied changes are highlighted in the table.

3.2. Methodology (FCN)

FCN [16] uses the identical convolutional network based on VGG-16 layers [8] and converts the fully connected layer to a convolutional layer. In the traditional CNN structure, the first five layers are convolutional layers, and the sixth and seventh layers are a one-dimensional vector with a length of 4096. The eighth layer is a one-dimensional vector with a length of 1,000, corresponding to 1,000 different probabilities of the category. Instead, FCN represents these three layers as a convolutional layer whose size (number of channels, width, height) is (4096, 1, 1), (4096, 1, 1), (1000, 1, 1), respectively. There does not seem to be any difference in numbers, but convolution is not the same concept, and calculation process as a fully-connected layer. The convolutional layer uses the weights and biases that CNN has pre-trained, but the difference is that the weights and offsets are their own scope and belong to their own convolution kernel. Therefore, all layers in the FCN network are convolutional layers and are called fully convolutional networks.

3.2.1. Upsampling

Upsampling is also called deconvolution at some points, as both of their operations are multiplication and addition. The upsampling of the accompanying factor f is a convolution operation on a fractional input with a step size of 1/f.

Backward convolution is called deconvolution. Meanwhile, the forward and backward propagation of upsampling can only be achieved by reversing the forward and backward propagation of the convolution. Consequently, it performs well no matter the optimization or backward propagation. Deconvolution can enlarge input size by learning parameters while training [20].

3.2.2. Skip Layers

Now that we have a 1/32 size, a 1/16 size, and a 1/8 size feature map, after upsampling the 1/32 size heatmap, the image restored by this operation is a convolutional kernel of conv5. The features are limited to the accuracy, so they cannot restore the features in the image well; consequently, forward propagation will be applied. Detailed information could be complemented by upsampling the conv4. Afterwards, conv3 does the exact same calculation as conv4. Finally, the process completes the restoration of the entire image. According to different strides with 32, 16, and 8, FCN splits into three kinds of results: FAC-32s, FCN-16, and FCN-8s [20]. Obviously, FCN-8s combined previous FCN-32s and FCN-16 results, which gives FCN-8s more accuracy. We decided to choose FCN-8s in our experiment.

3.2.3. Why choosing FCN

The reason why this research employs FCN are below:

- Traditional segmentation method segmented the image into similar sections, including coherent area-independent, low-level clues based on categories, such as pixel color or proximity. Semantic segmentation, on the other hand, assigned each pixel of the image to a semantic tag. This usually means classifying each pixel (for example: the pixel 1 belongs to the glass, the pixel 2 belongs to the hair). Pixel-level classification seems more efficient than cutting out each patch [32]. The FCN-8s model brings state-of-the-art performance on PASCAL VOC in 2011 and 2012 [20].
- No fully connected layer is used in this kind of architecture, which reduces the number of parameters and the computation time. In addition, the network works regardless of the original image size. It could be trained from small and big images, and it does not require any particular number of units at any stage. Additionally, all connections are local [20]. FCN brought great breakthroughs on semantic segmentation tasks when implemented on image segmentation [20].

It is tough to obtain millions of diverse images with ground truth segmentation labels, which could be difficult and mostly performed manually. However, it is convenient to use a pre-trained model for research purposes. Our method applies the FCN segment to the visible parts of faces from their specific circumstance and impediments, which is a fine-tuned system for face segmentation on PASCAL [20] on the data set of IARPA Janus CS2 [33] with 9,818 segmented faces achieved.

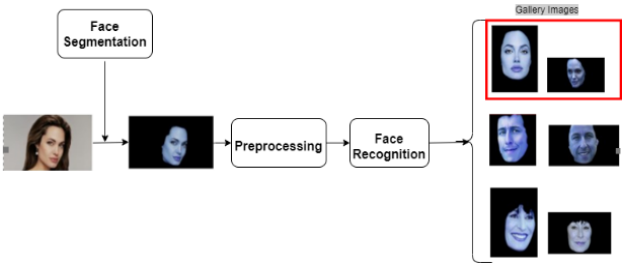


Figure 3: FCN-8s-VGG-16 System 2

This research has achieved an excellent segmentation performance that could be obtained with a standard FCN trained on plenty of rich and varied examples. FCN-8s combined previous FCN-32s and FCN-16s results, which gives FCN-8s more accuracy. We decided to choose FCN-8s in our experiment. Figure 3 shows the process of our experiment.

3.3. Methodology

3.3.1. Data selection

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [34] was dominant in giving this information to the general picture classification task. Moreover, analysts have made data sets accessible for object recognition [35]. Labeled Faces in the Wild (LFW) [36], as the most popular benchmark data set has dominated in the field of face recognition for many years. It is a

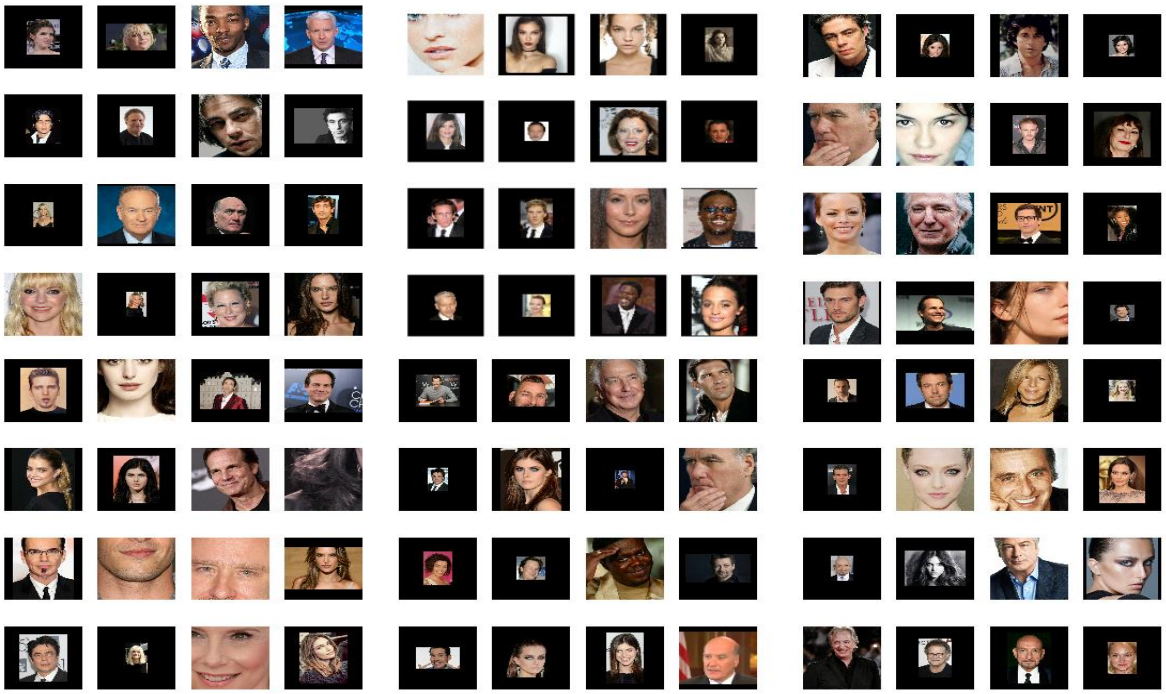


Figure 4: Gallery Images

huge scale open data set and the benchmark database of face photos intended for concentrating the issue of unconstrained face recognition, which has been widely applied in numerous research studies [10,12]. The data set contains more than 13,000 images of face pictures gathered from online.

Each face has been labeled according to the name of the individual envisioned. However, due to the shortage of obvious challenging face images, such as block face, tilt the head, wrinkle the eyebrows, and so on, it is not an ideal image set in terms of the purpose of the paper, which is to compare face recognition performance with and without segmentation. Furthermore, FRGC [37] MS-celeb-1M [38] and MOBIO [39] are other benchmark face data sets used to identify face images though various experiments with the same problem as LFW.

3.3.2. Applied Data Set

In this research, we decided to employ the Celebrity-Face-Recognition-Dataset, which consists of 1,100 famous celebrities with 8,000 images [40] in each. In this dataset, the total size of the data set is 172 GB with 800*800*3 pixels in each. Due to the limitations of our hard drive and memory, this research randomly and manually selected 85 individuals with 100 images for each as classes for the face recognition task.

Since accuracy difference along with and without segmentation face section from the background is in a significant index for the research, the images with obvious covers such as hats, accessories, or different poses and hairstyles, are our preferred selection among the 8,500 figures from the Celebrity-Face-Recognition dataset. Figure 4 shows partial data processed in our experiment.

3.3.3. Data Augmentation

Image recognition technology focuses on the work of few data sets and small variability in object images. This results in very few images that can train a particular CNN model, even with fine-tuning. One way to solve this problem is to augment the data, by randomly applying transformations, color perturbations, and other random transformations. Randomly rotating and flipping the images can make up the deficiency of diversity pixels of the model. In the experiment, the following values are applied:

- Brightness: Images are randomly changing brightness [-0.1, 0.1) [41].
- Flipping: Images are horizontally flipped from left to right.
- Scaling: The pixel values are scaled to the range [0, 1].
- Contrast: Images are randomly adding contrast to the range [0.2, 1.8].

The result shows below as Figure 5.

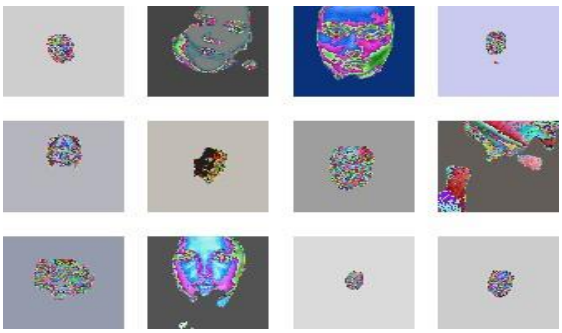


Figure 5: Pre-processing Result with FCN-8s Segmentation

3.3.4. Cross Validation

We found that in using the same data set, both training and model error estimation, the error estimation is mostly inaccurate, which is called the optimism of the model error estimation. To overcome this problem, cross-validation was proposed. According to Kohavi [42], cross-validation is a technology used to access predictive systems by dividing the original data set into a training set for training the parameters, with a test set for assessing the trained model. There are three more popular methods: leave-one-out, leave-P-out, and K-fold. Compared to the previous two technologies, K-fold only needs to calculate k times, which dramatically decreases code complexity. In K-fold cross-validation, the input data is split into k same size subsets, which is called folds as well. K-1 subset data will be trained into a system afterward, assessing the model on the subset that was not utilized for preparing. This procedure is repeated k times (the folds) until each subset has been used for evaluation purposes (and prohibited from training) each time finished. The results from k times calculation of K-folds can then be averaged (or joined) to receive a final estimation. The advantage of this method is that all data information is used for both training and evaluation, and each sub-dataset is used for validation exactly once. Based on the advantage of cross-validation, this paper has applied the 10-folds cross-validation on a data set.

3.3.5. Hardware Utility

Our implementation is based on Tensorflow Framework. AWS as a service platform that support the measurement of runtimes and performance. We implemented the experiment on the 16 GB memory of Nvidia Tesla 100V GPU.

4. Experimental Results

During this examination process, the implementation of the proposed method has been evaluated based on the percentage recognition rate of the identities on the testing data set. Initially, we set the learning rate as 0.001 for the CNN system, and a total of 10,000 steps are required for every k times calculation. Then for each step, 10 images (batch size) and 20 images (batch size) are sent for training through the network individually. At the same time, a 50% dropout was also been used in the experiment. Rectified Linear Unit (ReLU) as activation functions and Cross Entropy Loss function are applied to guide CNN training. The Adam algorithm calculates the gradient of entire data sets and updates values in the opposite direction to the gradients until a local minimum is found. Top-1 accuracy is applied to measure the performance of processes. This entire test experiment will be repeated ten times until the average of performance is obtained. The results of our experiment are used to compare with the performance of others. To check and analyze the implementation of the proposed method, 10 image samples from each category are collected into the test set.

According to the outstanding performance of our experiment, we took the best parameter as the final performance and compared it with the model consisting of original images. We performed various experiments by altering the image size. For example, we drafted the size of each picture after pre-processing as 350x350x3, 500x500x3, and 800x800x3. Among the data set, 90% of the data are the training set and 10% are the test set. 10,000 steps were applied for CNN training process ten times. Based on Top-1 error calculation, the result of the proposed system was acquired. According to whether the prediction is the same as the target category, the Top-1 error rate returns a Boolean value.

Table 2: Performance Comparison with Various Parameters

Input Image size	Numbers steps	Learning rate	Batch size	(VGG-16 System) Top1 Error	(Proposed FCN-8s-VGG-16 system) Top1 Error
350x350x3	10000	0.001	10	82.69%	92.15%
350x350x3	10000	0.001	20	84.34%	93.21%
500x500x3	10000	0.001	10	92.43%	97.25%
500x500x3	10000	0.001	20	93.54%	98.33%
800x800x3	10000	0.001	10	99.84%	99.57%
800x800x3	10000	0.001	20	99.88%	99.69%

Table 2 shows the comparison performance of various parameters. As seen from Table 2, the input size of 350x350x3 and 500x500x3 of the FCN-8s-VGG-16 system obtained a higher accuracy rate than the VGG-16 System which is trained without segmentation technology. Meanwhile, the accuracy rate of 800x800x3 input size achieved the closest result with the VGG-16 System. Furthermore, the running time has reduced on average 4.52% than the VGG-16 System in all three experiments owing to reduced pixel with segmentation technology. Figure 6 demonstrates the relationship between the batch size, image size and the system. We can conclude that as the batch size increased, both accuracy rate of VGG-16 System and FCN-8s-VGG-16 have increased slightly.

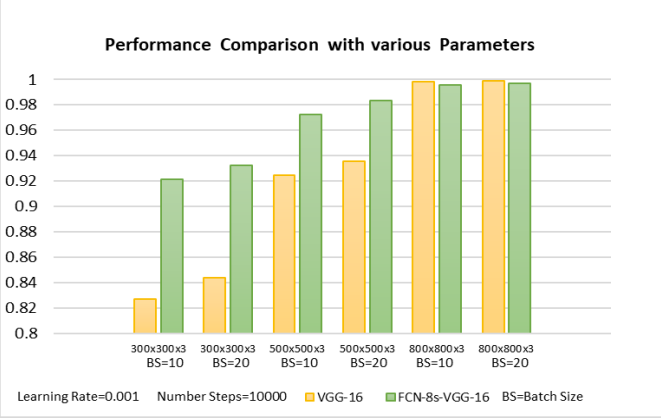


Figure 6: Performance Comparison with various Parameters

However, this will expand running time as well. Therefore, the batch size and running time could be a future trade off work for us. Above all, we could gather that image input size with high pixels will have a higher accuracy rate for recognition.

5. Conclusion and Future Work

In this paper, we proposed the FCN-8s-VGG-16 system to segment face images and recognize segmented images with high precision. Our face recognition system, FCN-8s-VGG-16 is composed of face image segmentation with FCN-8s [16] and

VGG-Face fine-tuning model recognition. Experiments were conducted on Celebrity-Face-Recognition dataset [40] for training and testing purposes. Compared with the original face recognition without segmentation accuracy, the effective classification rate using different parameters has been significantly improved from 92.15% to 99.69%, reaching a level of 82.69% to 99.88%. It also demonstrates that increasing the size of the input and batch number, while keeping the learning rate and number of steps constant, will improve the accuracy rate in certain situations. However, such observation is not completely ensured for other systems.

Accuracies on Top-1 in face recognition results means that some limitations happened in our algorithm, especially when applied to a wild data sets: First, even though FCN-8s has successfully segmented most face images, there were still a few images that did not achieve ideal results. For example, some small black areas occur after segmentation, which will affect the accuracy of performance. There are a few things that have triggered these failures. First, it takes three training processes to get the FCN-8s, which is not sensitive enough to combine the details of the image. This is because when the decoding is performed, that is, when the original image is restored, the label map of the input upsampling layer is too sparse. Secondly, FCN-8s does not consider the relationship between pixels during classification, and it lacks spatial consistency. Additionally, we only computed limited image input size, batch sizes, and learning rate, so it is very possible that we will receive a better accuracy rate if there are more parameters to choose from, with time allowed. Lastly, with the fine-tuning FCN model [16], the model will be more robust [43]. Only one recognition model processed in the experiment is limited. Some other CNN model has achieved an excellent performance in recent competition. For example, Google Net [44] is a winning architecture on ImageNet 2014. We will be applying it as our recognition system in future research for comparing with VGG-16. Besides, in reality, a human face is made of a 3D model; it's more realistic to segment and recognize images based on 3D faces. [45] has proposed a system of face recognition 3D based on segmentation by classifying fields of facial images before and after fusion of color and depth images. It brings us a potential research in face recognition study that we could investigate more about 3D image segmentation technologies instead of 2D.

Additionally, we will develop a mobile app such as [5], to complete the SFR-IoT architecture that could receive notification from our system though IoT devices when recognizing face in order to make contribution in real world application. In the future, data encoding and decoding will be applied to untrusted devices and servers to protect privacy data to make software more powerful and trustworthy [46].

Above all, the comparison reveals significant improvement in performance. Further experiments can be done by improving segmentation technology or by changing the recognition model.

References

[1] S. Zalke, P.M.-2015 I.C. On, U. 2015, Survey on implementation of enhanced uniform circular local binary pattern algorithm for sketch recognition, Oct. 2020.

[2] J. Cui, J. Li, Y.H.-2008 I.C. on, undefined 2008, "Research on Face Recognition Method based on Cognitive Mechanism," Ieeexplore.Ieee.Org, Oct. 2020.

[3] X. Dong, H. Huang, H.W.- ISCSCT, undefined 2010, "A comparative study of several face recognition algorithms based on pca."

[4] H. Yu, J. Yang, "A direct LDA algorithm for high-dimensional data — with application to face recognition," Pattern Recognition, **34**(10), 2067–2070, 2001, doi:10.1016/s0031-3203(00)00162-x.

[5] J. Bobulski, "Face recognition method with two-dimensional HMM," in Advances in Intelligent Systems and Computing, Springer Verlag: 317–325, 2016, doi:10.1007/978-3-319-26227-7_30.

[6] T. Ojala, M. Pietikäinen, D. Harwood, "A comparative study of texture measures with classification based on feature distributions," Pattern Recognition, **29**(1), 51–59, 1996, doi:10.1016/0031-3203(95)00067-4.

[7] S. Zalke, P.S. Mohod, "Survey on implementation of enhanced uniform circular local binary pattern algorithm for sketch recognition," in ICIIECS 2015 - 2015 IEEE International Conference on Innovations in Information, Embedded and Communication Systems, Institute of Electrical and Electronics Engineers Inc., 2015, doi:10.1109/ICIIECS.2015.7192917.

[8] S. Karen, Z. Andrew, "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION Karen," American Journal of Health-System Pharmacy, **75**(6), 398–406, 2018.

[9] A. Krizhevsky, I. Sutskever, G.E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, **60**(6), 84–90, 2017, doi:10.1145/3065386.

[10] O.M. Parkhi, A. Vedaldi, A. Zisserman, "Deep Face Recognition," in Procedings of the British Machine Vision Conference 2015, British Machine Vision Association: 41.1-41.12, 2015, doi:10.5244/C.29.41.

[11] D. Virmani, T. Sharma, M. Garg, GAPER: Gender, Age, Pose and Emotion Recognition Using Deep Neural Networks, Springer, Singapore: 287–297, 2021, doi:10.1007/978-981-15-5463-6_26.

[12] F. Schroff, D. Kalenichenko, J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07-12-June, 815–823, 2015, doi:10.1109/CVPR.2015.7298682.

[13] A. Aminian, M.S. Beni, "Face detection using color segmentation and RHT," in 3rd International Conference on Pattern Analysis and Image Analysis, IPRIA 2017, Institute of Electrical and Electronics Engineers Inc.: 128–132, 2017, doi:10.1109/PRIA.2017.7983032.

[14] S.D.F. Hilado, E.P. Dadios, R.C. Gustilo, "Face detection using neural networks with skin segmentation," in Proceedings of the 2011 IEEE 5th International Conference on Cybernetics and Intelligent Systems, CIS 2011, 261–265, 2011, doi:10.1109/ICCIS.2011.6070338.

[15] J.C. Klontz, A.K. Jain, "A case study of automated face recognition: The Boston marathon bombings suspects," Computer, **46**(11), 91–94, 2013, doi:10.1109/MC.2013.377.

[16] Y. Nirkin, I. Masi, A.T. Tuăn, T. Hassner, G. Medioni, "On face segmentation, face swapping, and face perception," in Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, Institute of Electrical and Electronics Engineers Inc.: 98–105, 2018, doi:10.1109/FG.2018.00024.

[17] K. Okamoto, K. Yanai, "An automatic calorie estimation system of food images on a smartphone," in MADiMa 2016 - Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, co-located with ACM Multimedia 2016, Association for Computing Machinery, Inc: 63–70, 2016, doi:10.1145/2986035.2986040.

[18] C. Rother, V. Kolmogorov, A. Blake, "'GrabCut' - Interactive foreground extraction using iterated graph cuts," in ACM Transactions on Graphics, 309–314, 2004, doi:10.1145/1015706.1015720.

[19] Y. Ben Fadhel, S. Ktata, T. Kraiem, "Cardiac scintigraphic images segmentation techniques," in 2nd International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2016, Institute of Electrical and Electronics Engineers Inc.: 364–369, 2016, doi:10.1109/ATSIP.2016.7523107.

[20] J. Long, E. Shelhamer, T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society: 431–440, 2015, doi:10.1109/CVPR.2015.7298965.

[21] S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixé, D. Cremers, L. Van Gool, "One-Shot Video Object Segmentation," Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-January, 5320–5329, 2016.

[22] S. Qin, S. Kim, R. Manduchi, "Automatic skin and hair masking using fully convolutional networks," in Proceedings - IEEE International Conference on Multimedia and Expo, IEEE Computer Society: 103–108, 2017, doi:10.1109/ICME.2017.8019339.

[23] LFW Face Database | Part Labels, Oct. 2020.

[24] D. Bussey, A. Glandon, L. Vidyaratne, M. Alam, K.M. Iftekharuddin, "Convolutional neural network transfer learning for robust face recognition in NAO humanoid robot," in 2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017 - Proceedings, Institute of Electrical and Electronics Engineers Inc.: 1–7, 2018, doi:10.1109/SSCI.2017.8285347.
[25] D. Yi, Z. Lei, S. Liao, S.Z. Li, "Learning Face Representation from Scratch," 2014.
[26] A. Raj, S. Gupta, N.K. Verma, "Face detection and recognition based on skin segmentation and CNN," in 11th International Conference on Industrial and Information Systems, ICIIS 2016 - Conference Proceedings, Institute of Electrical and Electronics Engineers Inc.: 54–59, 2016, doi:10.1109/ICIINFS.2016.8262907.
[27] STL-10 dataset, Oct. 2020.
[28] M. Coskun, A. Ucar, O. Yildirim, Y. Demir, "Face recognition based on convolutional neural network," in Proceedings of the International Conference on Modern Electrical and Energy Systems, MEES 2017, Institute of Electrical and Electronics Engineers Inc.: 376–379, 2017, doi:10.1109/MEES.2017.8248937.
[29] F. Gao, J. Liu, "Face recognition using segmentation technology," in Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019, Institute of Electrical and Electronics Engineers Inc.: 545–548, 2019, doi:10.1109/ICMLA.2019.00102.
[30] A. Babenko, A. Slesarev, A. Chigorin, V. Lempitsky, "Neural Codes for Image Retrieval," Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8689 LNCS(PART 1), 584–599, 2014.
[31] The PASCAL Visual Object Classes Homepage, Oct. 2020.
[32] G. Csurka, D. Larlus, F. Perronnin, "What is a good evaluation measure for semantic segmentation?," BMVC 2013 - Electronic Proceedings of the British Machine Vision Conference 2013, 2013, doi:10.5244/C.27.32.
[33] B.F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, A.K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07-12-June, 1931–1939, 2015, doi:10.1109/CVPR.2015.7298803.
[34] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, **115**(3), 211–252, 2014.
[35] T.Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, "Microsoft COCO: Common objects in context," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag: 740–755, 2014, doi:10.1007/978-3-319-10602-1_48.
[36] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Oct. 2020.
[37] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the Face Recognition Grand Challenge *, 2005.
[38] Y. Guo, L. Zhang, Y. Hu, X. He, J. Gao, "MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition," 2016.
[39] C. McCool, S. Marcel, A. Hadid, M. Pietikäinen, P. Matějka, J. Černocký, N. Poh, J. Kittler, A. Larcher, C. Lévy, D. Matrouf, J.F. Bonastre, P. Tresadern, T. Cootes, "Bi-modal person recognition on a mobile phone: Using mobile phone data," in Proceedings of the 2012 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2012, 635–640, 2012, doi:10.1109/ICMEW.2012.116.
[40] GitHub - prateekmehta59/Celebrity-Face-Recognition-Dataset: Dataset of around 800k images consisting of 1100 Famous Celebrities and an Unknown class to classify unknown faces, Oct. 2020.
[41] J.J. Lv, X.H. Shao, J.S. Huang, X.D. Zhou, X. Zhou, "Data augmentation for face recognition," Neurocomputing, **230**, 184–196, 2017, doi:10.1016/j.neucom.2016.12.025.
[42] R. Kohavi, A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, 1995.
[43] Y. Sun, D. Liang, X. Wang, X. Tang, "DeepID3: Face Recognition with Very Deep Neural Networks," 2015.
[44] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society: 1–9, 2015, doi:10.1109/CVPR.2015.7298594.
[45] M. Belahcene, A. Chouchane, M. Amin Benatia, M. Halitim, "3D and 2D face recognition based on image segmentation," in 2014 International Workshop on Computational Intelligence for Multimedia Understanding, IWCIM 2014, Institute of Electrical and Electronics Engineers Inc., 2014, doi:10.1109/IWCIM.2014.7008800.
[46] W. Xue, W. Hu, P. Gauranvaram, A. Seneviratne, S. Jha, "An Efficient Privacy-preserving IoT System for Face Recognition," in Proceedings - 2020 Workshop on Emerging Technologies for Security in IoT, ETSecIoT 2020, Institute of Electrical and Electronics Engineers Inc.: 7–11, 2020, doi:10.1109/ETSecIoT50046.2020.00006.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 45-51 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Business Intelligence for Generating Comprehensive Report in Electronic Completion and Handover

Fadhillah Moulita Andiani, Faizal Abid, Hendri, Abba Suganda Girsang*

Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

ARTICLE INFO

Article history:
Received: 05 June, 2020
Accepted: 08 October, 2020
Online: 08 November, 2020

Keywords:
Data Warehouse
Business Intelligence
Project Completion
Project Management System
Tree Model Structure

ABSTRACT

Project completion is a common and best practice for oil and gas and construction industry. It provides a comprehensive completion approach and gives total confidence to asset owner to operate the facility handed over by construction contractors. While the nature variation of asset hierarchy is unique from one type of asset to another, a project completion software must have prominent ability to adapt wide range of variations with unlimited level of hierarchy. One of the approaches to overcome this is implementing a tree-model concept to accommodate flexible hierarchy. Unfortunately, the package is loaded with complexity to retrieve data and takes longer join operation. This paper proposes a business intelligence approach to analyze and make an optimum reporting retrieval using data warehouse. This is implemented in 4 steps following Kimball methods. The objective of this paper is to generate model by using data warehouse starting the extract, transform and load process on the flexible tree model hierarchy. It can be used to generate report and comprehensive dashboard especially progress report of project and work schedule as needed in the oil and gas and construction industry.

1. Introduction

Project completion or commonly known as commission are one of critical activities for heavy industrial construction project. It gives total confidence for handing over the full facility from contractor to the operator or owner [1] . To support this activity, a project completion software is implemented. It is used to help the contractor's construction team to ensure that the facility is physically complete and made ready for handover to the commissioning team. Using the construction check sheets agreed by both parties (construction and commissioning), project management team can track, report and control all inspection and testing of workmanship and materials [2].

The barcoded, scanned documentation stored in the system helps to prove the complete and correct execution of all fabrication and installation work in accordance with the project specifications and design drawings. The phase of construction to Commissioning, eventually is continued to Commissioning to Operation provided that each phase requirements are fulfilled, completed, and verified by all related parties

The initial activity with respects to the system completions process is the installation and commissioning of the Electronic Completion and Handover (ECHO) as shown Figure1. An implementation plan should be produced to schedule all key activities necessary to deliver a commissioned PCMS.

The next process is to define the type of equipment or tag type deployed on the project, examples of tag types are: Compressor, Pump, Turbine, Filter, Heat Ex-changer, etc. When a Tagged Item is defined, check sheets can be allocated and generated in a form of Tag Type Matrix. Lead Discipline Engineers and Commissioning Engineers review the project's Piping and Instrument Diagram (P&ID)s and drawings to define by Discipline and assigned to the related folder, subsystem and eventually by system.

P & IDs structure are then represented in the ECHO system as shown Figure 2. Due to wide variety of both facilities and assets, ECHO is required to be flexible to adapt its level of hierarchy with no limitation, one project can consist of three (3) levels of hierarchy, another can be ten (10) levels as shown Figure 3. Most companies focuses in analysing its business mostly on structured data, and use a relational database [3]-[5]. To generate one report from that flexible hierarchy is such a tedious task, since the Structure Query Language (SQL) joint operation can be as many as required. One example of tree concept is illustrated as in Figure 4. ETL process is also used in some companies for processing data

*Corresponding Author: Abba Suganda Girsang, Jakarta, & agirsang@binus.edu

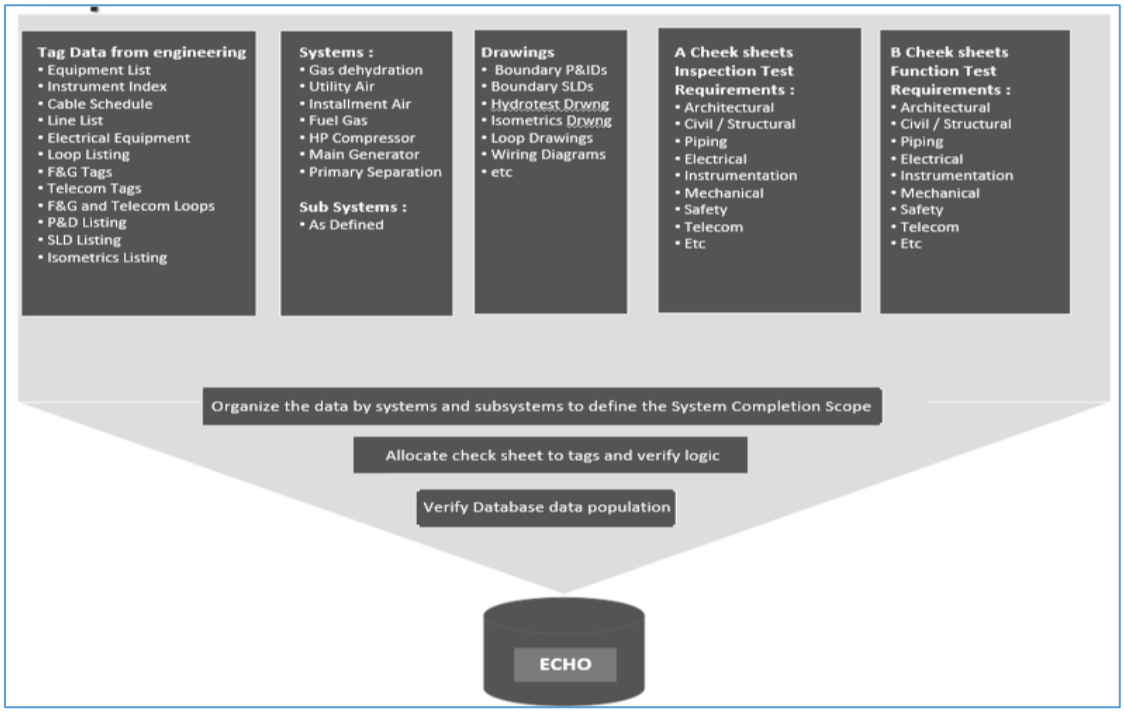


Figure 1: System Setup

warehouse, ETL processes extract data from various sources, transform the data to fit analytical needs and then load it into the database [6]. ETL Process as a part of the Data Warehouse implementation maybe a perfect solution for this and boost the performance of information retrieval of the ECHO system. An ETL tools can be used to extract data and uses business rules to transform the data into new formats [7]. It helps many companies to monitor its performance [8], [9].

This paper implements data warehouse system to the current system and propose a data warehouse model that can be used to make an online analytical processing through an interactive dashboard that can be drilled down, up and across [10]. The goals are to monitor the trend of the project completion and to get insight of the construction inter dependency between systems, disciplines and data-driven coordination among the contractors. It also can be used to obtain the strategy information for decision makers.

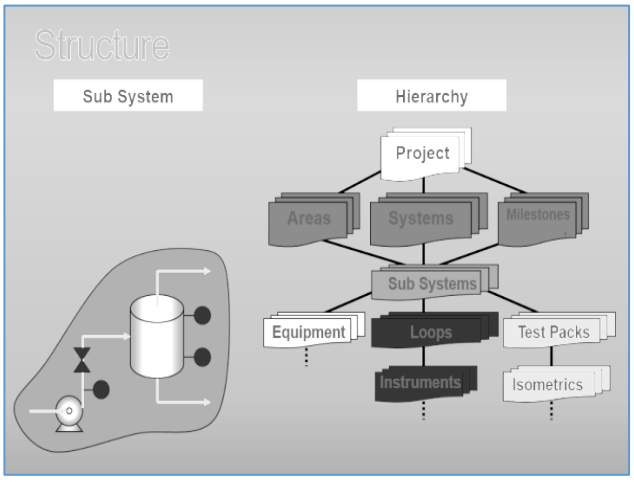


Figure 2: P&ID Structure in ECHO

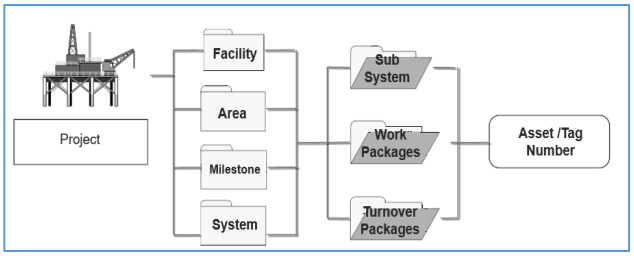


Figure 3: Project's level of hierarchy

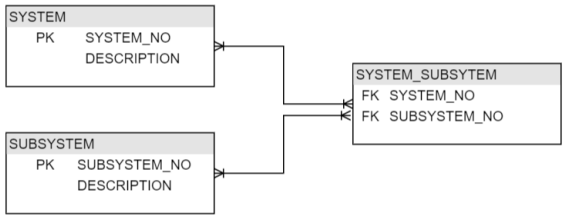
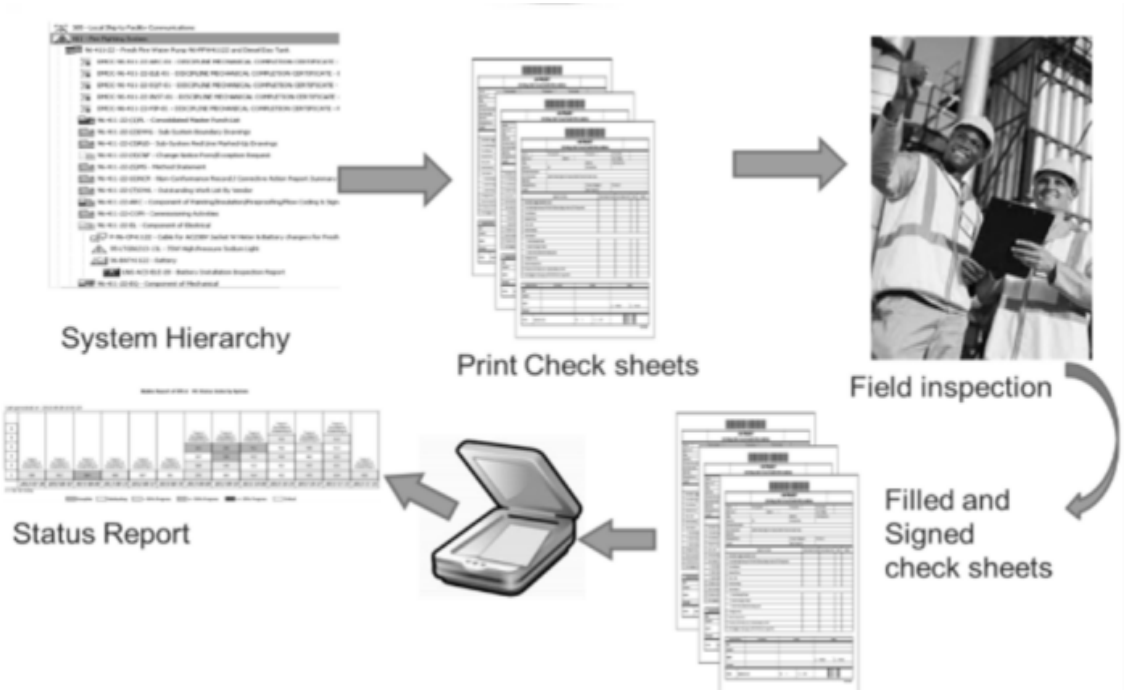
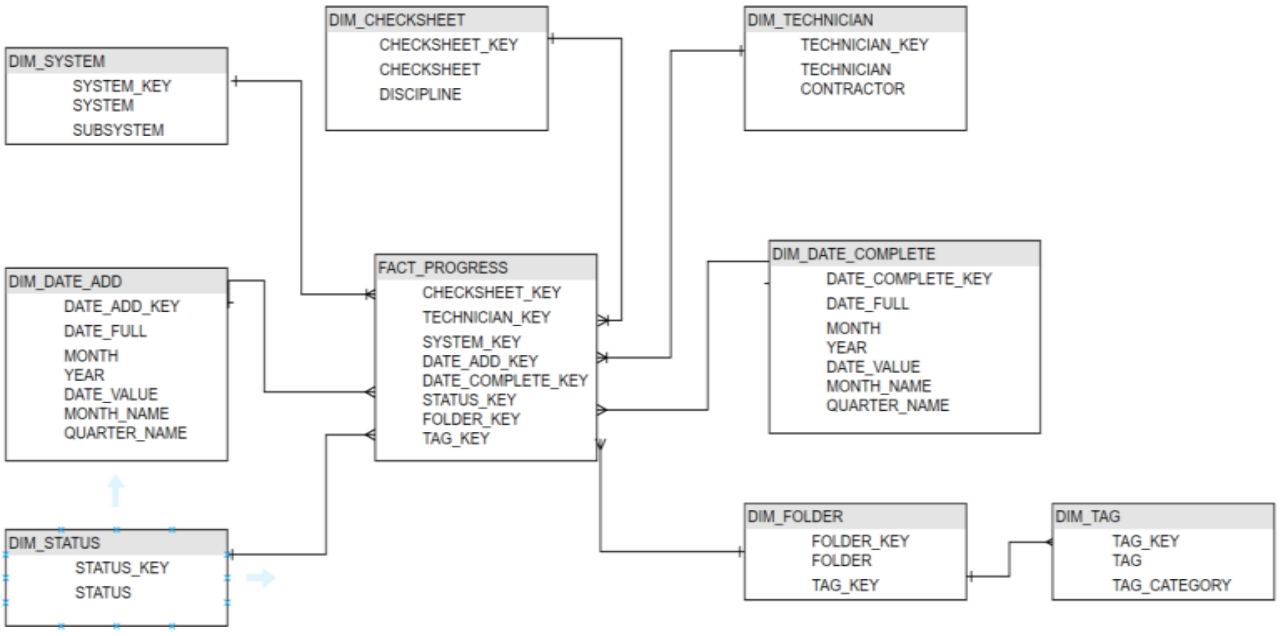


Figure 4: System and Subsystem Tree Structure

2. Literature Review

Since the amount of stored data is growing faster continuously, and the data processing brings many complications, the analytical databases (OLAP) based on the data warehouses rapidly gain popularity in analytical data processing [11]. The research of data warehouse is currently also developed by combining machine learning [12], [13]. The data warehouse is a preformatted data repository that consists of data collection from various format and structure in a particular way to expedite the information retrieval for analytical purposes [14] . Data warehouse is then loaded into a Business Intelligence (BI) platform which is then utilized as decision support technologies that aim to assist in a quick and data-driven decision making [15].





3. Proposed Method

As our guideline to develop a complete data warehouse system for ECHO, we are referring to industry widely known as Kimball four-step dimensional design process. Kimball presents a different vision of data warehouses. Kimball considers that the data warehouse be a set of consistent data marts and based on shared conformed dimensions.

Select the Business Process. The first step is to identify the business process. A business process is defined as a major operational process in an organization. Process in business is defined as a natural business activity in an organization supported by legacy source data-collection system (in this project database is used as data-collection). The business processes of this research are: 1. adding completion check sheet based on equipment's discipline into the project's scope of work and 2. Technician assigned to the project completes the work and turn the status of its check sheet to complete as shown Figure 5

Declare the Grain. The next step is to declare the grain. The goal of this step is to define the lowest detail on the fact table. The type of the fact table for this research is factless fact table which provide the status of one check sheet assigned to a tag or

equipment in a project hierarchy, system-wise and discipline-wise. It also shows the addition and its completion by technician from various contractors.

Identity the Dimension. With the business process selected and the declaration of the grain, we then observe and analyze the supporting information within the granularity of the data, which is called by a dimension, by defining it is a classification tables in original database that have correlation to each other into one table dimension. All the table in original database must be analyzed can classified into dimension tables. We observed that eight dimensions are identified, they are: dim_system, dim_checksheet, dim_technician, dim_date_add, dim_date_complete, dim_status, dim_date, dim_technican, dim_folder and dim_tag.

Identity the Fact. The last step is identifying the fact tables, the fact table normally store the transaction data, and it holds something that measurable. In our case, the fact table is indeed stores transaction data, the transaction of adding a completion check sheet into a project, and completing its work by assigned technician, by definition: A factless fact table is a fact table that does not have any measures. It is essentially an intersection of dimensions (it contains nothing but dimensional keys). There are two types of factless tables: One is for capturing an event, and one is for describing conditions, according to the definition above, we can conclude that the type of our fact table is factless. Figure6 shows the star scheme of the project completion progress where;

- system_key referred to DIM_SYSTEM,
- status_key referred to DIM_STATUS,
- date_add_key referred to DIM_DATE_ADD,
- checksheet_key referred to DIM_CHECKSHEET,
- technician_key referred to DIM_TECHNICIAN,
- date_complete_key referred to DIM_DATE_COMPLETE,
- folder_key referred to DIM_FOLDER,
- tag_key referred to DIM_SYSTEM,

They are respectively assigned into fact_progress as foreign keys.

4. Result and Discussion

In this section, we explain the ETL process for dimensions and for OLAP analysis. In brief, the process is getting data from csv exported from ECHO System, transformed as required then load them into dimensions and fact table, the transformation process is done by Pentaho Data Integration. ETL process consists converting some dimensions and fact process. It can be detailed as follows.

Dimension Add Date. This dimension is to store the addition check sheet date to the project as shown Table 1.

Table 1: Table structure of DIM_DATE_ADD

Field Name	Type	Description
DATE_ADD_KEY	INT	This is a surrogate key in a sequential number
DATE_FULL	DATE	Date with format: "dd-mm-yyyy"
MONTH	INT	Month index, i.e: 1,2,3,..12
YEAR	INT	Year value, i.e: 2015, 2016, … 2019
DATE_VALUE	INT	Date index in a month, i.e: 1,2,…30
MONTH_NAME	VARCHAR(20)	Month Name, i.e January, February, .. December
QUARTER_NAME	VARCHAR(2)	Quarter name, with the concatenation of Q and the number of Quarter, i.e: Q_1, Q2, .. Q4

Dimension Complete Date. This dimension is to store the complete date of check sheet to the project as shown Table 2.

Table 2: Table structure of DIM_DATE_COMPLETE

Field Name	Type	Description
DATE_COMPLETE_KEY	INT	This is a surrogate key in a sequential number
DATE_FULL	DATE	Date with format: "dd-mm-yyyy"
MONTH	INT	Month index, i.e: 1,2,3,..12
YEAR	INT	Year value, i.e: 2015, 2016, … 2019
DATE_VALUE	INT	Date index in a month, i.e: 1,2,…30
MONTH_NAME	VARCHAR (20)	Month Name, i.e January, February, .. December
QUARTER_NAME	VARCHAR (2)	Quarter name, with the concatenation of Q and the number of Quarter, i.e: Q1, Q2, .. Q4

The ETL process of Dim_Complete_data can be shown as shown in Figure 7.

Figure 7: Dimension Complete Date

Dimension Check Sheet. This dimension is to store the dimension of check sheet to the project as shown Table 3.

Table 3: Table structure of DIM_CHECKSHEET

Field Name	Type	Description
CHECKSHEET_KEY	INT	This is a surrogate key in a sequential number
DISCIPLINE	VARCHAR(20)	Code discipline of check sheet
CHECKSHEET	VARCHAR(95)	Description of discipline check sheet

The ETL process of Dim_CheckSheet can be shown as shown in Figure 8.

Figure 8: Dimension CheckSheet

Dimension Status. This dimension is to store list of status in the project as shown in Table 4

Table 4: Table structure of DIM_STATUS

Field Name	Type	Description
STATUS_KEY	INT	This is a surrogate key in a sequential number
STATUS	VARCHAR(20)	Detail of status, i.e: Outstanding, Completed, Accepted, etc.

Figure 9: Dimension Status

Dimension Tag. This dimension is to store the information tag in the project as shown Table 5

Table 5: Table structure of DIM_TAG

Field Name	Type	Description
TAG_KEY	INT	This is a surrogate key in a sequential number
TAG_CATEGORY	VARCHAR (20)	Consist of category tag, i.e: Building, Electrical cables, etc.
TAG	VARCHAR (100)	Detail of tag category, i.e: 1 phase transformer, AC Ups, etc.

The ETL process of Dim_Complete_data can be shown as shown in Figure 10.

Figure 10: Dimension Tag

Dimension Technician. This dimension is to store the information of technician and contractor in the project. It consists of:

Table 6: Table structure of DIM_TECHNICIAN

Field Name	Type	Description
TECHNICIAN_KEY	INT	This is a surrogate key in a sequential number
TECHNICIAN	VARCHAR (50)	Full name of technician
CONTRACTOR	VARCHAR (50)	Name of contractor

The ETL process of Dim_Technician can be shown as shown in Figure 11.

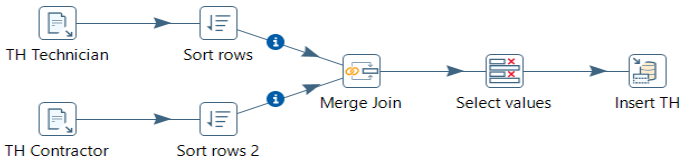


Figure 11: Dimension Technician

Table 7: Table structure of DIM_SYSTEM

Field Name	Type	Description
SYSTEM_KEY	INT	This is a surrogate key in a sequential number
SYSTEM	VARCHAR (20)	Description of system key, i.e: Water Spray System, Mooring Equipment
SUB_SYSTEM	VARCHAR (20)	Key of subsystem
SUBSYSTEM_DESC	VARCHAR (100)	Description of subsystem key, i.e : Mooring Equipment for Marine Terminal Control Shelter
TARGET_DATE	DATE	Target date with format: “dd-mm-yyyy”

The detail of Fact progress can be seen in Table 8.

Table 8: Table structure of FACT_PROGRESS

Field Name	Type	Description
CHECKSHEET_KEY	INT	cheecksheet_key is a key from dim_checksheet
TECHNICIAN_KEY	INT	technician_key is a key from dim_techncian
SYSTEM_KEY	INT	system_key is a key from dim_system
DATE_ADD_KEY	INT	date_add_key is a key from dim_date_add
DATE_COMPLETE_KEY	INT	date_complete_key is a key from dim_date_complete
STATUS_KEY	INT	status_key is a key from dim_status
FOLDER_KEY	INT	folder_key is a key from dim_folder
TAG_KEY	INT	tag_key is a key from dim_tag

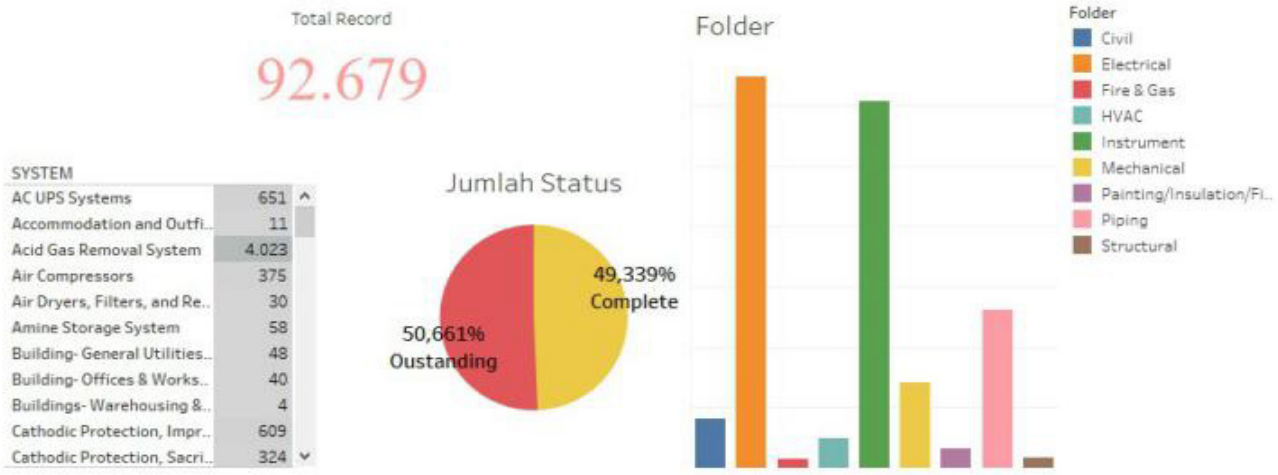


Figure 12: Dashboard Example

Report Generation. Once the star schema and its ETL Processes are defined, we then generate the report based on the new transformed table structure. One of the critical information required by the project is figuring the total number of check sheets required to be completed based on the project's system. We use Pentaho Report Generator as it is a dynamic and easy to use tool to generate dynamically from the database.

We can conclude in the project, that most of the work is in the Civil Onshore system which covered more than fifty percent, sixty six percent to be precise, and followed by the Dehydration System. Setting up the Dashboard. Beside report, dashboard can also be generated from tables in database. For data visualization and interactive dashboard, Tableau is used. Tableau's main products offered by Tableau are Tableau Desktop, Tableau Public, and Tableau Online, all offer Data Visual Creation and choice depends upon the type of work [16].

Figure 12 is an example dashboard that can be generated from all table in database. Bar chart shows how many folders that have been used. Bottom left is treemap, to select tenant name. And next to it is pie chart that shows which protocol is used. By clicking one of the charts, it will change the whole dashboard information based on the clicked condition. For example, if February is clicked tree map of tenant name will adjust which tenant sent bytes in February. The pie chart will do the same thing, it will show the protocol that is used in February.

From interactive dashboard, we instantly see the proportion of contractors assigned to a system as shown Figure 13. In this way we can see the interdependency between integrated project team with the contractors who are supporting the team on delivering their scope of work in an ordered manner.

5. Conclusion

By implementing data warehouse, ECHO with the dynamic tree table structure improves significantly to provide progress report and helps to provide informed decision making in managing the project completion by prioritizing on the work schedule and chasing the deadline by analyzing the best resource allocations.

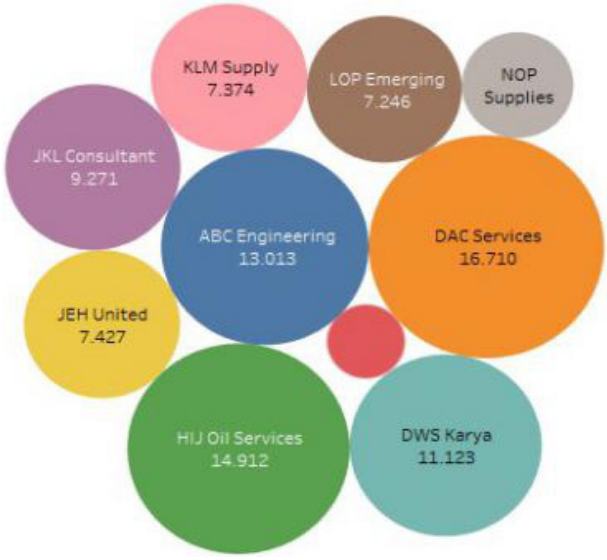


Figure 13: Progress Contractor Dashboard

For future research, the data warehouse then can be used to be analyzed with another method like projection on for another business process such as trend prediction, certificate completion, decision making on the resource allocation etc.

References

[1] Facilities Systems Completion Planning and Execution, American Petroleum Institute, **1**, 2013.

[2] S. Andriole, "Unstructured Data: The Other Side of Analytics," Forbes.[Online]. Available: Http://Www. Forbes. Com/Sites/Steveandriole/2015/03/05/the-Otherside-of-Analytics, 2015.

[3] N. Mali, S. Bojewar, "A Survey of ETL Tools," International Journal of Computer Techniques, **2**(5), 20–27, 2015.

[4] O. Andersen, C. Thomsen, K. Torp, "SimpleETL: ETL Processing by Simple Specifications.," in DOLAP, 2018.

[5] S. Kozielski, R. Wrembel, New trends in data warehousing and data analysis, Springer, 2009. https://doi.org/10.1007/978-0-387-87431-9

[6] H. Santos, V. Dantas, V. Furtado, P. Pinheiro, D.L. McGuinness, "From data to city indicators: A knowledge graph for supporting automatic generation of dashboards," in European Semantic Web Conference, Springer: 94–108, 2017, https://doi.org/10.1007/978-3-319-58451-5_7

[7] M. Breslin, "Data warehousing battle of the giants," Business Intelligence Journal, **7**, 6–20, 2004.
[8] R. Thiagarajan, J. Khuntia, A. Kathuria, T.J. V Saldanha, "An Integrated Model of Business Intelligence & Analytics Capabilities and Organizational Performance," Communications of the Association for Information Systems, **46**(1), 31, 2020. https://doi.org/10.17705/1CAIS.04631
[9] C. Sivamathi, S. Vijayarani, "A utility-based approach for business intelligence to discover beneficial itemsets with or without negative profit in retail business industry," International Journal of Business Intelligence and Data Mining, **16**(3), 361–380, 2020, https://doi.org/10.1504/ijbidm.2020.106136
[10] L. Yessad, A. Labiod, "Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault," in 2016 International Conference on System Reliability and Science (ICSRS), IEEE: 95–99, 2016, https://doi.org/10.1109/icsrs.2016.7815845
[11] R. Kimball, M. Ross, The data warehouse toolkit: the complete guide to dimensional modeling, John Wiley & Sons, 2011.
[12] M.A. Khan, S. Saqib, T. Alyas, A.U. Rehman, Y. Saeed, A. Zeb, M. Zareei, E.M. Mohamed, "Effective Demand Forecasting Model Using Business Intelligence Empowered With Machine Learning," IEEE Access, **8**, 116013–116023, 2020 https://doi.org/10.1109/access.2020.3003790
[13] W. Basri, "Examining the Impact of Artificial Intelligence (AI)-Assisted Social Media Marketing on the Performance of Small and Medium Enterprises: Toward Effective Business Management in the Saudi Arabian Context," International Journal of Computational Intelligence Systems, **13**(1), 142–152, 2020. https://doi.org/10.2991/ijcis.d.200127.002
[14] R. Kimball, "Factless fact tables," DBMS, **9**(10), 16–18, 1996.
[15] J.M. Villamarín-García, "Contributions from organisational collaboration to business intelligence solutions success," International Journal of Business Information Systems, **33**(1), 103–131, 2020. https://doi.org/10.1504/ijbis.2020.10012991
[16] M. Mussa, S. Souza, E. Freire, R. Cordeiro, H. Hora, "Business Intelligence in Education: An Application of Pentaho Software," Revista Produção e Desenvolvimento, **4**(3), 29–41, 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 52-62 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Development of Group-Based Differentiated Learning (GBDL) Models

Toto Ruhimat*,[1], Deni Darmawan[2]

[1]*Curriculum and Educational Technology, Department, Universitas Pendidikan Indonesia, Bandung, 40154, Indonesia*

[2]*Educational Technology and Communication Science, Department Universitas Pendidikan Indonesia, Bandung, 40154, Indonesia*

ARTICLE INFO

Article history:
Received: 17 August, 2020
Accepted: 15 October, 2020
Online: 08 November, 2020

Keywords:
Differentiated Group Learning
Thinking Ability
Social Ability

ABSTRACT

The ability to think and the social abilities of junior high school students is a potential ability developed through learning based on the optimization of differences in learning abilities. For this reason, appropriate learning design and implementation model is needed in realizing this, especially in one of the social group subjects. This study uses research and development methods that begin with a preliminary study, develops a draft model, tests, and experiments to test model validation. Data collection techniques with observation, experimentation, and interviews. The results showed that aspects of the model of differentiated learning design based on groups put forward the process of analyzing the characteristics of students about readiness, interests learning profile, flexibility for curriculum elements of content (content), process (process), and results (product). Learning objectives are developed in the form of measurable and specific abilities that describe student behavior. In particular, the learning design components include; (a) Analysis of characteristics and curriculum elements; (b) Learning objectives or competencies to be achieved; (c) Essential material; (d) Learning Process, and (e) Learning Evaluation. The Implementation Model is carried out through learning syntax, namely: orientation, general exploration, learning techniques, investigation, evaluation, and follow-up. Group learning model, in which students learn individually to collaborate in the first group after completing the second group, after the second group rejoining the first group.

1. Introduction

Improving quality in the National Education System should be in line with globalization, which has an impact on changes in social life, political science, and technology in various countries quickly. Globalization has several impacts, including free markets, cooperation and competition between countries, democracy, and the implementation of human rights. This is in line with [1], which focused on Intelligence, Globalization, Complex, and Multi-Level Society. These forces have implications for national education, particularly in the effectiveness and relevance of the education system. It also relates to the need for efficiency, innovative, and creative movements in improving education quality.

To realize the education quality, schools need to develop attitudes, abilities and provide essential knowledge and skills needed to live in a global society. Learning needs to be pursued based on an individual approach, though it does not neglect the students' social abilities (sociability), such as relationships with peers, cooperative, tolerance, and respect the opinion of others. The learning characteristics are in line with the concept of a democratic education that uses a scientific approach both in groups and individually [2], [3]. According to [4], individuals have differences in intelligence diversity, and student's potential needs attention [5].

Learning methods need to be developed based on the characteristics that accommodate student differences. Their implications need to be flexible and oriented towards the formation of diverse competencies in students. Learning should focus on optimizing students' abilities individually and prioritize their differences. Moreover, learning activities should be flexible and dynamic, democratic, fun, and active [6]. Essential material that needs to be developed in diverse learning can be classified based on the level of learning, such as the classification of facts,

*Corresponding Author: Toto Ruhimat, Indonesia University of Education, Street of Dr. Setiabudi, 229 Bandung, 40154, Phone (022)2013163, email: toto.ruhimat@upi.edu

concepts, principles, attitudes, and skills [7]. In the selection of subject matter, it is necessary to pay attention to the balance of ethics, logic, aesthetics, and kinesthetic. It is vital to ensure that the subject matter is developed as expected.

The preliminary study on learning with the above competencies in the Social Sciences established that most junior high school students felt the learning process was not optimal, especially the method used, development of learning materials, the use of time, and the utilization of media as learning resources. The learning process tends to be expository or informative, and therefore students are more positioned learning objects.

Learning should prioritize the needs and differences of students based on the desired outcomes. This relates to the aspects of readiness, interests, needs, and learning profiles of students [7]. These aspects can be used as a basis for student differences. The results of these aspects are useful in developing education programs and determining learning materials and approaches, such as group learning, which develops students' social abilities, including cooperation, tolerance, and understanding the opinions of others. Additionally, it provides opportunities for students to recognize and develop their abilities based on multiple intelligences [8]. In this context, [9] established that learning involves 4 variables, including presage, context, process, and product.

Based on the background description, it is necessary to research and develop learning models that optimize the thinking and social abilities of students corresponding to differences in individual potential. This means that the resulting learning model needs to appreciate, accommodate, and facilitate the potential of students individually without neglecting social abilities. Therefore, this study focuses on the following questions (1) What kind of group learning design model optimizes students' thinking and social abilities based on their potential? ; (2) What kind of learning implementation model optimizes students' thinking and social abilities based individual potential ?; and (3) What kind of learning environment standards can optimize students' thinking and social abilities based on individual potential?.

2. Theoretical Review

2.1. The Essence of Student Differences

Students are individuals with unique differences from each other. [2] identified individual differences based on the multiple intelligences. The pattern of multiple intelligences is a form of curriculum organization that can be independently developed and guided by each individual. It can as well be perceived as an essential ability of each individual, which needs to be developed early for future survival. The characteristics of gifted students[10] can be identified from the abilities possessed, such as (1) general intellectual, (2) specific academic aptitude, (3) creative or productive thinking, (4) leadership skills, (5) visual and performing arts, (6) psychomotor abilities. According to Renzulli, a minimum of 3 to 5 percent of the students in schools have this ability.

2.2. Learning Procedures for the Development of Social Ability

The development of effective social abilities can be carried out using group or cooperative learning. [11] established that the teaching model can be grouped into 4, including information processing, personal, social, and behavioral groups. All student activities in learning programs have to be democratic, meaning that the learning process should develop the ability as individuals and have a beneficial social ability, which is beneficial for students. According to [12], social ability [13], including interpersonal intelligence, involves understanding and interacting with others.

The group learning approach is an alternative that can be developed in this regard. The stages of group investigation learning include (1) Students encounter puzzling situation (planned or unplanned); (2) exploring reactions to the situation; (3) formulating study tasks and organize for study (problem definition, role, assignment); (4) independent and group study; (5) analyzing progress and process and recycle activity [14].

Stages of learning are developed in ways that demand problem-solving and inquiry in learning. In general, students are organized into democratic problem-solving groups that attack academic problems and are taught procedures and scientific methods of inquiry [15]. This is in line with the concept of the group investigation approach, which involves placing students in the context of a social system with democratic and cooperative processes and provides opportunities for individuals to learn optimally. According to [16] students are provided the opportunity to learn based on higher-order thinking processes such as analytical, hypothesis, and scientific, especially in the learning process through problem-solving and inquiry. Cooperative learning is organized to conduct student activities in small groups in collaboration between students in learning [12].

3. Research Method

This study used research and development methods in education [17]. The implementation started from a preliminary study, developing a draft model, and testing and conducting experiments for model validation. Data collection techniques used include observation, experimentation, and interviews, while Relevant, Observable, useful, Realizable, Understandable, and demand by teachers and students tests were used for analysis.

Table 1: Research Samples Size

No	Name of School	Status	Size
1	State Junior High School 15	Experiment class	46
2	State Junior High School 10	Experiment class	46
3	State Junior High School 11	Control Class	46
4	Nugraha Junior High School	Control Class	46
		Total Sample	184

There were several reasons for choosing and using the junior high school level, including 1) according to the level of cognitive ability development, junior high school students (between 11-15 years old) have learned to think scientifically or rationally; 2) based on the level of social ability development, junior high school students can be independent, collaborate, socialize, lead and be responsible in groups; 3) junior high school students can carry out group activities; 4) their curriculum contains social science subjects especially geography; 5) in the social science curriculum

objectives, students need to develop the ability to reason, understand and solve problems; 6) social sciences allow variety, flexibility, and modification of learning content, process or product; and 7) junior high school already has a Subject Teacher Forum. Assessment technique of thinking and social abilities referred to the [18] model. The analysis data was used throughquantitative techniques with correlations, t-tests, and percentages.

4. Result and Discussion

4.1. Differentiated Group Learning Model Design

From the testing results conducted during the Research & Development [19], a learning design model matrix for Differentiated Group learning is formulated as follows.

Table 2: Syntax of Learning Design Stage

Component	Development Aspects	Criteria
Learning Purpose	- Able to identify and study the patterns of economic activity based on land use and region - Able to identify the distribution of forestry in Indonesia - Able to identify obstacles of forestry preservation in Indonesia - Able to identify the responsibility of the government in preserving forestry in Indonesia - Able to identify the responsibility of the community in preserving forestry in Indonesia - Able to identify problems and obstacles of forestry in Indonesia	- The design of learning purpose in the form of competencies - The purpose is developed in stages and flexibility - The purposes are oriented towards general learning programs
Learning Material	- Patterns of economic activity based on land use and region - Distribution of forestry in Indonesia - Constraints on forest preservation in Indonesia - Responsibility of the government in preserving forestry in Indonesia - Responsibility of the community in preserving forestry in Indonesia - Forestry issues and constraints in Indonesia	- Facilitating the choices of subject matter - Use various sources - Problem-based content - The material tends to be sought, built, or compiled by students - Contextual
Learning Activity	Development of Learning Design with Stages - Orientations : o The teacher creates the initial conditions for learning Social Sciences o The teacher conveys topics or sub-problems to be discussed on the patterns of economic activity based on land use and region o The teacher implements apperception, and students are allowed to ask questions or answer related to patterns of economic activity based on land use and region o The teacher provides a stimulus to be active in studying, discussing, and learning individually or in groups - General exploration o The teacher conveys a general description of the patterns of economic activity based on land use and region, explained in general to the sub-topics discussed o The teacher provides stimulus, motivation, and attention to students - Mastery of Learning Techniques (If students have mastered the learning technique, then this stage the teacher does not need to explain again unless deemed necessary) o The teacher explains learning techniques individually and in groups (cooperative) o The teacher explains effective learning individually and cooperatively o The teacher explains the stages of problem-solving or inquiry or group study.	The design of learning activities applies to the following principles : 1. Competency-oriented: the design of student activities is aimed at achieving competence in learning purpose 2. Problem-based learning: the design of student activities should be based on problems and issues to be studied. 3. Individual learning in a group. Student activities are designed individually and collaborate in groups 4. Students understand individual and group learning techniques. The teacher designs the implementation of activities, students should understand the learning techniques to be pursued 5. Flexibility. Both individual and group studies should be clear to students but can be flexible and modifiable 6. Position of teacher activity: facilitator and resource person collaborating

Learning Activity	○ The teacher explains how to use the worksheets or find learning resources related to Social Sciences, especially the topics discussed ○ The students are already in their respective groups. - Investigation (This is the core stage that needs to be carried out individually and cooperatively in assessing or solving problems) ○ Using groups and using student worksheets, students formulate general problems on "what is the pattern of economic activity based on land use and territory in Indonesia"? ○ The teacher restates the sub-topics in the form of problems that should be solved by students as follows • How is the distribution of forestry in Indonesia? • What are the obstacles to forestry preservation in Indonesia? • What effort needs to be carried out by the government in preserving forestry in Indonesia? • What efforts need to be carried out by the community in preserving forestry in Indonesia? • What are the problems and obstacles of forestry in Indonesia? ○ After understanding a general problem, students individually choose one or more sub-topics to solve or study the problem in the group. ○ Continuously monitored, guided, and assessed by the teacher, students follow the stages of learning Scientific thinking (problem-solving) according to the worksheet guidelines ○ After individual studies and problem-solving are completed, the results of individual work are discussed in groups (with discussions equating concepts, complementing each data and information) ○ Students individually process the results of the study of other members in the group and individually answer other sub-questions. ○ Each student makes a report (portfolio) of answers to common problems and all sub-problems. - Evaluations (Assessment is carried out continuously in the learning process) • Provide an essay or filling-gap test • Each group under the teacher's guidance summarizes the results of the discussion • Collecting the results of individual and group work. - Follow-up • Give assignments or enrichment individually to students deemed necessary • Announce the topics to be discussed in the next meeting	
Learning Outcomes	Assessment Procedure The assessment is carried out continuously IInnitial, process, and final tests Types of Assessment Forms, The assessment uses observation, written, oral, and portfolio forms.	- The assessment design prioritizes to assess the learning process apart from the input and results. - Assessment is designed as a basis for improvement, enrichment, and diagnostics. - The results of the assessment as a modifiable elements. Portfolio assessment. - Group and individual activities

The Group-Based Differentiated Learning Design was developed based on the aspects of Learning syntax as follows; (a) Analysis of students' characteristics of interest, readiness, and learning profile; (b) Analysis of curriculum elements (content, process, and product) that can be modified; (c) Conditions of the learning environment; (d) Assessment is designed continuously and integrated; (e) Objectives formulated and developed in the form of abilities; (f) Development of learning activities refers to the stages of orientation, general exploration, mastery of learning techniques, investigation, evaluation, and follow-up; (g) The curriculum components in learning design prioritizes the components of objectives, the content of subject matter, learning activities, and evaluation of learning developed systematically and systemically; (h) The format of the learning design can be developed flexibly depending on the needs of developing learning programs and the applicable curriculum. This finding is in line with [16], which focused on the balancing acts of building positive relationships with students.

Student Grouping Design was carried out in 2 stages, the first and the second groups. The first group resolves general problems with individuals who collaborate, and the product is a general problem-solving portfolio. This finding is supported by[16], which stated that learning outcomes serve as references for preceding teaching and assessment activities that help students construct learning. Furthermore, students in the first group choose one or two sub problems to be studied more deeply in the second group whose members chose the same sub-problem. The product of this second group is the portfolio of individual studies on selected sub-problems. Once the second group is completed, students return to join the first group to study and formulate more profound conclusions on the general problems.

Table 3: Syntax of Learning Implementation

Procedure	Details	Teacher's Scenario
Orientation (1)	Creating conditions for effective learning	- It starts from the beginning of learning by creating a conducive classroom atmosphere for readiness to take lessons. - Direct the concentration of students and encourage them to be interested in learning. - As a teacher, show attitude and enthusiasm with full of motivation.
	Apperception	- Observe the students' initial abilities on the topic to be studied
		- Allow students to ask questions and express opinions related to the topic.
	Delivering learning purpose	- Deliver the topics or purpose to be discussed in learning.
General exploration (2)	Delivering general explanation	- Deliver to the students the general description of the topic with or without learning media.
	Conducting question and answer (dialogue) about the topics to be discussed	- Give stimulus to students to make them respond, ask questions (dialogue), or give opinions on topics to be discussed.
		- Invite students to understand or discuss the topics briefly on the problem to be solved.
Mastery of learning techniques (3)	(*this stage should not always be conducted in case students master the learning techniques to be used or carried out at previous meetings. The teacher's task is only to provide an operational understanding of the learning techniques*)	
	Explain learning techniques	- Explain the techniques to be used, operational procedures, and an overview of learning outcomes.
		- Give stimulus to students for them to ask questions related to the learning method used.
		- Give direction and guidance to students on how to learn in groups, the importance of collaboration and responsibilities in groups.
	Explain students worksheet	- Briefly discuss the operational use of students worksheet
	Explain group - individual learning techniques	- Briefly explain the individual learning in groups. - Form groups using the jigsaw method
Investigation (4)	(*this stage is the core learning activity and should be carried out individually by students collaborating in groups*)	
	Preparation of learning activities	- Observe that students are ready in the 1st group to start learning.

		- When students discuss worksheets, re-explain to students about the problems that should be answered individually or in groups.
	Conditioning problem solving and collaboration activities	- Motivate students to be serious in solving the problem individually in a group.
	Implement the learning techniques and the process of mastering lesson content	- Guide students individually in the 1st group to solve problems with scientific thinking procedures (formulate problems and sub-problems, collect data, study data, and formulate conclusions). Individual student work is written in portfolio sheet 1
		- Conditioning help students choose one sub-topic or problem to be studied with other groups (second group)
		- With the teacher's direction, students join the second group studying the same sub-topic or problem. The results of individual studies are written in portfolio sheet 2
	Monitor the learning process	- Walk around each group to observe, and give individual or group direction.
	Helping students difficulties	- Help students if they need learning resources that are necessary for the class.
	Process / performance evaluation	- Conduct a process assessment of the activities carried out in groups.
		- Remind them in case time for the activity is almost up.
		- After completing the study in the second group, direct students to rejoin the 1st group to answer general problems. Make the group study results in portfolio sheet 3
Evaluation and follow up (5)	Carry out the final test	- Give questions to several students as a final oral test
	Collect the results of individual and group work	- Collection of student work individually or in groups
	Deliver the topic of the next subject	- Deliver the topic to be studied in the next meeting

4.2. Implementation Model

Learning Implementation Model with Differentiated Group Learning in social science subjects at junior high schools can be described as follows; (a) Teacher and student activities have increased compared to conventional learning; (b) Procedures to be carried out in implementation include orientation, general exploration, mastery of learning techniques, investigation, evaluation, and follow-up; (c) The flexibility of curriculum elements (content, process, and results) is likely to be more dominant during the learning process; (d) Eliminating the weaknesses of both individual and group approaches; (e) Students are more confident.

Table 4: Syntax of Learning Implementation

Procedure	Details	Teacher's Scenario
Orientation (1)	Creating conditions for effective learning	- It starts from the beginning of learning by creating a conducive classroom atmosphere for readiness to take lessons. - Direct the concentration of students and encourage them to be interested in learning. - As a teacher, show attitude and enthusiasm with full of motivation.
	Apperception	- Observe the students' initial abilities on the topic to be studied
		- Allow students to ask questions and express opinions related to the topic.
	Delivering learning purpose	- Deliver the topics or purpose to be discussed in learning.
	Delivering general explanation	- Deliver to the students the general description of the topic with or without learning media.

General exploration (2)	Conducting question and answer (dialogue) about the topics to be discussed	- Give stimulus to students to make them respond, ask questions (dialogue), or give opinions on topics to be discussed.
		- Invite students to understand or discuss the topics briefly on the problem to be solved.
Mastery of learning techniques (3)	(*this stage should not always be conducted in case students master the learning techniques to be used or carried out at previous meetings. The teacher's task is only to provide an operational understanding of the learning techniques*)	
	Explain learning techniques	- Explain the techniques to be used, operational procedures, and an overview of learning outcomes.
		- Give stimulus to students for them to ask questions related to the learning method used.
		- Give direction and guidance to students on how to learn in groups, the importance of collaboration and responsibilities in groups.
	Explain students worksheet	- Briefly discuss the operational use of students worksheet
	Explain group - individual learning techniques	- Briefly explain the individual learning in groups. - Form groups using the jigsaw method
Investigation (4)	(*this stage is the core learning activity and should be carried out individually by students collaborating in groups*)	
	Preparation of learning activities	- Observe that students are ready in the 1st group to start learning.
		- When students discuss worksheets, re-explain to students about the problems that should be answered individually or in groups.
	Conditioning problem solving and collaboration activities	- Motivate students to be serious in solving the problem individually in a group.
	Implement the learning techniques and the process of mastering lesson content	- Guide students individually in the 1st group to solve problems with scientific thinking procedures (formulate problems and sub-problems, collect data, study data, and formulate conclusions). Individual student work is written in portfolio sheet 1
		- Conditioning help students choose one sub-topic or problem to be studied with other groups (second group)
		- With the teacher's direction, students join the second group studying the same sub-topic or problem. The results of individual studies are written in portfolio sheet 2
	Monitor the learning process	- Walk around each group to observe, and give individual or group direction.
	Helping students difficulties	- Help students if they need learning resources that are necessary for the class.
	Process / performance evaluation	- Conduct a process assessment of the activities carried out in groups.
		- Remind them in case time for the activity is almost up.
		- After completing the study in the second group, direct students to rejoin the 1st group to answer general problems. Make the group study results in portfolio sheet 3
	Carry out the final test	- Give questions to several students as a final oral test
	Collect the results of individual and group work	- Collection of student work individually or in groups
	Deliver the topic of the next subject	- Deliver the topic to be studied in the next meeting

4.3. Learning Environment Model

The environmental condition contributes to the effectiveness of learning. The scope of the learning environment that supports starts from the conditions of the group, classroom, school, and community. Environmental standards and learning conditions that need attention include; 1) pleasant atmosphere in class; 2) activity and motivation in groups and class; 3) democratic class atmosphere; 4) discipline in the classroom; 5) effective school management; and 6) conducive community conditions. This finding is in line with [16] which detailed the support for learning environments, including a) pre-lesson conferences for joint lesson planning, b) core issues for lesson designs, c) both elements, or d) another educational topic (control group), effects on the quality of collaborative exchange in lesson conferences, student teachers' competency gains, and instructional quality.

Learning Environment Models [20] established in the differentiated group shows the existence of media empowerment and maximum learning resources support its success. Students are guided to search for and study data and information related to the topic, based on[21]. During the experiment class, especially at the second meeting, the data and information were obtained not only from textbooks but also from various sources. In general, the information contained in geography textbooks at that time tended to be less appropriate to the latest developments, such as charts, distribution maps, growth charts, and factual data [22]. During the second meeting, students and teachers were gradually encouraged to complete data and information that suits the latest developments.

4.4. Model Testing Results in measuring Students' Thinking and Social abilities

- Cognitive Ability

The level of students' ability to solve issues is evident in the process, and the results of problem-solving described in Using Metaphors to Aid Student Meta-Learning [16]. The results of problem-solving are classified by categories, including (1) knowledge ability, such as remembering data and information, (2) the ability to understand data and information, and (3) analysis and synthesis aptitude. From the first to the third learning on the large-scale testing of students' abilities, there was progress towards a higher level. In the first learning, students' ability mainly involved knowledge (cognitive; memory), while the ability of analysis, synthesis, or application was very little. In the subsequent learning, the analysis ability gradually increased. Many students showed an increase in the analysis and application ability in the third learning. This finding was in line with[16], which established that critical thinking is essential for success in all human endeavors. The details are shown in the table below.

Table 5: Recapitulation of Student Cognitive Ability Results Based on Study Large Scale Learning Testing Results

Ability level	Learning Stage		
	The First	The Second	The Third
Knowledge	62.16 %	39.61 %	22.86 %
Understanding	28.4 %	37.83%	48.98 %
Analysis & application.	9.42 %	22.5 %	28.15 %

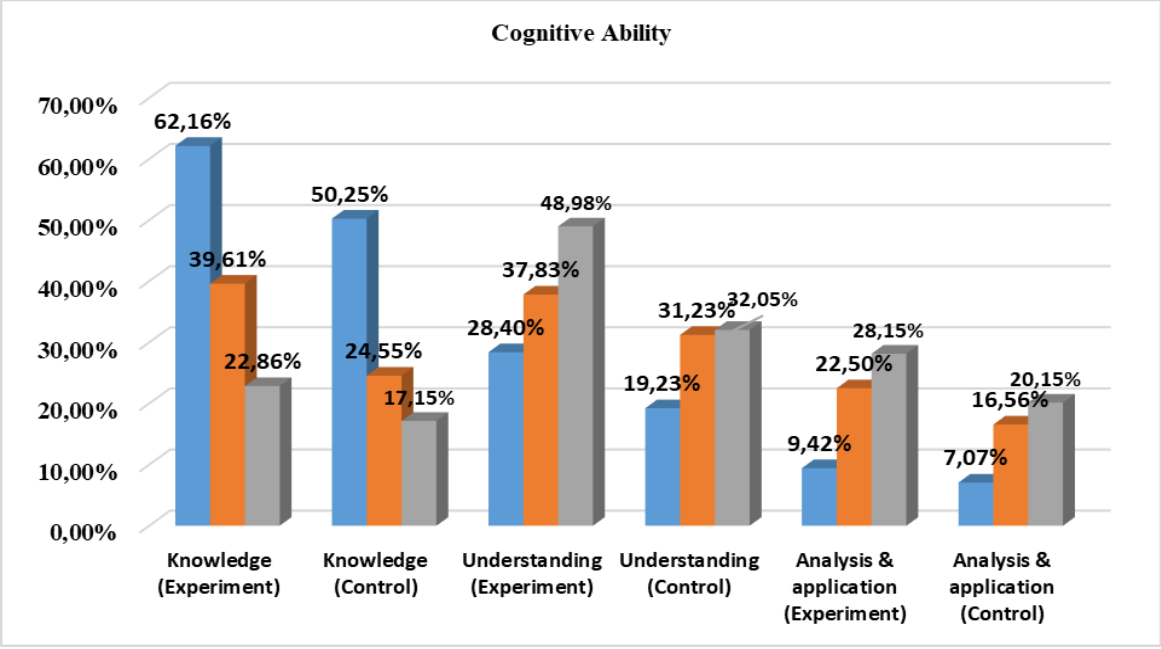


Figure 1: Student's Cognitive Ability Results Based on Experiment and Control Groups

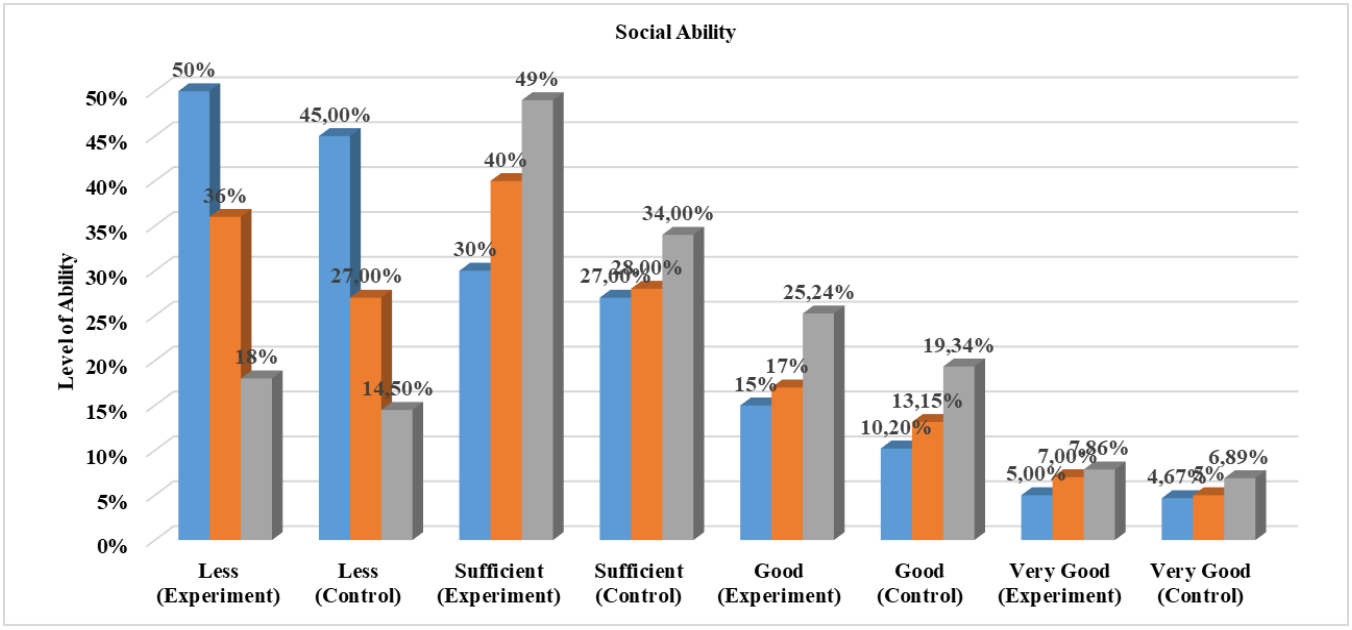


Figure 2: Student's Social Ability Results Based on Experiment and Control Groups

Table 6: Recapitulation of Students' Social Ability Results Based on the Study of Large Scale Learning Testing Results

Ability level	Learning Stage		
	The First	The Second	The Third
Less	50%	36%	18%
Sufficient	30%	40%	49%
Good	15%	17%	25.24%
Very Good	5	7	7.86

Table 7: Objective Test Results in Experiment and Control Groups

	N	Minimum	Maximum	Mean	Std.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Deviation Statistic	Statistic	Std. Error	Statistic	Std. Error
Pretest-Experiment	92	2.80	5.20	3.9641	.52927	-.063	.251	-.485	.498
Pretest-Control	92	2.80	5.00	3.7935	.61301	.404	.251	-.843	.498
Post Test-Experiment	92	6.00	8.40	7.4196	.50758	-.170	.251	-.683	.498
Posttest-Control	92	5.80	8.00	6.8717	.53807	-.030	.251	-.394	.498
School	92	1	2	1.50	.503.503	.000	.251	-2.045	.498
Gain-Experiment	92	.20	.71	.4891	.11282	-.600	.251	.153	.498
Gain-Control	92	.37	.74	.5693	.09251	-.283	.251	-.813	.498
Valid N (list wise)	92								

Table 8: Descriptive Test Result in Experimental and Control Groups

	N	Mean	Std.	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Deviation Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Pretest-Experiment	92	4.2239	.57919	.335	-.027	.251	-.295	.498
Pretest-Control	92	4.1185	.64124	.11	.145	.251	-.893	.498
Post Test-Experiment	92	7.2978	1.0059	1.012	-.604	.251	-.009	.498
Posttest-Control	92	5.9793	1.04729	1.097	-.264	.251	-1.049	.498
Gain-Experiment	92	.5319	.15943	.025	-.359	.251	.263	.498
Gain-Control	92	.3179	.16273	.026	-.117	.251	-.188	.498
Valid N (list wise)	92							

Table 5 and figure 1, shows that in the first learning, the results of problem-solving by the student (individual) were more on knowledge (remembering) (62.16%). The ability to analyze was only around 39%. This finding was in line with (Savage 2019), which focused on student achievement and [23] on high-level thinking and learning outcomes of learners who have been analyzed using N-Gain.

- Social Ability

The table above shows that in learning, the level of social ability of some students was less (50%). Students with excellent social abilities were only 15%. In contrast to the third learning, the level of social ability of students was high.

The table.7 above shows that the averages for the pretest, posttest, and gain for the experimental group were 3.9641; 7.4196, and 0.5693, while the control groups had 3.7935; 6.8717, and 0.4891 respectively. Based on these results, the average ability of the experimental group both at pretest and posttest was slightly higher than the control group. Another aspect that should be considered is data dissemination. Data (pretest, posttest, and gain) in the control class tended to be more spread than in the experimental group. The description of learning outcomes data is shown in the following table 8.

Analysis of learning outcomes through tests with description is almost the same as the tests with an objective form, showing the consistency of students' abilities in learning outcomes. Based on the size of the central tendency, the average ability of the experimental group was higher than the control group, both in the pretest, post-test, and gain.

This research and development were meant to produce a learning model that could serve students' differences and optimize individual learners' potential in thinking and social abilities. The essence of the model was learning based on students' and cooperative differences. Several assumptions underlie this learning, including 1) Individual development is a matter that needs to be considered in teaching, in this case, the position of students as individuals and social beings who need to socialize [24], [25]; 2) in learning it is necessary to develop democratic processes that can be carried out by students through scientific methods, [26]; 3) through group study, students can exchange knowledge, experience, and views, develop feelings, attitudes, and beliefs, working together and increasing motivation[14]; 4) 11-15 years old children, including those in the formal operation phase, have stages of intellectual development that need problem-solving techniques in learning (Piaget); 5) learning implementation needs to recognize and develop students 'abilities based on their multiple intelligence [13].

This learning model can be implemented in regular classes or the Class of Excellence. It can optimize students' abilities through acceleration, escalation, or enrichment, and can serve learning activities that are compatible with the diversity of students' interests. It can assist teachers to develop and improve students' abilities in regular classes, this result support from [27]. This is proven by students who participated in the learning model. Many students (> 50%) gained improvement and deepening from the ability of basic level knowledge to analyze and synthesize, [28], [29]. The formation of rational abilities can be developed through problem-solving activities in teaching and learning events adjusted to the level of student progress, this result support by [30]. The problem-solving activities by investigating (reviewing) data and information from the media and learning resources,[31] were appropriate and can be implemented to junior high school students. The investigative approach in learning is an integral part of problem-solving activities that implement the principles of democracy to the scientific method and inquiry as a learning process,[32]. The characteristics of learning activities developed further provide opportunities for individual learning that synergizes with groups to be carried out optimally according to student potential. This finding was in line with [33], which discussed how to support learning for measurable positive impacts on teaching quality and teacher morale. Therefore, it is necessary to implement the principles of democracy, competitive, and cooperative developed in the learning process, which leads to the formation of higher-order thinking skills and social abilities.

5. Conclusion and Recommendation

GBDL models which optimize students' thinking and social abilities based on an individual potential focus on aspects of competency; Problem based learning and individual learning in groups. Students understand the individual and group learning techniques, as well as the teacher design. The learning objectives are developed in the form of measurable and specific abilities describing student behavior. The design of the learning process is developed in stages systematically and follows several activities, including orientation, general exploration; practice (mastery) of learning techniques; investigation, evaluation, and; follow-up. Learning design components of GDBL include (a) Analysis of characteristics and curriculum elements; (b) Learning objectives or competencies to be achieved; (c) Essential material; (d) Learning Process; (e) Learning Evaluation. Learning implementation models of GBDL which optimizes students' thinking and social abilities are developed in the following stages, (a) Learning syntax procedures; (b) An integrated manner guided by student worksheet (SW) and the teacher; (c) The group approach is implemented into 2 models, the first and the second groups. The process requires students to learn individually, collaborating in groups; (d) Alternative effective learning strategies used in this learning include problem-solving, inquiry, and group study. Learning environment standards that optimize students' thinking and social abilities based on an individual potential focus on the standard conditions of the learning environment relating to students, teachers, and essential subject matter. This makes the learning environment to be oriented to the activities of the study and the creation of relevant learning resources.

The findings of the GBDL model are expected to be used by teachers when teaching with the group learning method. Especially in social science learning which refers to the use of natural learning resources at the junior high school level.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The author would like to express gratitude to the Ministry of Research and Technology of Indonesia that has provided research

grants of Post graduate School of the Universitas Pendidikan Indonesia, who has permitted to complete this research to provide benefits in the world of education.

References

[1] A. C. Vilasi, "Intelligence, Globalization, Complex and Multi-Level Society," Open J. Polit. Sci., **08**(1), 47–56, 2018, doi: 10.4236/ojps.2018.81004.

[2] J. A. de Ruiter, A. M. G. Poorthuis, and H. M. Y. Koomen, "Relevant classroom events for teachers: A study of student characteristics, student behaviors, and associated teacher emotions," Teach. Teach. Educ., **86**(2), p. 102899, 2019, doi: 10.1016/j.tate.2019.102899.

[3] V. B. Ngalim, "Harmonization of the Educational Sub-Systems of Cameroon: A Multicultural Perspective for Democratic Education," . Creative Education, **04**(1), 334–346, 2014.

[4] S. Mohammad and A. Jado, "The Level of Multiple Intelligences in Arabic Language Textbooks for Grades from (1 - 4) in Jordan in Light of Gardner ' s Theory," **08**(1), 1558–1572, 2015.

[5] M. R. Wadaani, "Teaching for Creativity as Human Development toward Self-Actualization : The Essence of Authentic Learning and Optimal Growth for All Students,"., **07**(5), 669–679, 2015.

[6] M. E. M. Bernardes, "Active Teaching Approach: Teaching and Learning Methods on Historical and Cultural Theory," Creat. Educ., **05**(10), 734–742, 2014, doi: 10.4236/ce.2014.510086.

[7] C. A. Tomlinson, "The differrentiated classroom, Responding to needs of all learner.," Alexandria, VA. USA.ASCD, 2019.

[8] Gardner.H., "Multiple Intelligences the Theory in Practice," New york Prentice-Hall Co., 1993.

[9] D. M. Biddle.B, "The Studi of Teaching," New York Holt, Renehart Winston., 2004.

[10] F. Hellen et al., "Working with Creativity of Gifted Students through Ludic Teaching,". Online Submission, **07**(2), 1641–1647, 2016.

[11] F. Egert, V. Dederer, and R. G. Fukkink, "The impact of in-service professional development on the quality of teacher-child interactions in early education and care : A meta- analysis," Educ. Res. Rev., **29**(11), 100309, 2020, doi: 10.1016/j.edurev.2019.100309.

[12] T. Stoltz et al., "Creativity in Gifted Education : Contributions from Vygotsky and Piaget,".,**05**(1), 64–70, 2015.

[13] S. S. Kavanagh, C. Monte-Sano, A. Reisman, B. Fogo, S. McGrew, and P. Cipparone, "Teaching content in practice: Investigating rehearsals of social studies discussions," Teach. Teach. Educ., **86**(2), 102863, 2019, doi: 10.1016/j.tate.2019.06.017.

[14] B. and W. M. Joyce, Models of Teaching, New Jersey Prentice Hall., 1996.

[15] C. C. Deneen, G. W. Fulmer, G. T. L. Brown, K. Tan, W. S. Leong, and H. Y. Tay, "Value, practice and proficiency: Teachers' complex relationship with assessment for learning," Teach. Teach. Educ., **80**(2), 39–47, 2019, doi: 10.1016/j.tate.2018.12.022.

[16] H. P. A. Boshuizen, H. Gruber, and J. Strasser, "Knowledge restructuring through case processing: The key to generalise expertise development theory across domains?," Educ. Res. Rev., **29**(1), 100310, 2020, doi: 10.1016/j.edurev.2020.100310.

[17] W. R. B. Gall, Meredith D., J.P. Gall, Educational Research: An Introduction, NY Pearson Educ. Inc., 2003.

[18] L. W. & K. D. R. A. Anderson, Taxonomy for Learning Teaching and Assessing, New York Addison Wesley Longman,Inc, 2002.

[19] M. Borg, R.W., Gall, Educational Research An Introduction. 5ed, Longman, Newyork, 2019.

[20] W. Zhao, "Paradigm of Foreign Language Teaching and Learning : A Perspective of Self-Regulated Learning Environment Construction," **05**(1), 167–177, 2016.

[21] N. M. Ardoin and A. W. Bowers, "Early childhood environmental education : A systematic review of the research literature," Educ. Res. Rev., **31** (7), 100353, 2020, doi: 10.1016/j.edurev.2020.100353.

[22] C. Kivunja, "Creative Engagement of Digital Learners with Gardner's Bodily-Kinesthetic Intelligence to Enhance Their Critical Thinking," Creat. Educ., **06**(6), 612–622, 2015, doi: 10.4236/ce.2015.66060.

[23] A. Ambarita, E. Liana, and C. Ertikanto, "Development Problem Based Learning to Increase Skill Thinking High Levels," Open J. Soc. Sci., **06**(10), 34–49, 2018, doi: 10.4236/jss.2018.610004.

[24] O. Costache, E. Becker, F. Staub, and T. Mainhard, "Using uncertainty as a learning opportunity during pre-lesson conferences in the teaching practicum," Teach. Teach. Educ., **86**(2), 102890, 2019, doi: 10.1016/j.tate.2019.102890.

[25] H. Taba, Curriculum Development: theory and practice, New York Harcourt, Brace world, Inc Atlanta., 1962.

[26] J. Dewey, How We Think, Heeath. Bost., 1992.

[27] M. D. Ilie et al., "Reviewing the research on instructional development programs for academics . Trying to tell a different story : A meta-analysis," Educ. Res. Rev., **30** (3), 100331, 2020, doi: 10.1016/j.edurev.2020.100331.

[28] Stamatios Papadakis and Michail Kalogiannakis, "A Research Synthesis of the Real Value of Self-Proclaimed Mobile Educational Applications for Young Children," Mob. Learn. Appl. Early Child. Educ., **07**(5),p. 19, 2020, doi: 10.4018/978-1-7998-1486-3.ch001.

[29] W. Pardjono, "Improving the Capability of Analysis, Synthesis, and Evaluation Through Problem Solving Learning," Yogyakarta Univ. Negeri Yogyakarta, 2009.

[30] M. A. Veldman, S. Doolaard, R. J. Bosker, and T. A. B. Snijders, "Young children working together . Cooperative learning effects on group work of children in Grade 1 of primary education," Learn. Instr., **67**(2), 101308, 2020, doi: 10.1016/j.learninstruc.2020.101308.

[31] J. M. Thomson, N. Foldnes, P. Henning, M. Njå, O. Judith, and K. Lundetræ, "Can children ' s instructional gameplay activity be used as a predictive indicator of reading skills ?," Learn. Instr., **68**(11), 101348, 2020, doi: 10.1016/j.learninstruc.2020.101348.

[32] S. Becker, P. Klein, A. Gößling, and J. Kuhn, "Using mobile devices to enhance inquiry-based learning processes ☆," Learn. Instr., **69**(6), 101350, 2020, doi: 10.1016/j.learninstruc.2020.101350.

[33] J. Gore, A. Lloyd, M. Smith, J. Bowe, H. Ellis, and D. Lubans, "Effects of professional development on the quality of teaching: Results from a randomised controlled trial of Quality Teaching Rounds," Teach. Teach. Educ., **68**(8), 99–113, 2017, doi: 10.1016/j.tate.2017.08.007.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 63-71 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

The COVID-19 Effect in Mexican SMEs

Abel García-Villagrán*, Patricia Cano-Olivos, José Luis Martínez-Flores, Diana Sánchez-Partida

Universidad Popular Autónoma del Estado de Puebla, UPAEP, Puebla, 72410, México

ARTICLE INFO

Article history:
Received: 27 August, 2020
Accepted: 15 October, 2020
Online: 08 November, 2020

Keywords:
COVID-19
Resilience
SME´s
Supply Chain

ABSTRACT

This article presents a literature review on how the COVID-19 pandemic changed the way of managing risks in the supply chains of Mexican SMEs. Research was carried out in the most important scientific research databases. The spread of the pandemic in 2020 has affected public health and the world´s companies, including micro, small and medium enterprises (SMEs). The impact on business activities is tremendous, and stringent government policy and response to curb the disease are necessary. SMEs are more vulnerable to internal and external events and depend on the capacity of their owner-manager to run the businesses. The Coronavirus is already considered a factor that requires attention on the behalf of Mexican companies. It is now difficult to imagine a world that isn't permanently changed, and organizations need to adapt. The normality of 2019 is not going to return, and Mexico has a fragile economy and a low-budget health system. The SME supply chains contain many links. At the moment of being struck by a crisis such as the present one, we are experiencing that disruption is likely to occur at the weakest link. Supply chain resilience is the adaptive capacity to prepare for unexpected events, respond to disruption, and recover. A Mexican SME that wishes to be resilient must know all the risks that can affect it. An opportunity for SMEs due to the coronavirus refers to the possibility for continuous innovation through the development of new or better products, processes, and business models. Resilience and risk thinking are life-saving phenomena for Mexican SME´s.

1. Introduction

The coronavirus (COVID-19) is a viral infection that emerged at the end of 2019 in Wuhan, China, and has spread to all regions of the planet. It is highly transmissible and causes severe respiratory problems. The signs and symptoms are somewhat similar to influenza and seasonal allergies. There is no clinically approved antiviral drug or vaccine available at this moment. Many companies are working on the development of an effective cure [1]. The coronavirus outbreak is a tragedy that is affecting millions of the world's inhabitants. The pandemic also had negative impacts on the global economy, industries, corporations, and small and medium-sized businesses [2]. Many thousands have died, others have lost friends or family members, and many businesses have ceased to exist [3].

The World Health Organization (WHO) declared the problem a health emergency and classified it as a pandemic at the international level. Containment measures were put in place, such as the closing of factories and offices, cancellation of trips, and a mandatory at-home quarantine was even imposed, leading to an unprecedented economic crisis [4]. In all countries, people are asked to choose privacy and health, that is, protecting their health and staying home to stop the spread. When the population trusts the public authorities, they do the right thing and do not need to be monitored to comply with the quarantine. A motivated, well-informed society is often much more powerful and effective than an ignorant, government-monitored population [5].

The methodology used for this article was a literature review. The theme was the risks in SMEs with the appearance of COVID-19 and its impact in Mexico. Research was carried out in the databases of: Google Scholar, SpringerLink, and Elsevier, among others. Related topics were searched for in books, theses, and lectures. Research started with search formulas and the selection of keywords such as SMEs, COVID-19, Risks, Resilience, and Supply chain.

This situation affects not only the health of the planet but also the structure of the world economic order. As a result, many

*Corresponding Author: Abel García-Villagrán, ingabelgarvill@gmail.com

economies are in crisis and on the verge of a recession [2]. Although policies to contain the expansion of the pandemic such as quarantines and temporary closures are necessary, they cause a reduction of international trade, interruptions in supply chains, lower productivity, contraction of economic activity, less investment, closure of businesses and loss of jobs [6]. Coronavirus crudely shows us the limits of the system in which we live, the inequalities that we have wanted to ignore, and the urgent need to revalue and strengthen the public, frequently and collectively [7]. Small and medium-sized enterprises (SMEs) are organizations with fewer than 250 employees, and are the backbone of the world's economy. SMEs have characteristics that support them when there is a crisis. Being small organizations helps them to be flexible when new opportunities arise [8]. COVID-19 is a highly infectious and lethal virus; people's sense of insecurity is significant and may include feelings of severe panic. Fear and anxiety probably force consumers to avoid purchases in shops, restaurants and entertainment facilities, changing their lifestyle and behaviour. SMEs are not immune to this current and future situation [9].

With the disruptions caused to the world's SMEs, a great deal of research has been carried out by various authors on how these companies could successfully face this global crisis. In the case of Eggers [8], he investigated how SMEs had faced crises and disasters before COVID-19 to turn them into opportunities. In the case of authors Omar, Ishak, & Jusoh [2], they investigated how the pandemic affected SMEs in Malaysia to identify strategies to survive risks, based on the experiences of business owners. Thorgren & Williams [10] carried out a study in several SMEs in Sweden. They analyzed what actions they took to survive and prevent a major disaster, looking for how they could have a "better" or "critical" long-term impact. In this way, researching to understand the risks that affect SMEs in Mexico helps us to expand our knowledge of how these companies manage disruptions, seeking to be resilient to this pandemic that has changed the normality of how the world works.

2. Methodology

To prepare this research, a literature review was carried out. The search was guided by risks in the supply chains in Mexican SMEs with the appearance of COVID-19. The information investigated consisted mainly of articles published between March and September 2020. The methodology combined a systematic literature review with the most relevant articles. The literature review was augmented by the use of online computerized search engines including Google Scholar (350 articles), SpringerLink (105 articles), and Elsevier (252 articles). Information was also sought in books, theses, and lectures, using Keywords such as SMEs in Mexico, COVID-19, Risks, Resilience, and Supply Chains. Approximately 700 sources were retrieved and analyzed. In a careful analysis, the articles that contained the best information were selected. Based on this review, the main risks in Mexican SMEs and their effects during the COVID-19 pandemic were found.or the preparation of this investigation, a literature review was carried out.

3. Supply Chain in Mexican SMEs

The focus of a supply chain is to provide the right product to the consumer, at the right cost, with the correct quality and quantity. This network includes the manufacturing of a product or service, suppliers, retailers, transportation, and customers on both sides [11]. Supply chains work the same way our bodies do to fight the COVID-19 virus. If we imagine the human body as a medieval city to be defended, the immune system would be the army with all its strategies to take care of the safety of the inhabitants. The first line of defence is the wall. In the case of our body, it is the skin. But, there are other possible entrances, such as doors, cracks, water outlets, which are essential for the city to function. In our body, those entrances that are related to the outside are the nose, eyes, ears, mouth, and small wounds, which are the places through which the virus can enter to attack and end the person's life [12]. A supply chain contains many links. At the moment of being struck by a crisis like the present one, the weakest link is the place where the disruption that causes the failure or stoppage of the flow of materials, causing a drop in production, customer dissatisfaction or economic losses. If not corrected on time, these affect the chain and may cause the closure or weakening of any of the companies involved.

Our society seeks efficiency and economic benefit more than safety, a mindset that must change after the current outbreak. Many companies have closed, disrupting trade in most industrial sectors. Entrepreneurs and brands face many short-term challenges in areas such as health, safety, supply chains, workforce, cash flow, changes in demand, sales, and marketing [13]. The pandemic has several characteristics: it is a global phenomenon, with impacts on public health and the economy. It is spread between people, but also in commerce. The world economy is connected by value chains and international movements of people, capital, goods, and services [3]. The prolonged closure of companies in the world economy reduces inventory stock in supply chains. Companies must minimize disruption and adapt quickly to this new business landscape [14].

For Latin America, an increase in unemployment and poverty is anticipated as a result of the global economic slowdown, tourism decline, reduced remittances from immigrants, the interruption of value chains, and the lack of international investment [6]. As to the effects of COVID-19 on the real economy, there are already projections that assume an employment crisis in Mexico. Forecasts for negative growth rate are estimated to be between -3.9% (the most optimistic) and -9.6% (the most pessimistic). Therefore, the extent of the crisis and the duration of recovery depend on strategies implemented by the Mexican government at its three levels [15]. The Coronavirus crisis is projected to be a considerable economic crisis and the greatest challenge since the Great Recession. Unlike the 2008 crisis, it is not in the financial markets, but is now in the real economy, consisting of companies, mainly SMEs and their workers. Demand is expected to affect durable goods more than non-durable products, especially the manufacturing sector [4].

One characteristic of the crisis is that people have been confined to their homes. The population is trying to determine the best way to protect themselves and their loved ones. Many are fearful of losing their jobs or work around people who may be infected. Each country has adopted different actions to manage stress at work and supports its citizens in different ways [13]. Right now, we are in the hands of the decisions that people and governments will be making in the coming weeks that will shape

the world for years to come. And not only health systems, but also the economy, politics, and culture. What world will we inhabit once the storm passes? Even when all of this endsand we survive, the world we live in is going to be different [5]. Mexico, like the rest of the world, is at a crucial moment, in which the response of government and society will shape the future of the country. The challenge is enormous for a country which, in addition to having great inequalities, has received COVID-19 in a situation of great economic fragility, with a fragmented health system, little investment and a government that has not shown itself able to understand the magnitude of this crisis [16].

Contagion in supply chains leads to direct supply shocks among the countries least affected by the pandemic. It is more difficult and expensive to acquire the industrial materials necessary for manufacturing, and the most affected countries sell these inputs at a higher price due to the shortage of units [4]. Ways of consumption and purchase are going to change. We are having supply versus demand shocks. Companies and individuals are changing their consumption habits as regards many products. The demand for non-essential consumer goods is falling, even without supply restrictions [3]. Even if small manufacturers close their businesses temporarily, they continue to face payments such as rent and bank loans, in addition to debts for supplies they had ordered before the pandemic [17]. Despite these difficulties, supply chains are not broken. The global pandemic has been an opportunity to integrate a new set of skills, strategies, and innovations to develop a new chain model, one that is more resilient, more flexible, and with much less dependence on the Asian market [18].

Supply chain risk management only applied to first-tier suppliers, leaving companies blind and vulnerable to crises affecting their lower-tier "invisible" suppliers. Reality shows us that all suppliers are significant, and interruptions at any level can cause disruptions throughout the chain [14].

4. SMEs and the COVID-19 pandemic

In a globalized economy, small and medium-sized enterprises (SMEs) are the main source of dynamism, innovation, and flexibility in developing countries, as well as in industrialized nations. The economic development and job creation of a country are a consequence of the proper functioning of SMEs [11]. The spread of the COVID-19 pandemic has affected various sectors of the economy, including the micro, small and medium enterprises, which are usually resilient when economic problems occur. Currently, SMEs are at the forefront of crises becausemany employees are limited to doing activities outside the home. The disruption of business activities reduces business performance. Therefore, companies make layoffs and threats of bankruptcy, or they are unable to survive [19].

SMEs face the liability of being small. If a company is small, it controls fewer resources, which makes it more vulnerable to internal and external events, such as the resignation of an employee, a decrease in its finances, a reduction in demand, or a global economic crisis. [8]. Small businesses depend on the intelligence and ability of the owner to run them properly. The owner is a fundamental part of the organization, can develop the supply chain, alliances, performance indicators, and business vision [11].

Survival is difficult for small businesses, even under the best of circumstances. SMEs have limited cash and few reserves. They generally do not have access to credit or capital markets. A temporary interruption can mean the closure of the business. It is not surprising that the restrictions imposed by governments to combat the spread of the pandemic have become a risk for many of them [17]. The effect of the coronavirus on the business activities of SMEs around the world is tremendous. Governments have to put strict policies in place to curb the disease. These decisions cause challenges for SMEs such as cash flow, closure of the operation, dismissal of workers, changes in business strategies, and the search for new sources and opportunities [2]. In recent weeks, purchases of cleaning products have increased, and the population is recycling more garbage by staying at home. At the same time, they accumulate goods, make panic purchases, and some leave the cities to go to the countryside [13].

SMEs have to respond on multiple fronts. They must work to protect the safety of their workers and take care of the operations they perform. They are having an impact on the supply chain and need to move forward in these times of crisis [20]. The most common concern in SMEs we heard was a need for assistance in covering fixed operating costs: rent and mortgage payments, but also utilities and insurance [17]. Micro and small companies are less productive than large companies and have few financial resources to face this crisis. At this time of slowdown and possible loss of jobs, they must be prioritized as compared to large corporations. Forced to close in order to contain the contagion of COVID-19, some companies have announced layoffs and seek a way to reduce their costs during the suspension of activities [7]. It is essential to take action to ensure that fewer jobs are lost and to distribute subsidies to companies and workers to protect the most vulnerable.

The virus does not discriminate, but its economic impact does, so it is necessary to guarantee policies that have the following three priorities:

- Guaranteeing minimum income for the people most vulnerable to the crisis.
- Providing employment protection.
- Encouraging consumption, demand, production, and economic growth.

In Mexico, we had a fragile economy and a low-budget health system even before the health emergency caused by COVID-19 [16]. The President of the Republic, on two different dates, presented the list of welfare programs that minimize the economic and social effect of the pandemic. On April 23, a decree establishing austerity policies for the Secretariats and entities of the Republic until December 31, 2020, was issued [15]. In the case of SMEs, they receive benefits with the following points:

a) 3 million credits are granted to individuals and small family businesses.

b) Two million jobs are created.

c) Fuel prices and taxes are not to be increased.

The government has a very important role in disaster recovery. It manages aid and assistance programs, which provide a basis for

a faster recovery, and the economy should return to normal [21]. Predictions show us that the contagions will continue. It is the government's responsibility to have an adequate policy for the containment of COVID-19, in addition to ensuring that the country has better precautionary measures and preparations before the reopening of its borders [22].

The job loss trend from the first case of COVID-19 detected in Mexico on February 29 is as follows [15]:

- As of March 13, there has been a drop in the number of jobs covered by the IMSS (Mexican Institute of Social Security).
- 346,878 jobs were lost from March 13 to April 6, of which 62% no longer have access to health services, and the remaining 37.7% lose access after seven weeks.
- 47.5% of the jobs lost from March 13 to April 6 occurred in medium-sized companies with 51 to 500 workers.

5. SMEs Disruptions.

No supply chain is infallible; they are subject to different factors that affect their regular operation. Among the possible reasons for this are the closure of plants, the lack of supplies, a lack of information, unforeseen situations such as natural disasters, and epidemics that not only curtail supply but also affect demand [23]. There is a great difference between large companies and SMEs. Large companies have plans for almost all eventualities. It is commonly accepted that SMEs can be disproportionately impacted by extreme events [24]. With technological advances in information processing, there is a greater dependence on networks in supply chains. Companies must incorporate actions to mitigate risks, guaranteeing the safety of the chains and being able to compete effectively [25].

Organizations must be flexible to change. They need to have strategies as to managing the volatility, and ambiguity of the world. Knowing how to manage business risks improves resilience and can be useful for any organization. By knowing the most significant risks, companies implement actions that allow them to act early [26]. The ISO 9001: 2015 standard explicitly addresses the identification and intervention of risks as a preventive mechanism of the management system. This norm helps eliminate potential causes, preventing their occurrence, and minimizing disruptions. Organizations must determine the risks and opportunities to fulfil the purpose of the quality system and thereby increase the trust and satisfaction of the interested parties [27].

Speaking of higher risks, we have to refer to the events of greatest danger for an organization, which are those of High impact / low probability. The reason is that such activities are unimaginable or of such rare incidence that they have never happened in recent memory, and they are not on the "radar" of risk managers. These so-called black swans are events that were thought impossible or that had never been imagined until they happened; experts mistakenly assume that the probability of a black swan is zero, when in fact it is not [28].

Any disruption has consequences on the performance of SMEs, whether in sales, production level, or profits. The authors divide disruptions into eight different phases: Preparedness, Disruptive Event, First Response, Initial Impact, Total Impact, Preparedness for Recovery, Recovery, and Long-Term Impact. The measurement of business resilience to disruptive events should be carried out before the event so we can act more quickly and with distinct guides to the problem [29]. For successfully managing disruptions, both pre-disruption and post-disruption approaches are necessary and need to be taken into consideration [30].

Disruptions allow organizations to have experiences that help them to have answers to future problems. Analyzing an incident helps to identify lessons learned and to understand supply chain risks. In this way, future outages will be managed efficiently, and partners will have to collaborate to obtain better results [31]. In recent years, the manufacturing of companies has been disrupted by earthquakes in New Zealand, air traffic closures due to terrorism in France, and volcanic eruptions in Iceland. Researchers are concerned about the increase in disruptions of the supply chain and its implications [32]. Many SMEs cannot survive after a disaster, so ensuring their survival during and after crises is essential. Governments should also learn from past disasters [33]. On March 11, 2011, a great earthquake occurred in Japan and a tsunami that destroyed many houses and buildings. This event caused a 15% reduction in industrial production in the following months. Although the heaviest damage was on the east coast of the Tohoku and Kanto regions, companies in other parts of Japan were affected by supply chain disruptions [34].

Historically, many diseases, such as the plague, syphilis, smallpox, or influenza, have caused millions of deaths and caused panic in the world. In recent decades, the appearance of new diseases has created alarm and anxiety across the globe, affecting public health and the world economy. Countries with these problems must carry out inspections of all people at international airports, seaports, and land crossings [35]. In addition to human losses, epidemics threaten the free movement of people and goods in global supply chains. Contagious events almost always include medical and financial problems. They spread through human networks that are very often closely linked to supply chains. They affect several countries and multiple industries simultaneously. The weakest and least prepared companies are the most damaged [28].

Cholera in Peru, SARS in Vietnam, Ebola in Africa, and now COVID-19, which emerged in 2019 in China, are examples of global diseases that affect the economy of SMEs and cause fear among the population. The spread of the Coronavirus from China and its expansion to northern Italy, Korea, and the rest of the world, affects the industry in various ways. Mexican SMEs could be affected by the impact of the Chinese slowdown. Another risk factor is the possible shortage of products that affects the prices of many goods worldwide, so the Coronavirus is already considered a factor that Mexican companies have to pay attention to [36]. It is hard to imagine a world that is not changing because of coronavirus. The success of scientists, together with the political and social leadership, will determine the future. It is time to think about what we can do to help. It is a battle that tests our mutual responsibilities, our strengths, and our compassion [37].

Transportation risks are also a potential cause of disruption for SMEs in the supply chain. However, exporters (packers and wholesalers) are more vulnerable than other nodes in the supply chain [38]. Logistics is the engine of world trade; safety in the transport of goods appears as a critical element in its consolidation.

The World Customs Organization (WCO) developed a regulatory framework, to adopt the necessary measures to guarantee the integrity of transport throughout the supply chain against theft, piracy or terrorism [39]. In Mexican SMEs, the transportation of goods is a crucial element for the success of operations. Unfortunately, trucking theft has become a problem for entrepreneurs and a disruption for supply chains. According to the report of the Executive Secretariat of the National Public Security System [40] from January to September 2018, more than 8,727 thefts of carriers were registered, which meant an increase of 4.10% compared to the previous year, and 87% of these events were violent. These facts directly affect the competitiveness of SMEs, and managers must think about how to deal with it because now it is a risk every time their trucks hit the road to deliver to their customers.

The first research on the impact of natural disasters on business appeared in the 1980s. The first was carried out in the United States after the 1989 Loma Prieta earthquake. SMEs are more vulnerable to natural disasters and can be affected in many ways. There are also interruptions related to public services such as electricity, water supply, sewage, lack of fuel, transportation, and telecommunications. Failures in public infrastructure make companies look for alternative logistical support that allows SMEs to function after disruption of this type [41]. Risks from natural disasters in Mexico are many; the National Center for Disaster Prevention (CENAPRED) [42] mentions in its report for 2017 that the estimated damages and losses amounted to 73,862 million pesos, a figure that was 400% higher per year, which impacted 0.34% on GDP. The phenomena that most impacted the country were those of a geological type, with an earthquake of a magnitude of 7.1 degrees on September 19, representing 90.7% of the losses, and hydrometeorological events caused by cyclonic activity in the Pacific and Atlantic, representing 8.2 % of disasters. Manufacturing SMEs that were close to the epicenters were affected directly or indirectly without being able to do anything about it.

It may be that each of the causes that initiates a disruption is unique. Still, they may have some aspects in common, so that understanding them helps orient collective action towards efficient and effective situations. What is lacking for SMEs to remain is to create conditions for permanence, growth, and transformation into mature companies [43]. These different causes are similar to what happened on January 19, 2019, in which teachers from the Mexican state of Michoacán blocked in six different points and for more than eighteen consecutive days the railroad tracks that connect the main ports of the Pacific: Lázaro Cárdenas and Mazatlan. Eleven states of the Republic were affected, and results included 252 stopped trains, 10 thousand immobilized containers, and 2.1 million tons of cargo affected, leaving many SMEs without supplies for several days [44].

Many lessons can be learned, but one of special concern is that most companies fail to recognize high-impact risks and as a consequence are not able to respond adequately [45]. After a disruptive event, companies must react and control the situation to stop the effects of the disruption as much as possible, and preserve the business structure and the most valuable assets [46]. Disruptions usually occur suddenly, and their impact is experienced over a long period, requiring much effort and sometimes the help of others [33].

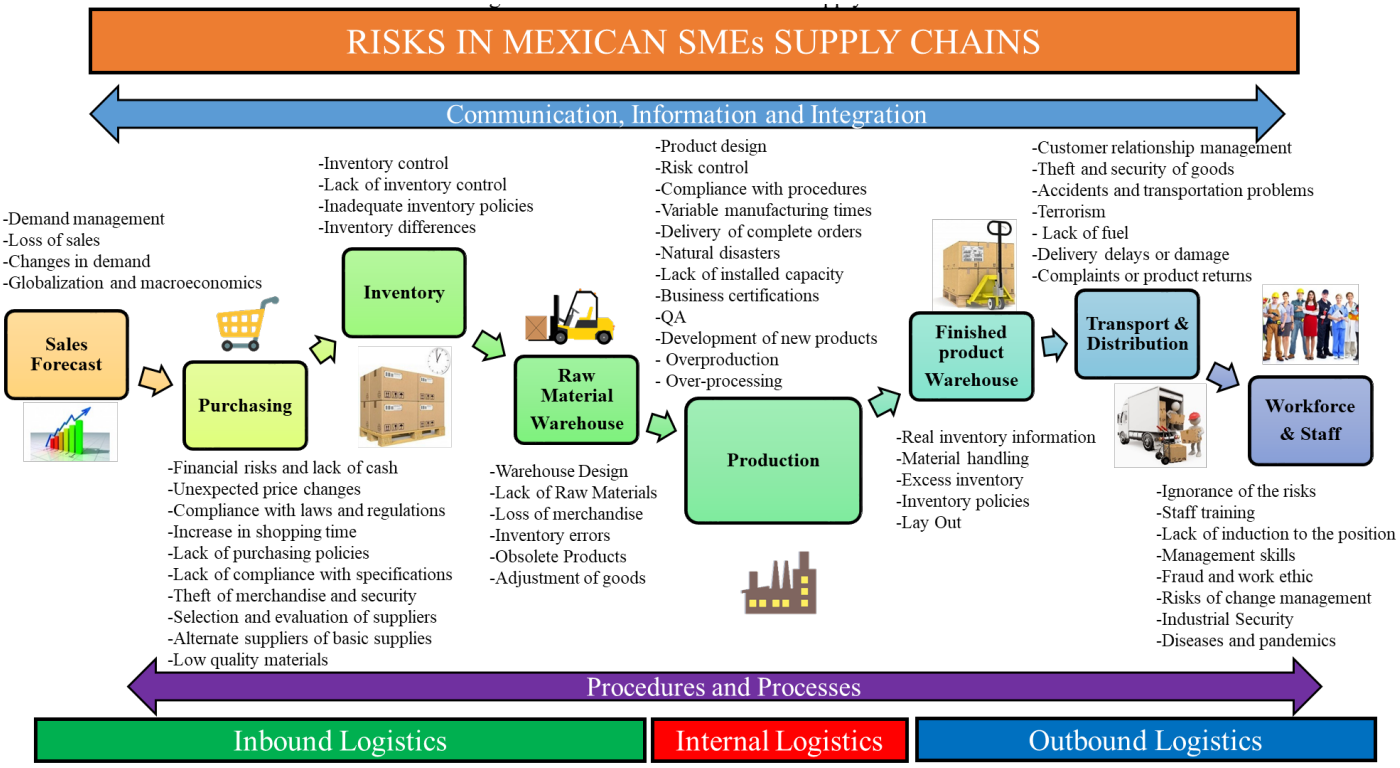
The design of a risk management strategy is the key to a well-functioning value chain. This strategy must identify the potential sources of risks and the appropriate actions to prevent them or reduce the impact of a disruption. As long as the risks are better known, entrepeneurs can identify and prioritize actions to have greater resilience, sustainability, and profitability [47]. SMEs face constant risks that negatively impact all areas, causing losses and endangering their survival in the market. Risk management in the different stages of an SME supply chain should not be an isolated program, but part of the business strategy. The magnitude of the efforts required to control them will depend on the complexity of the operational activities and the size of the company. Most of the time, Small Business owners do not pay attention to this issue because of their company's success in the market.

Risks are interconnected with each other. For example, factors such as natural disasters that affect the supplier increase the producer's risk of not receiving the raw material he needs. Logistics operators can be affected by lower demand, which translates into wasted production capacity with lower sales, increasing their risk of liquidity. Finally, distribution and transportation can be affected by the need to lower their inventories. It is essential for the operation and performance of SMEs' supply chain, to develop management strategies that identify the risks to which they are exposed, their probability of occurrence, and the effects they may have in a disruption such as COVID-19. In (Figure 1), a summary of the risks found in the literature review of each stage of the supply chain of SMEs in Mexico is made. The risks were classified into three areas: Inbound Logistics, Internal Logistics, and Outbound Logistics. In this way, it is easy to identify where outages may occur and take the necessary precautions.

6. Resilience in SME's in times of COVID-19.

In recent decades, resilience has become globally significant as the ability of a system to respond to change and continue to develop. This term has been used to describe and explain various life situations in different disciplines, and resilience is something that must be developed in the ecosystem, in humans, and therefore in organizations and specifically in supply chains [48]. The need for an "Antifragile" property is an imperative to ensure the business continuity and future success of SMEs in times of disruptive events. This concept originated from philosophy, and builds upon the Black Swan Theory, which describes unexpected and unpredictable events having a high impact in a world in which everyone believes that only white swans exist [9].

The word resilience comes from the English language and expresses the ability of a material to regain its original shape after being subjected to high pressure [49]. In the business world, resilience is understood as a crisis management and business continuity, as well as responding to all types of risks that organizations may face, ranging from cyberspace threats to natural disasters and many others [19]. Supply chain resilience is the adaptive capacity to prepare for unexpected events, respond to disruption, and recover from them by maintaining continuity of operations at the desired level of connectedness and control over structure and function [38].



The elements to measure business resilience are the adaptive capacity of the company, vulnerability, probability of occurrence of disruption, time, level of recovery, commitment, and responsibility of the supply chain [50]. Companies seek to become stronger and find the formula to continue despite adverse circumstances [46]. Therefore, they assume two primary missions:

1) Implementing strategies to face and overcome the adverse effects produced by those unexpected situations and,

2) Developing protective barriers to face eventual future circumstances.

Resilient companies are those that, in times of constant change such as economic-social crises or business globalization, perform better than others, and obtain benefits from adverse or unforeseen circumstances. Some companies are expected to emerge from intense crises, while others could weaken or even go bankrupt. In all these possibilities, those organizations that maintain a tendency to renew and perform better during the crisis are resilient [51]. An organization with these characteristics has a set of capabilities aimed at carrying out robust actions in the face of a specific condition generated by unexpected and powerful events that endanger the survival of the organization [49]. The majority of business resilience risks come from internal factors in the supply chain, thus requiring strong collaboration. The level of resilience risks and barriers varies in SMEs according to their size [19].

Many firms are struggling to survive during the pandemic, which focuses attention on the apparent failings of their business models and on how to build greater resilience in the future. There are new business opportunities [3]. Many questions that we could ask ourselves about the rapid closure of companies due to the pandemic have arisen. For example, how can we take better care of employees in such situations? Why weren't companies ready? How can companies and countries use this situation to improve their competitiveness? [13].

Diversified supply-chains across companies and geographies significantly reduce exposure. If firms are tied to single suppliers, the risk from disruptions should be carefully measured and contingency plans considered [14]. In most companies, inventories are stored for two to five weeks, with no additional deliveries. Each company has a different supply-chain strategy and ensures their delivery times. If the components do not arrive or their delivery is interrupted for a longer time, production stops. A resilient supply chain must align its strategy and operations to adapt to the risks that may affect it. Therefore, networking with providers is the key to resilience [20]. An SME that wishes to be resilient must carry out the following actions: recruit resilient personnel, define clear and precise positions and functions, have business strategies known by all members of the organization, establish quality systems, have an organizational culture with solid values, be attentive to economic trends, review consumer trends, and promote initiative and creativity among its staff [43].

Business certifications ensure that a business meets the requirements of governing industry standards. The pursuit of these certifications reinforces the commitment to continuous improvement, thus reducing risk in production and demand disruption and reducing trade barriers for firms. There is a positive impact of business certifications on firms' resilience. Certifications could be Hazard Analysis and Critical Control Point (HACCP), Good Manufacturing Practices (GMP), ISO 9000, among others [38]. Some Resilience Models have been used in SMEs worldwide. One of them is: "Healthy and Resilient Organization, HERO", a heuristic model that describes the functioning of healthy and resilient organizations. It is made up of three elements: Healthy organizational resources and practices, Healthy

employees, and Healthy corporate results. This model was created from the global context that refers to organizations that survive and adapt to crises, becoming more robust in the face of these negative experiences [52]. When evaluating risks in a Supply Chain, the SCOR (Supply Chain Operations Reference) reference model provides a common language in the areas of planning, manufacturing, distribution, and return. It includes a methodology that allows an analysis of the operations of the chain, helping them communicate company information, measure performance, and set improvement objectives. Being a reference model, it does not have a mathematical description or a specific method. Its usefulness lies in providing a standard of chain terms and processes to a model using performance indicators [25].

Using information technology during extreme interruptions can help people stay connected and ease their work situation. Lessons from the use of information technology by SMEs to support business continuity during COVID-19 help us minimize some of the risks of information flow throughout the supply chain [53]. The normality of December 2019 is not going to return. In many countries, the concept of omnichannel networks accelerated. People now work, buy, spend, and see their families digitally. Unexpectedly, COVID-19 has done more for digital acceleration than possibly anything seen in the last decade [18]. COVID-19 showed the weaknesses of the business model that existed. Some of the changes that are being experienced are: People involved in the physical distribution of goods can become infected with the virus and stop global value chains. International air travel is limited, but most ports and sea routes remain open. Social distancing and controls for the flow of people create delays. Companies and governments have a shortage of critical goods and services because foreign suppliers are looking for their products with local customers [3].

Some of the problems of lack of supply thanks to COVID-19 indicate that if there had been a previous study on these issues, many problems could have been foreseen: A lack of laws to control price increases due to a crisis; mapping the supply chain and ensuring certified suppliers and products. Government procurement of health products. International supply assessment, trade barriers, and market capacity [54].

Owners/managers must make the right decisions, which promote flexibility in SMEs and the ability to react rapidly in their markets. Another opportunity provided by COVID-19 refers to the possibility for continuous innovation through the development of new or better products, process and business models suggested by employees, customers and partners after the emergency. Also, SMEs could be redesigning and defining new strategies to meet the rising global demand for disinfectants, medical masks and other anti-Covid19 products [9].

In June the SoyLogístico Association of Mexico [55] shared some recommendations based on the experience of its partners and executives from different supply chains to return to the new normal. They issued a document called: "Ten Operational Tactical Actions towards the New Normal", in which some of the points that can support the logistical resilience and continuity of operations in Mexican SMEs are the following:

- Define health and safety protocols so that employees work safely.
- Analyze demand, align supply by adjusting sales and financial plans.
- Analyze the financial situation of the company, and also eliminate or reduce inventories.
- Review supply, and return to local suppliers instead of global suppliers.
- Define an operating strategy throughout the supply chain.
- Analyze the distribution of merchandise, review delivery restrictions, and ensure the requirements of distribution centres.
- Analyze the transport of goods and last-mile deliveries.
- Guarantee clear and continuous communication with collaborators, clients, suppliers, partners, and authorities.
- Have a leadership focused on the achievement of objectives and results of collaborators, seeking to "humanize" operations.
- Promote savings plans throughout the chain and define actions to digitize activities.

7. Conclusion

World political and economic scenarios in this century are leading us to collapse. Economic downturns, competition, corrupt governments, damage to the environment, natural catastrophes, and now global pandemics are leading the world into uncontrolled free fall. Disruptions are not only increasing but are also more potent; the possibility of bankruptcy or closure of companies is escalating. Resilient thinking is not a game; it is a life-saving phenomenon for organizations large and small [56]. Although viruses like COVID-19 have the potential to do much damage, they are very simple invaders. When a large number of them enter our body, our defences go into action to attack them and keep the person alive [12]. The global pandemic is striking Mexican supply chains, more than ever before, communication, timely information, innovation, best logistics practices, and the resilience of the most robust links help them to resist this time of crisis.

In the short term, the effects of the Coronavirus on the global economy are substantial. The temporary suspension of production and supply chains has also had an impact on the evolution of various financial indices. If the adverse effects of the epidemic persist, countries must adopt fiscal, monetary, and economic actions that support consumers and small businesses that are most affected. The COVID-19 emergency reminds the world that it must be prepared for a crisis with risks to the health of the population [6]. It was clear that SMEs' resilience requires partnerships and cooperation among firms, government, and other private organizations. It cannot be achieved by the business entity alone or by the government on its own [41].

In current conditions where demand is changing and customers are more interested in digital commerce, having the right information technologies is important for the sustainability of company operations. SMEs must make efforts to be part of the digital economic ecosystem and use social networks to promote products or businesses. By promoting their brands effectively, they create online communities, knowing the preferences of their

customers effectively and directly [57]. In these times of staying at home, consumer behavior has changed. Consumers are becoming more and more comfortable with digital technology and digital commerce. SMEs need to invest in data analysis technologies in order to know their customers effectively and offer them what they need [58]. The time has come for Mexican micro, small, and medium-sized companies to think about starting operations online. This option is the only way to guarantee their survival and development in the future. According to the Mexican Online Sales Association (AMVO), electronic commerce will grow by 59% due to the pandemic. If the benefits of electronic commerce are expected to be real in Mexican companies and consumers, some challenges have to be overcome, especially in logistics. The first is a lack of capacity of logistics infrastructure under normal conditions, and the second is the issue of the pandemic because demand continues to exceed expectations [59].

The Mexican government is in a race against time to implement a solid economic rescue plan. If it is not successful, the recession will be complicated, and access to financial markets will be restricted. The signs of President López Obrador are disturbing, placing the social and work life of citizens whose life has been altered by the pandemic at risk [16]. The priority must be to support individuals and small businesses urgently. Large companies are able to assume part of the cost of the crisis, especially those that are in sectors that may benefit. Companies with direct links to high government positions should be prevented from receiving financial support [7].

The challenges resulting from the COVID-19 are not disappearing quickly, and government policies for assisting small businesses and managing a return to some version of typical activities continue [17]. A global effort is needed for the production and distribution of medical equipment, test kits, and respirators. Rather than each country trying to achieve this locally, a coordinated effort could accelerate production and ensure that life-saving equipment is distributed fairly [5]. Recent articles debate what the future holds, considering questions such as how long social distancing needs to continue, the social and political consequences of lockdown, and implications for supply and demand chain structures [54].

Future expectations are anyone's guess. The COVID-19 crisis has consumers and companies in continuous uncertainty [4]. Articles published in the popular press or social media discuss how the crisis is changing the way we live and work, and this is driven mainly by small and medium-size organizations that create innovative solutions for problems. SMEs have a unique role in the macro-environment when it comes to creating a path forward [8].

Many changes have affected the supply chains of manufacturing SMEs in Mexico. One has to discover new ways of managing them if one wants to survive in the modern world context. The delivery times and availability of materials and goods are changing due to COVID-19 because no one is manufacturing them. The risks of diversifying with more suppliers globally or developing existing suppliers have to be considered. SMEs are increasing their inventory levels to face future disruptions. Electronic commerce is a new way of selling their products, forcing SMEs to strengthen electronic platforms and their information systems. The distribution, reception, and delivery of orders must be reinvented, so as to avoid any risk when inspecting or arranging products in the warehouses. What seemed unlikely to happen is happening now, and logistics managers, company staff, and researchers on these issues have to reinvent models to make companies resilient. It is necessary to consider every risk encountered throughout the supply chain.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors are very grateful for the support received from the Universidad Popular Autónoma del Estado de Puebla for this project.

References

[1] M. A.Shereen, S. Khan, A. Kazmi, N. Bashir, R. Siddique, "COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses", Journal of Advanced Research, 2020, doi.org/10.1016/j.jare.2020.03.005.

[2] A. R. C. Omar, S. Ishak, M. A. Jusoh, "The impact of Covid-19 Movement Control Order on SMEs' businesses and survival strategies", Geografia-Malaysian Journal of Society and Space, **16**(2), 2020, doi.org/10.17576/geo-2020-1602-11.

[3] R. Strange, "The 2020 Covid-19 pandemic and global value chains", Journal of Industrial and Business Economics, **47**, 455–465, 2020, doi.org/10.1007/s40812-020-00162-x.

[4] J. Cifuentes-Faura, "Crisis del coronavirus: impacto y medidas económicas en Europa y en el mundo", Espaço e Economia. Revista brasileira de geografia econômica, **18**, 2020, doi.org/10.4000/espacoeconomia.12874.

[5] Y. N. Harari, "El mundo después del coronavirus", Publicación mensual del Instituto Histórico Centroamericano, **39**(457), 2020.

[6] J. L. Clavellina Miller, M. I. Domínguez Rivas, "Implicaciones económicas de la pandemia por COVID-19 y opciones de política", Instituto Belisario Dominguez Senado de la República, **81**, 2020, bibliodigitalibd.senado.gob.mx/handle/123456789/4829.

[7] P. E. L. P. Y. Minimizar, S. E. S. E. Las, "El coronavirus no discrimina, las desigualdades sí", Oxfam Internacional, 2020.

[8] F. Eggers, "Masters of disasters? Challenges and opportunities for SMEs in times of crisis", Journal of Business Research, **116**, 199-208, 2020, doi.org/10.1016/j.jbusres.2020.05.025.

[9] V. Ciasullo Maria, A. Douglas, R. Montera, "Antifragility for handling a systemic crisis. Some reflections about SMEs in the COVID age", University of Salerno, The TQM Journal, 2020.

[10] S. Thorgren, T. A. Williams, "Staying alive during an unfolding crisis: How SMEs ward off impending disaster", Journal of Business Venturing Insights, **14**, 187, 2020, doi.org/10.1016/j.jbvi.2020.e00187.

[11] T. A. Chin, A. B. A. Hamid, A. Rasli, R. Baharun, "Adoption of supply chain management in SMEs", Procedia-Social and Behavioral Sciences, **65**, 614-619, 2012, doi.org/10.1016/j.sbspro.2012.11.173.

[12] L. D. Alvarez, "Cómo responden las defensas de nuestro organismo al COVID-19", Cienciorama Unam, 2020.

[13] N. Donthu, A. Gustafsson, "Effects of COVID-19 on business and research", Journal of Business Research, **117**, 284-289, 2020, doi.org/10.1016/j.jbusres.2020.06.008.

[14] B. McKenzie, "Beyond COVID-19: Supply Chain Resilience Holds Key to Recovery", Bakermckenzie, 2020.

[15] E. M. Rivero Cob, "Medidas para la recuperación económica ante la emergencia sanitaria Covid-19: Del crédito solidario al seguro de desempleo", Instituto Belisario Dominguez Senado de la República, 2020, bibliodigitalibd.senado.gob.mx/handle/123456789/4861.

[16] R. F. Nieva, "México y el coronavirus: pasividad gubernamental en una sociedad desigual", Análisis Carolina, **16**, 1, 2020, doi.org/10.33960/AC_16.2020.

[17] C. R. Neu, D. Gehlhaus, H. J. Shatz, "Preserving Small Businesses: Small-Business Owners Speak About Surviving the COVID-19 Pandemic", Rand Corporation, 2020.

[18] E. Escudero, "9 ejecutivos y gurús logísticos comparten sus aprendizajes ante el Covid-19", Revista Énfasis Logística, 2020.

[19] S. Hadi, "New Perspective on the Resilience of SMEs Proactive, Adaptive, Reactive from Business Turbulence: A Systematic Review", Journal of Xi'an

University of Architecture & Technology, **12**, 1265-1275, 2020, doi.org/10.37896/JXAT12.05/1524.
[20] A. Majid, "Pakistan's Supply Chain Resilience", Humanities Commons, 2020, dx.doi.org/10.17613/4dr4-6g0.
[21] S. Pathak, M. M. Ahmad, "Flood recovery capacities of the manufacturing SMEs from floods: A case study in Pathumthani province, Thailand", International Journal of Disaster Risk Reduction, **18**, 197-205, 2016, doi.org/10.1016/j.ijdrr.2016.07.001.
[22] R. O. Ogundokun, A. F. Lukman, G. B. Kibria, J. B. Awotunde, B. B. Aladeitan, "Predictive modelling of COVID-19 confirmed cases in Nigeria", Infectious Disease Modelling, **5**, 543-548, 2020, doi.org/10.1016/j.idm.2020.08.003.
[23] A. S. Altamiranda, "Revisión de modelos para la toma de decisiones para el control de fallas en cadenas de suministro", Gestión, Competitividad e Innovación, **3,** 2, 2016.
[24] B. Sullivan Taylor, L. Branicki, "Creating resilient SMEs: why one size might not fit all", International Journal of Production Research, **49**(18), 5565-5579, 2011, doi.org/10.1080/00207543.2011.563837.
[25] F. A. Pereira Souza, Modelo de referencia para el dimensionamiento de riesgos operacionales en la cadena de suministros caso: sector mármol, Master's Thesis, Universidad Autónoma de Occidente, 2015.
[26] Coso, "Enterprise Risk Management Integrating with Strategy and Performance", Committee of Sponsoring Organizations of the Treadway Commission, Executive Summary, 2017.
[27] A. M. Alzate Ibañez, "ISO 9001: 2015 base para la sostenibilidad de las organizaciones en países emergentes", Revista venezolana de gerencia, **22**(80), 576-592, 2017, redalyc.org/articulo.oa?id=290/29055967003
[28] Y. Sheffi, El poder de la Resiliencia, Temas Grupo Editorial SRL., 2016.
[29] R. Sanchis, R. Poler, "Medición de la resiliencia empresarial ante eventos disruptivos una revisión del estado del arte" in 2011, V International conference on Industrial Engineering and Industrial Management, 104-113, 2011.
[30] A. Robles, R. Severson, Improving supply chain resilience for a fashion retailer: a qualitative and quantitative study, Ph. D Thesis, 2016.
[31] E. Revilla, M. J. Saenz, "The impact of risk management on the frequency of supply chain disruptions: A configurational approach", International Journal of Operations & Production Management, **37**(5), 557-576, 2017, doi.org/10.1108/IJOPM-03-2016-0129.
[32] S. Chopra, M. Sodhi, "Reducing the risk of supply chain disruptions", MIT Sloan management review, **55**(3), 72-80, 2014, openaccess.city.ac.uk/id/eprint/14261.
[33] Z. Auzzir, R. Haigh, D. Amaratunga, "Impacts of disaster to SMEs in Malaysia", Procedia Engineering, **212**, 1131-1138, 2018, doi.org/10.1016/j.proeng.2018.01.146.
[34] A. Kondo, "The effects of supply chain disruptions caused by the Great East Japan Earthquake on workers", Japan and the World Economy, **47**, 40-50, 2018, doi.org/10.1016/j.japwor.2018.03.007.
[35] C. Maguiña Vargas, "Enfermedad por el virus del Ébola", Revista Medica Herediana, **26**(3), 195-201, 2015.
[36] Y. Valles, ¿Cómo impactará el coronavirus a las empresas mexicanas?, ExpertoPyme, 2020.
[37] H. Holden Thorp, "Time to pull together", Science, **367**(6484), 1282, 2020, doi.org/:10.1126/science.abb7518.
[38] I. Ali, S. Nagalingam, B. Gurd, "Building resilience in SMEs of perishable product supply chains: enablers, barriers and risks", Production Planning & Control, **28**(15), 1236-1250, 2017, doi.org/10.1080/09537287.2017.1362487.
[39] C. Kirby, N. Brosa, La logística como factor de competitividad de las Pymes en las Américas, Banco Interamericano de Desarrollo, 2011.
[40] SESNSP, Informe del Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública México, Secretaria de Seguridad Pública, 2018.
[41] G. Samantha, "The Impact of Natural Disasters on Micro, Small and Medium Enterprises (MSMEs): A Case Study on 2016 Flood Event in Western Sri Lanka", Procedia Engineering, **212**, 744-751, 2018, doi.org/10.1016/j.proeng.2018.01.096.
[42] CENAPRED, Resumen ejecutivo del impacto socioeconómico de los principales desastres en 2017, Centro Nacional de Prevención de Desastres México, 2017.
[43] C. L. Camacho, A. T. Palacios, F. V. Mendoza, J. R. L. Sánchez, "La resiliencia estratégica y las crisis en las PyMes", Semantic Scholar, 2018.
[44] E. Duarte, Bloqueo Ferroviario Provoca Mayor irritación empresarial, Revista T-21, 2019.
[45] E. Revilla, M. J. Sáenz, ¿Sabemos gestionar las disrupciones que afectan a nuestra cadena de suministros?, Strategia, **25**, 72-73, 2012.
[46] G. L. Laverde Verástegui, H. A. Rivera Rodríguez, La disrupción: El punto de partida de la resiliencia o del fracaso empresarial (Reflexiones desde la ingeniería), Revista Espacios, **38**(07), 2017.
[47] A. Calatayud, J. A. Ketterer, Gestión integral de riesgos para cadenas de valor, División de Mercados de Capital e Instituciones Financieras, 1-57, 2016.
[48] A. M. Cotte Urrutia, Cadenas de suministro resilientes, Masters Thesis Universidad Militar Nueva Granada, 2018.
[49] C. Medina Salgado, "La resiliencia y su empleo en las organizaciones", Universidad Autónoma Metropolitana, Gestión y estrategia, 2012, hdl.handle.net/11191/2968.
[50] P. Cordero Cortés, J. F. Núñez Espinoza, O. Hernández Romero, O. A. Arana Coronado, "Factores resilientes en micro y pequeñas empresas rurales", Agricultura, sociedad y desarrollo, **11**(4), 481-498, 2014.
[51] B. J. G. Hernández, J. R. Molina, H. C. Islas, L. E. G. Escorza, A. C. García, M. M. M. Leal, "Desarrollo de un modelo de resiliencia para la incorporación de tecnologías de información y comunicación en las pequeñas y medianas empresas del sector textil Mexicano", Strategy, Technology & Society, **4**(1), 2018.
[52] R. A. V. Rivera, "Modelos de Resiliencia Organizacional en las Pymes y su Impacto en la Cadena de Abastecimiento", Revista Ingeniería, Matemáticas y Ciencias de la Información, **4**(8), 2017, dx.doi.org/10.21017/rimci.2017.v4.n8.a35.
[53] T. Papadopoulos, K. N. Baltas, M. E. Balta, "The use of digital technologies by small and medium enterprises during COVID-19: Implications for theory and practice", International Journal of Information Management, 102192, 2020, doi.org/10.1016/j.ijinfomgt.2020.102192.
[54] L. Knight, J. Meehan, E. Tapinos, L. Menzies, A. Pfeiffer, "Researching the future of purchasing and supply management: The purpose and potential of scenarios", Journal of Purchasing and Supply Management, 100624, 2020, doi.org/10.1016/j.pursup.2020.100624.
[55] Soy Logístico, 10 Acciones Táctico-Operativas rumbo a la nueva normalidad, La visión desde la logística, #SoyLogisticoAsociación, 2020.
[56] H. Vaca, F. Contreras, "Resiliencia estratégica y valor compartido", in 2018 III Congreso: Ciencia, Sociedad e Investigación Universitaria, 2018, repositorio.pucesa.edu.ec/handle/123456789/2611.
[57] M. Indriastuti, K. Fuad, "Impact of Covid-19 on Digital Transformation and Sustainability in Small and Medium Enterprises (SMEs): A Conceptual Framework" in 2020 Conference on Complex, Intelligent, and Software Intensive Systems, 471-476, 2020, doi.org/10.1007/978-3-030-50454-0_48.
[58] Y. K. Dwivedi, D. L. Hughes, C. Coombs, I. Constantiou, Y. Duan, J. S. Edwards, R. Raman, "Impact of COVID-19 pandemic on information management research and practice: Transforming education, work and life", International Journal of Information Management, 102211, 2020, doi.org/10.1016/j.ijinfomgt.2020.102211.
[59] A. D. L. A. R. Ruíz, "Emergencia Sanitaria y Transacciones electrónicas: COVID–19 Caso México", Perfiles de las Ciencias Sociales, **8**(15), 2020.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 72-95 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Automated Extraction of Heavyweight and Lightweight Models of Urban Features from LiDAR Point Clouds by Specialized Web-Software

Sergiy Kostrikov[1,2,*], Rostyslav Pudlo[2], Dmytro Bubnov[2], Vladimir Vasiliev[2], Yury Fedyay[2]

[1]*Department of Human Geography and Regional Studies, School of Geology, Geography, Recreation and Tourism, V. N. Karazin Kharkiv National University, 4 Svobody Sq., Kharkiv, 61022, Ukraine*

[2]*EOS Data Analytics, 31 Alchevskyh St., Kharkiv, 61002, Ukraine*

ARTICLE INFO

Article history:
Received: 28 August, 2020
Accepted: 08 October, 2020
Online: 08 November, 2020

Keywords:
LiDAR
AFE
Phased Methodological Flowchart
High Polyhedral Modeling
Low Polyhedral Modeling
Heavyweight Models
Lightweight Models
Web-Software
ELiT Server
ELiT Geoportal

ABSTRACT

3D city modeling may be considered as one of the key applications, that are provided by the Automated Feature Extraction (AFE) techniques from LiDAR data. The authors attempt to prove that with growing availability of LiDAR surveying methods the resulted 3D city models become the most significant modeled features for any urban environment. Our paper represents the conceptual multifunctional approach within the AFE frameworks, that has been introduced through consequent steps of the phased methodological flowchart with its two branches: High Polyhedral Modeling (HPM) and the Low Polyhedral Modeling (LPM) of buildings. Both branches result in the heavyweight models, and in the lightweight ones, correspondingly. The research purpose of this paper is to outline our multifunctional approach (functionalities of Building Extraction, Building Extraction in Rural Areas, Change Detection, and Digital Elevation Model Generation) to the fully automated extraction of urban features, and present our original contributions to the relevant algorithmic solutions within both HPM, and LPM pipelines, as well as represent desktop, web-, and cloud-based software elaborated for these intentions. Original enhancements and optimizations of the adopted AFE-techniques have been bounded to the phases of the methodological flowchart, while some derivative results have been presented not only as the software description, but also in the discussion chapter. Joint implementation of various functionalities in a web-based application (the Server) is presented with several interface samples of research in urban block and district scopes, while a cloud-based application (the Geoportal) is an Internet-toolbox for solutions in the scope of a whole city.

1. Introduction

This text is a significantly changed extension of the work firstly presented as a conference paper [1].

The global world has been already transferring into the information society for several recent decades. It is quite acceptable to consider *a Geographic Information System – GIS* as one of the core tools of such transfer together with the relevant technologies of remote sensing, including *LiDAR (Light Detection and Ranging)* technique. It has been remarkable for few recent decades that exactly this period also has been featured by the continuing urbanization process, that still takes place nowadays in many developing countries. Numerous facts and phenomena indicate, that we may face the largest urban growth wave within the whole human history, which also concurs with prompt development of information technologies, remote sensing, computer sciences, geoinformatics, and *GIS-platforms / modules*. All listed entities could and should be involved in resolving of those drastic problems arisen within the urbanized areas, for example, by considering these territories as the *hierarchical urban geosystems* and making the corresponding decision support recommendations [2, 3]. Thus, even several highly challenging issues relevant to the Smart City development can be met in this way [4]. Moreover, the rapid growth of both remote sensing technique and GIS-technological involvement in surveying environmental consequences of the urbanization process has been

*Corresponding Author: Sergiy Kostrikov, sergiy.kostrikov@karazin.ua

clear evident over the past few decades [5], while quite a few both approaches, and methods, as well as user interfaces have been developed [6-15].

Urban features can be accepted as the core constituents of any virtual presentation for each digitally simulated city. A *3D City Model* is the key formalized entity of the mentioned city content, and it enables effective and accurate 3D modeling of the urbanized environment with respects to housing sets and infrastructures [16-23]. Within the urban studies perspectives it can be possible to accept a 3D object / feature model as a work-flow key issue of the GIS output completed from the urban remote sensing input. This assumption may be even more evident, while the discrete objects are considered as those modeled results, which are intended for a customized solution within the framework of one only, or several urban applications. In this meaning 3D modeling appears to become a key component within the common geoinformation paradigm [24]. It is considered be possible to accept a 3D city model of urban area as that entity, which corresponding environmental analog is placed within 3D urban environment described by routine urban features and structures with buildings as the dominant objects among them.

Exactly for a couple of latest decades LiDAR (Lidar) data collecting processing technique has become an alternative (to areal imageries) data source for generating a 3D representation of natural landscapes and housing environments as well as the overall human-environment intercourse [25, 26]. Being able to collect directly the accurate *3D point clouds* of various density over urban territories, the LiDAR technique proposes an effective and beneficial data source in this meaning. By the way, this is illustrated further in this text.

There are Airborne (ALS), Terrestrial (Mobile - MLS), and UAV-LS (Unmanned Aerial Vehicle Laser Scanning) LiDAR techniques within the common remote sensing technology, that measures distances on the base of the time intervals between the laser signal transmitting / receiving. All ALS / MLS /UAV-LS mapping are surveying and mapping tools based on hardware platforms and software solutions [27-30]. The ALS lidar normally completes low-average density measurements of the underlying topographic surface, that result in the georeferenced, but unstructured set of points – lidar Point Clouds, while the MLS technique usually captures the walls – with a high resolution of façade details for buildings of different sizes. The UAV-LS lidar accomplishes *high density measurements* and can provide surveys that may be considered as hybrid ones, since it combines those output data, which can be provided by both ALS, and MLS Lidars.

The *automated feature extraction* (AFE) methods are the derivatives of ALS / MLS / UAV-LS Lidar remote technique. AFE-procedures produce 3D city models mentioned above, that are positioned in three dimensions just because of the AFE-method implementation. Since a common three-dimensional urban model may be a subject of more, than one hundred applications in various industrial domains [24], the AFE-methodology itself can be hardly overvalued. Municipal management, urban emergency services, noise and other hazard mapping, visibility analysis city infrastructure inventory, population and energy demand estimation with building models are those key urban applications, in which 3D models obtained from LiDAR data through AFE procedures are in an evident growing marketing demand according to the obvious reasons. To meet this demand requirements an AFE-model should be of high accuracy [31], what in some cases implies *a hybrid data source*. The latter means either a "LiDAR + areal image" hybrid [32], or additional involvement of MLS / UAV-LS data processing procedures for detailed extractions of building façades [33].

The *main research purpose* of this text is to outline with the phased methodological flowchart our conceptual multifunctional approach to the fully automated extraction of urban features, and present the original contribution to relevant algorithmic solutions within the modeling pipeline, as well as represent the web- and cloud-based software elaborated for these intentions.

2. Previous Works Done due to Building Detection, Extraction, and 3D Reconstruction from LiDAR Data

Fully Automated Feature Extraction, building Point Cloud segmentation, rooftop modeling, and 3D reconstruction of buildings, all have been among the main topics of thematic discussions in the remote sensing community through the relevant papers and on forums of various scales especially for the latest decade [26, 32, 34-42].

Now, AFE is still a vitally crucial part of what is being done and what researchers and other professionals are attempting to do better within the LiDAR surveying / processing domain. How has it further been progressed with its technique recently? We may consider as a key issue for efficient AFE-results a provision of a bridge between MLS / UAV-LS lidars, from one side, and ALS lidar, from another, one and vice-versa. Without any further explanations just in this paper section, we may only express a somewhat trivial idea, according to which *the composite models* of urban features extracted may be considered as the most effective ones [43].

2.1. Overall AFE issues

Because of ALS / MLS / UAV-LS joint surveying methods the automated feature extraction as a generating procedure for 3D urban models has become an evident alternative to the urban photogrammetry. It is a well-known fact, that besides various direct processing of aerial images, the Urban Remote Sensing (URS) (e.g., multispectral, hyperspectral ones [44]) traditionally provides the 3D building model generation from airborne-mobile-drone photogrammetric point clouds, but the lidar surveying / processing approach deals with similar dataset structures. Moreover, the novel sensor technology with its lower cost in comparison with previous hardware has expedited Lidar involvement in urban studies. Significant pros for such solution may be that fact, according to which ALS / MLS / UAV-LS sensors can deliver point datasets with densities of huge range (from one up to several thousand points per sq. meter). Even with the lower edge of this density range, it is feasible to detect urban structures, their approximate boundaries, and other various man-made features. Those models can be generated, which correctly resemble both wall, and roof structures. Many relevant methods for ALS / MLS / UAV-LS surveyed data processing have been proposed due to building extraction, detection, and 3D reconstruction [45-54].

Once we already commonly classified the automated feature extraction approaches on the base of an input data source [1]. The first approach means processing the high-resolution airborne imageries with additional including of *digital elevation models* (DEM) into an AFE-pipeline [7, 55]. Although some significant results have been obtained, the "only aerial imagery" approach may be considered as that one, which does not perform well enough in the densely built-up urban areas, primarily because of shadows, landscape gaps, and unsatisfactory contrasts in various urban configurations. According to this, the AFE procedures based exclusively on the first approach may not be fully reliable for the robust practical usage [33, 56]. The second approach straightforwardly employs LiDAR data and produces improved AFE-results, if compared to the imagery-only methods [31, 35, 37, 56]. What is more, this second approach implies an employment of very different techniques for processing ALS point clouds [57-59]. Methods applied within the third approach combine in a common case both aerial imageries, and all kinds of LIDAR data (ALS / MLS / UAV-LS) so that to employ the complementary information from all data sources [33, 60].

The AFE complete algorithmic content implies the ground and vegetation detection, while the man-made feature AFE-technique has to use either single-, or multi-return ALS / MSL / UAV-LS range and intensity information with application of the various thematic algorithms, e.g., as: neural networks [61]; *RANSAC* (the *RANdom SAmple Consensus* algorithm) approach for extraction of feature plains [62] with its key modifications [63, 64], that have been successfully employed by the Polygonal Surface Reconstruction method (the *PolyFit*) for feature reconstruction [65]; 3D Standard / Randomized Hough Transform (SHT / RHT) methodology that generally consists of three main steps: building points' detection, detection of building planes, and these planes' refinement [66-68]; implementation of knowledge-based entities [69]; the multi-scale approach [70].

Besides all algorithmic solutions mentioned above it may be reasonable to emphasize the hierarchical terrain recovery algorithm, that may robustly provide distinguishing between ground and non-ground points within an input point cloud by implementing the "adaptive and robust filtering" method [50]. Within this approach it is necessary to consider the whole range of input data to evaluate a DEM of high resolution for further feature extraction steps using this relevant hierarchical strategy. Thus, road linear features can be identified by classifying signal intensity and elevation data. Not only discrete building boxes, but also network features can be detected and extracted then. For example, man-made features of the road networks can be derived using a customized transformation technique, and then validated with road lines and topographic shapes obtained from an initial Lidar point cloud. Further one can obtain the attributes of road segments such as their widths, lengths, and slopes by computing some derivative information and enhancing existing metadata in this way. Other man-made features, firstly, building models are created with the higher level of accuracy, which would correspond to various Levels of Detail (LODs), beginning from LOD1 simplified box models and up to the models with internal partitions [71]. City Geography Markup Language (*CityGML*) is employed as a geoinformation data standard for presentation of the geometry and geographical data in digital models related to city buildings.

We have already presented *the High Polyhedral* models *(HPM)* of buildings and *the Low Polyhedral* ones *(LPM)* [1]. These two categories not only are the components of the authors' Lidar data processing methodology but also they are mentioned in the surveying section of this text only because HPM / LPM definitions reflect two significant mainstreams in the existing automated feature extraction. Obviously, these two dominant trends can be named with other titles by different authors.

Mostly a literature survey already completed above concerns exactly the HPM issues. The latter imply the generation of building models consisting of numerous polyhedrons, and because of this the relevant modeled entities can be accepted as "heavy ones" – *the heavyweight (HW) models*, which are produced by *the High Polyhedral Modeling approach*. It means, an HPM-building model may be generated from up to more, than one hundred thousand of Lidar points. It may be reasonable to state that the HPM-procedures are primarily based on the *Lidar point classification* that one, which is not directly associated with clustering, while the LPM-operations – *on the Lidar point cloud segmentation* through *clustering.*

The common work-flow of the HPM building model generation can be outlined as follows on the base of those literature sources that we have already referred to.

At its *first stage*, the building footprints (building base boundaries) are detected by segmenting DEM data obtained from LiDAR for two general classes: *ground* class and *non-ground* ones. The bare ground as a grid is delineated upon this step. A well-known, so called "sequential linking technique" is often suggested to reconstruct building footprints into regular polygons. These polygons then are improved so that to reach the cartographical standards [52, 72].

The so-called *prismatic models* are generated for those urban structures, which roofs are flat, and *polyhedral models* are created for those structures, which roofs are non-flat, at the *second stage.* At last, at the *third stage*, the vertical wall rectification operations should be applied, if there are sufficient MSL / UAV-LS or other correcting data in a relevant geodatabase for processing of this *area-of-interest* (AOI).

These three introduced stages may overlap almost any LiDAR data processing workflow. Most urban attributes of these building models are obtained from ASL / MSL / UAV-LS data. All corresponding HPM AFE-algorithms, that conclude the three stages workflow referred to above, should be tested using several geodatabases of varying earth surface type, vegetation coverage type, urban area type and LiDAR point density. Afterwards, normally the most effective algorithm should be chosen.

If we complete the general summary even for several overviewed HPM AFE-algorithmic results, this summary may demonstrate that in many urban territories the derivative DEMs accumulate most topographic details and remove non-ground features reliably enough. The transportation infrastructural network features are also depicted mainly satisfactorily even within densely built-up city districts. The extracted building footprints demonstrate to have enough positioning accuracy. Their estimated values may be equal to the accuracy obtained from data surveyed in field monitoring, while this traditional

surveying technique in many cases is a routine procedure of Lidar processed result accuracy evaluation [73, 74].

2.2. Detection, Segmentation, and Reconstruction of Building Roofs

The relevant modeling software tools can provide by a Point Cloud segmentation and clustering procedures of building detection and extraction the models of low-rise buildings preferably through rural areas and suburbs [1]. Simulating procedures stay within the Low Polyhedral Modeling approach, which is based on procedures of planar segmentation of Lidar point clouds rather, than on their classification (the case of HPM). The LPM building models produced are composed of not many polyhedral facets, and the number of points intended for a single model generation is limited approximately by a range from five and up to thirty thousand points maximum. Thus, the low polyhedral models can be considered as *the lightweight (LW) ones*. Overall low polyhedral modeling frameworks adapted to our LPM methodological approach proceed from the series of seminal papers in roof segmentation and reconstruction, therefore we have named this algorithmic technique as the *SAS*-methodology (the title has been abbreviated according to the authors of the initial approach) [35, 37, 75-77].

Methodologically we can reasonably define HPM as the automated extraction of *a whole building*, while LPM – the automated extraction of *this building roof.* A *segmentation* procedure of roof plains may be a key one within a corresponding AFE-pipeline titled for this section of our paper. This data-driven procedure was initially adapted for Lidar data from imagery processing [78]. Segmentation normally starts with applying *clustering methods* to Lidar point clouds [37, 76, 77].

The automated extraction of a whole building can be normally fulfilled by three similar sub-procedures outlined above for roofs, i.e., *building detection*, *building segmentation*, and *building reconstruction* [50, 59, 70, 79]. Although, contrary to roof extraction, all three sub-procedures, which are related to a whole building, may not be evidently less distinguishable. The case is a completely automated process of a whole building extraction may not yet be reliable enough from a practical point of view, because of the great complexity of actual urban environment with tremendous variety of its configurations. Thus, a whole building fully automated processing may need to pass a longer distance to provide wider usage of available supplementary data sources, for instance, city ground plans or municipal architecture schemes, so that to significantly enhance an ultimate processed result.

Somewhat more simplified methods, which provide roof extraction as equal to roof detection, are based on a *Digital Surface Model (DSM).* This model contrary to a DEM includes not only the ground, but the discrete features also. According to existing references, the DSM is calculated using an imagery and feature pyramids [80]. The finalized surface is refined then on the base of local adaptive regularizations. The roof detection step is grounded on the fact that a roof should be higher, than the neighboring ground – the topographic surface. This is normally estimated applying tools of the mathematical morphology to this DSM. The sliding window size technique involves the input information about the maximum roof size in its existing geographical extent.

This "only DSM involved" method has been further specified, when the building roofs are segmented depending on their estimated complexity, and in the same way finally reconstructed [81]. Usually, two types of parametric models are used for simple building roofs in those cases, when a building possesses either a flat, or a symmetrically sloped roof [50, 82]. Within other variants prismatic models are applied for complex separate roof structures or to connected building roof sets.

Two general classes of the roof extraction approaches are either the *data-driven technique* (e.g., *SaS*-methodology already mentioned above, and this technique also known as *a generic of polyhedral technique*), or the *model-driven one* (also known as *a parametric technique*). The latter implies some assumptions about topological and geometrical properties of a whole building model, generally, and due to a roof model, particularly. Two methodologies can contrast one to another in a categorical perspective: the data-driven method can be accepted as the *geometrical approach* (due to creating the roof geometry from the points of a given cloud), and the *topological one*. It is necessary to emphasize, that there is no a clear boundary between these two approaches [83]. Some researchers report the Hough transform technique within the model-driven segmentation methods [37], while other scientists recall it as one from the key data driven approaches together with region growing and *RANSAC* [66]. By the way, in the opposite case some fully automated approach has been clearly defined as a model-driven one, which applied for the 3D model reconstruction with prototypical roof templates (*CityGML LOD2*) [84].

The invariant moment technique has been applied for generation of the roof template library according to various building classes [46]. It is a quite known fact, that the model-driven approach may fail, when the inhomogeneity of feature distribution within a point cloud leads to biased parameters. Therefore, extracted models of complicated, not ordinary building roofs are produced by using data-driven algorithms. These algorithms are normally based on segmented intersecting planes and operate with triangulated point clouds [35-38, 62-68, 85].

It is a commonly known fact for the data driven approach, that most critical errors occur upon the determination of a total building roof outline, when trees are placed near a building box [62, 64, 66-68]. To avoid this issue various algorithmic approaches have been modified independently based on the detection and outlining of planar facets, e.g. [46]. A facet plane is normally determined by the point clustering procedure. A roof scheme is outlined by a connected component analysis. A key related assumption is that all the geometric model boundaries for roof segments are either parallel, or perpendicular to the main building disposition.

Thus, it is necessary to emphasize once more, that it may be not reasonable at all to distinguish whole building modeling from its roof reconstruction, while only ALS data are involved. It can be the same or, at least, very similar procedure, if only façades are not generated from MSL / UAV-LS data sources. Also, this procedure may be both within the HPM, and the LPM frameworks. In the mentioned context various researches suggested a boundary-based building / its roof extraction and corresponding TIN-based modeling and reconstruction [50, 86]. The buildings and their roofs

are reconstructed within this approach by triangulating each point cluster of identified facet candidates and clustering those fragmentary triangles into small patches, concluding piecewise planar fragments. Finally, the reciprocal intersections of the summarized planar facets are employed for extraction of building corners and for definitions of corresponding extracted feature orientations. A quite resembling technique have been used in our high polyhedral modeling for the building extraction [1].

It may sound reasonable to conclude this literature review by mentioning again, slightly more in details, the Hough transform and *RANSAC* algorithms and making references to so-called "global solutions to building extraction and reconstruction" [87]. Most earlier reports of the Hough transform (HT) applications mainly concerned 2D mapping of point clouds, until the "extending generalized HT" was proposed that provided detection of 3D features in a point cloud [88]. Standard and Randomized Hough Transform methods are ones among the most popular techniques of plane detection [68], and it can be employed for the determination of more, or less precise roof plane parameters – the triplets selected are accepted as possible roof planes [89]. Comparison of one of the Hough methods – RHT with RANSAC demonstrates a significant advantage of the RHT in the processing accuracy versus computing time tradeoff [68].

The classic RANSAC (FB81 according to [65]), as it was already mentioned above extracts primitives from point datasets [62]. The high-quality geometric instances extracted are a subject for further polygonal surface reconstruction of various features with 3D geometry. Nonetheless there are still no proved evidence that either Hough, or RANSAC optimizations have been successfully applied for modeling urban areas [68], but even optimized RANSAC methods may lead to the numerous false plains [37]. Probably it would be right to affirm that RANSAC is that approach, which has faced almost the greatest number of its optimizing solutions among other point cloud detecting / segmenting methods. Even early optimizations of this algorithms provided the removal of up to a half the outliers resulted from a given data set [90]. Probably the best overview of this algorithm improvements together with one more original optimization has been given in [63].

Despite all optimizations followed by rising algorithmic efficiencies, both HT, and RANSAC still hardly could be applied to more, or less significant urbanized areas, e.g. to the city district scope. To meet such challenge among number of other purposes sophisticated methods have been elaborated within the paradigm of "global solutions to building segmentation and reconstruction" [87, 91]. Roof segmentation and reconstruction have been consequently developed within the frameworks of this methodology. Both these two stages should be recognized as a solution of the appropriate energy functions' minimization problem. Firstly, after an initial segmentation completed, every Lidar point has been accurately assigned to its optimal plane by minimization of a global energy function [87]. It was named "a global solution", because the method could define multiple roof plains concurrently. After it, on the base of the segmented in this way roof planes another "global solution" has been developed and applied already for the roof reconstruction stage. The building box has been partitioned into volumetric cells, what allows to construct the roofs of the sustainable topology and the correct geometry [91].

2.3. Generalizing LiDAR-Based Solutions in 3D Building Modelling

We have finalized with building reconstruction reviewing AFE steps in the previous subsection of the text, while shortly mentioning in subsection 2.1 processing integrated data sources. According to number of references, reconstruction of buildings can be accepted as a complicated procedure of the digital presentation generation for those physical urban features, that can be extracted from Point Clouds and transformed into effectively structured 3D models with various attributes, while the quality of these models should be evaluated further [58, 86, 92, 93]. Generalizing or *hybrid* solutions in the mentioned extent mean not only the data fusion involvement [94, 95], but also using this basis for *footprint extraction* in the automated mode and its *boundary regularization*, since both operations are the basis for a roof reconstruction, and even for wall raising, when MSL data are not available.

Four following groups of authors almost independently developed some interesting *hybrid* approach lied within the following workflow of five major algorithmic steps [37, 38, 96-98]: A density-based algorithm of clustering for the individual building segmentation, which begins with footprint delineation; A rectified boundary-tracing algorithm is applied; A hybrid method for footprint planar patches segmentation developed; these methods select so-called "seed points" in the parametric space and generate the regions in usual (spatial) space; A boundary regularization approach that examines point outliers; Finally, reconstructing procedures are accomplished due to the topological and geometrical information about building roofs and using the intersections of footprint planar patches.

This hybrid approach proceeds from that fact, according to which there are two main classes of methods applied to trace building footprint boundaries: *raster-based* method and *vector-based* one. By the raster-based technique, the point clouds are usually transformed into a regular grid, after it the image processing methods are employed to indicate, trace, delineate, and regularize the boundary and footprint edges [99, 100]. Upon the vector-based approach, that was developed quite long before, a simple string delimits the exterior boundary, while other parameters, representing the inner boundaries, are extracted from raw LiDAR data [101].

3. Original Methodological Approach

3.1. Phased Methodological Flowchart

While introducing the methodological basis of our complete R&D cycle (from raw ALS data processing to solutions of use cases with building models), we stand on that broadly accepted point of view, according to which LiDAR remote sensing can be defined as the research and technological approach primarily used to obtain for further processing the information about the topographic surface, vegetation, and various features of the human infrastructure at a certain distance from an observer's point (buildings, bridges, roads, powerlines, etc.) [25-27, 31].

Due to the options and the necessities to cover large areas in the cities the ALS surveys obtain data from very different urban configurations, that can be distinguished, initially, by building

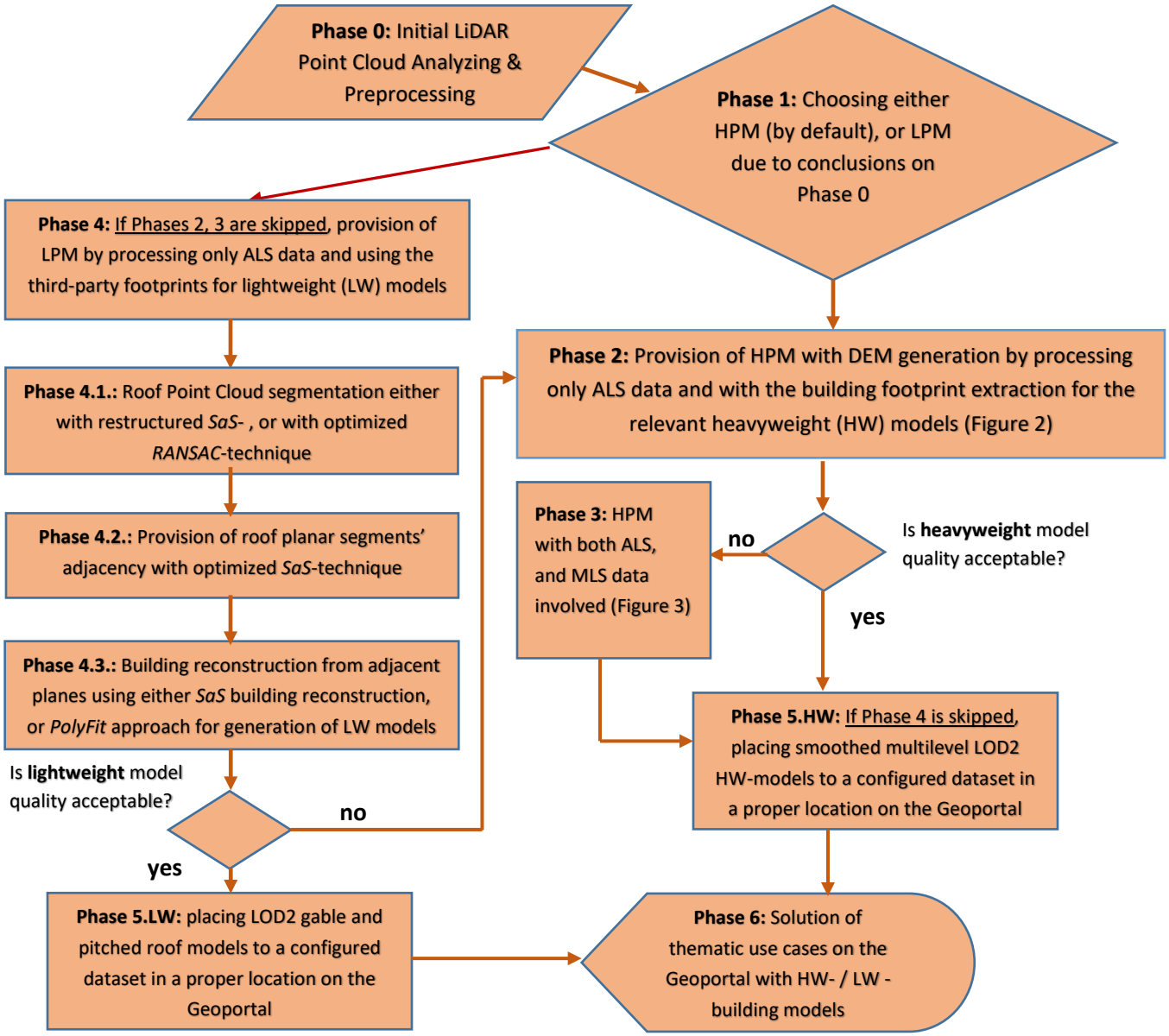


Figure 1: Phased flowchart of the overall methodological approach

geometries and by densities of sets of buildings. Probably two key types of the mentioned urban configurations are *high-rise buildings* of city downtowns and other central districts, on the one hand, and *low-rise buildings* of city suburbs / outskirts and neighboring rural areas, on the other hand. Two methodologically different AFE-techniques outlined in the previous chapter of literature review – the High Polyhedral Modeling of buildings and the Low Polyhedral Modeling – should be applied to these two urban configuration types exclusively – HPM to high-rise building sets, and LPM – to low-rise ones. We can emphasize as a strong point of our overall R&D approach just this joint employment of these two technically different methods within the united building detection, extraction, and modeling methodology, what is resulted further in the relevant software elaboration with two different tools – *Building Extraction (BE*, for a case of HPM), and *Building Extraction Rural Area* (*BERA*, for a case of LPM). Some common features of both approaches and relevant software tools have been already presented in some of our previous publications, although we have not made yet an emphasis on an allocation of both HPM, and LPM within a unified workflow [1, 4, 43, 102]. Since in many cases Airborne LiDAR survey relies on existing territorial regularities of urban areas, it is often provided consequently – from city central parts to outskirts. According to this, if presenting our general approach as *a phased methodological flowchart*, it would be reasonable to accept the HPM as *Previous AFE-Phase* and the LPM - as *Subsequent AFE-Phase* (Figure 1):

It is shown in Figure 1, that an activity diagram combines both HPM, and LPM issues, while an input is a raw LiDAR Point Cloud (*Phase 0). Analyzing & Preprocessing* phase implies evaluating point densities, survey induced and filtering induced errors, as well as checking a for a point cloud proper georeferencing. Two key

factors, that determine a solution to be made on *Phase 1* (either HPM, or LPM chosen) are the territories of specific urban configurations, that are overlapped by .LAS files, and if the third-party (*OpenStreetMap, Microsoft*, etc.) footprints are available for a given area.

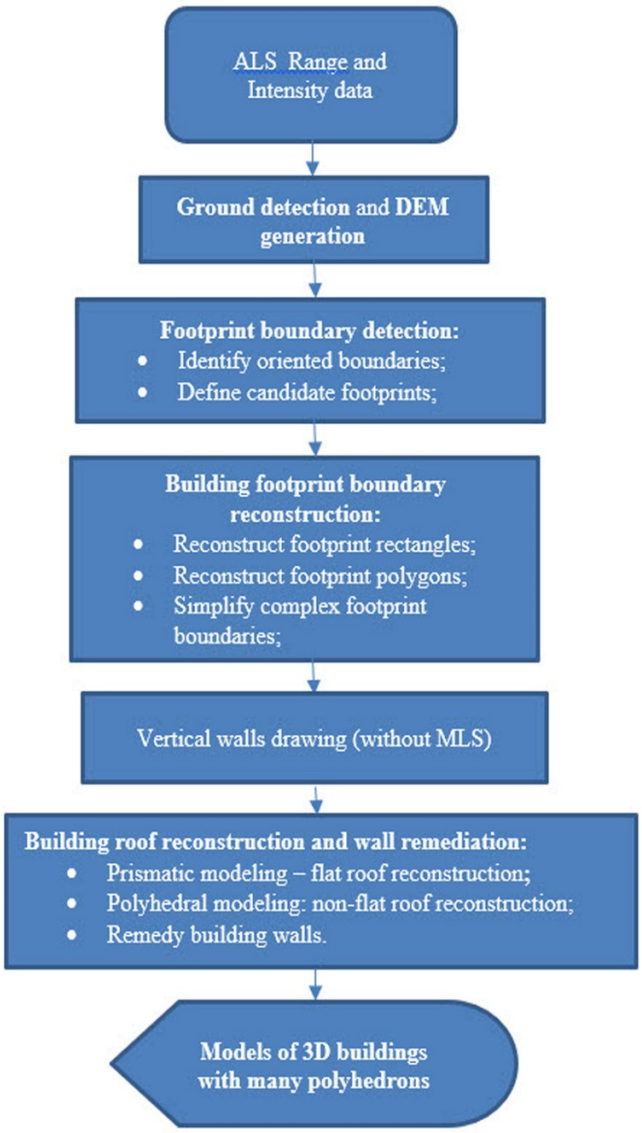
Phase 2 implies a provision of the phased flowchart HPM-branch by processing only ALS data with extraction of the original footprints and with generation of HW-building models, which surfaces consist of numerous polyhedral facets. The output model quality is estimated by several algorithmic evaluating parameters, and in the web-software UI (a user interface of the *Detailed Viewer* tool) by comparison of building surfaces' allocation in 3D space with their neighboring Lidar points. A procedure of such evaluation is illustrated further in this text. If model quality is accepted as either good, or satisfactory, it may be necessary to go directly to *Phase 5.HW*, since both *Phase 3* (involvement of MSL data in addition to ALS ones), and *Phase 4* (a whole LPM-branch) have to be skipped in this phased flowchart scenario. *Phase 3* is accomplished, if MLS data employment can substantially contribute to the quality of HPM-building models with not only precise reconstruction of their facades, but also with increasing of overall sustainability of the high polyhedral model of a building by adding supplementary facets of its minor surfaces. *Phase 5.HW* is also a further target, if acceptable model quality is obtained on *Phase 3*, and, in general, if the most of HW-models are of acceptable quality. Then on *Phase 5.HW* these models are place to a configured dataset to a proper location on our Geoportal. A digital elevation model is generated within the HPM branch only.

Phase 4 provides the LPM-workflow, which starts from *Phase 4.1*, within which surface normal determination as an initial step of roof surface reconstruction within the *SaS*-approach, is accomplished. It is followed by a roof point cloud segmentation either in the updated *SaS*-methodological frameworks, or by a segmentation provided with the optimized *RANSAC*-technique. Both these choices are described further in the text. *Phase 4.2* represents preliminary solutions for the LPM-building reconstruction stage with our original contribution to building a matrix of roof planar segment adjacency. *Phase 4.3* finalizes the building reconstruction stage by combining a building from geometric primitives extracted from a point cloud on the previous stages. If the most of LW-models of buildings with gable and pitched roofs are of acceptable quality, upon *Phase 5.LW* they are placed to a configured dataset of models to the Geoportal (it is defined further in this text). If the quality of the LPM output results is not acceptable, the HPM-pipeline is attempted as an alternative, and a repeated workflow starts again from *Phase 2.* The LPM branch of our overall methodological approach does not support the DEM generation functionality.

An accomplishment of the overall approach is normally finalized on *Stage 6* by solutions with HW- / LW- models of the thematic use cases on the Geoportal, what is briefly illustrated further in our paper.

Optimal LiDAR point density values that were empirically defined for processing techniques involved on *Phases 2* and *4* are between 5 and 140 ALS points per sq. m. If MSL / UAV-LS data are added to our HPM AFE-pipeline, these data are thinned out to acceptable values by sophisticated thinning algorithms, although *Phase 3* deals with up to several hundred of MSL points per sq. m.

A sketch of our core algorithmic workflow of High Polyhedral Modeling is depicted in two flowcharts below (Figures 2, 3). It has been elaborated within the frameworks of the integrated *Building Extraction (BE) / Change Detection (CD) / Digital Elevation Model-Generation (DEM-G)* pipeline of ALS / MLS / UAV-LS data processing. The complete algorithmic workflow depicted on the first flowchart (Figure 2) intends to extract both topographic surface, and urban features from ALS data arrays only, what corresponds to *Phase 2* of the overall methodological flowchart (Figure 1).



3.2. High Polyhedral Modeling for Building Extraction – Heavyweight Models

Thus, a flowchart depicted in Figure 2 does explain the thematic content of the *Phase 2* block of the phased

methodological diagram (Figure 1). Detailed consideration of this phase shows that initially the necessary preprocessing is accomplished within the first algorithmic block (*ALS Range…*). Then all Lidar points are separated for those, that belong to ground, and for other ones, that belong to non-ground features (block *Ground detection…*).

Delineation of the original building footprints (as a valuable alternative to the third-party ones) is a crucially important component of the HPM pipeline (*Phases 1-3, 5* Figure 1), and it is completed within the third block of the HPM ALS algorithmic flowchart (*Footprint boundary detection*) – Figure 2. The footprints extracted are a self-sufficient entity for a whole HPM-ALS pipeline of *Phase 2*, and the following block (*Building footprint boundary reconstruction*) predefines modeling of a whole building, but the preliminary to this reconstruction procedures of footprint boundary optimization and more precise allocation have to be accomplished according to three sub-blocks of this block indicated by the bulleted records (*Reconstruct footprint rectangles, Reconstruct footprint polygons, Simplify complex footprint boundaries*). Thus, through processing upon this fourth algorithmic block building footprints are determined as quadrangles, rectangles, or routine polygons. Then "pretended" walls are arisen from the defined boundaries of footprints (*Vertical walls…*block).

With the next algorithmic block (*Building roof reconstruction….*) the rooftops of buildings may be raised from these delineated boundaries and then corrected by the data of the same ALS point cloud used upon all five previous blocks of this flowchart (Figure 2). A building which has a flat roof is modeled in prismatic geometries (a bulleted record for a subblock *Prismatic modeling – flat roof reconstruction*). If a building possesses some complicated shape of its rooftop, it is modeled as a polyhedron (a bulleted record for a subblock *Polyhedral modeling: non-flat roof reconstruction*). As it was already mentioned above "HPM" means that initially reconstructed building facets consist of many polyhedrons and represent HW-models contrary to that solution ("LPM"), while LW-models can consist of few polygons only. According to the understandable reasons, HW-models are normally constructed and visualized as comparably heavier entities (from 20 to 150 thousand of points are processed per model). Thus, mandatory smoothing and noise removing should be evidently provided. For these aims our original contribution to a Delaunay refinement algorithm [100] has been employed. The corresponding "covering Delaunay TIN" is involved to algorithmic sub-blocks *Polyhedral modeling: non-flat roof reconstruction*, then – to a subblock *Remedy building walls.*

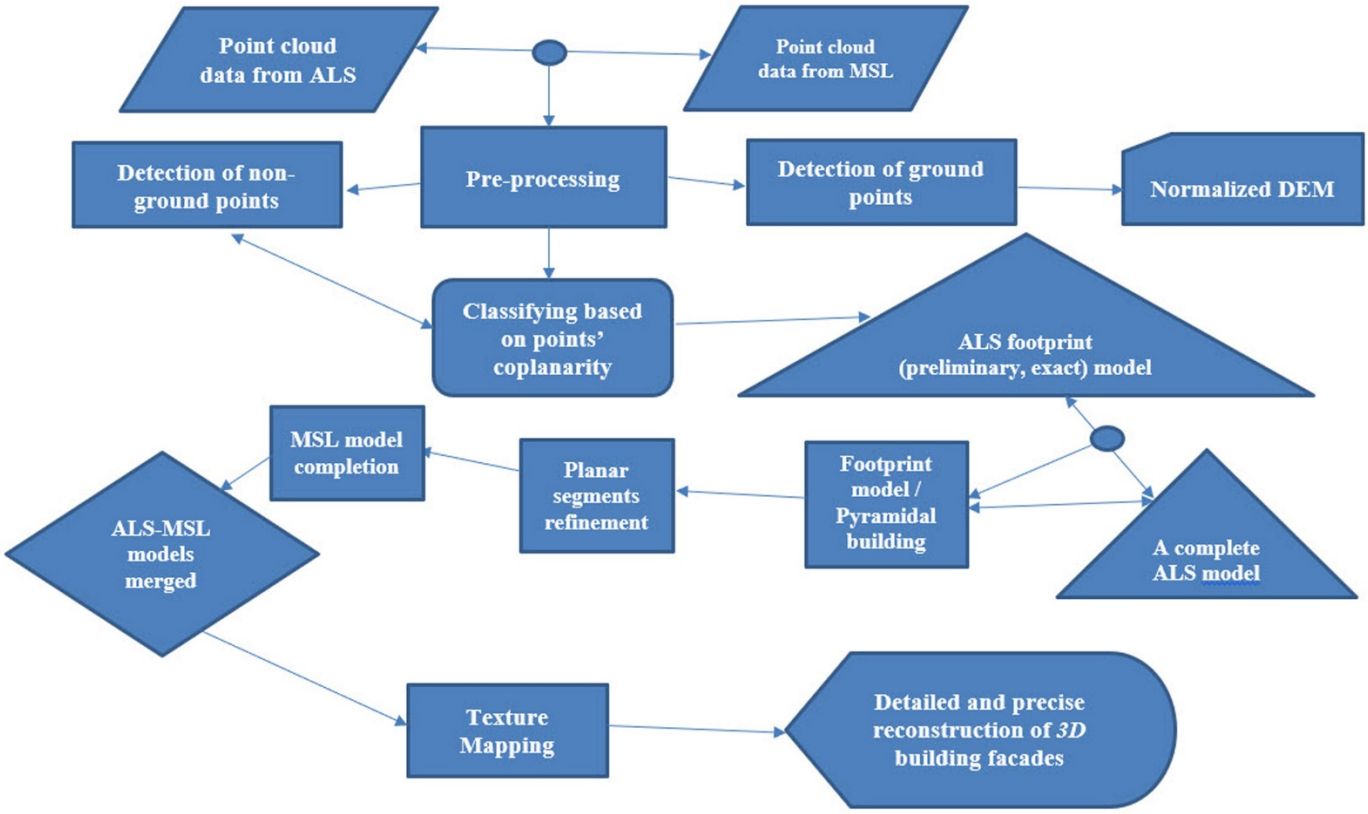
In the presented way we have just introduced the thematic content of *Phase 2*, which is in our methodological flowchart (Figure 1). As it has been emphasized above, if the output model quality (the finalized algorithmic block *Models of 3D buildings with many polyhedrons*) for the results based on the ALS data only is not acceptable *Phase 3* of the phased methodological diagram should be involved.

The initial input data for the HPM ALS / MLS / UAV-LS algorithm, which is core content of *Phase 3*, and which flowchart is presented in Figure 3, are as follows:

- The points that have been received by ALS (**x**, **y**, **z** coordinates and RGB color attributes) Lidar survey. It is geoprocessed as a set of 3D point layers;
- The points that have been received by either MSL, or by UAV-LS scanning, or by both (**x**, **y**, **z** coordinates and RGB color attributes). It is also geoprocessed as a set of 3D points layers;
- The regular earth surface (a grid layer - a DEM), that has been generated upon the first algorithmic steps with existing gaps within those locations, where ground data are absent, is also accepted as an initial input for further processing;
- The smoothed regular earth surface (a refined grid layer) as a DEM with gaps removed by chosen algorithmic procedures, therefore this surface is a continuous one.

On the base of points' distance to a smoothed DEM the HPM ALS / MLS algorithm of *Phase 3* (Figure 3) separates raw Lidar points into two categories, as it is done in those procedures completed by the HPM ALS algorithm of *Phase 2* (Figure 2). The first one contains the ground points that form "a ground level of this building footprint". The second category contains non-ground points that are clustered with this building footprint. The relevant point cluster normally represents a single building or a tree. The upper blocks of the ALS / MLS algorithmic flowchart display all preliminary pre-processing / classifying steps that are provided (*Detection of non-ground* and *ground points*; *Pre-processing; Classifying based on points' coplanarity*; *Delineation of preliminary footprints*) (Figure 3). On further classifying steps any Lidar point is assigned to be either a building point, or a non-building one. According to this binary labeling the primary footprint extraction is implemented as a technique of obtaining building footprint polygons from ALS data exclusively, thus this procedure is completely the same in both algorithmic workflows (Figures 2 and 3). This operation is normally performed in two steps: generating *preliminary footprints* from *existing grid gaps* and extracting the *exact finalized footprints* from these preliminary ones. The preliminary footprints are extracted as *No Data* holes in a grid. It is reasonable to examine this extraction slightly more in details in comparison with other components of both algorithmic workflows. It is the content of the flowchart block *ALS footprint (preliminary, exact) model* (Figure 1). The following modeling-extracting steps are performed for preliminary footprints:

1) An initial grid is divided into the blocks of a size not more than 4'000'000 cells. Neighboring blocks possess an intersecting part of the size, that is specified by the *Max building size* parameter. After this step each building block is processed detachedly;
2) Each cell either is marked as a ground (where the surface grid contains some value), or not classified at all (where a grid has no data);
3) Thus, assuming *expandCount = (params-> MaxCorridorSize / (2 * cellSize) + 1*; We expand the ground class on *expandCount* units in the urban metrics. Afterwards we expand *No Data* class on *expandCount***;** This step separates single long buildings in their ribbon sets connected by urban corridors;
4) Assuming *expandCount = params-> MaxCorniceSize / cellSize + 1*, we expand *No Data* class on *expandCount* to join semi-attached buildings;



5) Consequently, we find gaps in a grid through 1)-4) and each of these gaps retrieves its own index. In this way, only those holes that are completely isolated, i.e. surrounded by the bare ground, are expected to be found;
6) Afterwards, we should check each of gaps and store only those ones, that possess an area larger, than a certain threshold value specified in the preferences;

Thus, we obtain the areal boundary for each gap considered to be a preliminary footprint. The *exact footprints* are generated proceeding from the preliminary ones and from some supplementary information of this point cloud. For every "exact footprint" several following steps are provided.

1) All the points delineated by a geographical extent of some boundary (its size is the algorithmic parameter, which is equal to *MaxCorniceSize*) are selected for processing;
2) All Lidar points, that are above the ground surface, but lower than *2 m* above are classified as the *ground points*;
3) After this, through all points the triangulated network is being built. Within this network all edges, which possess at least one junction, that has been classified as the ground, should be deleted.
4) This triangulated network is divided for several connected segments by removing all the edges, that are longer than *MinDistanceBetweenBuildings.* All network segments that have an area more, than a minimal building footprint value (that is the algorithmic parameter), are assumed to be *the footprints* or *building parcels*;
5) The next algorithmic step is *filtering trees*. There are three possible customized options in this procedure:
a) Tree filtration is *turned off*. In such a case this step is skipped;
b) Applying *preliminary filtration*. The information about *trees* or *non/trees* is obtained from the classified source Lidar file;
c) Applying to a point cloud our own classifying technique.

Upon this fifth algorithmic step, the triangulated network is normally built through all points that are located inside those previously created footprints (preliminary ones). Then all vertical edges associated with this extracted from a point cloud entity are removed. The edge is supposed to be vertical, if its horizontal slope is more, than a certain threshold – a minimal angle, that is a parameter of the algorithm. After this procedure completed, the triangulated network is partitioned for several connected components. Then the points from components, which area is less than *MinBuildingPartArea*, are designated as non-building points, and are removed from the footprint entity.

6) At the next algorithmic step, a new triangulated network is built through the points that belong to building parcels only; moreover, those vertical edges, which 2D length is longer than *MinWallSize* (the key algorithmic parameter) should be removed. The jointed polyhedrons with their aggregated areas, which are more than minimal building parcel area, are accepted as *the footprints*.
7) The final step of the exact footprints extracting is expanding of their preliminary templates obtained through 1)-6). This step is necessary, since some building and infrastructural units, that are above the ground or above other constructive parts, are filtered out on the previous algorithmic step. To

expand a footprint template the iterative algorithm is applied. Upon every step of this algorithm those points, that have not been classified yet as a constructive building or the ground, are classified as *a building* in a case, when this point is nearer to a footprint, than some assigned threshold distance. Afterwards, all these points are added to a footprint, and the new expanded footprints are built as an external boundary, that crosses all points that were previously classified as building points.

After all *exact footprints* are extracted from the preliminary ones (footprint templates), each of them should be checked for intersection with a source preliminary footprint so that to avoid topology errors. All those ones that do no intersect have to be filtered out. At the end of the block *ALS footprint (preliminary*, *exact) model* all Lidar points, that are located inside the examined parcels, and that do intersect, should be marked as the *building points* (in this way avoiding an extraction of other footprints that may be located within the same area). Finally, we accept the exact footprints as some *planar parcels*. The algorithmic step of their refinement is completed at the next algorithmic block *A complete ALS model*. The procedure of the planar segment refinement is quite lengthy therefore it may be a subject of description in another text. Thus, summarizing all stated above, we should emphasize that a preliminary footprint is an entity, which is extrapolated through the ground points without filling corresponding gaps at perspective footprint locations. The exact footprints are built proceeding from preliminary ones and by providing outlined algorithmic steps.

Just as in the HPM ALS algorithmic pipeline of *Phase 2* (Figure 2) our update of a Delaunay refinement algorithm has been employed. The relevant "covering Delaunay TIN" creation is completed in the flowchart blocks of *Planar segment* refinement, *Footprint model / Pyramidal model*, *A complete ALS-model* (Figure 3). All algorithmic blocks mentioned above are provided for processing ALS data only and reconstructing only *HPM-building roofs* and some other supplementary constructive components.

The HPM AFE results of both *Phase 2*, and *Phase 3* obtained by the *Building Extraction* software tool is a set of building models, each of them, as a rule, consists of few façades. The textures may be placed on these façades. Each of these heavyweight models is stored by *.OBJ /.B3DM* ("Batched 3D Model") /*.KML* formats, but in the relevant software environment the main inner format is *.OBJ.* Each HW-model has its six mandatory components:

- *A footprint*. It represents a smoothed *2D* polygon, which allocates a building footprint;
- *A pyramidal model* that is *not draped*. This model consists of several horizontal polygons at different levels. From every layer a vertical building wall is dropped to the previous level. From the lowest level the walls are dropped to the ground level;
- *A draped pyramidal model*. This model is the previous one, but each of its polygons has *a texture*. ALS points are only used for creating textures due to quasi-horizontal polygons ("roofs"). As far as creating textures for vertical polygons ("walls") is concerned, preferably the MSL points should be applied to support a desirable level of details (*LOD*);
- *A complete model*. This model is generated either by using of the detailed TIN techniques on the Airborne Lidar points only (*A complete ALS model* block – Figure 3), or by combining ALS results with the wall segments reconstructed on the base of MSL data and obtaining a combined model (*ALS-MLS models merged*);
- *Texture Mapping:* draping textures on an ALS-MLS merged model;
- *Detailed and precise* reconstruction and visualization of the finalized model with 3D building façades.

The reconstructing operation for building façades is exposed by the *MLS model completion* block of the flowchart in Figure 3. The latter provides a necessary "noise clipping" procedure for building wall surfaces.Since if the distance threshold value, that means a metric length from a given Lidar point to an extracted façade, is either lower, or approximately equal to the points' density value, the reconstructed planar surface may be a set of outliers, e.g., small peaks. Exactly in this case a procedure of noise clipping should be provided.

As it has been aforementioned already, the ALS models, on the one hand, and the MLS models, on the other hand, are merged at a certain point of the flowchart in Figure 3, just after the *Planar segments refinement* and the *MSL model completion* blocks, and before the *Texture mapping* block. This block means the penultimate algorithmic step – texture draping on a merged model. In this way, for each modely component, every polygon in the pyramidal model, the textures are being built. All those Lidar points that are located near the texture extent can be found. These points are projected on the roof / façade plane, and the coloring of the texture pixel is initiated. All predeccor algorithmic blocks in the flowchart of are ended by the *Detailed 3D reconstruction of building façades* one. Finally, the results are being delivered into *.OBJ* and *.B3DM* formats, and a composite high polyhedral model can be displayed by the relevant visualizing software tools developed by the authors (Figure 4). It can be seen, that the HPM AFE results in numerous polygonal segments, which represent quite a rough surface, only while zoomed in. Even a model of numerous polyhedrons can be impressively displayed by our provided visualizing technique.

After completing the HPM-algorithmic workflow, a user obtains building models for display either in the *ElitCore* desktop application, or in the web-based *ELiT* (EOS LiDAR Tool) *Viewer* with three levels of detail (*LODs*): *LOD 0* represents a *model footprint* as its projection on the plane; *LOD 1* is a 3D object that exposes a building as a set of prisms - *a pyramidal model*;

LOD 2 displays smoothed multilevel models in minute details (Figure 5). We have expressed an idea in the introduction to this text, that the composite (ALS / MLS-UAV-LS) model of extracted features may be the most effective one, primarily, with respect to user' various applications in many industrial domains. Such a typically smoothed multilevel model of *LOD2*, a HW-model, may be like follows in the *ElitCore* desktop UI (a user interface) (Figure 5). Thus, despite visualizing rough building surfaces, if significantly zoomed in, while applying to the LiDAR data intentionally selected and refined by the ALS / MLS algorithmic workflow on *Phase 3* of the overall methodological flowchart (Figure 3).

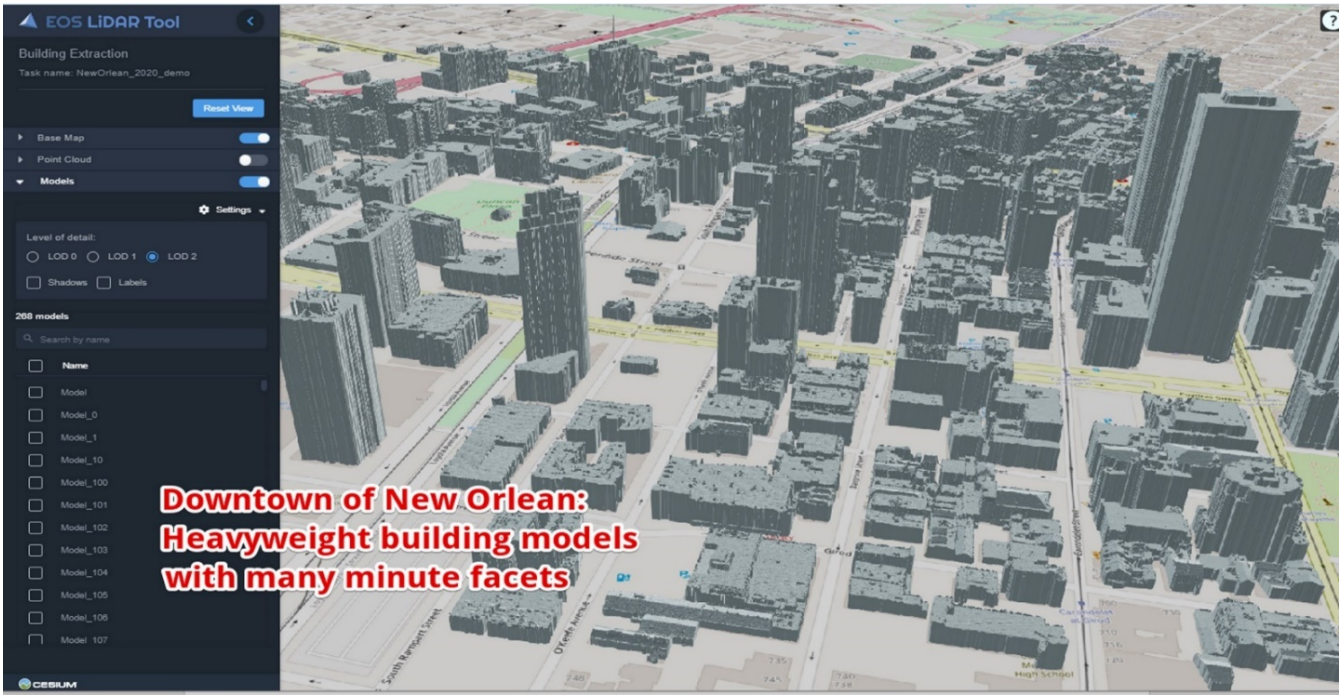


Figure 4: Results of the HPM ALS / MLS workflow displayed in the viewer of the web-based applications – building roofs of many polygonal segments. *The downtown of New Orlean, USA* (a dataset from the *EOS LIDAR Tool* landing page: https://eos.com/eos-lidar)

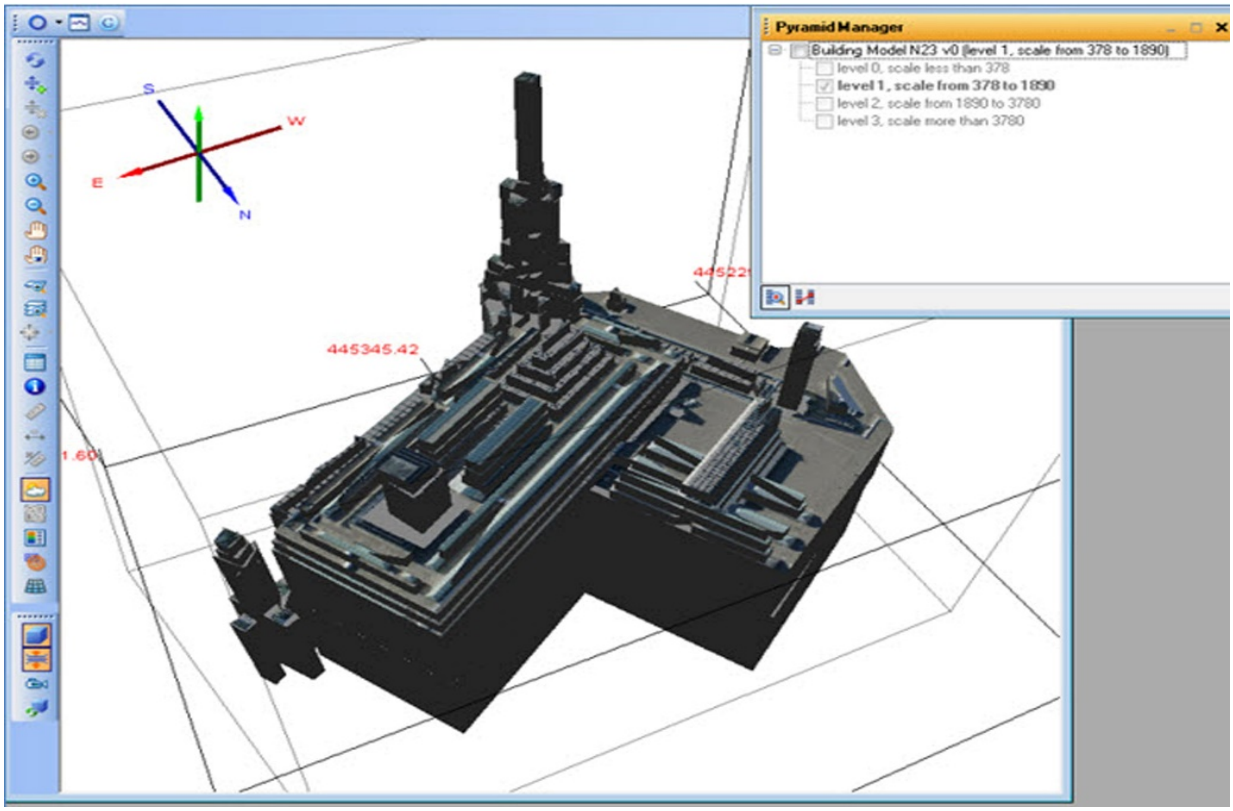


Figure 5: Model of one the historical buildings from the Ottawa downtown in the *ElitCore* desktop user interface

The presented methods of the HPM on *Phases 2, 3* is a completely original AFE-methodology elaborated by the authors. It has been implemented in *Building Extraction (BE)* (Figures 4, 5) and *Change Detection (CD)* (Figure 6) functionalities of our both desktop, and web-based applications.

3.3. Low Polyhedral Modeling for Building Extraction in Rural Areas – Lightweight Models

3.3.1. Common issues

Oppositely to the High Polyhedral Modeling technique our another AFE-approach is the Low Polyhedral Modeling method,

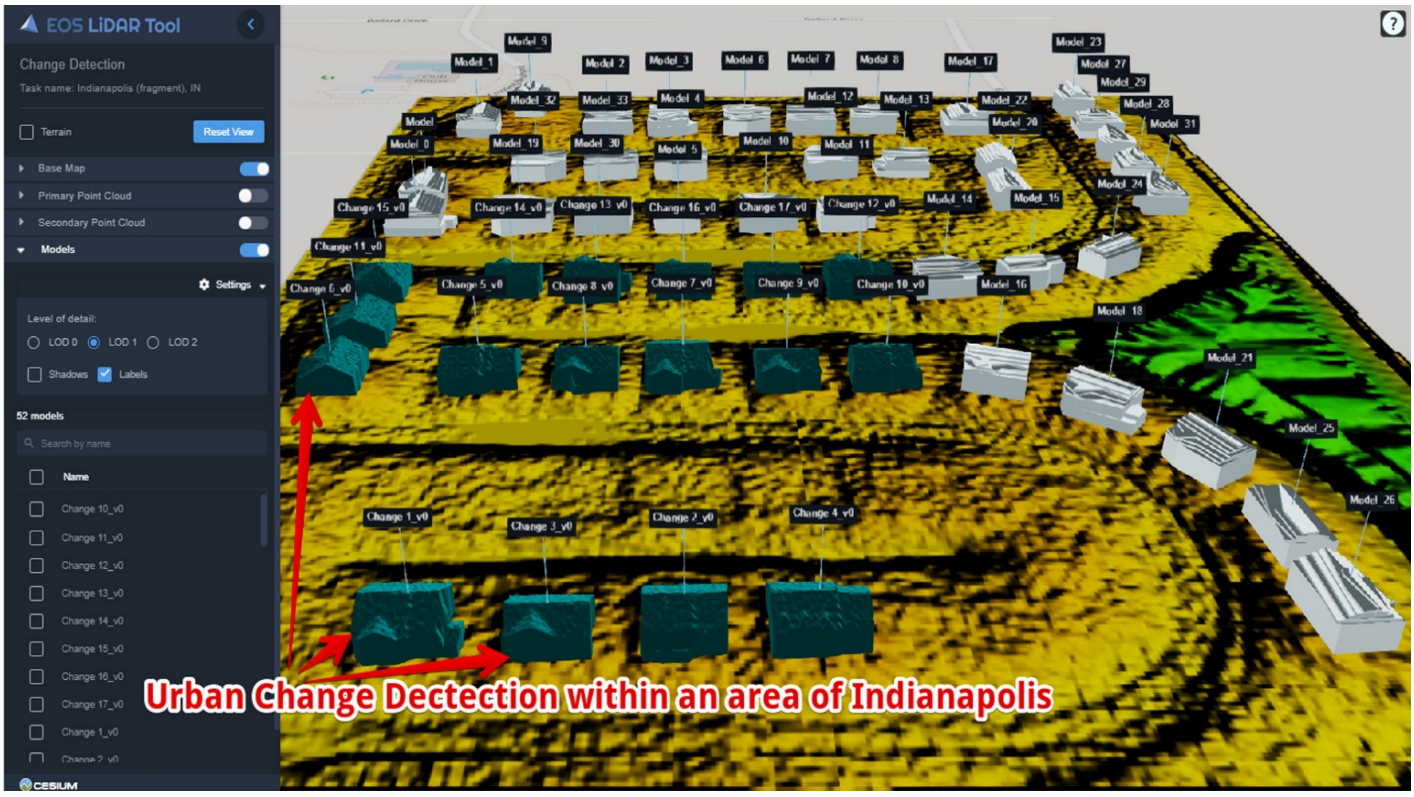


Figure 6: Results of the CD functionality application: new buildings (labeled as *Changes* in dark green color) as urban changes that appeared within one year's time. Models of both new, and old buildings (labeled as *Models* in light grey color) are placed on a DEM generated. *An area of Indianapolis, Indiana, USA* (a dataset from the *EOS LIDAR Tool* landing page: https://eos.com/eos-lidar)

On the finalizing building reconstruction stage theoretical basics of the *PolyFit* [65] have been effectively implemented in the applied solutions with web-software (*Phase 4.3*).

We have mentioned already, that initially the LPM is intended to extract low-rise buildings in either rural areas, or city suburbs. Nonetheless, it has been successfully applied to various urban configuration even within central parcels of the large city areas. Because of the segmenting / clustering procedures that drastically decrease a number of polyhedrons as constituents of a model extracted, such model is titled as "low polyhedral" one. We emphasized already, that these models are extracted, reconstructed, and visualized as the *LW-models* (approximately from 5 to 40 thousand of Lidar points processed per one entity). After completion these models are composed of only few polygonal planar segments. Models are reconstructed in this way, and each of them represents a lightweight modeled feature as a final solution.

Building reconstruction of low-rise constructive features is implemented as a finalizing procedure of building modeling (*Phase 4.3*). This algorithmic stage begins with the adjacency matrix of roof plains generation, that exposes the connectivity of the delineated planar segments. Our original contribution to the formation of the plane adjacency matrix is described in the one of the following subsections of this text.

Both roof interior, and exterior vertexes are determined. Topologically consistent and geometrically correct building models are obtained through implementation of the extended boundary regularization approach. The latter is based on multiple parallel and perpendicular line pairs, and just this approach expedites an achievement of the reliable building models. The model precision can be effectively tuned, despite it understandably still depends on the data precision. In any case the model precision may meet the strict customer requirements. The low polyhedral model implemented corresponds to *LOD 2* of the City GML standards [17, 19, 22]. The following visual presents the main content and some attribute information of this LPM algorithmic result in the interface of our cloud-based software (Figure 5). A number of acceptable models among all LW-models generated is a criterion of going either to *Phase 5.LW* with model output to the Geoportal, or *Phase 2* so that to repeat processing and modeling by shifting to the HPM branch (Figure 1).

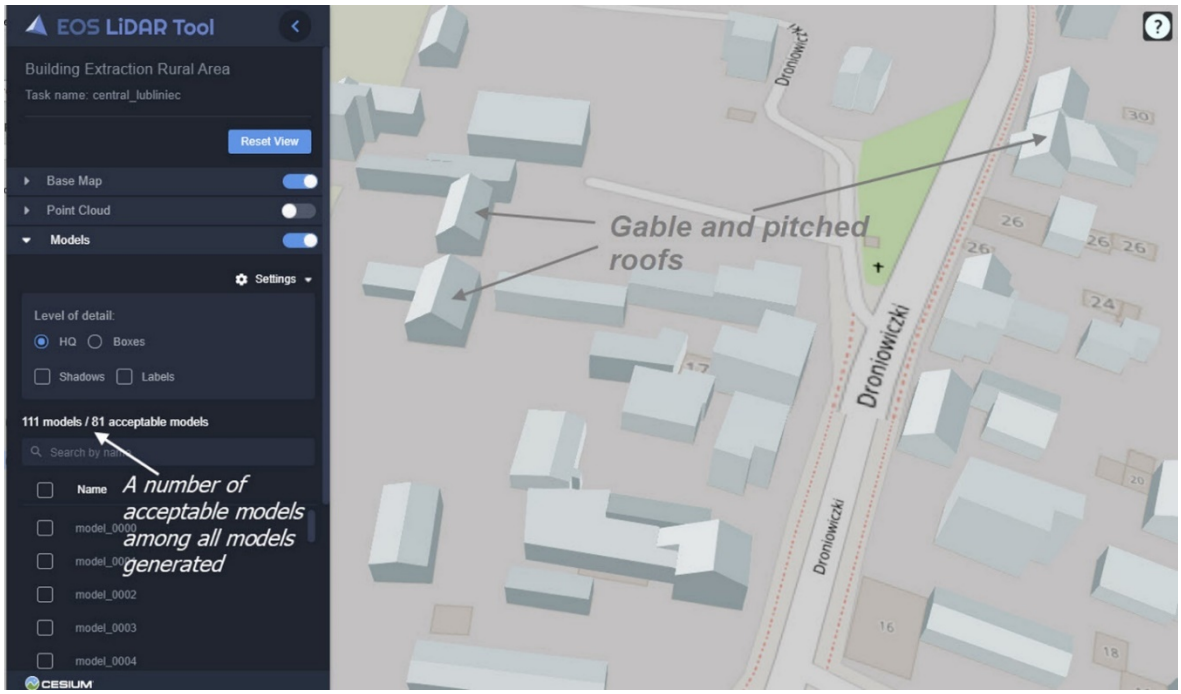


Figure 7: The LW-models of buildings with gable and pitched roofs (pointed to by grey arrows) in a small town. *The city of Lubliniec, Poland* (a dataset from the *EOS LIDAR Tool* landing page: https://eos.com/eos-lidar)

3.3.2. Enhancing SaS-segmentation by restructuring its pipeline

It is an initial content of *Phase 4.1* of the phased methodological flowchart (Figure 1).

2. To Assign unclassified points to the clusters. An addition of formerly unassigned points to a proper cluster might be a key procedure for various configurations of roof individual segment delineation. Some of these points might not be assigned to any cluster due to noise impact (a roof segment under a tree), or because of neighborhood nonplanarity on a seam between two or more roof slopes.

3. To Separate parallel planes inside each cluster. Surface normal vectors are being clustered initially. A clustered entity with the same normal vector is created. Thus, primarily a cluster is a set of planes, and a distribution of parallel planes into separate clusters inside this primary cluster takes place just on this step. The parameter ***D*** (from Item 1) is being computed for all points, and the latter are clustered. The output is a set of planes with the same normal vector, but on the different distance from the center of coordinates.

4. To Split coplanar clusters using the Voronoi neighborhood. A segmented plane may combine those roof segments, that are not connected spatially, but other segments are between them. Such coplanar planes are separated on this step. Commonly, these Items 5 and 6 correspond to the Separation of Parallel and Coplanar Planes step in [37].

5. To Remove small clusters (optionally – using the Voronoi neighborhood). Small area segments can be left after separation of parallel and coplanar clusters. These features should be eliminated as the noise and clustering errors.

6. To Remove the near-vertical clusters, because the modeling technique does not use them due to their unreliability.

It is reasonable to emphasize intentionally that the *Phase 4.1* initial content steps 1-6 can be completed only for a point cloud of some robust density, that may lie within the earlier mentioned range of 5-140 ALS points *per sq. m.*

3.3.3. Providing RANSAC enhancements

It is another content of *Phase 4.1*, and the optimized *RANSAC*-technique can be employed for a roof plane segmentation as an alternative to the *SaS*-approach, which has been somewhat restructured in comparison with an original issue in Item 1-6 in the previous subsection. We have already mentioned in the reviewing section of this text, that the RANSAC algorithm is able to extract a manifold of geometric primitives with different types of their shapes. RANSAC can deal with large number of outliers in the initial data, since this resampling method uses the smallest number of points necessary for estimation a given geometric primitive [62-64, 89, 103, 104]. Thus, the corresponding geometric primitives are obtained, if they approximate definite majority of points.

Some debatable issues present in the initially attempted *SaS*-segmentation, which appeared to be drastically sensitive to various outliers in the point clouds like, e.g., overhanging (above a building) trees. Because of this and due to the Lidar induced errors usually caused by sharp changes in the heights for the points belonging to the same feature, we have finally employed *RANSAC* with some editing of its general scheme. It may assist in meeting the challenges mentioned. For example, when the most points do not have the planar neighborhood, while the normal surface vectors, that are nonetheless found for the point minority, are located randomly, and this does not allow to delineate a corresponding point cluster for plane extraction.

In the outlined case, the *RANSAC* shape extraction strategy gives an opportunity to delineate a planar segment even facing the

challenges mentioned, because the sufficient condition is a randomly drawn point set, that can be quite feasible for this drawing, and employed then for constructing planar segment primitives. 2D planar points and 3D non-planar points can be separated ones from others using this technique. Figure 8 shows the example of a building model, which planes have been segmented with the optimized *RANSAC* from a point cloud with numerous outliers. Then this model has been refined and efficiently reconstructed, and even an overhanging tree, that is comparable in its size with a building, has been removed.

We have selected a following number of the basic *RANSAC* enhancements practicable for the ALS point cloud segmentation and implemented it in the relevant algorithmic pipeline embedded in our web-based software:

- *Number of iterations* upon the planar facet-candidate selection *is not a constant*, but it depends on a qualitative value that is the *Best Current Candidate Plane Index (BCCPI)*, a total number of Lidar points involved, and number of attempts completed yet. The purpose of the BCCPI value introduction is somewhat resembling to the RANSAC "score of the shape" [63], that also provides formalized estimation for the candidates of planar facets.
- Any BCCPI value is defined as *number of points within the delta-neighborhood of a plane minus the "penalty charges" for the point dispersion.* Thus, the less is point scattering from a detected plane, the higher is a qualitative

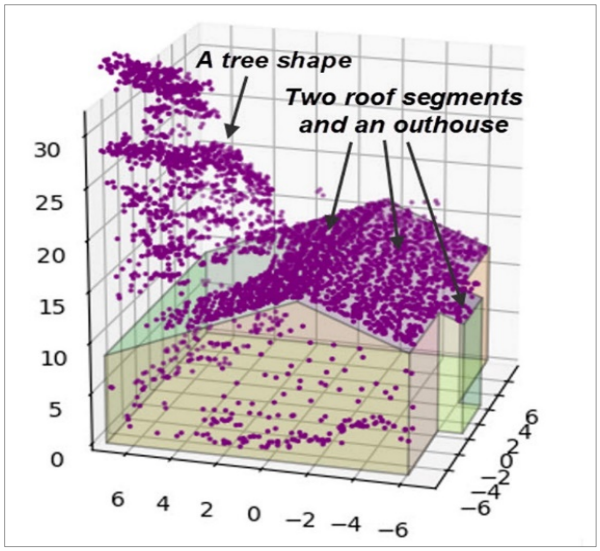


Figure 8: A building model reconstructed from the plains detected and segmented by the optimized *RANSAC*. A model is visualized by the *Matplotlib* python library. A building located in *the city of Lubliniec, Poland*

value of a candidate plane; the bigger is the scattering, the lower is a value mentioned. This technique reasonably allows to prefer those candidates, which approximate well even a relatively small point cluster, but not those random plains, that cross a whole extent of a model and overlap a big number of points in its delta-neighborhood. It is quite significant especially for the *gable roofs* with minor angles between its two slopes, or for *pitched roofs* with small angles within each pair of this roof slopes.

in point clouds with low data noise, but instead, a threshold for the BCCPI values may be intentionally decreased for overlapped strips resulted upon the airborne laser scanning, when a flight changes surveying direction.

- According to the improved *RANSAC* basics the point sampling for a plane-candidate may proceed from a whole point cloud with associated normal vectors, and the technique output is a set of planar primitives with corresponding point sets [105]. Such procedural content may cause some significant inaccuracy itself, while attempting to detect a plane, which may cross a whole cloud as explained above. Instead, we introduce *the point sampling from a randomly chosen limited neighborhood.* On the one hand, with this selection we still have the chances to approximate a big planar facet, if one does exist, on the other hand, we are expected to approximate some relatively small planar segment precisely enough, if a big one does not actually exist. If sampling neighborhood reduction is not provided, the probability that all three points, *a point triple*, would approximate (a key criterion of plane detection) one small planar facet has a low value.
- A point triple should be checked for *the degeneracy*: if an area of their triangle is too small, then the probability of processing error is high enough, and the whole computation cycle may fail. Therefore, *such triple is rejected*, and another one is sampled.

The overall structure of our optimizations of *RANSAC* is presented in the following *Algorithm Pseudo-Code* (Figure 9):

1. Main cycle:

```
working_set := input_points
result_planes := []
WHILE len(working_set) < threshold:
  plane_parameters, remaining_points := fit_plane_ransac(working_set)
   result_planes.append(plane_parameters)
   working_set := remaining_points
```

2. Selecting candidate plane function (fit_plane_ransac)

```
best_sample_quality := 0
best_sample_size := 0
best_sample := None
iteration := 0
WHILE probability(best_sample_size, points, iteration) <
quality_threshold:
  IF iteration >= iteration_threshold:
    reduce_dispersion_influence_on_quality_estimation()
```

```
  IF check_quality_limit_reached()
    BREAK
  iteration := 0
 p1, p2, p3 := select_candidate_points(points)
 iteration := iteration + 1
 IF NOT check_candidate_area(p1, p2, p3):
  CONTINUE
 candidate_sample, candidate_quality := get_best_candidate(p1, p2,
p3, points)
 IF candidate_quality > best_sample_quality:
  best_sample_quality := candidate_quality
  best_sample_size := size(candidate_sample)
  best_sample := candidate_sample
```

```
3. Assessing candidate quality function – BCCPI value
(get_best_candidate)
plane_parameters, sample_points := fit_plane(p1, p2, p3, points)
points_distance := get_distance_to_plane(plane_parameters,
sample_points)
distance_dispersion := mean(points_distance)
quality := quality_influence * (len(sample_points) / len(points)) + (1 -
quality_influence) * (1 - dispersion / plane capture_distance)
```

Figure 9: The Pseudocode of the optimized *RANSAC*

It has been determined that the optimal initial value of quality_influence may be 0.4, while the iteration step in reduce_dispersion_influence_on_quality_estimation is 0.1.

3.3.4. Optimizing SaS-reconstruction by Voronoi Neighborhood

One of our key contributions to the SAS-methodology corresponds to *Phase 4.2* of the phased methodological diagram (Figure 1). It consists in the refinement of the optimized adjacency matrix obtainment on the preliminary stage of building reconstruction. This matrix is the most significant issue for the delineation of the adjacent planar segments in the model generated. Thus, we have enhanced the *SaS*-workflow by *the extensive use of the Voronoi neighborhood* for computation of the roof segment cluster adjacency on the building *reconstruction stage.* Although the authors of the original workflow referred to the Voronoi diagram only on the point cloud *segmentation step* – the Voronoi neighborhood V_p has been employed for the surface normal computation only [37, 76]. Applying to the Voronoi diagram on the building reconstruction stage, we remove both horizontal, and vertical gaps in the processed data, as well as mitigate the nonhomogeneous point density of a primary point cloud.

The Voronoi diagram has been applied for the roof cluster adjacency determination and for separation of coplanar clusters, while the limited Voronoi diagram has been used for avoiding the side effects of the cluster adjacency determination. Also, the Voronoi diagram has been applied for the reliable identification of some traditional architectural constituents such as building awnings and or building overhands [102].

The Voronoi neighborhood for the roof cluster adjacency determination means the obtainment of the planar segment optimized adjacency, on the condition that these segments are preliminary delineated. The authors of the *SaS*-methodology use the Voronoi diagram while providing eigenvalue analysis, and each point of neighbors are being delineated. Nonetheless upon computing the cluster adjacency these authors apply a routine distance between all pairs of points using the following formula [37, P. 1562]:

$$d\,(P,Q) = min\,(d\,(p_i,q_j))\ \forall p_i \in P;\ \forall q_j \in Q, \qquad (1)$$

where: $\boldsymbol{d(p_i,q_j)}$ is a distance between any pair of points $\boldsymbol{p}_i$ and $\boldsymbol{q}_j$, which belongs to two different clusters $\boldsymbol{P}$ and $\boldsymbol{Q}$ correspondingly. The problematic issue of the *SaS*-approach like presented with (1) proceeds from a case of nonhomogeneous point cloud density that is a subject for clustering and segmenting. E.g., a sparse point cloud, which is also worsened by the faults of surveying technique, reasoned the case, according to which clustered points appear far from the boundary of a cluster. In such a case a distance value (from (1)) fails to be proved, if it is checked by a threshold parameter. Consequently, this may be a reason for the errored adjacency determination and for the consequent wrong model reconstruction. Nonetheless, it may be not so wrong to conclude, that because there are no Lidar points between two delineated planar segments, which would belong to other clusters, these two segments are sooner adjacent, than not and their seeming "non-connectivity" might be caused by the gaps and outliers in Lidar data only. This problem may be resolved by the adjacency determination with the Voronoi neighborhood.

The possible solution can be based on the following assumption. Even in a case, when two points lie far from each other, and there are no those points between them, which may belong to the third cluster (besides a given pair of clusters), and the Voronoi cells of these two points have common edges, then this pair of points can be determined by their Voronoi neighbors. In another case, when there are points of one more cluster between a given pair of points, these points are not considered as the adjacent ones even, if they pass a comparative test by applying a threshold parameter large enough.

On the base of all mentioned above and referring to an existing relevant example [106], we assume to consider as *adjacent ones* only those *clusters*, which points are *neighbors* within the Voronoi neighborhood. The only criterion is a case, when their Voronoi cells possess a *common edge*. Thus, we both solve a problem of data scarcity, gaps / outliers, and find a solution for a sparse point cloud. What is more, we get rid of a necessity of the enlarged threshold value introduction, but implement its definition by interpolating technique instead, since this value should be both big enough (not to remove the actual adjacent clusters), and small enough (not to define non-adjacent clusters as adjacent ones). In this way, we can significantly increase the applicability of our approach, and decrease its narrowness, e.g., when it strongly depends on point density and on the equitability or at least on the similarity of spatial distribution of the points that belong to two clusters. Moreover, the search of the Voronoi neighbors can be completed faster with such approach, because each point has a computed list of neighbors to be checked for their spatial acceptability. Thus, the overall algorithmic efficiency can be increased. It is only the first aspect of the overall LPM optimizing. The second one, that is not a less important solution, is introduced in the following paragraph.

While arranging the matrix adjacency in the frameworks of the *SaS*-technique, we define the distance between two clusters of planar points ***d* (*P*, *Q*)**, as the minimum of all point possible combinations between two clusters distance (1). Then, $m = |P|$ is a point number in the first cluster, and $n = |Q|$ - a point number in the second cluster. Hence, a whole number of possible combinations to be computed in such a case is the $m \cdot n$ value, the product of the point numbers.

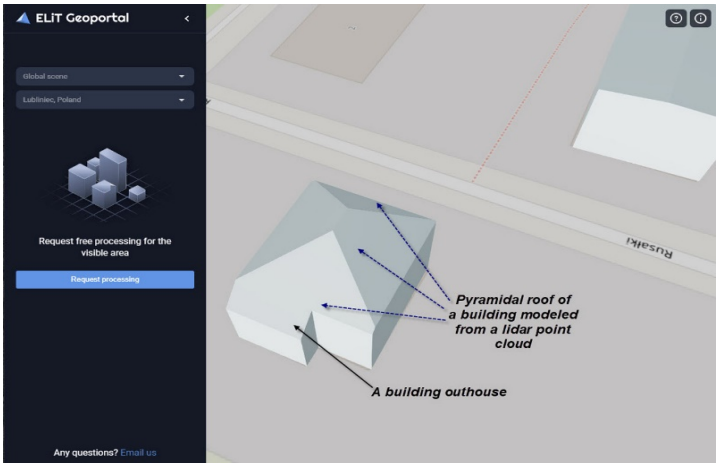


Figure 10: A lightweight model of a building with the pyramidal roof and an outhouse reconstructed at the *ElIT Geoportal* in the location of *the city of Lubliniec, Poland*

At this algorithmic step we suggest an alternative solution: to extend each of two clusters of planar points up to their mutual intersection, while a relevant linear segment of the cluster intersection is generated. Then we compute the distance between two clusters as the minimal distance of all possible measured combinations between points belonging to each of two clusters, from one side, and this linear segment of a cluster intersection, from another one. The statistical significance of difference between two mentioned values can be estimated. A combination number in this case becomes substantially fewer: the $m + n$ value only, the sum, but not the product. It defines the significant simplification of the overall LPM algorithmic complexity from the quadratic complexity $O(n^2)$ (a case of the *point product* that defines number of combinations) to the linear complexity $O(n)$ (a case of the *point sum*, that defines number of combinations). Decreased complexity provides the overall enhancement of the LPM-algorithmic efficiency.

It is evident, that the better a roof segment matrix is optimized, the more sustainable number of geometric primitives (e.g., vertices) is necessary for the robust building roof reconstruction. The introduced update of the adjacency matrix computation provides the definition of some threshold levels between two clusters of coplanar points within a footprint boundary, which is accepted as a geometric analogue of the roof edge. These clusters should never intersect, and this circumstance causes an indefinite exit of computing guaranteed, if the traditional approach is attempted to provide. Otherwise, the updated LPM technique, in its turn, completes the modeled topological sustainability and geometric correctness of a building footprint and of its roof as shown in Figure 10. It displays a model example from a fragment of one from modeled *CityGML LOD2* locations placed at our internet-resource – *ELiT (EOS LiDAR Tool) Geoportal* (http://elit-portal.eos.com/). The presented lightweight model has been selected for display, just because the relevant point cloud possesses quite a few non-intersecting planar segments, that is why the routine *SaS*-algorithm of reconstruction has failed while processing this data, but its update has succeeded – Figure10.

Phase 4.1 and partially *Phase 4.2* examined above deal with the point cloud segmentation stage in an overall processing workflow. Correspondingly, our prime LPM-concern has been a set of the roof segmentation operations, as it has been previously discussed for these two phases of the methodological diagram. *Phase 4.3* provides the finalized stage of building modeling – its reconstruction, while *Phase 4.2* grounds some substantial premises for it. We consider the urban feature reconstruction stage as presentation of building geometry and topology within a certain *LOD* with a defined number of feature segments. That is why we have attempted to implement in *Phase 4.3* one of the most interesting among other relevant theoretical technique – the polyhedral (polygonal) surface reconstruction (the *PolyFit* approach already mentioned in the literature review) [65, 107].

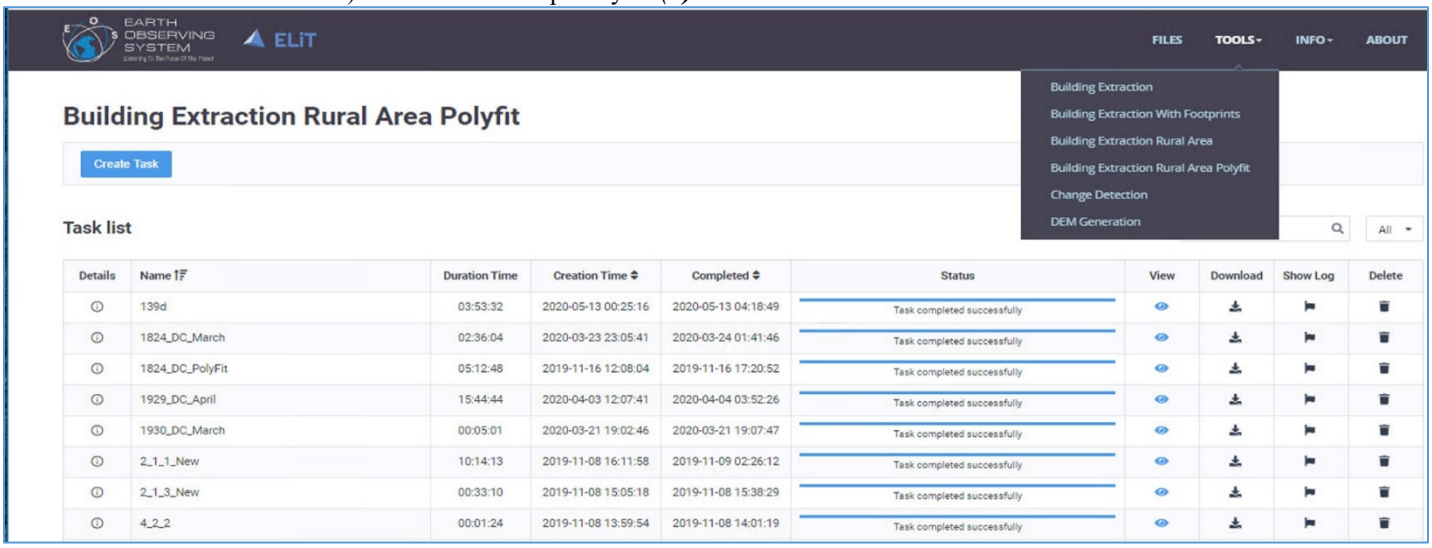


Figure 11: The *PolyFit* page selected among a list of other AFE-techniques in the menu *TOOLS* of the web-based software interface

This approach demonstrates, that reconstruction of geometric primitives from a point cloud can be treated as a binary labeling problem (what is, by the way, close enough to "global solutions" in the point cloud processing [87, 91]). The method proceeds from the hypothesizing and selecting strategy and provides intersecting the building planes already segmented, then seeking for a necessary combination of intersected planes and getting a manifold polygonal surface model. Since in its segmentation stage the method is strongly based exactly on the *RANSAC* algorithm for plane detection, then by enhancing *RANSAC* plane primitive extraction and refinement, what has been introduced in this text above, we understandably contribute to making the *PolyFit* solution more robust.

Therefore, constructing a whole pipeline for the *PolyFit* implementation in our web-based software, we attempted to employ on a relevant segmentation step both the original advances of the *SaS*-approach, and possibilities of enhanced *RANSAC*. Thus, the *PolyFit* appears to become one more AFE-option in the relevant menu option of our web-software package (Figure 11).

4. Results: Software Elaborated for AFE Purposes

The applied software solutions, that can be elaborated on the base of the presented multifunctional methodological approach to the generation of heavyweight and lightweight AFE-models (Figure 1) can be introduced as the robust results of this research. These software solutions are completed with the functionalities, which can provide the efficient support and realization of all three groups of advantages referred to above.

4.1. Desktop and Web-Based Applications

The ***ELiT*** (**E**OS **Li**dar **T**ool) Web-Based software is both a separate web-based (network) generator - *ELiT Server* (a landing demo-page available at https://eos.com/eos-lidar/), and an integrated component of EOS Platform-as-a-Service software - *ELiT Cloud.* The latter is the basis for *ELIT Geoportal* - a type of web portal used to find, access, and process LiDAR geospatial information, both primary, and resulted one, as well as provide the associated geographic services (display, editing, analysis, etc.) though the Web (http://elit-portal.eos.com/). Few examples of the *ELiT* Server interface have been already presented above (Figures 4, 6,7, 10, 11).

The congeneric to these two products is our desktop *ElitCore* package, which possesses even somewhat broader functionality, but can produce only the heavyweight polyhedral models (Figure 5). The *Building Extraction* functionality implemented in the standalone *ElitCore* provides detection, extraction, and reconstruction of heavyweight models according to our general methodological workflow of the HPM AFE (Figure 1) and its algorithmic content of *Phases 2* and *3* (Figures 2 and 3). It normally results in a DEM ("a bare ground class" delineated from an initial point cloud) and a set of urban features which are the discontinuous objects.

All three mentioned software products are based on the sophisticated HPM / LPM solutions based on those complicated algorithms for Lidar data processing that have been introduced in the previous section of this text. Both *ELiT Server*, which has been implemented as a distributed information system, and *ELiT Geoportal* usually operate from the Web browser installed on a user's workstation.

ELIT 3D city models as many other 3D models appear to be predominantly used for display; however, they are being increasingly employed in number of subject areas and for a large range of tasks beyond the visualization. In our previous publications, we sought to understand and present several use cases regarding their presentation not only on the *ELiT* Server, but also on the *ELiT* Geoportal with further utilization of these City GML LOD1 models for some tasks within and beyond model display [4, 102].

Our web-based software implies provision of the following *ELiT* tools as the functionalities of the distributed information system (these tools are shown in Figure 11 above):

- The *Building Extraction (BE)* tool (a sub-page *Building Extraction* of the *Tools* page – Figure 11) provides the HPM building detection, extraction, and reconstruction through various sub-procedures pursuing the final sub-procedure of building reconstruction for high-rise buildings, preferably in city downtowns. Thus, the *BE*-functionality accomplishes the generation of heavyweight models, which have been already grounded and presented above (Figures 2-4, 6). The building extraction functionality is that option, according to which any single building construction may be detected, extracted, and reconstructed as a heavyweight model with further visualization by the open source *Cesium 3DTiles* library with a certain *LOD* and with relevant spatial, geometric, and semantic attributes. Key components of the massive urban environment can be simulated as the heavyweight models with minor details (Figure 12);

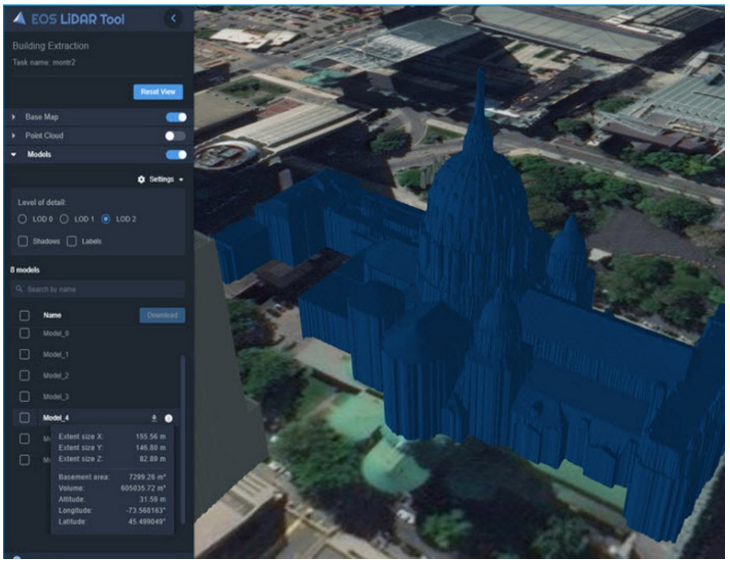


Figure 12: A historical building in *a downtown of Ottawa, Canada*, visualized in the *ELiT Server* interface

- The *Building Extraction with Footprints (BEwFP)* tool (a sub-page *Building Extraction with Footprints* of the *Tools* page – Figure 11) accomplishes the same HPM building extraction as the *BE*-tool producing the heavyweight building models. Oppositely to the previous tool, the *BEwFP*-functionality involves the third-party data – building footprints (global maps as *Open Street Maps*, *ESRI Imageries,*

Microsoft footprints, etc.) instead of accomplishing its own *ELiT* footprint delineation described in 3.2 Section of this text. Normally the footprints input is an archived ESRI *SHP* vector file;

- The *Building Extraction Rural Area* tool (*BERA* - a sub-page *Building Extraction Rural Area* located on the *Tools* page – Figure 11) provides "a classical pipeline", by which we implement the *SaS* point cloud segmentation / building reconstruction methods [37] of low polyhedral modeling introduced in *Phase 4.3* of the methodological flowchart (Figure 1). According to the *SaS*-pipeline and to its optimization described in subsections 3.3.2 and 3.3.4 the *BERA* tool accomplishes the hierarchical segmentation of a Lidar point cloud, refinement, and separation of extracted planes with further building reconstruction in rural areas and urban suburbs. As it has been already mentioned above (Subsection 2.2), the generated lightweight building models consist of not so many facets, and number of points involved in a single model generation is limited by a number from five to thirty thousand. Necessary limitation may be met by adaptive thinning at the cost of modeled details, if an initial value ranks over this quantity. The correspondence of a resulted model to an initial point cloud can be visually assessed for a conclusion, if it is modeled accurately or not, and if it is allowed to go to *Phase 5.LW* , or it is necessary to go back to *Phase 2*, and provide HPM instead of with LPM (Figure 1). This quality estimation is provided with the *ELiT Detailed Viewer* (Figure 13). In case of acceptable quality, the point locations should not deviate on more than a certain distance threshold value from the modeled surfaces. An exact threshold is introduced, basing on the point cloud characteristics, e.g., on its density.
- The *Building Extraction Rural Area Polyfit* tool (*BERA-P* - a sub-page *Building Extraction Rural Area Polyfit* located on the *Tools* page – Figure 7) implements somewhat another algorithmic pipeline in comparison with the *BERA* tool. The *RANSAC* optimized clustering and segmentation is used on the initial building modeling (subsection 3.3.3). The hypothesizing and sampling strategy of the *PolyFit* basics [65] together with its finalized reconstruction of polyhedral models is applied to concluding building modeling. Our optimization of the *RANSAC* segmentation described above is embedded in the relevant workflow, that fulfils the *BERA-P* tool. Its processing may take somewhat longer time, than the "ordinary" *BERA* tool does, but usually provides an increase of number of topologically correct and geometrically precise models among all output models in a dataset.
- The *Change Detection* tool (*CD* - a sub-page *Change Detection*, which is on the *Tools* page – Figure 11) automatically detects block-, district-, and city- scope alterations in an AOI selected. Changes in urban architectural morphology usually happen through city areas over some significant period, if only it is not any catastrophic event of either environmental, or social demolition. The *ELiT* CD tool detects locations of changes in positions and shapes of buildings and infrastructures as 3D models with additional spatial information. Usually two Point Clouds (the temporally primary, and the temporally secondary ones) are compared. It is not *BERA*, but the *BE*-functionality is the only one that is involved in preparing two resulted modeled sets. The difference between two input point clouds is computed as the *BE*-output of the high polyhedral modeling (Figure 6).

All those already listed *ELiT* functionalities are supported by the relevant web-API and by the Front-end web-interface for routine procedures of uploading, processing, and analyzing the initial data and for further downloading derivative data. The sets of modeled *.OBJ* and. *B3DM* entities (the latter are *Cesium 3D*

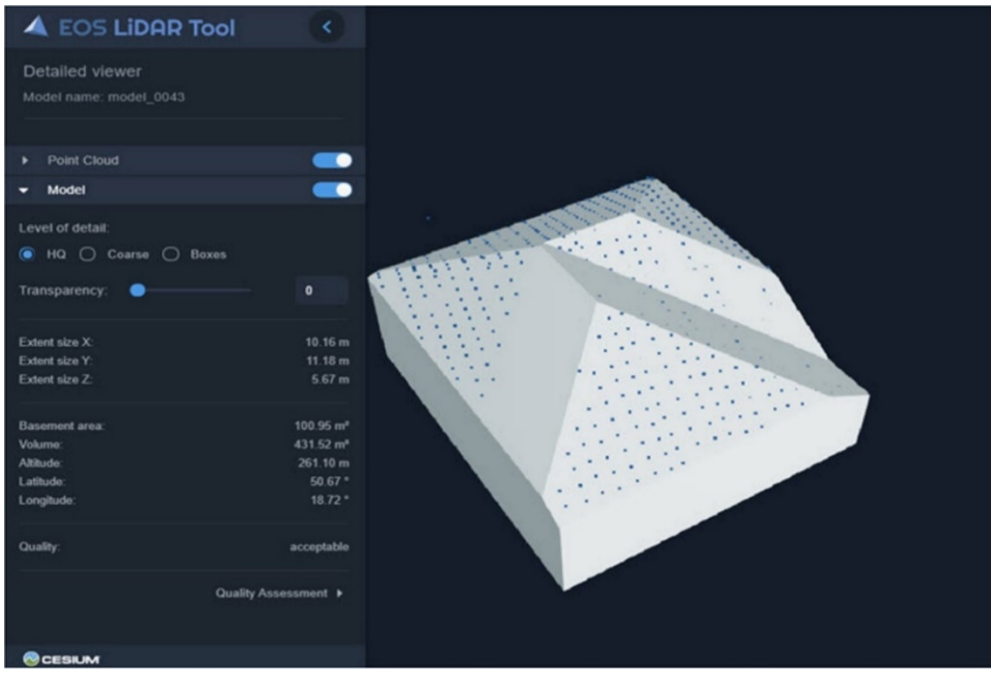


Figure 13: The *Detailed Viewer* Tool for examining accuracy of a lightweight model according to point cloud content

Tiles component) have to be delivered to 3D models, and can be stored and visualized in *.COLLADA, .KML,* and *glTF* formats.

According to the content of *3D Tiles* specification a *.glTF* is employed by *.B3DM* as *its payload* for delivering 3D building geometry. If data processed by *ELiT Server*, the results of all functionalities, *BE, BEwFP, BERA, BERA-P*, and *Change Detection,* are displayed by the *ELiT Viewer.*

- The *DEM Generation* tool (*DEM-G* - a sub-page of the *Tools* page Figure 11) runs a generation of a field of topographic heights making *a DEM / DSM grid* (Figure 17). It means we can create a gridded surface from sample data, what is commonly referred to as the interpolation. According to the contemporary basics of topographic modeling the most accurate uniformly spaced DEM matrix can be compiled directly either from photogrammetric processing of remote sensing imageries, or by processing results of LiDAR surveys [8]. In other words, a series of irregularly spaced height points are acquired, from which uniformly spaced elevation benchmarks are interpolated. A created DEM we understand as a synonym of a digital terrain model, which simulates the bare earth surface with uniformly spaced *z*-values of heights within a DEM neighborhood. Our automated processing for generation of a DEM of high precision provides modeling topography of various genetic types, e.g., sand-eolian (as shown in Figure 14), glacial, fluvial, volcanic, topography of human impact, etc. Results of the *DEM-G* functionality are visualized in *ELiT Viewer* by the *TerrainProvider* option of the open source *Cesium JavaScript* library (Figure 14).

4.2. Cloud-Based Application – ELiT Geoportal

A common definition of the *EliT Geoportal* (*EGP*) as a *cloud-based* application has been already done in the previous subsection of this text. We already explained in one of our previous publication the necessity of a geoportal as a robust solution for the web-GIS software promotion [102]. *EGP* is a kind of web portal used to transform occasional visitors of the EOS LiDAR web-site into the warm leads. The *EGP* site can be used to find and access the derivative Lidar information and relevant services for number of thematic use cases, that can be resolved through geospatial modeling, visualization, and GIS-analysis.

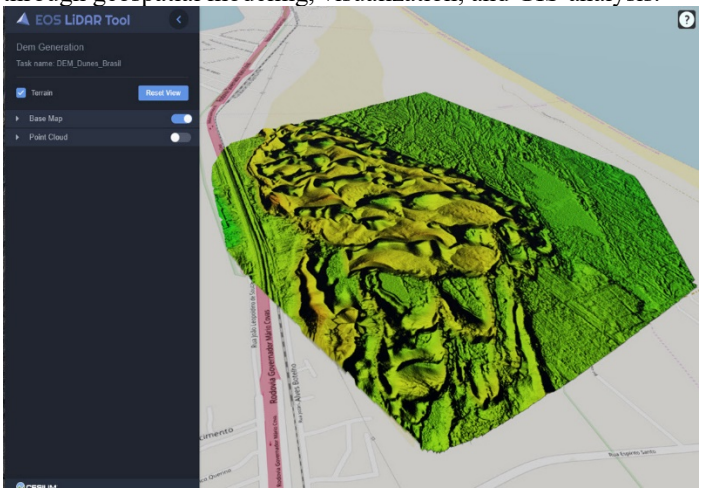


Figure 14: Topography of sand dunes in Brazil visualized in the *ELiT Viewer* interface (a dataset from the *EOS LIDAR Tool* landing page: https://eos.com/eos-lidar)

Thus, the Geoportal can be also defined as the Internet-toolbox for resolving local, regional, national, and even globalized urban problems, i.e., by performing at very different scales. A routine Web-based map-server software usually allows only backups of mapping results. Nonetheless, for various domains of Urban Studies, which can be provided with our research and modeling technique [1], we do need a Geoportal Internet-toolbox associated with a web-based Geodatabase, that would allow us to sort, query, update, and delete both different features and data associated with these features. Such architectural and structural web-software scheme differs towards its more efficiency if compared with solutions of a common case. Within these solutions some web-based GIS application generator is developed, and it provides on the base of the routine database scheme the number of web scripts. Then it deals with a map interface and with a database, as well as with provision of interactive maps defined by a user [19, 108, 109]. Otherwise, the EGP may support an effective integration of a Web-based GIS interface with a Geodatabase chosen within a web-architecture. Spatial references and 3D regional layers are automatically chosen for a given territory. What is more, semi-automatic generation of Web 3D Scenes according to a selected AOI can be provided. Owing to this architectural scheme, series of thematic EGP use cases are provided (http://elit-portal.eos.com/): *Population estimation in an AOI based on 3D building models; Energy consumption estimation based on 3D building models; Visibility analysis in urban environment for various city development scenarios' evaluation and comparison.*

5. Discussion

Choosing either the high polyhedral modeling, or the low polyhedral one with corresponding generation of either HW-, or LW-models may be selected as introductory topic of the discussion, which summarizes our research.

As explained above, in a common case, *Phase 1* of the methodological flowchart (Figure 1) suggests either one, or another option according to environmental and urban configurations of a selected AOI.

The LW-models are built exclusively on the base of ALS surveys together with input of the third-party footprints (Figure 7). The lightweight model accuracy exclusively depends on the segmentation quality and the reconstruction robustness. What is more, exactly clustering / segmenting procedures significantly define consequent LM-results. Application of the originally optimized *RANSAC* (Figure 9) has resulted not only in removal of

a point cloud outlier caused by a tree (Figure 8), but also in predetermined appearance of a building outhouse on the reconstruction stage in a finalized model, what cannot be guaranteed with the routine approach of segmentation (Figure 10).

The comparative evaluation of optimized clustering/segmenting technique with both restructuring *SaS*-pipeline (subsection 3.3.2), and with enhancing *RANSAC* (subsection 3.3.3) has been provided on the base of several ALS data surveys accomplished over the territory of the city of Lubliniec, Poland, in 2017-2018 (Figures 7, 8, 9). Table 1 is produced to compare the modeled results of the methods mentioned. Three different values of the key *RANSAC* parameter from Figure 9, *plane capture distance (PCD)*, are used in *meters*, evaluating the position uncertainty for each Lidar point. Consequent iterations are completed for *PCD=0.1*, *0.2*, and *0.3*. The model quality is evaluated both by a visual comparison with a processed point cloud (as in Figure 13), and by customized estimating of algorithmic parameters. Modeling warnings designate partially correct models that have some flaws.

Table 1: Comparison of the modeled results depending on clustering and segmenting method (a LiDAR point dataset on Lubliniec, Poland)

#	Method of clustering / segmentation	Pro-cessing time	Total number of models	Correct models	In-correct models	Modeling warnings
1	Original *SaS*	8 h 45 min	934	534	289	111
2	Original *RANSAC*	12 h 11 min	934	577	248	109
3	*SaS* with restructured pipeline	11 h 24 min	934	588	243	103
5	Optimized *RANSAC* with *PCD=**0.1***	15 h 57 min	934	662	167	105
6	Optimized *RANSAC* with *PCD=**0.2***	10 h 42 min	934	677	147	111
6	Optimized *RANSAC* with *PCD=**0.3***	7 h 42 min	934	699	133	102

Shown in Table 1 are the comparative statistics, which has stipulated the necessity development of the *BERA PolyFit* building modeling tool described above. *The total number of models* column has the same value for all records, since it is a number of building footprints obtained from a data provider for this AOI. Another almost constant column for all methods is number of partially correct models (*Modeling warnings*). The table shows, that for all other records the evidently best results are demonstrated by the optimized *RANSAC* with *PCD=0.3* (Method *# 6*). The accuracy of the segmented roof planes follows from a selected clustering-segmenting method thus it does predefine either a success, or not on the stage of reconstruction. Therefore, this technique has been selected for the *BERA-P* tool. Briefly the most efficient *RANSAC* method according to Table 1 can be illustrated in the following way, if examining only one model (*# 43*) from a dataset, what from Table (1) (Figure 15).

Two different iterations of Method # 6, which present resembling results (the same number of up to seven face candidates), are illustrated, because it is the only method, that produces an acceptable number of similar results through various iterations. Lidar point as color-coded segments have ratios for each coded class. Ratio numerator is the number of break-line (non-planar points that do not belong to this segment due to the *PCD* value) points, while ratio denominator – planar points (they satisfy the *PCD* threshold). Methods *##1-6* (Table 1) have been compared with respect to similarity of various iterations like in Figure 15. Method *# 6* meets criterion of the best efficiency in this respect. The following illustration shows that this technique is the only tool (among all other from Table 1), that can provide the clustering-segmenting premises for reconstruction of a problematic roof slope, that belongs to model # 21 in the considered Lubliniec dataset (Figure 16). All other five techniques missed this roof slope.

The finalized illustration of this text demonstrates an increase for up to 27% (*41 acceptable among 62 all models* versus *30 / 62 ratio*) in a number of acceptable models for a small urban area in Washington, D.C. by providing Method *# 6* as a software solution (Figure 17).

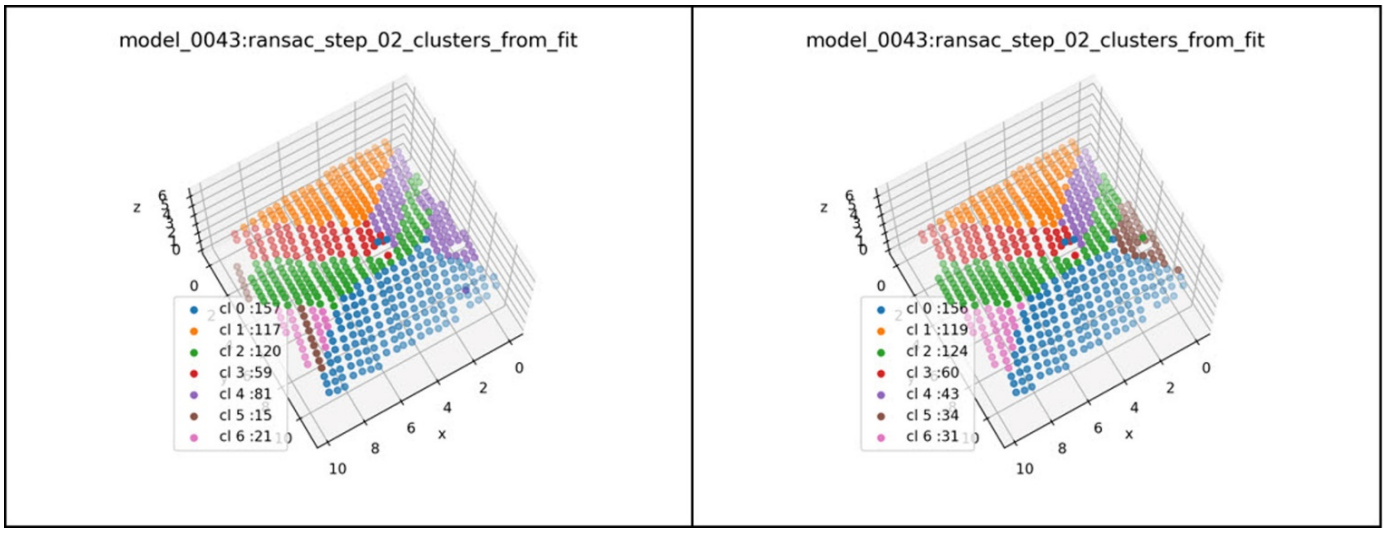


Figure 15: Two different clustering / segmenting iterations of Method # 6 (Table 1) to a point cloud of model # 43, that demonstrate resembling results

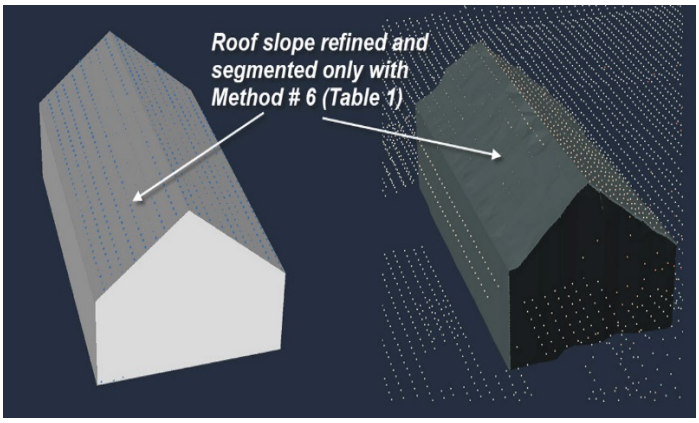


Figure 16: The roof slope pointed by arrows segmented by Method # *6* for further reconstruction. A CAD model built from a Google imagery (to the right) is compared with a point cloud with quite weak correlation for this building segment, while *ELiT* LW-model reconstructed is to the left

6. Conclusion

Our paper focuses on the conceptual multifunctional (functionalities of *Building Extraction*, *Building Extraction* in *Rural Areas*, *Change Detection*, and *DEM Generation*) approach within the AFE frameworks, that has been introduced through consequent steps of the phased methodological flowchart. In general, a workflow presented that has been completely automated for building detection, extraction and modeling by processing ALS and MLS point clouds. Two approaches of the high polyhedral modeling, and the low polyhedral one, have been proposed as two phased branches of a united overall workflow. These two modeling mainstreams have been compared on the base of their resulted models: the heavyweight models generated together with their original footprints within the HPM frameworks, and the lightweight models that employ the third-party footprints – with the LPM technique. With modeled results illustrated we have attempted to prove that the HPM branch can be appropriate for high-rise building reconstruction with either flat, or pitched roofs in densely build-up urban areas. The LPM branch – for modeling low-rise buildings with gable, or pitched roofs in either rural areas, or in city suburbs, as well as in small towns, i.e., in areas with thinned build-up.

A plot of our overall AFE workflow has been presented as the content of the methodological flowchart phases, while its *high polyhedral modeling* branch has been described by the relevant core algorithmic pipeline separately for phases of ALS and MLS data processing. The ALS block of footprint extraction and delineation has been presented in details. Visual samples of HPM-results processed by the *Building Extraction* and the *Change Detection* desktop and web-software tools have been provided.

The *low polyhedral modeling* has been explained as the technique consisting of two consequent stages: the point cloud clustering / segmenting stage, and the stage of building reconstruction. Our methodological flowchart has presented detailed phases of the LPM implementation. The *SaS*, *RANSAC*, and *PloyFit* approaches have been adopted, and selectively enhanced for implementation in the web-software as *Building Extraction Rural Areas* and *BERA-PolyFit* tools:

SaS-segmentation has been enhanced by restructuring its pipeline; *RANSAC*-segmentation has been optimized by both the introduction of the key BCCPI parameter, and by the adaptive alterations of rules due to the plane candidate "penalty charges", as well as by introducing the point sampling from a randomly chosen limited neighborhood; the optimized *RANSAC*-segmentation method with a certain parameter has been selected as the most efficient one among six compared methods on the base of results obtained for our key dataset;

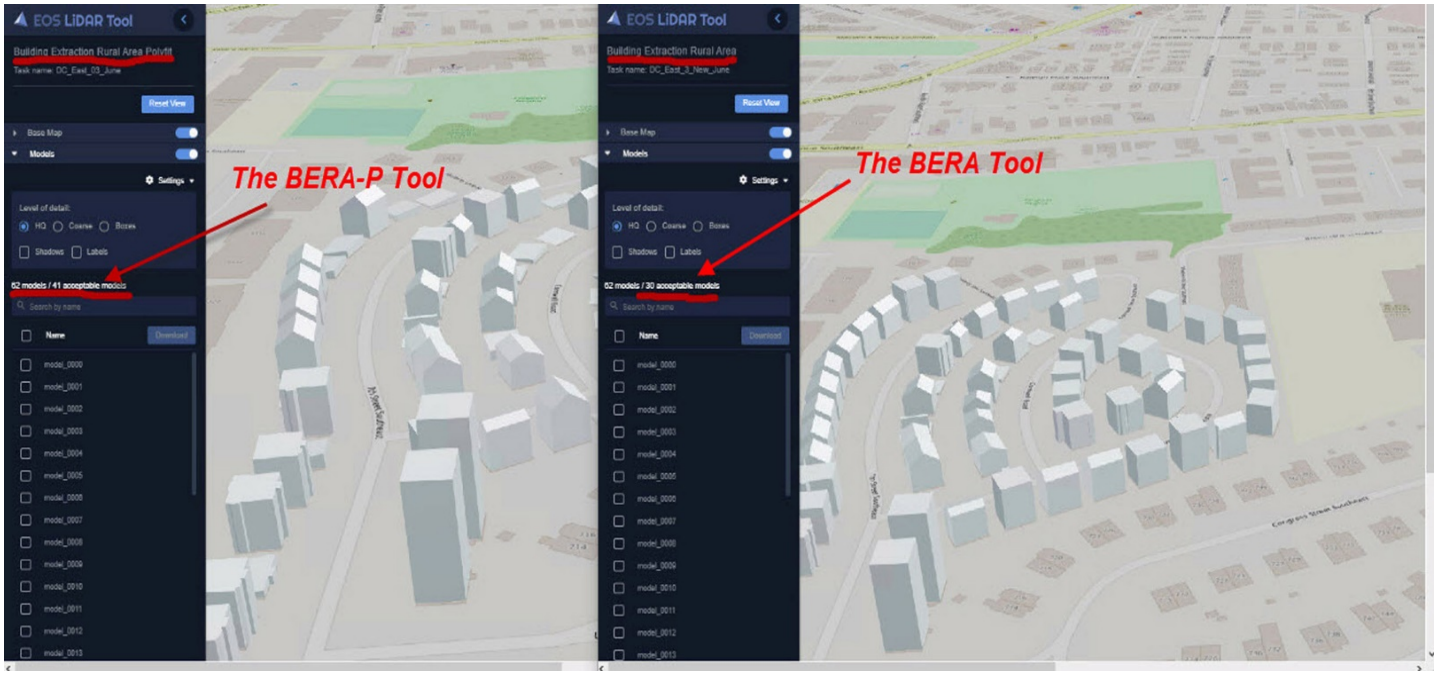


Figure 17: Comparison of modeled results, the number of acceptable models, produced by the *BERA Polyfit* Tool (with Method # *6* in clustering / segmenting algorithmic core - the left visual), on one hand, and by the *BERA* Tool (the right visual), on the other hand. *A small urban parcel in Washington, D.C., USA*

SaS-reconstruction has been optimized by the Voronoi neighborhood introduction, and thus obtaining the enhancement of overall algorithmic efficiency and the better optimization of the roof plane adjacency matrix. The Voronoi diagram can be accepted as a natural neighborhood without the necessity to specify its size. This property can be crucial for employing Voronoi neighbors within the reconstruction stage, while in the original methodology its usage is strictly limited to the stage of segmentation;

PolyFit approach has been implemented in the *BERA-P* web-software tool and proved to be as the most efficient one on the reconstruction stage if based on the optimized *RANSAC*. This issue has finalized our contributions to the low polyhedral modeling.

Besides all mentioned above functionalities the graphic user interface of the original family of software products for LiDAR data processing has been introduced as desktop, web-based, and cloud-based solutions: *ElitCore*, *ELiT Sever*, and *ELiT Geoportal*, correspondingly.

As far as future works are concerned the suggested algorithmic optimization and enhancements should be exercised through much more urban configurations, e.g., on those locations already placed on the *ELiT Geoportal.* Routine GIS-data consumers may employ the *ELiT Server* to search, access, and process their own Lidar data by the multifunctional tools in both block, and district scopes, and the data delivered to them by a provider. In such case, the *ELiT Geoportal*, as an Internet-toolbox, may serve an increasingly significant role in sharing derivative LiDAR information on the structure and spatiotemporal development of large human settlements, i.e., provide solution in a whole city scope.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research and development have been sponsored by the Noosphere Ventures that is an international asset management Company.

References

[1] S. Kostrikov, R. Pudlo, D. Bubnov, A. Kostrikova, "Studying of urban features by the multifunctional approach to LiDAR data processing," in Joint Urban Remote Sensing Event JURSE 2019: New Methodologies for urban investigation through remote sensing. UBS, France, 2019. https://ieeexplore.ieee.org/document/8809063. IEEE Xplore Digital Library, 2019. http://dx.doi.org/10.1109/JURSE.2019.8809063

[2] M. Batty, B. Hutchinson, Systems Analysis in Urban Policy Making and Planning, Plenum Press, 1982.

[3] S.V. Kostrikov, L.M. Niemets, K.Y.Sehida, K.A. Sehida, C. Morar, "Geoinformation approach to the urban geographic system research (cases studies of Kharkiv region," Visnyk of V. N. Karazin Kharkiv National University, series "Geology. Geography. Ecology", **49**(1), 107-124, 2018. https://doi.org/10.26565/2410-7360-2018-49-09

[4] S.V. Kostrikov, "Urban remote sensing with LiDAR for the Smart City concept implementation," Visnyk of V. N. Karazin Kharkiv National University, series "Geology. Geography. Ecology", **50** (2), 101-124, 2019. https://doi.org/10.26565/2410-7360-2019-50-08

[5] T. Esch, W. Heldens, A. Hirner, "The Global Urban Footprint," in Urban Remote Sensing. CRC Press, 34-43, 2018.

[6] M. Batty, D. Chapman, S. Evans, M. Haklay, S. Kueppers, N. Shiode, A. Smith, P.M. Torrens, "Visualizing the city: communicating urban design to planners and decision-makers," Center for Advanced Spatial Analysis, Working Paper Series, 2000. [Online]. Available: http://www.casa.ucl.ac.uk/visualcities.pdf

[7] C.J. Bellman, M.R. Shortis, "A classification approach to finding buildings in large scale aerial photographs," ASPRS Archives, **XXXV**(B3), 337–342, 2004.

[8] D.F. Maune (Editor), Digital Elevation Model Technologies and Applications: The DEM Users Manual, 2nd Edition, ASPRS, 2007.

[9] F. Leberl, "Human Habitat Data in 3D for the Internet," in Computer Vision, Imaging and Computer Graphics. Theory and Applications. VISIGRAPP 2009. Communications in Computer and Information Science, **68**, Springer, Berlin, 2010. DOI https://doi.org/10.1007/978-3-642-11840-11

[10] T. Esch, M. Marconcini, D. Marmanis, J. Zeidler, S. Elsayed, A Metz, A. Müller, S. Dech, "Dimensioning urbanization—An advanced procedure for characterizing human settlement properties and patterns using spatial network analysis," Appl. Geogr., **55**, 212–228, 2014. http://dx.doi.org/10.1016/j.apgeog.2014.09.009

[11] D. Potere, A. Schneider, S. Angel, D.L Civco, "Mapping urban areas on a global scale: Which of the eight maps now available is more accurate?," Int. J. Remote Sens. **30**(24), 6531–6558, 2009. http://dx.doi.org/10.1080/0143116090312113 4

[12] J. Congote, A. Moreno, L. Kabongo, J.L. Pérez, R. San José, O. Ruiz, "Web based hybrid volumetric visualization of urban GIS data," in Usage, Usability, and Utility of 3D City Models - European COST Action TU0801, EDP Sciences, Nantes, France, 2012. http://dx.doi.org/10.1051/3u3d/201203001

[13] J. Richards, Remote Sensing Digital Image Analysis, Springer, Heidelberg, Germany, 2013.

[14] S. Yasumoto, A. Jones, K. Yano, T. Nakaya, "Virtual city models for assessing environmental equity of access to sunlight: a case study of Kyoto, Japan," Int. J. Geogr. Inf. Sci., **26**, 1-13, 2012. http://dx.doi.org/10.1080/13658816.2011.570268

[15] I. Delikostidis, J. Engel, B. Retsios, C.P.J.M. van Elzakker, M.J. Kraak, J. Döllner, "Increasing the usability of pedestrian navigation interfaces by means of Landmark visibility analysis," J. Navig. **66**, 523–537, 2013. http://dx.doi.org/10.1017/S0373463313000209

[16] W. Förstner, "3D City Models: Automatic and Semiautomatic Acquisition Methods," in Proceedings Photogrammetric Week, University of Stuttgart, 291-303, 1999.

[17] J. Döllner, T.H. Kolbe, F. Liecke, T. Sgouros, K. Teichmann, "The virtual 3D city model of Berlin – managing, integrating and communicating complex urban environment," in Proceedings of the 25th International Symposium on Urban Data Management, 15-17 May 2006, UDMS, Alborg, Denmark, 2006.

[18] W. Ribarsky, T. Wasilewski, N. Faust, "From Urban Terrain Models to Visible Cities," IEEE Comput. Graph. Appl., **22**(4), 10-15, 2002. http://dx.doi.org/10.1109/MCG.2002.1016692

[19] A. Stadler, C. Nagel, G. König, T.H. Kolbe, "Making interoperability persistent: A 3D geodatabase based on CityGML," in Proceedings of the 3rd International Workshop on 3D Geo-Information, Seoul, Korea. Lecture Notes in Geoinformation & Cartography, Springer Verlag, 2009. http://dx.doi.org/10.1007/978-3-540-87395-2_11

[20] D. Fisher-Gewirtzman, "3D models as a platform for urban analysis and studies on human perception of space," in Usage, Usability, and Utility of 3D City Models - - European COST Action TU0801, EDP Sciences: Nantes, France, 2012. [Online]. Available: http://3u3d.edpsciences.org. https://doi.org/10.1051/3u3d/201201001

[21] K. Sanjaya, F. Henning, K.R. Purba, "3D LIDAR City Model Application and Marketing Plan Development," in 2017 International Conference on Soft Computing, Intelligent Systems and Information Technology (ICSIIT), 238-242, 2012. https://doi.org/10.1109/ICSIIT.2017.50

[22] T.H. Kolbe, "Representing and Exchanging 3D City Models with CityGML," in Proceedings of the 3rd International Workshop on 3D Geo-Information, Seoul, Korea. Lecture Notes in Geoinformation & Cartography, Springer Verlag, 2009. http://dx.doi.org/10.1007/978-3-540-87395-2_2

[23] S. Zlatanova, J. Beetz, "3D spatial information infrastructure: The case of Port Rotterdam," In Usage, Usability, and Utility of 3D City Models - European COST Action TU0801, EDP Sciences, Nantes, France, 2012. http://dx.doi.org/10.1051/3u3d/201203010

[24] F. Biljecki, J. Stoter, H. Ledoux, S. Zlatanova, and A. Çöltekin, "Applications of 3D city models: state of the art review," ISPRS Int. J. of Geo-Inf., **4**, 2842–2889, 2015. http://dx.doi.org/10.3390/ijgi4042842

[25] E. Baltsavias, "A comparison between photogrammetry and laser scanning,", ISPRS J. Photogramm. Remote Sens., **54**(2-3), 83-94, 1999. http://dx.doi.org/10.1016/S0924-2716(99)00014-3

[26] P. Dong, Q. Chen (Editors), LiDAR Remote Sensing and Applications, Boca Raton, FL, CRC Press, 2018. http://dx.doi.org/10.4324/9781351233354

[27] E.P. Baltsavias, "Airborne laser scanning: Existing systems and firms and other resources," ISPRS J. Photogramm. **54**, 164–198, 1999. http://dx.doi.org/10.1016/S0924-2716(99)00016-7

[28] Y. Lin, J. Hyyppa, A. Jaakkola, "Mini-UAV-LS-Borne LIDAR for Fine-Scale Mapping," IEEE Geosci. Remote Sens. **8**, 426–430, 2011. http://dx.doi.org/10.1109/LGRS.2010.2079913

[29] B. S. Yang, Z. Wei, Q. Q. Li, J. Li, "Automated extraction of street-scene objects from mobile LiDAR point clouds," Int. J. Remote Sens., **33**(18), 5839-5861, 2012. http://dx.doi.org/10.1080/01431161.2012.674229

[30] Y. Wang, L. Cheng, Y. Chen, Y. Wu, M. Li, "Building Point Detection from Vehicle-Borne LiDAR Data Based on Voxel Group and Horizontal Hollow Analysis," Remote Sens., **8**, 419, 2016. http://dx.doi.org/10.3390/rs8050419

[31] R. Wang, "3D building modeling using images and lidar: a review," Int. J. of Image and Data Fusion, **4**(4), 273–292, 2013. http://dx.doi.org/10.1080/19479832.2013.811124

[32] M. Awrangjeb, C. Zhang, C. Fraser, "Automatic extraction of building roofs using LIDAR data and multispectral imagery," ISPRS J. Photogramm. and Remote Sens., **83**, 1-18, 2013. http://dx.doi.org/10.1016/j.isprsjprs.2013.05.006

[33] B. Yaung, Z. Wei, Q. Li, J. Li, "Semiautomated building façade footprint extraction from mobile LiDAR point clouds," IEEE Geoscince and Remote Sens. L., **10**(4), 766-770, 2013. https://doi.org/10.1109/LGRS.2012.2222342

[34] M. Awrangjeb, C. Fraser, "Automatic segmentation of raw LIDAR data for extraction of building roofs," Int. J. Remote Sens., **6**, 3716–3751, 2014. http://dx.doi.org/10.3390/rs6053716

[35] J. Shan, A. Sampath, "Urban DEM generation from raw LiDAR data: a labeling algorithm and its performance," J. Photogramm. Eng. Remote Sens., **75**, 427-442, 2009. https://doi.org/10.14358/PERS.71.2.217

[36] N. Haala, M. Kada, "An update on automatic 3D building reconstruction," ISPRS J. of Photogram. Remote Sens., **65**, 570-580, 2010. http://dx.doi.org/10.1016/j.isprsjprs.2010.09.006

[37] A. Sampath, J. Shan, "Segmentation and reconstruction of polyhedral building roofs from aerial LIDAR point clouds," IEEE T, Geosci. Remote Sensing, **3**, 1554–1567, 2010. http://dx.doi.org/10.1109/TGRS.2009.2030180

[38] A. Habib, R. Zhai, C. Kim, "Generation of complex polyhedral building models by integrating stereo-aerial imagery and LiDAR data," J. Photogram. Eng. Remote Sens. 75, 609-623, 2010. http://dx.doi.org/10.1016/j.isprsjprs.2010.09.006

[39] P. Kumari, W.E. Carter, R.L. Shrestha, "Adjustment of systematic errors in ALS data through surface matching," Adv. Space Res., **47**, 1851–1864, 2011. http://dx.doi.org/10.1016/j.asr.2010.12.015

[40] M.Y. Belkhouche, B. Buckles, "Iterative TIN-based automatic filtering of sparse LiDAR data," IEEE Geosci. Remote. Sens. Lett., **2**(3), 231-240, 2012. http://dx.doi.org/10.1080/01431161.2010.515266

[41] D. Chen, L. Zhang, R. Liu, "Urban building roof segmentation from airborne LiDAR point clouds," Int. J. Remote Sens., **33**, 6497–6515, 2012. http://dx.doi.org/10.1080/01431161.2012.690083

[42] S. Sun, C. Salvaggio, "Complex Building Roof Detection and Strict Description From LIDAR Data and Orthorectified Aerial Imagery," in 2012 IEEE Geoscience and Remote Sensing Symposium, 5466- 5469, 2012. http://dx.doi.org/10.1109/IGARSS.2012.6352369

[43] S. Kostrikov, R. Pudlo, A. Kostrikova, "Three key EOS LiDAR Tool functionalities for Urban Studies," in 39th Asian Conference on Remote Sensing (ACRS 2018): Remote Sensing Enabling Prosperity, Proceedings of a meeting held 15-19 October 2018, Kuala Lumpur, Malaysia, AARS – Curran Associates, Inc., **3**, 1676-1685, 2019.

[44] J.R. Jensen, "Remote Sensing of the Environment: An Earth Resource Perspective (2nd edition)," Prentice Hall, Upper Saddle River, NJ, 2006.

[45] A. Gruen, "TOBAGO—A semi-automated approach for the generation of 3-D building models," ISPRS J. Photogramm. Remote Sens., **53**, 108–118, 1998. http://dx.doi.org/10.1016/S0924-2716(97)00034-8

[46] H.-G. Maas, G. Vosselman, "Two algorithms for extracting building models from raw laser altimetry data," ISPRS J. Photogramm. Remote Sens., **54**, 153–163, 1999. http://dx.doi.org/10.1016/S0924-2716(99)00004-0

[47] N. Haala, C. Brenner, "Extraction of buildings and trees in urban environments" ISPRS J. Photogramm. Remote Sens., **54**, 130–137, 1999. http://dx.doi.org/10.1016/S0924-2716(99)00010-6

[48] G. Priestnall, J. Jaafar, A. Duncan, "Extracting urban features from LiDAR digital surface models", **24**(2), 65-78, 2000. https://doi.org/10.1016/S0198-9715(99)00047-2

[49] U. Stilla, U. Soerge, U. Thoennessen, "Potential and limits of InSAR data for building reconstruction in built-up areas," ISPRS J. Photogramm. Remote Sens. **58**, 113–123, 2003. http://dx.doi.org/10.1016/S0924-2716(03)00021-2

[50] Y. Hu, "Automated Extraction of DTM, roads, and buildings using Airborne lidar," PhD Thesis, University of Calgary, 2007.

[51] C. Alexander, S. Smith-Voysey, C. Jarvis, K. Tansey, "Integrating building footprints and LiDAR elevation data to classify roof structures and visualize buildings," Comput. Environ. Urban Syst., **33**, 285-292, 2009. http://dx.doi.org/10.1016/j.compenvurbsys.2009.01.009

[52] A.K. Aijazi, P. Checchin, L. Trassoudaine, "Automatic detection and feature estimation of windows in 3D urban point clouds exploiting façade symmetry and temporal correspondences," Int. J. Remote Sens., **35**, 7726–7748, 2014. http://dx.doi.org/10.1080/01431161.2014.975420

[53] E. Orthuber, E., J. Avbelj, "3D building reconstruction from Lidar point clouds by adaptive dual contouring," ISPRS Ann. Photogram. Remote Sens., II-3/W4, PIA15+HRIGI15, Joint ISPRS conference 2015, 25–27 March, Munich, Germany, 2015.

[54] T. Freeland, B. Heung, D.V. Burley, G. Clark, A. Knudby, "Automated feature extraction for prospection and analysis of monumental earthworks from aerial LiDAR in the Kingdom of Tonga," J. Archaeol. Sci., **69**, 64-74, 2016. https://doi.org/10.1016/j.jas.2016.04.011

[55] C. Brenner, "Towards fully automatic generation of city models," Int. Arch. Photogram., Remote Sens., **33**, 1-8, 2000.

[56] M. Romano, "Innovation in LiDAR processing technology," Photogramm. Eng. Remote Sens., **70**, 1202-1212, 2004.

[57] G. Sithole, G. Vosselman, "Filtering of airborne laser scanner data based on segmented point clouds," Int. Arch. Photogram. Remote Sens. , **34,** 66-71, 2005.

[58] M. Kada, L. McKinley, "3D building reconstruction from LiDAR based on a cell decomposition approach," Int. Arch. Photogramm., Remote Sens. Sensing, **38**, 47-52, 2009.

[59] J.X. Zhang, X.G. Lin, X.G. Ning, "SVM-based classification of segmented airborne LIDAR point clouds in urban areas," Remote Sens., **5**, 3749–377. 2013. http://dx.doi.org/10.3390/rs5083749

[60] C. Flener, M. Vaaja, A. Jaakkola, A. Krooks, H. Kaartinen, A. Kukko, E. Kasvi, H. Hyyppä, J. Hyyppä, P. Alho, "Seamless Mapping of River Channels at High Resolution Using Mobile LiDAR and UAV-LS-Photography," Int. J. Remote Sens., **5**, 6382-6407, 2013. http://dx.doi.org/10.3390/rs5126382

[61] C. Liu, B. Shi, X. Yang, N. Li, H. Wu, "Automatic buildings extraction from LiDAR data in urban area by neural oscillator network of visual cortex," IEEE J-STARS, **6**, 2008–2019, 2013. http://dx.doi.org/10.1109/JSTARS.2012.2234726

[62] M.A. Fischler, R.C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Communications of the ACM, **24**(6), 381–395, 1981. http://dx.doi.org/10.1145/358669.358692

[63] R. Schnabel, R. Wahl, R. Klein, "Efficient RANSAC for point-cloud shape detection," in Computer graphics forum, volume 26, 214–226, 2007. http://dx.doi.org/10.1111/j.1467-8659.2007.01016.x

[64] B. Li, R. Schnabel, J. Shiyao, J., R. Klein, "Variational surface approximation and model selection," Computer Graphics Forum (Proc. Pacific Graphics) 28, 7, 1985—1994, 2009. https://doi.org/10.1111/j.1467-8659.2009.01577.x

[65] L. Nan, P. Wonka, "PolyFit: Polygonal surface reconstruction from point clouds," in 2017 IEEE International Conference on Computer Vision (ICCV), 2017. http://dx.doi.org/10.1109/ICCV.2017.258

[66] L. Xu, E. Oja, Kultanen, "A new Curve Detection Method: Randomized Hough Transform (RHT)," Pattern Recognit. Lett., **11**, 331–338, 1990. https://doi.org/10.1016/0167-8655(90)90042-Z

[67] D. Borrmann, J. Elseberg, K. Lingemann, A. Niichter, "The 3D hough transform for plane detection in point clouds: a review and a new accumulator design," 3D Res., **2**, 1-13, 2011. http://dx.doi.org/10.1007/3DRes.02(2011)3

[68] E. Maltezos, C. Ioannids, "Automatic extraction of building roofs from Airborne LiDAR data applying and extended 3D randomized Hough transform," ISPRS Ann. Photogramm. Remote Sens., **III-3**, 209-216, 2016.http://dx.doi.org/10.5194/isprsannals-III-3-209-2016

[69] J. Susaki, "Knowledge-based modeling of buildings in dense urban areas by combining airborne LiDAR data and aerial images," Remote Sens., **5**, 5944–5963, 2013. http://dx.doi.org/10.3390/rs5115944

[70] L. Zhang, T. Xu, J. Zhang, "Building extraction based on multiscale segmentation," in 5th International Congress on Image and Signal Processing (CISP), October 16–18, 2012. http://dx.doi.org/10.1109/CISP.2012.6470035

[71] J. Cheng, Y. Deng, Q. Du, "Mapping between BIM models and 3D City models of different levels of detail," in Proceedings of the 13th International Conference on Construction Applications of Virtual Reality, 30-31 October 2013, London, UK, 2013.

[72] N.H. Arachchige, S.N. Perera, H.G. Maas, "Automatic processing of mobile laser scanner point clouds for building façade detection," ISPRS International Archive of Photogrammetry and Remote Sensing in Spatial Information Science, **XXXIX-B5**, 187–192, 2012. http://dx.doi.org/10.5194/isprsarchives-XXXIX-B5-187-2012

[73] R. Brinkman, C. O'Neill, "LiDAR and photogrammetric mapping," The Military Engineer, **5**, 56–57, 2000.
[74] M. Hodgson, P. Bresnahan, "Accuracy of airborne lidar-derived elevation: Empirical assessment and error budget," J. Photogram. Eng. Remote Sens., **70**, 331–339, 2004. http://dx.doi.org/10.14358%2FPERS.70.3.331
[75] A. Sampath, J. Shan, "Building boundary tracing and regularization from airborne LiDAR point clouds," J. Photogramm. Eng. Remote Sens., **73**(7), 2007, 805–812. http://dx.doi.org/10.14358/PERS.73.7.805
[76] J. Shan, A. Sampath, "Building extraction from LiDAR point clouds based on clustering techniques," in Topographic Laser Ranging and Scanning: Principles and Processing, Boca Raton, FL: CRC Press, 423–446, 2008. http://dx.doi.org/10.1201/9781420051438.ch15
[77] A. Sampath and J. Shan, "Building roof segmentation and reconstruction from lidar point clouds using clustering techniques," Int. Arch. Photogramm., Remote Sens., **XXXVII**(B3a), 279-284, 2008.
[78] R. C. Gonzalez, R.E. Woods, Digital Image Processing, Englewood Cliffs, NJ: Prentice-Hall, 2002
[79] K. Zhang, D. Whitman, "Comparison of three algorithms for filtering airborne LiDAR data," Photogramm. Eng. & Remote Sens., **71**, 313–324, 2005. http://dx.doi.org/10.14358/PERS.71.3.313
[80] F. Rottensteiner, "Using the Dempster-Shafer method for the fusion of LiDAR data and multispectral images for building detection," Inf. Fusion, **6**(4), 283–300, 2005. https://doi.org/10.1016/j.inffus.2004.06.004
[81] F. Rottensteiner, "Results of the ISPRS benchmark on urban object detection and 3D building reconstruction," ISPRS J. Photogram. Remote Sens., **93**(7), 256–271, 2014. http://dx.doi.org/10.1016/j.isprsjprs.2013.10.004
[82] L. Xu, D. Kong, X. Li, "On-the-fly extraction of polyhedral buildings from airborne LiDAR data," IEEE Geosci. Remote Sens. Lett., 1946–1950, **11**(11), 2014. https://doi.org/10.1109/LGRS.2014.2314458
[83] F. Tarsha-Kurdi, P. Landes, M. Grussenmeyer, M. Koehl, "Model-driven and data-driven approaches using LiDAR data: Analysis and comparison," in ISPRS Workshop, Photogrammetric Image Analysis (PIA07), Sep 2007, Munich, Germany, 87-92, 2007.
[84] A. Henn, G. Groger, V. Stroh, L. Plumer, "Model driven reconstruction of roofs from sparse LIDAR point clouds," ISPRS J. Photogram. Remote Sens., **76**, 17-29, 2013. http://dx.doi.org/10.1016/j.isprsjprs.2012.11.004+
[85] K. Kim, J. Shan, "Building footprints extraction of dense residential areas from LiDAR data local planarity," in 2011 Annual Conference of the American Society for Photogrammetry and Remote Sensing, Milwaukee, Wisconsin, 193-198, 2011.
[86] V. Verma, R. Kumar, S. Hsu, "3D building detection and modeling from aerial LIDAR data," in 2006 IEEE computer society conference on computer vision and pattern recognition, 17-22 June 2006, New York, NY, 2006. http://dx.doi.org/10.1109/CVPR.2006.12
[87] J. Yan, J. Shan, W. Jiang, "A global optimization approach to roof segmentation from airborne LiDAR point clouds," ISPRS J. Photogramm. Remote Sens., **94**, 183–193, 2014. http://dx.doi.org/10.1016/j.isprsjprs.2014.04.022
[88] K. Khoshelham, "Extending generalized Hough transform to detect 3D objects in laser range data," IAPRS, **XXXVI,** Part 3 / W52, 407-412, 2007.
[89] F. Tarsha-Kurdi, T. Landes, P. Grussenmeyer, "Hough-transform and extended RANSAC algorithms for automatic detection of 3D building roof planes from LiDAR data," IAPRS, **XXXVI,** Part 3 / W52, 206-210, 2007.
[90] G. Roth, M.D. Levine, "Extracting geometric primitives," Comput. Vis. Image Underst., **58**(1), 1-22, 1993. http://dx.doi.org/10.1006/cviu.1993.1029
[91] J. Yan, W. Jiang, J. Shan, "A global solution to topological reconstruction of building roof models from airborne LiDAR point clouds," ISPRS Ann. Photogram., Remote Sens. **3**, 379–386, 2016. http://dx.doi.org/10.5194/isprsannals-III-3-379-2016
[92] H. Huang, C. Brenner, M. Sester, "A generative statistical approach to automatic 3D building roof reconstruction from laser scanning data," ISPRS J. Photogramm. Remote Sens., **79**:29–43. 2013. http://dx.doi.org/10.1016/j.isprsjprs.2013.02.004
[93] R.J. You, B.C. Lin, "A quality prediction method for building model reconstruction using LiDAR data and topographic maps," IEEE Trans. Geosc., **49**(9), 2011. https://doi.org/10.1109/TGRS.2011.2128326
[94] F. Rottensteiner, C. Briese, "Automatic generation of building models from LiDAR data and the integration of aerial images," Int. Arch. Photogram. Remote Sens., **34**(3/W13), 2003.
[95] L. Cheng, L. Tong, Y. Chen, W. Zhang, J. Shan, Y. Liu, M. Li, "Integration of LiDAR data and optical multi-view images for 3D reconstruction of building roofs," Opt. Laser. Eng, **51**(4), 493–502, 2013. http://dx.doi.org/10.1016/j.optlaseng.2012.10.010
[96] I. Stroud, Boundary representation modelling techniques, London: Springer-Verlag London, 2006.
[97] Z. Lari, A. Habib, "An adaptive approach for the segmentation and extraction of planar and linear/cylindrical features from laser scanning data," ISPRS J. Photogram. Remote Sens., **93**(7), 192–212, 2014. http://dx.doi.org/10.1016/j.isprsjprs.2013.12.001
[98] R. Cao, Y. Zhang, X. Liu, Z. Zhao, "3D building roof reconstruction from airborne LiDAR point clouds: a framework based on a spatial database," Int. J. Geog. Inf. Sci., **16**, 1-22, 2017. http://dx.doi.org/10.1080/13658816.2017.1301456
[99] P. Gamba, B. Houshmand, "Joint analysis of SAR, LIDAR and aerial imagery for simultaneous extraction of land cover, DTM and 3D shape of buildings," Int. J. Remote Sens., **23**(20), 4439-4450, 2002. http://dx.doi.org/10.1080/01431160110114952
[100] J. Shewchuk, "Delaunay refinement algorithms for triangular mesh generation," *Comput. Geom.*, **47**(1), 741-778, 2014. http://dx.doi.org/10.1016/S0925-7721(01)00047-5
[101] R. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," Remote Sens. Env., 37(1), 35-46, 1991. http://dx.doi.org/10.1016/0034-4257(91)90048-B
[102] S.V. Kostrikov, D.Y. Bubnov, R.A. Pudlo, "Urban environment 3D studies by automated feature extraction from LiDAR point clouds," Visnyk of V.N. Karazin Kharkiv National University, series "Geology. Geography. Ecology", **52,** 156-181, 2020. https://doi.org/10.26565/2410-7360-2020-52-12
[103] J. Matas, O. Chum, "Randomized RANSAC with $T_{d,d}$ test," Image and Vision Computing, **22**(10), 837-842, 2004. https://doi.org/10.1016/j.imavis.2004.02.009
[104] M. Li, L. Nan, N. Smith, P. Wonka, "Reconstructing building mass models from UAV-LS images," Comput. Graph., **54**(C), 84–93, 2016. http://dx.doi.org/10.1016/j.cag.2015.07.004
[105] B. Xu, W. Jiang, J. Shan, J. Zhang, L. Li, "Investigation on the Weighted RANSAC Approaches for Building Roof Plane Segmentation from LiDAR Point Clouds," Remote Sens., **8**(1), 5, 2016. https://doi.org/10.3390/rs8010005
[106] R. Tse, Ch. Gold, D. Kidner, "Using the Delaunay Triangulation/ Voronoi Diagram to extract Building Information from Raw LIDAR Data," in 4th International Symposium on Voronoi Diagrams in Science and Engineering (ISVD 2007). IEEE Xplore Digital Library, 2007. http://dx.doi.org/10.1109/ISVD.2007.40
[107] L. Nan, C. Jiang, B. Ghanem, P. Wonka, "Template assembly for detailed urban reconstruction," Computer Graphics Forum, **34**(2), 217–228, 2015. http://dx.doi.org/10.1111/cgf.12554
[108] E. Aoidh, M. Bertolotto, D.C. Wilson, "Understand geospatial interests by visualising map interaction behavior," Inf. Vis., **7**(3), 275-286, 2008. https://doi.org/10.1057%2FIVS.2008.24
[109] D. Eum, T. Minoura, "WebSiteGen: Web-based database application generator," IEICE Trans. Inf. Syst., **E86-D**(6), 112-126, 2003.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 96-107 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Using TOST in Teaching Operating Systems and Concurrent Programming Concepts

Tzanko Golemanov*, Emilia Golemanova

Department of Computer Systems and Technologies, University of Ruse, Ruse, 7020, Bulgaria

ARTICLE INFO

Article history:
Received: 30 July, 2020
Accepted: 15 October, 2020
Online: 08 November, 2020

Keywords:
Operating Systems
Concurrent Programming
Teaching Tools

ABSTRACT

The paper is aimed as a concise and relatively self-contained description of the educational environment TOST, used in teaching and learning Operating Systems basics such as Processes, Multiprogramming, Timesharing, Scheduling strategies, and Memory management. TOST also aids education in some important IT concepts such as Deadlock, Mutual exclusion, and Concurrent processes synchronization. The presented integrated environment allows the students to develop and run programs in two simple programming languages, and at the same time, the data in the main system tables can be monitored. The paper consists of a description of TOST system, the features of the built-in programming languages, and demonstrations of teaching the basic Operating Systems principles. In addition, some of the well-known concurrent processing problems are solved to illustrate TOST usage in parallel programming teaching.

1. Introduction

This paper is an extension of work originally presented in 29th Annual Conference of the European Association for Education in Electrical and Information Engineering (EAEEIE) [1].

Operating Systems (OS) is one of the most fundamental courses in Computer Engineering, Computer Science, and Information Systems curricula. At the same time, students are faced with a lot of problems in learning OS basics, due to not fully understanding the wide range of techniques, strategies, and architectures involved in OS modules. The complexity and depth of concepts in this area requires a careful and detailed explanation from the educator in order to reach a better apprehending. Developing concurrent programs in an operating system environment (Unix, Linux, etc.) asks for students to already have prerequisite skills in both the C-programming and the operating system. Usually, such level of experience is difficult to be obtained at an undergraduate degree, thus a preferable approach would be to use a convenient and effective teaching tool that helps the students in their uneasy endeavor to apprehend and master the OS concepts.

We have assigned the widely spread tools for studying OS into the following two groups:

- Instructional Operating Systems
- Visual OS simulators

The systems from the first group (MINIX [2], Nachos [3], Xinu [4], Pintos [5], GeekOS [6]) run on real hardware and are too complicated to install and use. The main disadvantage of the instructional OS is the requirement students to be familiar with system-level programming in C, C++, Java or assembly language, and to be able to study large program codes (more than 4K program lines of Geek-OS and 15K of MINIX). Most of these systems do not have graphical visualizations, which makes them not very suitable for teaching purposes. In such cases, students spend a lot of time learning the systems themselves and how to use them properly.

In contrast, the visual tools from the second group (SOsim [7], SchedulerSim [8], CPU Sim [9]) are easier to use and convenient, however, they present rather restricted and pre-specified abilities (e.g. simulation of just one OS module or subsystem). For example, the TROJAN [10] simulator helps in teaching multiprocessor organization, the cache utilization, and the network traffic issues, while SIME [11] visualizations focus on memory management, but without shared memory concepts. Additionally, some of these systems (Proc OS [12], Alg OS [13]) do not connect theoretical concepts to a running code. Others present simulations of only one example problem (e.g. the Producer/Consumer problem [14]).

The authors propose a newly-developed by them OS teaching system named ***TOST*** *(Teaching in Operating Systems Tool)*. It has

*Corresponding Author: Tzanko Golemanov, Ruse University, Ruse, Bulgaria, tgolemanov@uni-ruse.bg

been created in response to the requirement of the educational tool to be straightforward to use, yet powerful enough, allowing the students to not only experiment with OS settings, architectures, and strategies but also to create and run any concurrent programs they want. This imposes the need for TOST to be developed as an integrated environment consisting of a code editor, a high-level language compiler, a virtual processor emulator, an interface for setting and modifying OS parameters, and a visualization of the content of the main OS tables.

The structure of the paper is as follows: Section 2 is an introduction to TOST giving an overview of the TOST interface, the grammars of the built-in programming languages, the manners to develop and execute concurrent processes, as well as the ways to set and modify system options, and to monitor the system tables' data. Section 3 is devoted to teaching Operating System basics, such as Process Management, Process Scheduling, and Memory Management. Section 4 describes the use of TOST in teaching Concurrent Programming concepts, specifically Mutual Exclusion (Dekker's algorithm, Peterson's algorithm, Test_And_Set instruction, and Semaphores), Interprocess Synchronization (Sleeping Barber problem), and Deadlock (Dining Philosophers problem). In Section 5, a student evaluation is presented. We conclude in Section 6 and discuss some further improvements to TOST.

2. The TOST System

The main objective of an educational tool in OS is to give students an inside view of the operating system. At the same time, in our opinion, the tool should meet the following basic requirements:

- to operate on an accessible hardware and software platform
- to support multiprogramming and concurrent processes
- to be with a well-known interface
- to support well-known programming languages with concurrency features
- to allow students the experimentation with the basic OS parameters and options.

TOST has been developed with exactly this idea in mind. It allows students to develop and execute concurrent processes, while at the same time they can change the basic settings of the OS (CPU scheduling, Quantum size, Memory management, Virtual memory strategies, etc.) and to monitor the main system tables (Processes, Blocked, Ready, Semaphores, etc.).

The proposed integrated environment includes:

- a multi-tasking, and a multi-window operating system
- editors and compilers for two simple programming languages
- a virtual processor emulator.

The last TOST version is developed with Embarcadero® Delphi within Embarcadero RAD Studio 10.3.3 Rio. It has no installation procedure, the executable file of the tool is very small in size (less than one MB), and it operates on any MS Windows environment. TOST is easy to learn and use, and allows students to become familiar with the capabilities and the limitations of the built-in programming languages within a short time (about an hour).

The first TOST environment is introduced in [15], while the focus of [1] is on using TOST in teaching mutual exclusion, synchronization, and deadlock. This paper expands the application area of TOST, describing its usage for teaching and learning OS concepts.

2.1. Programming in TOST

A general view of the current TOST main menu is depicted in Figure 1.

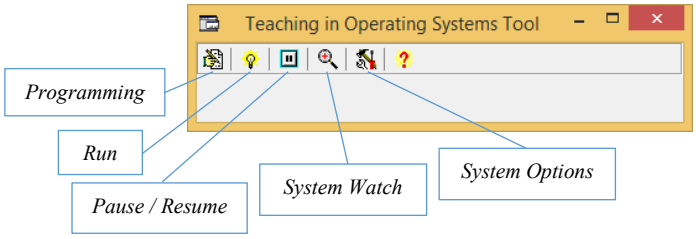


Figure 1: TOST interface

Using TOST, students can develop and execute concurrent programs in two simple programming languages - PASCAL-style and C-style. The grammars of these languages, consist of roughly twenty grammar productions only, which are presented in Figure 2 and Figure 3 in Extended BNF (ISO/IEC 14977).

```
program    = "void" "main" "(" ")" body
body       = [ { type identifier { "," identifier } } ] block
type       = "int" | "char" | "bool" | "shared" | "semaphore"
block      = "{" statement { ";" statement } "}"
statement  = block | assignment | if | while | read | write | lock | unlock |
             init | wait | signal
assignment = identifier "=" expression
expression = expr [ ( "<" | ">" | "==" ) expr ]
expr       = term { ( "+" | "-" | "or" ) term }
term       = factor { ( "*" | "/" | "and" ) factor }
factor     = [ "not" ] ( identifier | constant | random | "(" expression
             ")" )
constant   = number | """" char """" | "#" ascii_code | "true" | "false"
random     = "random" "(" expression ")"
if         = "if" "(" expression ")" statement
while      = "while" "(" expression ")" statement
read       = "cin" ">>" identifier
write      = "cout" "<<" expression { "<<" expression }
lock       = "lock" "(" identifier ")"
unlock     = "unlock" "(" identifier ")"
init       = "init" "(" identifier "," expression ")"
wait       = "wait" "(" identifier ")"
signal     = "signal" "(" identifier ")"
```

Figure 2: C-style language grammar productions

```
program    = [ "program" identifier ";" ] body "."
body       = [ "var" { identifier { "," identifier } ":" type ";" } ] block
type       = "integer" | "char" | "boolean" | "shared" | "semaphore"
block      = "begin" statement { ";" statement } "end"
statement  = block | assignment | if | while | read | write | lock | unlock |
             init | wait | signal
assignment = identifier ":=" expression
expression = expr { ( "<" | ">" | "=" ) expr }
expr       = term { ( "+" | "-" | "or" ) term }
term       = factor { ( "*" | "/" | "and" ) factor }
```

```
factor    = [ "not" ] ( identifier | constant | random | "(" expression
            ")" )
constant  = number | "'" char "'" | "#" ascii_code | "true" | "false"
random    = "random" "(" expression ")"
if        = "if" expression "then" statement
while     = "while" expression "do" statement
read      = "read" "(" identifier ")"
write     = "write" "(" expression { "," expression } ")"
lock      = "lock" "(" identifier ")"
unlock    = "unlock" "(" identifier ")"
init      = "init" "(" identifier "," expression ")"
wait      = "wait" "(" identifier ")"
signal    = "signal" "(" identifier ")"
```

Figure 3: PASCAL-style language grammar productions

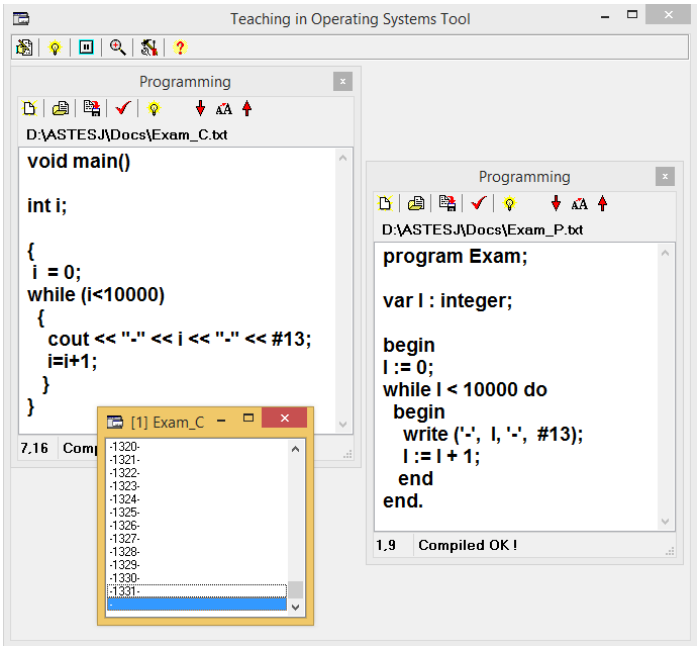


Figure 4: Process editing and executing in TOST

In the text bellow we will use C-style only for the programs' definition.

The programming languages in TOST are exclusively for learning purposes and while they are quite simple, there are some features that need to be addressed:

2.1.1. Types of variables

Only a few standard simple data types *integer*, *char* and *boolean* are available for local variables, nevertheless, some data types for concurrent programming purposes are also provided. Variables from the ***shared*** type can be declared in a TOST program. These variables become common for all concurrent processes running in the system. A shared variable is unified, i.e. it can accept any integer, char or boolean values, as well as any data can be extracted from it. As a result, passing through such variable can also be used in order to convert data from one type to another.

Another type of data used in parallel programming is the ***semaphore***. In TOST we implement the original Dijkstra's semaphores concept [16]. A semaphore variable ***S*** consists of two properties - ***S.value*** and ***S.queue***. The property S.value is a non-negative integer, while S.queue is a FIFO list of identifiers of the blocked processes (process ID).

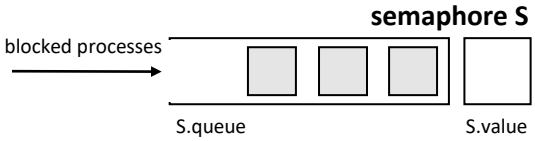


Figure 5: Semaphore's structure

2.1.2. Statements

Although TOST languages maintain only basic statements like assignment, conditional, loop and input/output, some statements for concurrent programming purposes are also available.

The semaphore operations ***init***, ***wait***, and ***signal***, which are defined below, are functions from the OS kernel:

init (S, value)

```
S.value = value
```

wait(S)

```
if (S.value > 0)
  decrement S.value;
else
  block this process and add it's ID in S.queue
```

signal (S)

```
if (S.queue is empty)
     increment S.value;
else
     unblock the first process from S.queue
```

Figure 6: Semaphore operations

Students can investigate the advantages and disadvantages of using the special assembly language instruction ***Test_and_Set*** [17] through the statements ***Lock*** (CS) and ***Unlock*** (CS), where CS (***C****ritical* ***S****ection*) is a boolean shared variable.

Function ***random***(*expression*) which returns a random integer value from *0* to *expression-1* can also be used in some synchronization primitives' implementations.

2.2. Processes Execution Management in TOST

After the successful compilation, an object code file is generated (with COD extension) and it is ready to be run in the environment. The compiled processes can be started from the ***Run*** at the main menu (as well as from the editing windows).

Figure 7 presents the general system view where in the ***Run*** dialog box students can select an object file to be executed, and set the initial parameters of the running process.

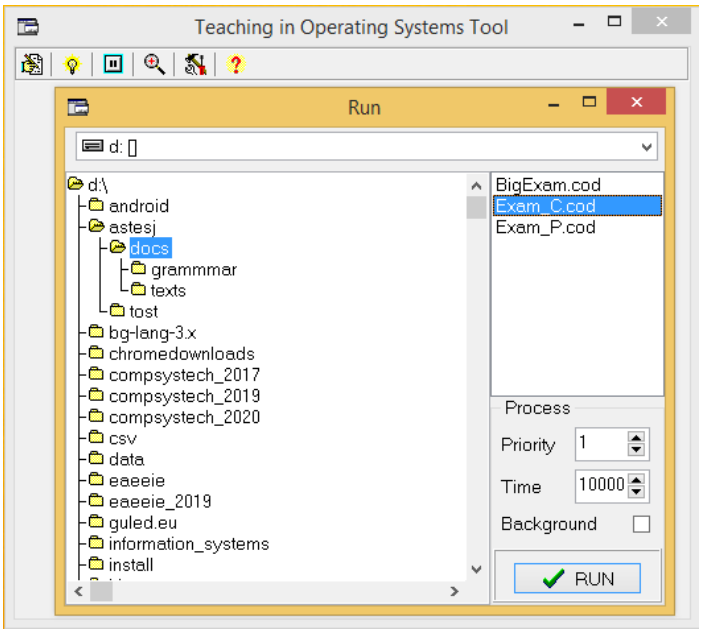


Figure 7: Process running

Priority spin-edit box sets the initial process priority while the total process execution time (used in some scheduling strategies) can be entered in ***Time***. By checking ***Background*** the students can specify the process visual mode (execution in a separate window or behind the scenes). Pressing the ***RUN*** button multiple times will cause the start of several concurrent processes.

If detailed monitoring of running processes is required, a temporary "freeze" can be performed via ***Pause / Resume*** at the main menu.

2.3. TOST System Options

Some basic OS parameters can be set and modified while several processes are executed in a concurrent mode.

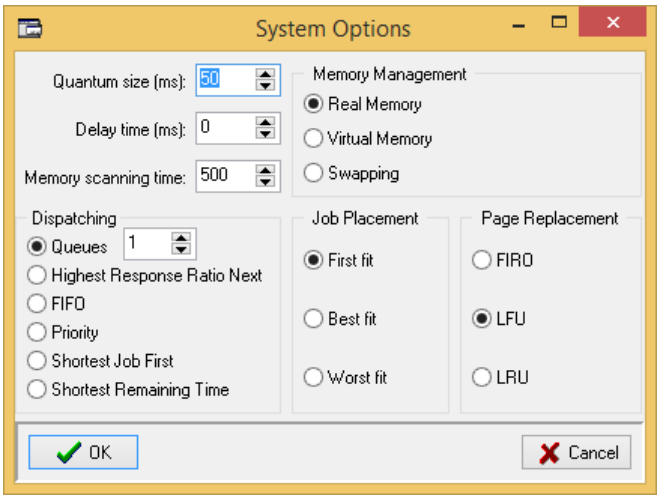


Figure 8: System Options dialog box

In System Options dialog box (Figure 8) students can modify:

- ***Quantum Size***: system time sharing quantum size in ms
- ***Delay time***: processes speed control
- ***Memory scanning time***: time interval of RAM scanning
- ***Dispatching***: current scheduling strategy
- ***Memory Management***: current memory management strategy – *Real Memory*, *Virtual Paging* or *Swapping*
- ***Job Placement***: jobs placement strategy in real memory mode
- ***Page Replacement***: page replacement strategy in virtual memory mode

These modifications of system parameters allow real-time monitoring of how each of them affects the processes executed.

2.4. TOST System Watch

For achieving an OS inside view, while several processes are executed in parallel mode, students can monitor the information dynamics in selected system tables (Figure 9): ***Processes***, ***Ready***, ***Blocked***, ***Memory Allocation*** map and ***Semaphores*** info. The information from each of the system tables is presented in a separate window, as it is shown in Figure 10.

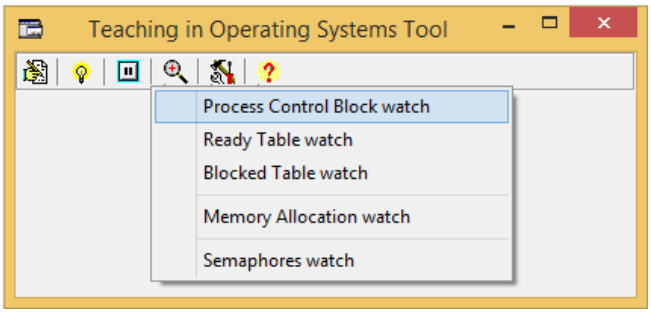


Figure 9: System tables Watch

2.4.1. Processes

The ***Processes*** table (Figure 10) contains the current ***Process Control Block*** of each of the existing concurrent processes.

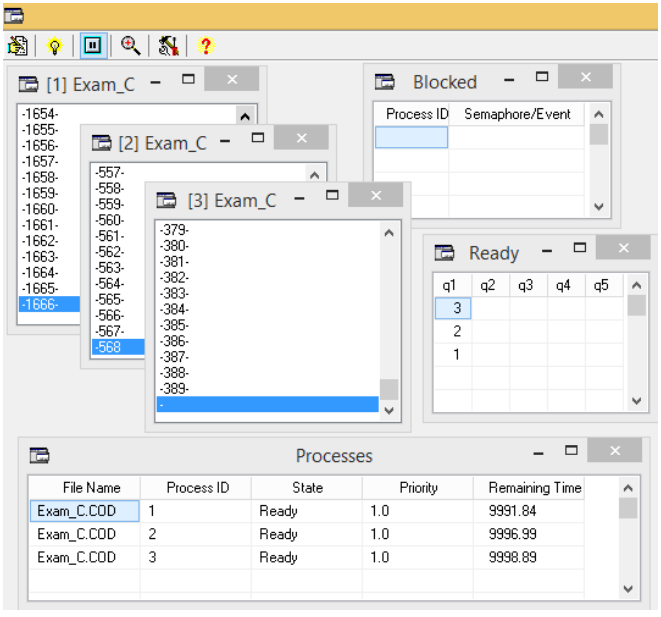


Figure 10: Processes execution and system tables monitoring

The information is presented in five fields:

- ***File Name***: the name of the compiled object file
- ***Process ID***: an integer unique identifier of the process
- ***State***: current process status – *Ready*, *Running* or *Blocked*
- ***Priority***: a real value determining the process priority
- ***Remaining Time***: the time remaining until the process completion.

In this table, students can manipulate the priority or the remaining time of the selected process.

2.4.2. Ready table

The ***Ready*** table contains IDs of all processes with status "Ready". It is possible (according the settings in System Options) to have up to 5 priority queues. Thus, students can experiment with scheduling strategies like ***Round Robin*** (one queue only) or ***Multilevel Feedback Queue*** [18] for the separation of the processes with different behavior (with more I/O operations or with more computations).

2.4.3. Blocked table

The identifiers of blocked processes and the names of semaphores/events, related to their blocking are presented in the ***Blocked*** table. In fact, these are FIFO-queues with blocked processes of all semaphores used. The information in this table is especially helpful when students have to monitor semaphores usage in processes ***synchronization***, ***mutual exclusion*** and ***deadlock*** prevention cases.

2.4.4. Memory Allocation map

Important information about memory allocation and referencing is given by ***Memory Allocation Watch***. When this submenu item is selected, the system periodically scans the TOST RAM and displays results in a separate window. This feature allows students to monitor a map of the memory blocks, where the occupied blocks are marked with the process' ID. The detailed referencing of the pages in the virtual memory is also displayed.

2.4.5. Semaphores information

Semaphores Watch submenu item can be selected for monitoring all semaphore variables' current values in a separate window. This helps students to master the mechanism of using both binary and counting semaphores.

3. Teaching Operating System Basics

3.1. Teaching Process Management

A process [2] is basically a program in execution. During its existence, the process goes through several different states, and the basic three are ***Ready***, ***Running***, and ***Blocked***.

Once students run several processes, they can monitor (as shown in Figure 10) their life cycles in the *Processes* (ID, the current State and the state changing), *Ready* (ID list of all ready processes), and *Blocked* (ID list of all blocked processes, and the event each process is waiting for) tables.

3.2. Teaching Process Scheduling

The act of choosing which of all the ready processes should be moved to the running state is not a trivial task [19]. It is known as Process Scheduling and is performed by OS Dispatcher.

There are several goals of the process scheduling system:

- to maximize the fairness according to the processes priorities
- to keep the CPU busy at all time
- to maximize throughput and to deliver minimum response time for all processes
- to minimize resource starvation
- to support a lot of interactive users
- to minimize system overheads (OS CPU time, OS RAM).

In fact, very often these goals contradict with each other, so the Dispatcher should implement an appropriate compromise [20]. Through experimentation, students are expected to gain understanding in the scheduling strategies and to be able to choose the most appropriate one, depending on the user's needs and objectives. In TOST students can change (at any time) the current scheduling strategy using ***Dispatching*** radio-buttons in ***System Options*** (see Figure 8).

Scheduling disciplines of the two general categories are supported in TOST:

Non-Preemptive Scheduling: once the OS assigns the CPU to a process, the process does not release the CPU until it has finished:

- ***FIFO:*** Simplest scheduling algorithm with only one queue (see queue q1 in *Ready*). New processes are appended in q1 where context switches only occur upon running process' termination. Then the top placed process in q1 becomes *Running*. Students can monitor:
  - scheduling overheads are minimal
  - throughput can be low
  - no starvation
  - waiting time and response time can be high
  - no prioritization occurs
  - interactive users are not allowed
- ***Priority:*** Whenever a scheduling event occurs, the *Processes* table will be searched for the process with the maximum priority number (the most important one), which will be the next one to be scheduled for execution. Students can monitor:
  - scheduling overheads are low
  - throughput can be low
  - starvation can occur in a busy system with many high-priority processes in the Ready table

- waiting time and response time can be high
- interactive users are not allowed

- ***Shortest Job First (SJF)***: Dispatcher selects the process with the least estimated processing time to be executed. The estimated processing time in TOST can be set while process is initialized (Figure 7). Students can monitor how the algorithm increases throughput, and how the starvation can become a problem, in a system with many short processes in the *Ready* table.
- ***Highest Response Ratio Next (HRRN):*** is a dynamic priority discipline proposed by Brinch Hansen to mitigate the problem of the large processes starvation. When CPU becomes free, Dispatcher calculates the priorities of all ready processes according to:

  Priority = 1 + waiting time / processing time

  and selects the high priority process to be executed. Students can monitor the priority calculations (in *Processes* table) which gives preference to short processes, but with each subsequent recalculation the priority of the large ones grows and they can be selected as well.

Preemptive Scheduling: Dispatcher can interrupt a running process any time in middle of the execution and the process is returned back in the *Ready* table:

- ***Shortest Remaining Time (SRT):*** is a modification of SJF with a preemption. When a new shortest process arrives Dispatcher interrupts the current process and assigns the CPU to the new one. Students can monitor that this discipline achieves maximum throughput in most cases, and the starvation is possible for large processes where a lot of small processes arrive.
- ***Queues:*** By specifying the number of queues (Figure 8), students can set two strategies:
  Queues (1): Round-robin (RR): is a FIFO modification with the process preempting where *Ready* table is a queue. Dispatcher assigns a small time interval (Quantum) per process, then interrupts and appends it at the end of the *Ready* queue. When students set *Delay time* in *System Settings* they can monitor processes preempting in the *Ready* table in detail. When students run processes with different priorities, they can monitor how the more important processes are with proportionally larger quantum and get a better service, and the starvation can never occur.
  Queues (2-5): Multilevel feedback queues: In this case, processes are grouped according to their behavior - interactive processes vs batch processes, and processes with more I/O vs. computational processes. In order to obtain a balance in system's resources utilization Dispatcher has to provide better service for interactive and I/O processes. Students can monitor in detail how processes pass into the multilevel set of priority queues (from 1 to 5) in *Ready* table.

3.3. Teaching Memory Management

In general, when discussing the topic of memory management [2], [19] the focus is on the virtual memory organization. However, before reaching this point, students need to understand in detail what the issues are in using the real memory organization. Only then they can comprehend why this kind of organization has been replaced everywhere by the virtual one.

3.3.1. Real Memory

Real memory organization is a classical (one of the oldest) memory allocation model according to which the process is placed in a consecutive memory blocks [21]. Each process takes up exactly as much memory as it needs, and due to the fact that the memory blocks have sequential addresses, process loading, executing and the memory releasing are extremely fast. Then *"Where is the problem?"* comes to be the next instructor's question towards the students.

Students are instructed to develop and try to run several large processes. Later, they run five small processes and with using the TOST feature ***Memory Allocation Watch,*** monitor the memory map with real memory organization. Then students kill the first, second and fourth process, so these memory blocks are freed. Figure 11 shows how the memory blocks occupied by a process are marked with its identifier.

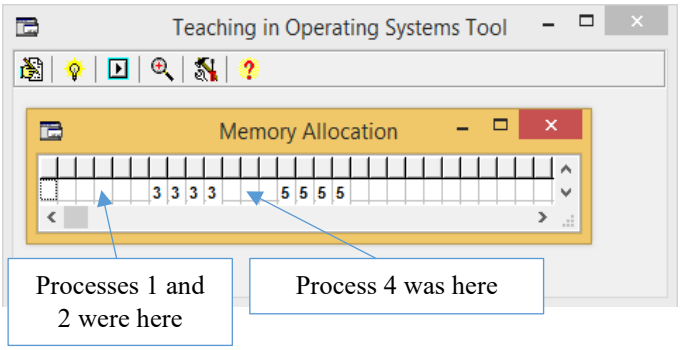


Figure 11: Memory allocation map

Following these instructions, students quickly realize the deficiencies of the real memory:

- A process cannot be started if it is even one byte larger than the available physical memory.
- There is a low CPU utilization, due to the low degree of multiprogramming (number of ready processes loaded in physical memory).
- The occurrence of memory fragmentation: In high dynamics of started and finished processes, the free memory area is broken into small parts, in which none of the ready processes can be allocated. Thus, a lot of memory blocks can remain unusable.

In order to minimize the problem of memory fragmentation, the placement strategies ***First-fit***, ***Best-fit*** and ***Worst-fit*** can be applied (Figure 8). To assess their usefulness students can run a new process and monitor its placement in the memory.

3.3.2. Virtual Memory

Although the problem of memory fragmentation can be solved, other serious problems of real memory remain: small maximum process size and low degree of multiprogramming.

Virtual memory organization allows each process to be divided into blocks (***pages***) while only a small part of the code and data (that are currently needed) are located in the physical memory. The physical memory is divided into blocks, which are called ***frames***. Figure 12 illustrates a moment in the execution of a large process where ***Virtual Memory*** as *Memory management* is set (Figure 8).

At that stage, students can monitor the occupation and references to the blocks (pages) in progress, with the memory scanned at a certain interval (0.5s for example). When the process starts, only a small part of its blocks are loaded - those that are needed at the beginning. Later, the OS loads additional blocks into the memory, those which are needed by the process but are not available. The blocks that the process refers to during the scan interval are colored red.

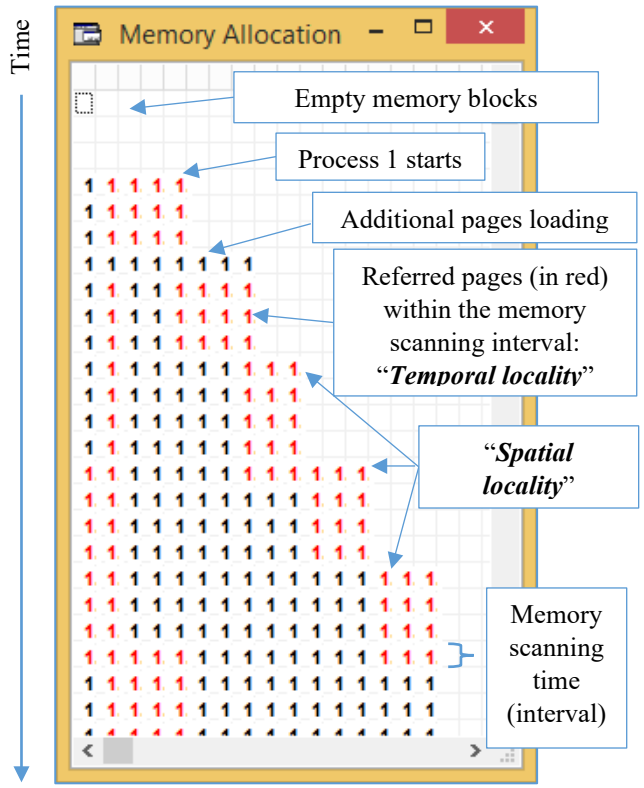


Figure 12: Memory allocation and pages referencing

Memory allocation map allows the instructor to clarify the ***principles of the locality*** [22] - ***Temporal locality*** and ***Spatial locality***.

Principles of the locality can be used for anticipatory paging (swap prefetch) to increase processes' execution speed. These principles are useful when the optimal replacement strategy has to be chosen.

In TOST students can set (Figure 8) and examine various replacement strategies such as *Least Recently Used* (***LRU***), *Least Frequently Used* (***LFU***) and ***FIFO***.

4. Teaching Concurrent Programming Concepts

Even if the students are familiar with the theoretical aspects of parallel programming, in order to gain intuition about the subject, it is vital to be provided with examples of incorrect operation of competitive processes. Therefore, our approach is to firstly develop concurrent programs without any protections and then, analyze the faulty actions of the processes. Consequently, modify the programs in order to achieve correct results. In addition, to strengthen this knowledge, we solve and program some of the well-known toy-problems in concurrent processing.

Representative examples of using TOST in teaching ***Mutual Exclusion***, ***Interprocess Synchronization*** and ***Deadlock*** are considered in the next sections.

4.1. Teaching Mutual Exclusion

The parts of the concurrent processes where they access a shared resource are called ***Critical Sections*** (CS) [23]. Critical Sections must be protected by synchronization primitives, assuring only one of the processes can be in the CS.

Figure 13 presents programs with no protection of the critical section. Students develop and execute an initialization process (init.txt), and two concurrent processes (p1.txt and p2.txt), incrementing the shared variable C 1000 times. After finishing the processes, the students can see the final result value of C = 1839, instead of the correct value of 2000.

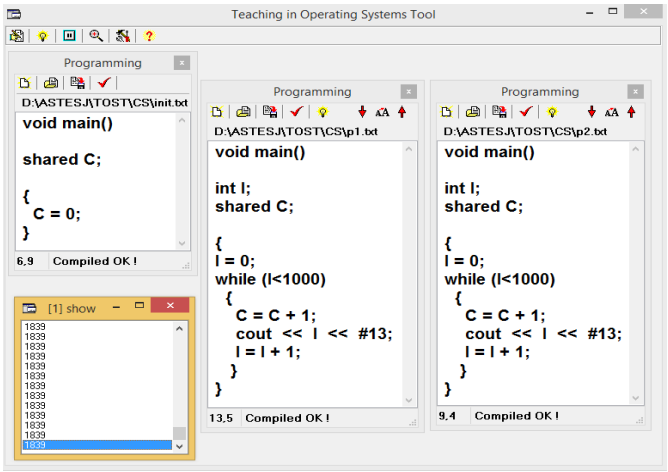


Figure 13: Concurrent processes with Critical Section

After examining the reasons for the wrong result, the students transform the programs gradually, seeking the correct solution for synchronization primitives. Thus, they examine program versions with one, two, and three shared variables, and at last, they come to the consideration of the ***Dekker's algorithm*** (Figure 14).

The ***Peterson's algorithm*** [24] is another well-known algorithm that students can try and compare with Dekker's algorithm. A simple solution of CS protection by the Peterson's algorithm is given in Figure 15.

In Figure 16 a simple example of CS protection, using primitives ***Lock*** and ***Unlock***, based on the hardware instruction ***Test_And_Set***, is presented.

The last example considered is the use of semaphores to protect the critical section (Figure 17).

```
void main()
shared C, ReqP1, ReqP2, Turn ;
{
 C = 0;
 ReqP1 = false;
 ReqP1 = false;
 Turn = 1;
}
```

```
void main()
int I;
shared C, ReqP1, ReqP2, Turn ;
{
I = 0;
while ( I < 1000 )
 {
  ReqP1 = true;
  while ( ReqP2 == true )
   If  (Turn  == 2)
    {
     ReqP1 = false;
       while (Turn  == 2) ;
     ReqP1 = true;
    };
        C = C + 1;
  ReqP1 = false;
  Turn  = 2;
  I = I + 1;
 }
}
```

(a) Initialization process (b) Process 1

Figure 14: Critical Section protection by Dekker's algorithm

```
void main()
shared C, ReqP1, ReqP2, Turn;
{
 C = 0;
 ReqP1 = false;
 ReqP1 = false;
}
```

```
void main()
int I;
shared C, ReqP1, ReqP2, Turn ;
{
I = 0;
while ( I < 1000 )
 {
  ReqP1 = true;
  Turn  = 2;
  while (ReqP2 and (Turn == 2)) ;
        C = C + 1;
  ReqP1 = false;
  I = I + 1;
 }
}
```

(a) Initialization process (b) Process 1

Figure 15: Critical Section protection by Peterson's algorithm

```
void main()
shared C, CS;
{
 C = 0;
 CS = false;
}
```

```
void main()
int I;
shared C, CS ;
{
I = 0;
while ( I < 1000 )
 {
  Lock (CS);
        C = C + 1;
  Unlock (CS);
  I = I + 1;
 }
}
```

(a) Initialization process (b) Process 1

Figure 16: Critical Section protection by Test_And_Set instruction

```
void main()
semaphore  CS;
shared  C;
{
 C = 0;
 init (CS, 1);
}
```

```
void main()
semaphore  CS;
shared  C;
int  I;
{
 I = 0;
 while (I<1000)
  {
   wait (CS);
     C = C + 1;
   signal (CS);
   I = I + 1;
  }
}
```

Figure 17: Critical Section protection by a semaphore

Figure 18 presents the contents of two main OS tables (Processes and Blocked) during the concurrent execution of three identical processes with the source code of Figure 17. Processes *P1_S* and *P2_S* are blocked, while the ready process *P2_3* is inside the critical section. Students also have the possibility of monitoring the semaphore variable CS in a separate window

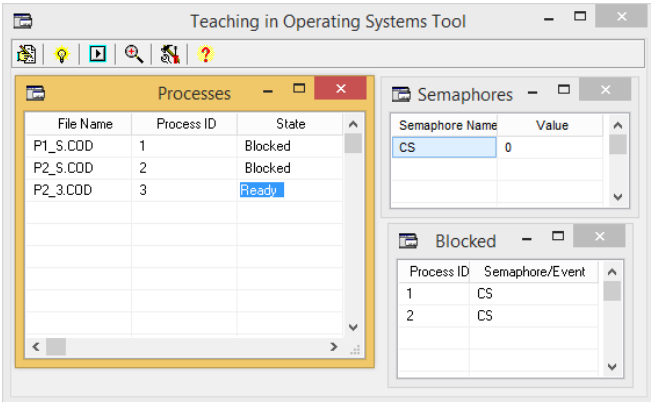


Figure 18: Monitoring TOST system tables

After finishing the processes, the value of the shared variable C is correct and equal to 3000.

4.2. Teaching Interprocess Synchronization

A good illustration of the Interprocess Synchronization is the. ***Sleeping Barber problem*** [25] depicted in Figure 19. As soon as the Barber or Client process reaches a certain point in the execution, it must stop and wait for an event to occur.

Figure 19: Sleeping Barber problem

The problem can be stated as follows: In a barbershop, there is a barber chair and an area for the waiting clients with five chairs. The algorithms of the concurrent processes of the Barber and the Client include the following steps:

Barber:

1) *The barber signals that he is ready to work (free).*
2) *The barber sits on the barber's chair and sleeps, waiting for a new client (indoor bell).*
3) *The barber gets up and waits for the client to sit on the barber's chair.*
4) *The barber begins the haircut and signals when he completes the work.*
5) *The barber waits until the client leaves the barbershop (outdoor bell).*
6) *Go to 1.*

```
void main()
semaphore Chairs, Barber,
          Client, Sit, Finish,
          Out;
{
   init ( Chairs, 5 );
   init ( Barber, 0 );
   init ( Client, 0 );
   init ( Sit, 0 );
   init ( Finish, 0 );
   init ( Out, 0 );
}
```

There are 5 chairs available
The barber is not available
No new client
No one is sitting in barber's chair
The barber has not finished work
The client has not come out

(a) Initialization

```
void main()
semaphore Chairs, Barber,
          Client, Sit, Finish,
          Out;
{
   while (true)
  {
    signal ( Barber );
    wait ( Client );
    wait ( Sit );
    signal (Finish );
    wait ( Out );
  }
}
```

(b) Barber

```
void main()
semaphore Chairs, Barber,
          Client, Sit, Finish,
          Out;
{
   wait ( Chairs );
   signal ( Client );
   wait ( Barber );
   signal ( Chairs );
   signal ( Sit );
   wait ( Finish );
   signal (Out );
}
```

(c) Client

Figure 20: Sleeping Barber solution

Client:

1) *The client checks if there is a free chair in the waiting room and eventually waits until a chair is available.*
2) *The client signals that there is a new client (input bell).*
3) *The client is waiting for the barber to be free.*
4) *The client frees the chair in the waiting room.*
5) *The client sits on the barber's chair and signals that he is ready for a haircut.*
6) *The client is waiting for the barber to complete the haircut.*
7) *The client gets out of the barber's chair and leaves the barbershop (output bell).*

The first thing students have to do is to identify processes synchronization events and implement one semaphore variable for each of the controlled events. Then they develop three programs - Initialization (for setting semaphore values), Barber, and Client (Figure 20).

The implementation of the synchronization of one barber's process and eight clients' processes are depicted in Figure 21.

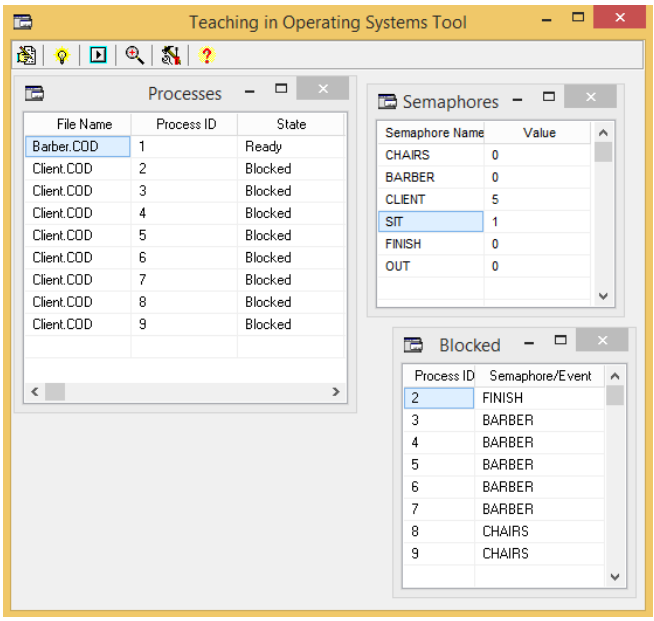


Figure 21: Sleeping Barber execution monitoring

According to the information in TOST system tables:

- the barber is busy (Semaphores: *Barber = 0*)
- a client is seated in the barber chair (Semaphores: *Sit = 1*) and waiting for the barber to complete the haircut (Blocked: *Finish*)
- five new clients are in the waiting room (Semaphores: *Clients = 5*) and all the chairs are occupied (Semaphores: *Chairs = 0*)
- all five new clients are waiting for the barber (Blocked: *Barber*)
- two clients are waiting for a chair outside of the barbershop (Blocked: *Chairs*).

4.3. Teaching Deadlock

The ***Dining Philosophers*** toy-example [26], depicted in Figure 22, is often used to analyze the synchronization and deadlock problems in the concurrent processes execution, as well as to demonstrate the approaches for solving them.

The students solve this problem by developing five programs (for the philosophers) while the correct access (mutual exclusion) to the shared objects (forks) is implemented by semaphores. The example programs for the first two philosophers are presented in Figure 23.

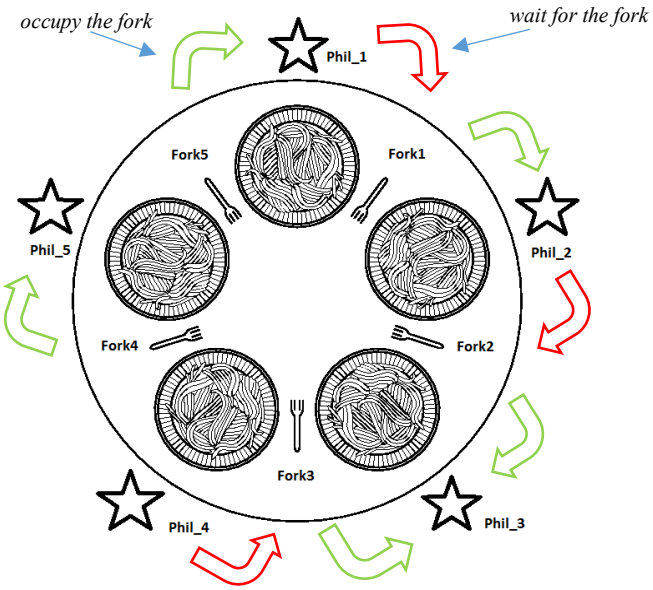


Figure 22: Dining Philosophers problem

```
void main()
semaphore FORK5, FORK1 ;
{
while ( true )
 {
    cout << "T" << #13;
   wait ( FORK5 );
   wait ( FORK1 );
    cout << "E" << #13;
   signal ( FORK1 );
   signal ( FORK5 );
 }
}
```

```
void main()
semaphore FORK1, FORK2 ;
{
while ( true )
 {
    cout << "T" << #13;
   wait ( FORK1 );
   wait ( FORK2 );
    cout << "E" << #13;
   signal ( FORK1 );
   signal ( FORK2 );
 }
}
```

Figure 23: Phil_1 and Phil_2 program samples

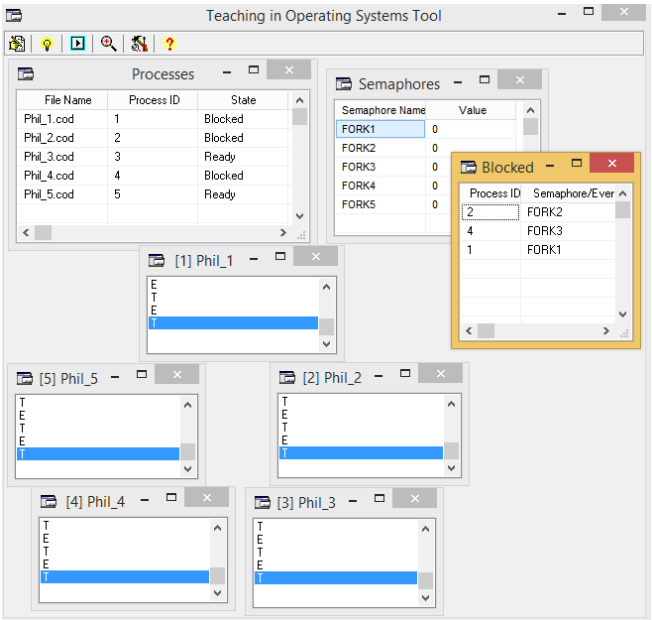


Figure 24: Dining Philosophers execution

TOST allows students to monitor (Figure 24) how philosophers are waiting for the forks (in *Processes* and *Blocked*) and how they take and put the forks on the table (in *Semaphores*).

Another aim of the Dining Philosophers problem is to demonstrate the prevention of ***Deadlock***.

Students can monitor how in the case of all philosophers taking a fork with their right hand, while expecting the release of a second fork, a *Deadlock* occurs (Figure 25 and Figure 26).

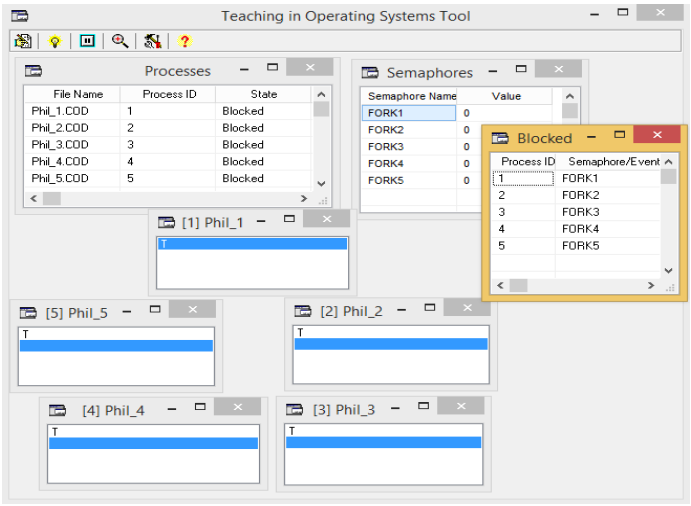


Figure 25: Dining Philosophers Deadlock example

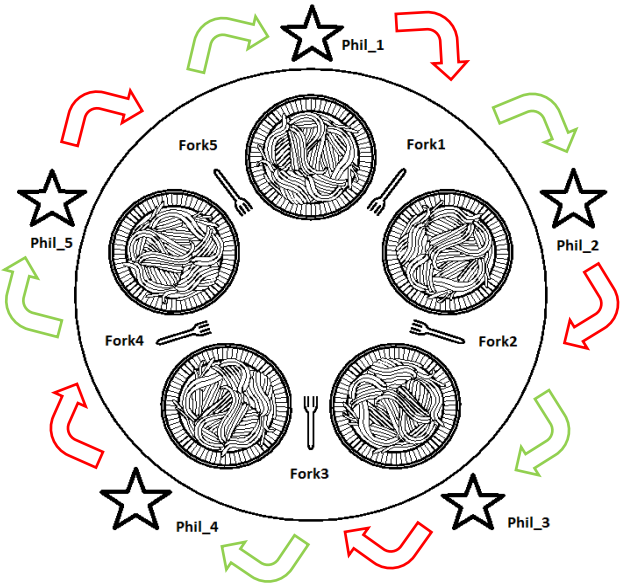


Figure 26: Dining Philosophers Deadlock example (Circular Wait)

Students have to find a solution and save philosophers from starvation dead. One idea is to break the fourth necessary condition for Deadlock - ***Circular Wait*** [27]. Students apply this approach by reprogramming Phil_1 (Figure 27) and introducing the general requirement for all processes to request resources in **ascending order only**.

```
void main()
semaphore FORK5, FORK1 ;
{
while ( true )
  {
      cout << "T" << #13;
     wait ( FORK5 );
     wait ( FORK1 );
       cout << "E" << #13;
     signal ( FORK1 );
     signal ( FORK5 );
   }
}
```

```
void main()
semaphore FORK5, FORK1 ;
{
while ( true )
  {
      cout << "T" << #13;
     wait ( FORK1 );
     wait ( FORK5 );
       cout << "E" << #13;
     signal ( FORK1 );
     signal ( FORK5 );
   }
}
```

Figure 27: Phil_1 program – **before** and **after** changes

5. Students' Opinion

Trying to improve the quality and usefulness of the integrated environment TOST, moreover to get better results from its usage, we also rely on end-users' opinions. For this reason, in a span of two consecutive years, we conduct qualitative surveys with students at the end of the course. The survey used a 5-point Likert scale and is based on a relatively small sample of 131 bachelor students. The summary data for user acceptance of the presented tool are in Table 1.

Table 1: Opinion Survey Results

Survey questions	**A**	**B**	**C**	**D**	**E**
I like TOST environment	2	13	30	55	31
TOST is easy to learn	0	3	15	40	73
I think that programming in TOST is complex	75	29	11	12	4
TOST is visually very attractive	7	22	26	61	15
Overall, I am satisfied with the TOST approach to studying Operating systems	5	11	15	41	59
Overall, I am satisfied with the performance of the TOST environment	9	9	12	21	80
The installation process of TOST is quick and easy	0	0	3	5	123
TOST often freeze, crash, or does not behave as expected	106	9	11	4	1
Overall, I am satisfied with the experimenting with TOST example solutions in understanding concepts related to Mutual Exclusion, Synchronization, Deadlock, and Operating system in general	4	2	3	111	11
I think the interface of the TOST environment is user-friendly	8	5	39	57	22
TOST is successful and effective in stimulating Operating system algorithms and strategies	0	0	0	95	36

A: *strongly disagree*
B: *disagree*
C: *undecided*
D: *agree*
E: *strongly agree*

Based on the data presented, it can be assumed that the responses are positive. According to the majority, the TOST environment is easy to learn and use, user-friendly, useful, and supportive for their understanding of the fundamental concepts of operating systems and the problems concerning concurrent programming.

6. Conclusions and Further Research

Learning the concept and basic principles of Operating systems poses a great challenge to Computer Systems and Technologies students. In this paper, the authors present an integrated environment named TOST, specially designed to be used in Operating systems and Concurrent programming courses at Ruse University. A description of the TOST environment and a demonstration of how it can be very effectively used in teaching Operating systems principles are discussed, as well as some of the TOST features that address the learning of Mutual Exclusion, Interprocess Synchronization, and Deadlock.

The TOST includes a time-sharing operating system with compilers for two simple and easy to use programming languages. There are shared variables and semaphores variables for concurrent algorithms development. The ability to monitor the real-time information from basic OS tables allow students to gain a better apprehending of the algorithms, concepts, and theories behind the design, construction, and insights of operating systems.

The presented environment might be extended in the following directions:

- developing a built-in file system with a set of file allocation methods and management strategies;
- developing a spooling system for mediation between a process and slow peripheral devices;
- implementing complex data types and subroutines for the programming languages integrated.

We have been using TOST for several years in the Department of Computer Systems and Technologies at Ruse University and the student's feedback and results are positive. The lack of any installation procedure, the small size, and the 'copyleft' license allow TOST to be obtained and used without limitation and with full inclusivity.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

This work/paper is supported by project 20-FEEA-01 "Methods and tools for multimedia content analysis and automated document and big data processing", funded by the Research Fund of the "Angel Kanchev" University of Ruse.

References

[1] T. Golemanov, E. Golemanova, "Using TOST in Teaching Mutual Exclusion, Synchronization, and Deadlock," in 29th Annual Conference of the European Association for Education in Electrical and Information Engineering (EAEEIE) 2019, Ruse, Bulgaria, 2019, doi:10.1109/EAEEIE46886.2019.9000441.
[2] A.S. Tanenbaum, A.S. Woodhull, Operating Systems Design and Implementation (Second ed.), 2011.
[3] W. Christopher, and T.A. S. Procter, "The Nachos instructional operating

system," in 1993 Winter USENIX Conference, San Diego, California, 1993, doi:10.5555/1267303.1267307.
[4] D. Comer, Operating System Design: The XINU Approach, Prentice-Hall, 1984.
[5] B. Pfaff, G. Romano A. and Back, "The Pintos Instructional Operating System Kernel," in SIGCSE 2009, Chattanooga, TN, 2009, doi:10.1145/1539024.1509023.
[6] D. Hovemeyer, J.K. Hollingsworth, B. Bhattacharjee, "Running on the bare metal with GeekOS," ACM SIGCSE Bulletin, **36**(1), 2004, doi:10.1145/1028174.
[7] L.P. Maia, F.B. Machado, P.J.A. C., "A Constructivist Framework for Operating Systems Education: a Pedagogic Proposal Using the SOsim," in ITiCSE'05, Monte de Caparica, Portugal, 2005, doi:10.1145/1067445.1067505.
[8] T.W. Chan, "A Software Tool in Java for Teaching CPU Scheduling," Journal of Computing Sciences in Colleges, **19**(4), 257–263, 2004, doi:10.5555/1050231.1050269.
[9] D. Skrien, "CPU Sim 3.1: A tool for simulating computer architectures for computer organization classes," Journal on Educational Resources in Computing (JERIC), **1**(4), 2001, doi:10.1145/514144.514731.
[10] D. Park, R.H. Saavedra, "Trojan: A high-performance simulator for shared memory architectures," in Proceedings of the 29th Annual Simulation Symposium. IEEE, 1996, doi:10.5555/829528.831210.
[11] A.R. Lopes, D.A. de Souza, J.R.. de Carvalho, W.O. de Silva, S.V.L. P., "SIME: Memory simulator for the teaching of operating systems," in Computers in Education (SIIE), 2012 International Symposium on. IEEE, 2012.
[12] P. Gayet, B. Bradu, "Procos: A real-time process simulator coupled to the control system," in Proceedings of ICALEPCS, Kobe, Japan: 794–796, 2009.
[13] A. Garmpis, "Design and development of a web-based interactive software tool for teaching operating systems," Journal of Information Technology Education, **10**(10), 1–17, 2011, doi:10.28945/1357.
[14] B. Mustafa, "Visualizing the modern operating system: simulation experiments supporting enhanced learning," in Proceedings of the 2011 Conference on Information Technology Education. ACM, 209–214, 2011, doi:10.1145/2047594.2047650.
[15] T. Golemanov, E. Golemanova, "A Teaching in Operating Systems Tool," in CompSysTech'06, 2006.
[16] E.W. Dijkstra, "Cooperating sequential processes (EWD-123)," Center for American History, University of Texas at Austin, 1965, doi:10.5555/1102034.
[17] M. Herlihy, "Wait-free synchronization," ACM Trans. Program. Lang. Syst. **13**(1), 1991, doi:10.1145/114005.102808.
[18] A.C. Arpaci-Dusseau, Remzi H.; Arpaci-Dusseau, Operating Systems: Three Easy Pieces, Arpaci-Dusseau Books, 2014.
[19] A. Silberschatz, P.B. Galvin, G. Gagne, Operating System Concepts. 9., John Wiley & Sons, Inc., 2013.
[20] P. Krzyzanowski, Process Scheduling: Who gets to run next?, 2015.
[21] D. Samanta, Classic Data Structures, Prentice Hall India Pvt, 2004.
[22] S. William., Computer organization and architecture : designing for performance (8th ed.), Prentice Hall, Upper Saddle River, NJ, 2010.
[23] M. Raynal, Concurrent Programming: Algorithms, Principles, and Foundations, Springer Science & Business Media, 2012.
[24] G.L. Peterson, "Myths About the Mutual Exclusion Problem," Information Processing Letters, **12(3)**, 115–116, 1981, doi:10.1016/0020-0190(81)90106-X.
[25] H.M. Deitel, P.J. Deitel, D.R. Choffnes, Operating systems, Pearson/Prentice Hall, 2004.
[26] C.A.R. Hoare, Communicating Sequential Processes, Prentice Hall International, 2004.
[27] J. Coffman, Edward G., M.J. Elphick, A. Shoshani, "System Deadlocks," ACM Computing Surveys, 1971, doi:10.1145/356586.356588.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 108-112 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Technology Adoption in Education: A Systematic Literature Review

Kayode Emmanuel Oyetade*, Tranos Zuva, Anneke Harmse

Vaal University of Technology, ICT, 1930, South Africa

ARTICLE INFO

Article history:
Received: 11 August, 2020
Accepted: 09 September, 2020
Online: 08 November, 2020

Keywords:
Technology Adoption
Systematic Literature Review
Education
M-Learning
E-Learning
COVID-19

ABSTRACT

Technology is advancing faster today than ever before with evidence of its impact in all facets of our lives. With the spread of the novel COVID-19 pandemic across the world, schools were closed as part of lockdown measures to contain the virus thereby disrupting academic curricula. Academic institutions leveraged ICTs to virtually engage students and teachers. Technology adoption will become a new reality for teaching and learning processes. However, the choice of adopting and not adopting this technology is based on an individual's decision on the benefits or risks in using this technology. The objective of this study is to find topical and relevant studies that have been conducted on technology adoption in education. A systematic literature review was adopted to classify and evaluate articles that fits pre-specified selection criteria using Google Scholar and IEEE databases. 132 papers were found to match the search criteria and filtered to 17 articles using applied exclusion criteria. Current research highlights the trends in technology adoption and provide empirical evidence of applications that have been used to implement technology in educational settings. This research also adds to new literature on COVID-19 in relation to its effect on academic curriculum across the world. Future research will investigate areas that can be expanded and improved on to leverage the benefits of ICTs to limit the adverse impact on the performance of teachers and students in the event of a disruption to academic activities.

1. Introduction

Technology adoption is the choice of an individual to acquire and use an innovation. Technology is advancing faster today than ever before with developments such as cloud computing, robotics, Internet of Things (IoT), Artificial Intelligence (AI), Machine Learning (ML), etc., opening fresh opportunities for value creation and contributing to our life experiences. There is a consensus in both industry and academia on the growing importance of technology as a robust means for survival in today's fast-paced and knowledge-driven world [1]. Evidence abound of its success and impact on teaching and learning via documented research studies [2-4].

Moreover, adopting or rejecting a new technology is based on an individual's perception of the benefits of using the tool to achieve a goal [5, 6]. Although technologies are advancing, the decision not to adopt has been categorised by poor innovation implementation, high cost, high failure rates, maintenance, poor technological planning, inadequate stakeholder participation [6-8]. An understanding of the factors affecting an individual's choice to adopt technology is essential both for stakeholders using these technologies as well as creators and producers of such technologies. Several theories and models exist to allow us understand the acceptance or rejection of a technology by an individual: This includes but not limited to Technology Adoption Model (TAM), Innovation Diffusion Theory (IDT), Unified Theory of Acceptance and Use of Technology (UTAUT), etc. [9].

The present research aims to fulfil the objective: To provide a summary of studies to determine the level of technology adoption in an academic setting. On this basis, a systematic literature review approach was employed to achieve this objective. Google Scholar and IEEE databases were searched using the keyword 'technology adoption and education or learning'. One hundred and thirty-two articles were initially obtained from the search. Inclusion and exclusion criteria were applied which reduced the list to 17 articles. These 17 articles were further analysed to fulfil the purpose of this research.

The next section presents the research method for this study. The result section provides the outcome of the literature review

*Corresponding Author: Kayode Emmanuel Oyetade, VUT, +27846514453 & kayoyetade@gmail.com

and the analysis of the selected literature. This is followed by the conclusion and possible area of future research.

2. Methodology

This section provides an overview of the research method used in the present research. The approach stipulated by [10] to conduct systematic literature was adopted.

2.1. Data sources and selection criteria

This paper reviews the adoption of various technology used in schools for teaching and learning. Therefore, the combination of the following keywords was applied: "Technology Adoption AND education OR learning". The title and abstract were checked for the occurrence of the keywords to ensure results are for articles focused on the study's research objective. English language was used to conduct this search to obtain a global perspective on the topic as well as avoid narrow searches. We limited our search to peer-reviewed texts available in online published journals and conference proceedings from Google Scholar and IEEE electronic databases.

The following inclusion criteria were deemed relevant for the present research:

- Studies on technology adoption application in education
- Studies that engage in any kind of technologies to support learning
- Studies written in English

The following exclusion criteria to remove articles not deemed fit for the current study:

- Studies that are not published as peer-reviewed journal articles and conference papers
- Studies not originally written in English language
- Studies whose full text is not available online and
- Studies discussing technology adoption but not with an educational focus.

3. Result

As presented in Figure 1, the search and refinement process followed six steps.

Step 1: Database search was conducted using the above-mentioned keywords. A total of 132 articles was realised. Step 2 focused on the application of the selection criteria by checking to remove duplicates due to the search using dual databases resulting in 130 articles. Step 3 reviewed the abstract to determine whether the article focused on the study's research objective and relevance to technology adoption. This allowed the study to remove 81 articles and left with 49 articles.

In step 4, full-text review was done on the 49 literatures for eligibility to provide a clear understanding of the research objectives. This allowed the study to exclude 8 articles which do not fit with reason resulting in 41 articles. Further qualitative review for quality assurance to ensure appropriateness and relevance to the research study was applied on these 41 articles. This allowed this study to arrive at a final list of 17 articles to be considered for review.

3.1. Review of the Selected Literature

One of the well-known success areas of technology adoption is the usage of m-learning in education. One of the motivations is the provision and easy access to learning content [11].

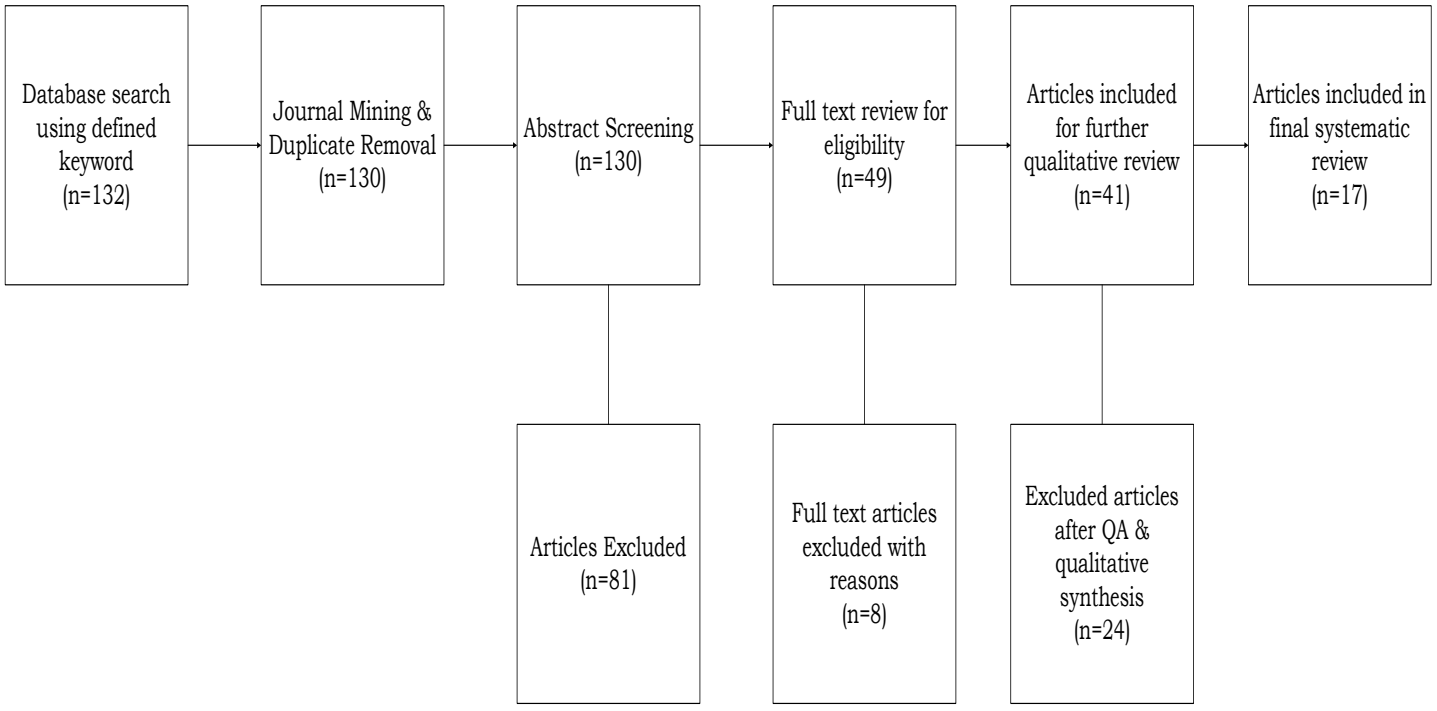


Figure 1: Steps to selection criteria and final outcome

A study by [12] used TAM to analyse factors that influence learner's technology adoption. They found that the factors economics, usefulness, suitability, ease of use, enjoyment, and social influence impact learners' adoption of m-learning to facilitate and promote learning. Likewise, [13] explored the capacity and acceptance of mobile instant messaging, an m-learning platform to increase student performance. They suggest that device portability, communication cost, collective capabilities, and learner control of the device are key factors of flexible learning in digital environments.

Evaluating the factors and effectiveness of e-learning adoption on student learning, [14] adopted the Interactivity Distance Education Web Learning (IDEWL), an enhanced online course creative and conferencing tool that integrates other add-on tools like online survey and e-blackboard for their study. There survey found e-learning effective with the enhancement of the interactivity modules in the students learning process.

On the other hand, [15] reported that the compulsory usage factor is essential for overall adoption of the e-learning system (ELS), either in a compulsory or voluntary setting. This is due to its effectiveness in promoting ELS as a core learning platform for students as well as discourage them from using other means to receive an education.

Similarly, [16] measured the factors that affect learners' adoption of blended ELS (BELS) and the perceived satisfaction of BELS according to their learning styles and gender differences. They report learning style did influence adoption of technology to provide flexible learning and improve understanding. In another related study, [17] examined the relationship between student's learning satisfaction and their use of video in a Moodle environment. He found that the use of video significantly impacted student's learning performance and satisfaction.

In answering student's willingness to learn using e-learning, [18] opines that designers of e-learning sites must ensure ease of navigating the web interface during the learning process. This designers must also ensure the quality of contents uploaded on the websites to enhance students learning while exploring college students learning attitudes with respect to their adoption of English e-tutoring websites. Using TAM as a theoretical framework for their study. There result showed that most users of English e-tutoring sites found online sources more effective and offer greater ease than traditional resources.

Furthermore, [19] conducted a study to examine key factors influencing the adoption of iPads using factors from extended TAM to design and deliver course content by academics. They found that University academics found iPads useful and easy to use with a high level of compatibility with their work.

Using the tablet devices in the classrooms, [20] reports that tablets impacts teaching and learning practices. This tablet devices provides a shift from the traditional way students learn to the media-rich and interactive environments. They therefore recommend stakeholders introduce educational and technical support to assist both educators' and learners' understanding of the benefit of these technologies in education.

Despite the growth and positive outcomes of technology adoption in education, [21] found that males and females do not differ in their belief about the usefulness of e-books. Usage of e-book in South African higher education corresponds to the effort and the external pressure exerted on students to promote the acceptance of e-books.

On mobile technology adoption for learning, [22] explored the correlation between students' ICT skills and factors that impact mobile technology adoption using the TAM framework. They found that basic and advanced ICT skills in mobile technology both played important roles in the intention to adopt m-learning.

Using mobile technology to enhance the convenience of teaching the English Language in Taiwan, [23] extended the factors of TAM to explain factors that affect the adoption of PDA. There result showed that extended TAM was effective in predicting and explaining the acceptance of English mobile learning.

To facilitate teaching and learning processes using the Learning Management Systems (LMSs) by University staffs, [24] used TAM extensions to identify factors that influence LMS usage behaviour and their underlying causal relationships. The analysis of data reveal that the proposed factors were significant predictors of faculty staff attitude towards LMSs. This confirms the validity of the extended TAM in determining users' technology acceptance.

In testing the effectiveness of a Blackboard (Bb) course management system (CMS), [25] used an experimental and control group to conduct his test on English writing instruction in and outside of the Bb environment. The experimental group achieved encouraging results than the control group thereby contributing to learning English writing using Bb. They also identified attitude, perceived usefulness, perceived ease of use, and technical support affect the adoption of CMS.

In [26], authors studied the effect of experience on students' intention to adopt webcast and found that students with previous experience in educational webcasts are most likely to adopt webcast. Ease of use and usefulness were found to be important factors in determining students' BI to use webcasts for learning purposes. They recommended that Webcast developers should pay attention to learners' intrinsic motivations and make them feel comfortable when using the environment for learning.

In conclusion, [27] extended the TAM model to explain the adoption and usage behaviour of social media users in higher education. It was found that subjective norm, perceived usefulness, perceived playfulness (happiness), and perceived ease of use strongly predict students' social media usage behaviour. Their findings suggest that social media use in classrooms should continue in higher institutions. This is because the intention to use translate to the actual use of this ICT tools.

The above review shows technology adoption has the potential to be used as an effective educational platform either directly or as an alternative tool for teaching and learning process.

4. Results of the Review

From the 17 articles examined, this section presents other findings from the above review and discusses the outcomes of the issues related to technology adoption as follows:

4.1. Publication History and Frequency

Figure 2 shows the history and frequency of the identified articles selected for review. We observe the number of articles remains reasonably constant (2006-2012) with a noticeable peak in 2015. The low volume of publications between 2016 and 2019 could be related to our exclusion criteria which may have led to the omission of recent literatures. However, the existence of up-to-date papers lends itself to continuing research efforts and trends of technology adoption in education.

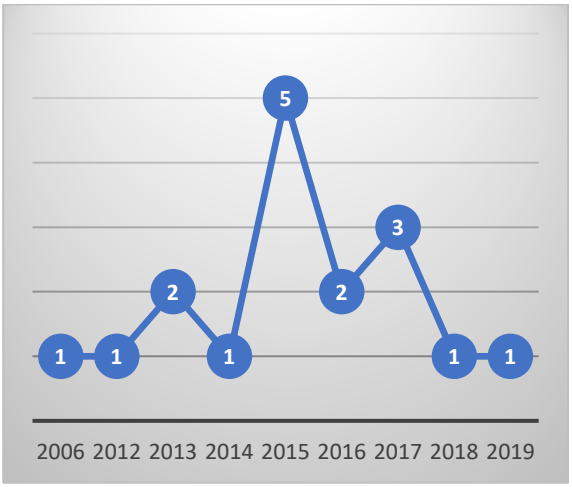


Figure 2: Publication History and Frequency

4.2. Publication Type

From the review of identified literature, Figure 3 presents the result of the articles reviewed by the publication type. It was observed that most of the articles (88%, 15 articles) were obtained from published journals and the remainder (12%, 2 articles) are conference proceedings. The different sources of articles for the current research eliminate possible publication bias.

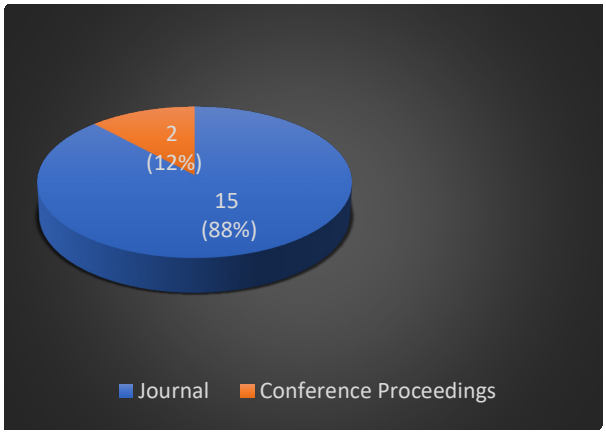


Figure 3: Publication Type

4.3. Word Cloud of technology adoption tools

As evidenced by the articles reviewed, Figure 4 shows the various technology adoption tools adopted for teaching and learning in education. M-learning and e-learning are mostly used technology. Other tools are classified into these two categories.

Figure 4: Word Cloud of Technology Adoption Tools

Even though, this research followed the steps for conducting a systematic review as specified by [10], it is not without limitations. For instance, the review results are limited to the search keywords used, and thus some studies may not have been included. However, we attempted to compensate for this by also relying on related recent literature reviews.

5. Future Research

Future research will investigate areas that can be expanded and improved on to increase the predictive performance of the Technology Adoption in education to support student's academic achievements.

6. Conclusion

The potential of technology has been identified in the above reviewed literature [14-17]. However, these technologies were blended with face to face interactions [25]. With the recent wave of the COVID-19 pandemic and the closure of schools, many schools were forced to seek alternative and virtual means of keeping the curriculum going using appropriate and connective technologies. This implies that technology adoption will gradually become a new normal in education as countries are gradually opening economic and academic activities that were affected by the COVID-19 pandemic.

Although a range of technologies is used in education, e-learning and m-learning are the major tools adopted by many institutions to enhance teaching and learning as identified in the reviewed literature. However, despite the usage of technology adoption as reported in this study, [28-29] reports present findings on the challenges faced with the adoption of technologies as related to COVID-19. This includes affordability and availability of connectivity to the internet to access academic resources while most educators lack the technical support needed to use this ICTs.

The current study implemented a systematic literature review approach to provide a general picture of the usage of technology in education. Google scholar and IEEE databases were searched for articles published in online journals or conference proceedings from 2006 to 2019. The keyword Technology Adoption AND

education OR learning was which resulted in an outcome of 132 articles. This was reduced to 17 articles for in-depth analysis using defined selection criteria. The researchers specifically conducted an analysis of contributions considering the research objectives that covered the year of publication, publication type, technology adoption platforms employed in the research study.

The findings highlight that technology is a relevant tool in schools to deliver the learning needs of 21st century students as learners prepare to return to schools with the lifting of barriers to academic activities. It is believed that the current COVID-19 challenge will change academic curricula and lead to the prevalent usage of technology in the future.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The present research has been prepared by Kayode Emmanuel under the supervision of Dr Anneke Harmse and Prof Tranos Zuva and supported by the Department of ICT at the Vaal University of Technology, Vaal, South Africa.

References

[1] A. Hidalgo, J. Albors, "Innovation management techniques and tools: a review from theory and practice," R&d Management **38**, 113-127, 2008, doi:10.1111/j.1467-9310.2008.00503.x.

[2] A. R. A. Arokiasamy, "A systematic review approach of mobile technology adoption in higher education,". Economics, Management Sustainability **2**, 48-55, 2017, doi:10.14254/jems.2017.2-2.5.

[3] R. Scherer, F. Siddiq, J. Tondeur, "The technology acceptance model (TAM): A meta-analytic structural equation modelling approach to explaining teachers' adoption of digital technology in education", Computers Education **128**, 13-35, 2019, doi:10.1016/j.compedu.2018.09.009.

[4] K. E. Oyetade, A. Harmse, T. Zuva, "Technology adoption factors in education: A review," in 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems, (IEEE, Durban, 2020, doi:10.1109/icABCD49160.2020.9183830.

[5] M. C. M. O. Nemoto, E. P. G. d. Vasconcellos, R. Nelson, "The adoption of new technology: conceptual model and application,". Journal of technology management innovation **5**, 95-107, 2010, doi: 10.4067/S0718-27242010000400008.

[6] C. Plewa, I. Troshani, A. Francis, G. Rampersad, "Technology adoption and performance impact in innovation domains," Industrial Management & Data Systems **112**, 748 – 765, 2012, doi:l0.1108/02635571211232316.

[7] E. D. Adamides, N. Karacapilidis, "Information technology support for the knowledge and social processes of innovation management," Technovation **26**, 50-59, 2006, doi:10.1016/j.technovation.2004.07.019.

[8] L. Dooley, D. O'Sullivan, "Developing a software infrastructure to support systemic innovation through effective management," Technovation **23**, 689-704, 2003, doi:10.1016/S0166-4972(02)00020-2.

[9] K. E. Oyetade, T. Zuva, A. Harmse, "A review of the determinant factors of technology adoption" in Computer Science On-line Conference (Springer) **1226**, 274-286, 2020, doi:10.1007/978-3-030-51974-2_26.

[10] A. Siddaway, "What is a systematic literature review and how do I do one" University of Stirling **1**, 1-13, 2014, Available online: https://pdfs.semanticscholar.org/2214/2c9cb17b4baab118767e497c93806d741461.pdf (accessed on 20 Sep. 20).

[11] S. Jeremy, K. E. Oyetade, K. G. Fomunyam, Operational Use of Mobile Learning: Understanding and Retention of Curriculum Content, Apple Academic Press, 2019, https://www.appleacademicpress.com/title.php?id=9781771887854.

[12] M. Sarrab, I. Al Shibli, N. Badursha, "An empirical study of factors driving the adoption of mobile learning in Omani higher education," International Review of Research in Open Distributed Learning, **17**, 331-349, 2016, doi: 10.19173/irrodl.v17i4.2614.

[13] A. Bere, P. Rambe, "An empirical analysis of the determinants of mobile instant messaging appropriation in university learning," Journal of Computing in Higher Education, **28**, 172-198, 2016. doi: 10.1007/s12528-016-9112-2.

[14] A. T. Baharin, H. Lateh, S. S. Nathan, H. mohd Nawawi, "Evaluating effectiveness of IDEWL using technology acceptance model," in 2015 Procedia-Social Behavioral Sciences (Elsevier), **171**, 897-904, 2015, doi: 10.1016/j.sbspro.2015.01.207.

[15] Y. C. Lee, "An empirical investigation into factors influencing the adoption of an e-learning system," Online information review, 2006, doi: 10.1108/14684520610706406.

[16] A. Al-Azawei, P. Parslow, K. Lundqvist, "Investigating the effect of learning styles in a blended e-learning system: An extension of the technology acceptance model (TAM)," Australasian Journal of Educational Technology **33**, 2017, doi: 10.14742/ajet.2758.

[17] J. T. Nagy, "Evaluation of online video usage and learning satisfaction: An extension of the technology acceptance model," International Review of Research in Open Distributed Learning, **19**, 2018, doi: 10.19173/irrodl.v19i1.2886.

[18] P. J. B. Tan, "An empirical study of how the learning attitudes of college students toward English e-tutoring websites affect site sustainability" Sustainability, **11**, 1748, 2019, doi: 10.3390/su11061748

[19] M. S. Lane, A. Stagg, "University staff adoption of iPads: An empirical study using an extended TAM model," Australasian Journal of Information Systems **18**, 2014, doi: 10.3127/ajis.v18i3.876.

[20] H. Montrieux, R. Vanderlinde, T. Schellens, L. De Marez, "Teaching and learning with mobile technology: A qualitative explorative study about the introduction of tablet devices in secondary education," PloS one, **10**, 2015, doi: 10.1371/journal.pone.0144008.

[21] D. K. Maduku," Factors of E-book use intentions: Perspectives of students in a developing country," Perspectives on Global Development and Technology, **14**, 597-618, doi:10.1163/15691497-12341364.

[22] K. Mac Callum, L. Jeffrey, "The influence of students' ICT skills and their adoption of mobile learning,". Australasian Journal of Educational Technology **29**, 2013, doi: 10.14742/ajet.298.

[23] C.-C. Chang, C.-F. Yan, J.-S. Tseng, "Perceived convenience in an extended technology acceptance model: Mobile technology and English learning for college students," Australasian Journal of Educational Technology **28**, 2012, doi: 10.14742/ajet.818.

[24] N. Fathema, D. Shannon, M. Ross, "Expanding the Technology Acceptance Model (TAM) to examine faculty use of Learning Management Systems (LMSs) in higher education institutions," Journal of Online Learning Teaching, **11**, 2015, https://jolt.merlot.org/Vol11no2/Fathema_0615.pdf.

[25] Y.-R. Tsai, "Applying the Technology Acceptance Model (TAM) to explore the effects of a course management system (CMS)-assisted EFL writing instruction," Calico Journal, **32**, 153-171, 2015, doi:10.1558/calico.v32i1.25961.

[26] M. N. Giannakos, P. Vlamos, "Educational webcasts' acceptance: Empirical examination and the role of experience," British Journal of Educational Technology, **44**, 125-143, 2013, doi:10.1111/j.1467-8535.2011.01279.x.

[27] D. Z. Dumpit, C. J. Fernandez, "Analysis of the use of social media in Higher Education Institutions (HEIs) using the Technology Acceptance Model," International Journal of Educational Technology in Higher Education, **14**, 5, 2017.

[28] T. E. Africa, "The Effect of Covid-19 on Education in Africa and its Implication for the Use of Technology," 1-81, 2020, doi:10.5281/zenodo.4018774.

[29] E. J. Sintema, "Effect of COVID-19 on the Performance of Grade 12 Students: Implications for STEM Education," EURASIA Journal of Mathematics, Science and Technology Education, **16**, 1-6, 2020, doi: 10.29333/ejmste/7893.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 113-117 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Understanding the usage, Modifications, Limitations and Criticisms of Technology Acceptance Model (TAM)

William Ratjeana Malatji*, Rene Van Eck, Tranos Zuva

School of Applied and Computer Science, Department of ICT, Vaal University of Technology, Vanderbijlpark, 1911, South Africa

A R T I C L E I N F O

Article history:
Received: 23 July, 2020
Accepted: 14 August, 2020
Online: 08 November, 2020

Keywords:
TAM
User acceptance
Perceived usefulness
Perceived ease of use
Behavioural intentions
Criticisms
Limitations

A B S T R A C T

Technology Acceptance Model (TAM) is broadly accepted and has proved applicable in identifying consumers' willingness to utilize information and communication technology (ICT). The theory proposes that Perceived Ease of Use (PEOU) and Perceived Usefulness (PU) are actually determining factors of individual attitudes, while attitude is a determining factor of Behavioural Intention (BI) and Behavioural Intention (BI) influences usage. It is necessary to understand the usage and modifications that have been made to Technology Acceptance Model (TAM) since user acceptance and confidence are of great importance for advance improvement and successful implementation of any new technology. Numerous frameworks and models have been designed and created to describe user acceptance of modern innovations. These models introduced factors that contribute towards user acceptance. This is a review paper on understanding the usage, modifications, limitations, and criticisms of Technology Acceptance Model (TAM). The model has been utilized to measure new different technologies for usage as well as adoption. The literature indicates that the modification of the model was mostly the addition or removal of variables and in some cases the addition of moderators or mediators. The model has limitations identified in literature such as the problem of reliably quantifying behaviour in an observed investigation. Moreover, there are notable criticisms identified in literature such as TAM's incapability of noticing other issues, for example, cost and structural imperatives that pushes users to adopt an innovation. TAM will continue to be accepted and modified according to the successful application of any new technology. This study can be used to enhance users' knowledge of usage, modifications, limitations, and criticisms of Technology Acceptance Model (TAM).

1. Introduction

Academics and practitioners have developed an interest in recognizing the factors that affect users' acceptance or denial of modern innovation [1]-[5]. Providing an answer to this question might assist them to improve techniques for assessing, predicting, and designing the feedback of customers to the new innovations [6]. Technology Acceptance Model (TAM) and frameworks were implemented in an expansive range of domains to comprehend and foretell users' behaviour for instance fasting, education, elections, etc. [7].

Numerous researchers in the discipline of technology acceptance, improved models, and frameworks to evaluate the utilization of modern innovations. These models introduced factors that contribute towards user acceptance [4]. TAM has been repeatedly studied and being extended to the two important upgrades named the TAM2 [8] and the Unified Theory of Acceptance and Use of Technology (UTAUT) [9]. TAM3 which includes the effects of use and perceived risk on system utilization in the circumstance of e-commerce was proposed by [10]. Numerous studies have utilized these traditional systems to carry out their studies and the remaining joint past models [11]. Great innovation and implementation may be planned and improved but if human beings do not partake and utilize it, the project is unsuccessful, therefore, user acceptance is not a deniable key of any continuous application and improvement of any innovation and implementation [4].

This paper aims to determine and understand usage and modifications, limitations, and criticisms of Technology

* Corresponding Author: William Ratjeana Malatji, villywr@gmail.com

Acceptance Model (TAM) for IT acceptance. User acceptance and confidence are of great importance for further improvement and to the successful application of any new innovation. Due to above-mentioned reason, it is necessary to determine and understand the usage and modifications that were made to Technology Acceptance Model (TAM). Therefore this paper is arranged as follows; Background, TAM criticisms and limitations, and conclusion.

2. Background

Growing the interest of customers' response toward Information Technology (IT) increased the significance of theories and models that foretell and describe information technology acceptance and utilization [4]. Researchers accepted and used theories and models such as Theory of Reasoned Action (TRA), Social Cognitive theory (SCT), Technology Acceptance Model (TAM), Theory of Interpersonal Behaviour (TIB) etc. to evaluate the usage of new innovations [4]. Several studies utilized these original frameworks to carry out their researches; however, in this paper TAM is the main focus.

2.1. Technology Acceptance Model

Technology Acceptance Model is one of the greatest leading extensions of [12]'s Theory of Reasoned Action (TRA) in the related works. In the 1980s TAM was created, given that employees were not utilizing Information Technologies made accessible to them [13]. Its founders highlighted that the way to multiplying the use was to initially multiply acceptance of information technology, which might be evaluated by questioning persons regarding their future intentions to utilize information technology [14]. Recognizing elements that formed a person's intentions might enable organisations to modify those factors so that they can improve acceptance and therefore multiply IT utilization [14].

An earlier TAM study found that three variables were required to describe, foretell and manage acceptance [15]. To reach this model, the founders adopted TRA, a universal social-psychological or behavioural theory which was confirmed suitable for comprehending different behaviours for example; exercise, condom use and voting [15]. As it was usual to adapt such theory to new circumstances, an introductory study was carried out to discover what could be the proper variables to add on comprehending IT utilization behaviour [12]. The variables chosen and created the first model of TAM are described in Table 1.

According to [16] TAM substitutes numerous attitude measures of the TRA with multiple technology acceptance measures namely, ease of use and usefulness. Moreover, TAM and TRA, which both contain powerful behavioural components, presume that when an individual develops a willingness to take action, that individual will act freely without any limitation. It is necessary to expand TAM to incorporate variables that are responsible for changing methods and that this can be attained by adopting the innovation Model into TAM [17].

2.2. Usage and Modifications of TAM

Numerous researchers have used TAM to carry out their studies. TAM was used by [18] to investigate the learning effectiveness of utilizing IT as a learning podium. The findings of the study highlighted that PEOU, PU, and perceived playfulness might all be utilized to efficiently foretell the learning behavioural intention of scholars (the variables are explained in detail in Table 1). The theory of planned behaviour and the expectation-confirmation model was used for two hundred and seven mobile information system users to investigate their behavioural intention to continue utilizing the system [19]. The findings of the study highlighted that consumer approval, PU, and perceived playfulness remained the main components for consumers to carry on utilizing the system.

TAM was used by [19] to examine Chinese clients' behavioural intention about instant messenger. The conclusion was reached that PU and perceived playfulness directly affect clients' attitudes. Using the theory of planned behaviour, it was found that social norm and perceived behavioural control might as well have a direct influence on BI. Moreover, [20] carried out a study on the utilization of emails and file processing software by one hundred and twenty workers in IBM Canada Laboratory. They discovered that workers' PU, PEOU, and software utilization were positively and significantly correlated.

TAM was used by [21] to study the behavioural intention (BI) of scholars in the higher institutions of learning and utilize a web-based application evaluation framework as an instrument to assess their specialised capability in e-book creation. The study discovered that participants' computer self-efficacy positively influence PEOU and PU. Participants' PU and PEOU also had considerable and direct influences on their intention to utilize the framework. Thus, when a person possesses greater computer self-efficacy with a greater point of view of the usefulness and ease of use of the web-based valuation framework, that individual will be more willing to utilize the framework.

TAM theory went through numerous changes. For example, an upgraded named TAM2 erased attitude (ATT) variable from the model, which initially arbitrated some of the effects of PU and PEOU. TAM2 also included a variable meant to capture the social influence that forces end-users to positively assess and welcome information technology, named subjective norm (SN) [9]. The connection amongst variables which were hypothesised from the theory that proposes and motivates the combination of TAM and DIT are:

- Social influence negatively impacts complexity of understanding.
- Social influence positively impact triability.

Lately, an impressive attempt to merge the information technology acceptance writings brought about the UTAUT (defined in Table 2), a theory with clear similarity to TAM [9]. UTAUT includes PU into a performance expectancy construct, PEOU into effort expectancy and SN as social influence (the terms are defined and explained later in Table 1). Modern to UTAUT, but not to the information technology acceptance study generally is the modelling of easing circumstances as one determining factor of behavioural intention [22]. UTAUT is a modern but capable theory; Prior experiments of UTAUT described a magnificent 70 percent of the variance in behavioural intention and at most 50 percent in actual use [22]. It has been embraced by current studies in healthcare, this include:

- The Modern Portfolio Theory (MPT): Independent of TAM, [23] designed the technology model and matching human being as part of national science foundation-funded dissertation research. The modern portfolio theory is fully explained in Scherer (2005)'s text.
- The Hedonic-Motivation System Adoption Model (HMSAM): TAM has been successful in describing multiple systems utilization (i.e. E-Learning, web portals, learning management system, etc.) [24, 25]. Nevertheless, TAM is not preferably relevant to describe the adoption of completely inherent or hedonic frameworks (e.g. learning for pleasure, music, online games etc.). HMSAM was developed to better the understanding of the Hedonic-Motivation System Adoption Model [26].
- Extended TAM: Numerous studies suggested an expansion of original TAM by including variables that are external on it with an intention of investigating the outcomes of factors that are external on user attitude, actual use and behavioural intention of an innovation [24, 25]. This model became implemented in acceptance of health care innovations [27].

Technology Acceptance Model and its improved versions are not appropriate for all applications since they exclude main constructs such as perceived risk[28]. Therefore TAM has also turned out to be an accepted model that can be expanded and adjusted [28, 29].

The table below lists and defines the variables in TAM and related models since TAM theory has undergone numerous changes.

Table 1: TAM variables and related models

Variable	Definition	Models that include the variable
Behaviour Use (BU).	One particular behaviour of interest made by persons with great consideration to a certain information system.	TRA/TPB, TAM, TAM2, UTAUT.
Behavioural Intention (BI)	A person's inspiration or intention to apply an effort to carry out the objective behaviour.	TRA/TPB, TAM2, TAM, UTAUT
Attitude (ATT)	A person's evaluative decision of the objective behaviour on a certain dimension.	TAM, TRA/TPB
Perceived Ease of Use (PEOU)	A being's point of view that utilizing an IT framework will be effortless	TAM, TAM2
Perceived Usefulness (PU)	A being's point of view that utilizing IT framework will improve job performance	TAM, TAM2
Subjective Norm (SN)	A person's point of view of the degree to which significant people favour or are not in favour of the objective behaviour	TAM2,TRA /TPB
Perceived Behavioural Control (PBC)	A being's point of view regarding how simple or hard it might be to make the objective the behaviour of elements that hinder or ease the behaviour, or of the amount of control that a person has above performing the behaviour	TPB
Effort Expectancy	(Refer PEOU)	UTAUT
Performance Expectancy	(Refer PU)	UTAUT
Social Influence	(Refer SN)	UTAUT
Facilitating Conditions	(Refer PBC)	UTAUT
Behavioural beliefs, normative believes and control beliefs	A person's point of view regarding particular positive or negative results of applying objective behaviour, particular groups or individuals who promote or depress the behaviour and particular factors or situations simplify behaviour or make it more complex	TRA/TPB

2.3. Summary of TAM and Modified versions

Technology Acceptance Model has been modified and the table below lists the summary of TAM and modified versions together with their aims.

Table 2: Summary of TAM and modified versions aims

TAM and modified versions	Model Aim	Source
TAM	TAM was created to understand the natural sequence joining external variables to its user acceptance and natural utilization in an establishment. TAM helps in comprehending the background of perceived ease of use.	[30]
TAM2	TAM2 intends to comprehend perceived usefulness in order to create working place involvements that would increase user acceptance of new framework.	[9]
UTAUT	UTAUT shapes a unified model as a suitable instrument for managers requiring to measure the probability of success for modern technology outlines.	[9]
TAM3	TAM3 offers a combined model with significance assigned on perceived usefulness and perceived ease of use so that it can address how managers and people who make decisions can make informed decisions regarding interventions	[10]

3. TAM criticisms and limitations

TAM has been cited by numerous researchers due to its simplicity without taking into consideration, an actual application in their researches. In this section, some of the TAM criticisms and limitations are presented considering, utilization of the model technology-related studies in various disciplines.

3.1. TAM criticisms

TAM has been broadly criticized, regardless of how it is being frequently used, leading the founders to try to re-explain it

numerously. TAM criticisms as a "theory" incorporate contentious heuristic value, restricted descriptive and foretelling capability, criticism triviality and short of any practical value [31]. TAM changed researchers' attention by moving it away from other significant research matters and it has caused the impression of progression in accumulating knowledge[31]. Moreover, various researchers have independently tried to extend TAM so that it can adapt to the continually changing information technology environments which resulted in a state of theoretical havoc and uncertainty [31].

Generally, TAM is based on a being's utilization of a computer, with an abstract idea of PU and disregarding essential social methods of Information Systems (IS) development and application, with no question that "where many technologies are actually best and social after-effects of using Information Systems (IS)" [31]. The system of PU and PEOU fail to notice other issues, for example, cost and structural imperatives that push users to adopt a new innovation [32].

According to [33] both TAM and TAM2 take responsibility for at most 40% of a technological framework's utilization. Some researchers gone through the limitations of TAM implementation under organisation setting into details, and indicated that better predictive capability can be reached even with the easy, simply applicable model when exact first screening methods are implemented [34].

Studies conducted by [35-37] found that PEOU is unlikely to become a determining factor of ATT and intention to use. Moreover, [34] reported the same results when researching on the adoption of Blockchain technology.

The study carried out by [38] revealed that PEOU has no impact on the adoption of multimedia online innovations for Malaysian small and medium-sized enterprises (SMEs). Similar findings were outlined by Hong Kong when evaluating small and medium-size enterprises (SMEs) [34].

3.2. TAM limitations

Several conditions must be taken into account so that a theoretical framework can be applied and researchers must be aware of the numerous limitations that exist. To better comprehend the factors that encourage improved utilization of information technology (IT), it is compulsory to have a broad theoretical and application knowledge of structures and frameworks through which IT utilization is investigated [39]. One of the restrictions of TAM relates to the variable that refer to the behaviour of clients, which is necessarily assessed over subjective measures like behavioural intention and interpersonal influence [39]. However, social influence as the subjective norm refers to when an individual is influenced by "word of mouth" from a fellow worker [39].

The second limitation of TAM is that underlines of behaviour is unable to be quantified reliably in an observed research, owing to various multiple subjective elements, for instance, values and norms of societies, individual features and personality attributes [40, 41]. Consequently, a relative, friend argument that could influence technology use through demanding social influence is extremely fabricated [40, 41] . Even though it might be correct in theory or for individual utilization of innovation, the concept might not be credible or precise in a working environment.

4. Conclusion

In the research area, many models and frameworks were proposed by researchers to explore the user acceptance behaviour on information technology and information systems. In the 1980s TAM was created, given that employees were not utilizing Information Technologies (ITs) made accessible to them. Numerous researchers in the discipline of technology acceptance, improved theories and frameworks to evaluate the utilization of recent innovations and these frameworks introduced factors that can influence user acceptance. This led TAM theory to go through numerous modifications. Writers and researchers have criticized the TAM model in terms of its variables. This study aims to determine and understand the usage and modifications of Technology Acceptance Model (TAM) for IT acceptance.

Acknowledgement

This review paper has been prepared by William Ratjeana Malatji under the supervision of Dr Van Eck and Prof Zuva and sustained by the Department of ICT at the Vaal University of Technology Johannesburg (Vanderbijlpark) in South Africa.

References

[1] H. Taherdoost, S. Sahibuddin, N. Jalaliyoon, "Smart Card Technology: Awareness and Satisfaction " Journal of Computing, **4**(6), 128-132, 2012.
[2] H. Taherdoost, N. Jalaliyoon, M. Namayandeh, A. Forghani, M. Zamani, "Adoption Framework Expansion based on the Computer Ethics' Related Research Models and Ethical Scenarios Analysis," International Conference on Economics, Business and Management, **2**(5), 219-223, 2010.
[3] H. Taherdoost, S. Sahibuddin, "How Security Issues Can Influence on Usafe of Electronic Service," Advanced in Information Science and Computer Engineering, 310-316, 2015.
[4] H. Taherdoost, G. Hamta, G. Ahoora, R. Club, A. Hamta, C. Tablokar, "Importance of Technology Acceptance Assessment for Successful Implementation and Development of New Technologies," Global Journal of Engineering Sciences, **1**(3), 2019.
[5] H. Taherdoost, M. Masrom, "An Examination of Smart Card Technology Acceptance Using Adoption Model," ITI 2009 31st Conf. on Information Technology Interfaces, 329-334, 2009.
[6] A. Dillon, M. Morris, "User Acceptance of Information Technology: Theories and Models," Annual Review of Information Science and Technology **31**(4), 3-32, 1996.
[7] H. Taherdoost, S. Sahibuddin, N. Jalaliyoon, "Smart Card Security," *T*echnology and Adoption International Journal of Security, **5**(2), 2011.
[8] V. Venkatesh, F.D. Davis, "A theoretical extension of technology acceptance model: four longitudinal field studies," Management Science, **46**(2), 186-204, 2000.
[9] V. Venkatesh, M. G. Morris, G.B. Davis, F.D. Davis, "User acceptance of information technology: Toward unified view. ," MIS Quaterly, **27**(3), 425-478, 2003.
[10] V. Venkatesh, H. Bala, "Technology acceptance model 3 and research agenda on interventions," Decision Sciences, **39**(2), 273-315, 2008.
[11] H. Taherdoost, "A Review paper of Technology Acceptance and Adoption Models and Theories," Procedia Manufacturing, **22**(1), 960-967, 2018.
[12] I. Ajzen, M. Fishbein, "Understanding attitude and predicting social behaviour.," Engelwood Cliffs. NJ: Prentice- Hall, 1980.
[13] F. D. Davis, "Perceived usefulness, perceived ease of use and user acceptance of information technology," MIS Quaterly, **13**(1), 319-39, 1989.
[14] R. J. Holden, B.A. Karsh, "theoretical model of health information technology usage behaviour with implications for patient safety," Behav Inf Technolo, **28**(1), 21-38, 2009.
[15] M. Fishbein, I. Ajzen, "Belief, attitude, intention, and behavior: An introduction to theory and research," Reading, MA: Addision-Wesley, 1975.
[16] R. P. Bagozzi, F.D. Davis, P.R. Warshaw, "Development and test of a theory of technological learning and usage," Human Relations, **45**(7), 660-686, 1992.

[17] P. Legris, J. Ingham, P. Collerette, "Why do people use information technology? A critical review of the technology acceptance model," Information & Management, **40**(3), 191– 204, 2003.
[18] M. J. Sanchez-Franco, "WebCT-The quasimoderating effect of perceived affective quality on an extending Technology Acceptance Model," Computers & Education, **54**(1), 37-46, 2010, doi.org/10.1016/j.compedu.2009.07.005.
[19] B. Kim, " An empirical investigation of mobile data service continuance: Incorporating the theory of planned behavior into the expectation–confirmation model.," Expert Systems with Applications, **37**(10), 7033-7039, 2010, doi.org/10.1016/j.eswa.2010.03.015.
[20] F. D. Davis, R.P. Bagozzi, P.R. Warshaw, " User acceptance of computer technology: A comparison of two theorical models. ," Management Science, **35**(8), 982-1003, 1989, doi.org/10.1287/mnsc.35.8.982
[21] S. Liao, J.C. Hong, M.H. Wen, Y.C. Pan, Y.N. Wu, "Applying Technology Acceptance Model (TAM) to explore Users' Behavioral Intention to Adopt a Performance Assessment System for E-book Production " EURASIA Journal of Mathematics, **14**(10), 2018, doi.org/10.29333/ejmste/9357.
[22] S. Taylor, P. Todd, "Assessing IT usage: a test of competing models," Inform Syst Res, **6**(1), 1995.
[23] M. J. Scherer, "Living in the state of stucK, ", Fourth Edition ed. Cambridge MA: Brookline Books, 2005.
[24] N. Fathema, K. Sutton, "Factors influencing faculty members' Learning Management Systems adoption behavior: An analysis using the Technology Acceptance Model," International Journal of Trends in Economics Management & Technology, **2**(1), 20-28, 2013.
[25] N. Fathema, D. Shannon, M. Ross, "Expanding the Technology Acceptance Model (TAM) to examine faculty use of Learning Management Systems (LMS)," Journal of Online Learning and Teaching, **11**(2), 210-233, 2015.
[26] B. P. Lowry, J. Gaskin, N.W. Twyman, T.L. Roberts, Bryan, "Taking fun and games seriously: Proposing the hedonicmotivation system adoption model (HMSAM)," Journal of the Association for Information Systems, **14**(11), 617-671, 2013.
[27] H. Nadri, B. Rahimi, N.A. Loft, S. Mahnaz, A. Garavand, S. Hadi, "Factors Affecting Acceptance of Hospital Information Systems Based on Etended Technology Acceptance Model. A case Study in Three Paraclinical Departments," Applied Clinical Informatics, **9**(2), 238-247, 2018.
[28] J. Wu, " Developing an explorative model for SaaS adoption," *Expert* Systems with Applications, **38**(1), 15057-15064, 2011.
[29] C. López-Nicolás, F.J. Molina-Castillo, H., Bouwman, "An assessment of advanced mobile services acceptance: Contributions from TAM and diffusion theory models," Information & Management, **5**(1), 2008.
[30] V. Venkatesh, F.D. Davis, " A model of the antecendents of perceived ease of use: Development and test," Decision Sciences, **27**(3), 451-481, 1989.
[31] I. Benbasat, H. Barki, "Quo vadis, TAM?," Journal of the Association for Information Systems, **8**(4),211, 2007.
[32] B. Lunceford, "Reconsidering Technology Adoption and Resistance: Observations of a Semi-Luddite," Eplorations in Media Ecology, **8**(1), 29-47, 2009.
[33] Legris, "Assessing Tam Model," Journal of Computing, 191, 2003.
[34] J. Li, "Blockchain technology adoption: Examining the Fundamental Drivers," in Proceedings of the 2nd International Conference on Management Science and Industrial Engineering, 253-260, 2020.
[35] P. J. Hu, P. K. Chau, O.R. Sheng, "Examining the tehnoogy acceptance model using physician acceptance of telemedicine technology," Journal of Management Information Systems, **16**(2), 91-112, 1999.
[36] J. Wu, S. Wang, "What drives mobile commerce?: An empirical evaluation of the revised technology acceptance model," Information & Management, vol. **42**(5), 719-729, 2005.
[37] T. P. Pikkarainen, K.H. Karjaluoto, "Consumer acceptance of online banking: An extension of the Technology Acceptance Model," Internet ResearchElectronic Networking Applications and Policy, **14**(3), 224-235, 2004.
[38] D. J. Okafor, M. Nico, B.B. Azman, "The influence of perceived ease of use and perceived usefulness on the intention to use suggested online advertising workflow," Canadian Journal of Science ang Technology, **6**(14), 2016.
[39] L. M. Maruping, H. Bala, V. Venkatesh, S.A. Brown, "Going beyond intention: Intergrating behavioral intergration into thr unified theory of acceptance and use of technology," Journal of the Association for Information Science and Technology, **68**(3), 2016.
[40] M. C. Ang, T. Ramayah, H. Amin, "A theory of planned behavior perspective on hiring Malaysians with disabilities. Equality, Diversity and Inclusion," An International Journal, **34**(3), 186-200, 2015.
[41] Y. Shan, K.W. King,"The effects of interpersonal tie strength and subjective norms on consumers' brand-related eWOM referral intentions," Journal of Interactive Advertising, **15**(1), 16-27, 2015.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 118-126 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

The Impact Assessment of the Errors in Determining the Mass and Zero Lift-Drag Coefficient on the Aircraft's Performance Data

Klyagin Viktor Anatolievich, Laushin Dmitry Andreevich*

Department Aircraft Design and Certification, Moscow Aviation Institute (National Research University), Moscow, 125993, Russia

ARTICLE INFO

Article history:
Received: 30 July, 2020
Accepted: 01 October, 2020
Online: 08 November, 2020

Keywords:
Uncertainties
Risks
Flying qualities and performances
Influence coefficients

ABSTRACT

The adoption of technical decisions in the development of a new aircraft is carried out under the condition of insufficient information on the mass - dimensional and aerodynamic characteristics that will be achieved during the project. Insufficient information on the input parameters (weight and aerodynamic characteristics) leads to uncertainties in the values of the flying qualities and performances. When creating a new aircraft, the chief designer and specialists of the design bureaus establish certain reserves (limits) for mass and aerodynamic characteristics. The allocation of reserves for the characteristics of aircraft elements is carried out, as a rule, on the basis of experience gained from previous projects, and to a certain extent subjective. At the same time, it is not possible to quantify the risk of non-fulfillment of the tactical and technical task in terms of performance characteristics. Knowing the contribution of the uncertainties of the input parameters (for the values of which a margin is allocated) to the uncertainty of the final values of the flight characteristics will allow the person who makes the decision to reasonably establish reserves in the design of the aircraft and take a more reasonable risk. In this paper, the authors describe an approach that allows for a quantitative assessment of the impact assessment of the errors in determining the aircraft's characteristics on the tactical and technical task in terms of the aircraft's performance data. Using the example of the mass and zero lift-drag coefficient, analytical dependencies are derived that allow a quantitative assessment of the effect. To verify the obtained analytical dependencies, the calculation of weight coefficients was made for several aviation complexes of operational-tactical aviation. The analysis of the sensitivity of the aircraft's performance data to the parameters under consideration is carried out.

1. Introduction

When developing new aircraft, the most important decisions that determine the future fate of the project are made at the initial stages of work in conditions of uncertainty about the impact of these decisions on the performance of the – tactical and technical task (TTT). Mistakes in the early stages of design tend to cost more money and time for their elimination at the subsequent stages of the aircraft, and sometimes on the correctness of decisions depends on the feasibility of the project as a whole.

The uncertainty of the performance of the tactical and technical assignment in terms of flight characteristics at the stage of internal and external design is associated with many factors. For example, a delay in the receipt from related enterprises of reliable data on the characteristics of aircraft systems (such as engine, target equipment, etc.), insufficient knowledge of some physical laws, which is most pronounced when new technical solutions are introduced, and others.

When predicting the level of performance characteristics of a new aircraft, they rely on indicators of its technical perfection, which will be achieved as a result of the project. These include the relative mass of the structure and systems, specific engine thrust, specific fuel consumption, aerodynamic quality, and others. With the development of aviation, these indicators are constantly improving.

The choice of a reasonable level of technical excellence, and therefore TTT in the design of a new aircraft, is an important task. Excessive overestimation of the requirements for its weight and

*Corresponding Author: Laushin Dmitry Andreevich, 125993, Moscow, Volokolamskoe shosse, 4, Moscow Aviation Institute (National Research University), Email: laushin.d@yandex.ru; maksmai33@gmail.com

aerodynamic characteristics can lead not only to non-fulfillment of the promising requirements for flight performance, but also deterioration of the flight characteristics of the designed aircraft relative to the current level already achieved due to the mismatch of the characteristics of its elements. At the same time, the use of previously developed technical solutions and technologies, in which there is no risk, does not make it possible to ensure the required increase in flight performance and, as a consequence, efficiency in the development of new generation aircraft. Therefore, at the initial design stage, when applying certain technical solutions, it is necessary to have a clear idea of the real possibilities of performing the performance characteristics specified in the TTT.

In accordance with the existing design methodology, when creating a new aircraft, the chief designer and specialists of the design bureaus establish certain reserves (limits) for mass and aerodynamic characteristics. The reason for the allocation of such reserves is that at the initial design stage there is no exact idea of the final value of mass and aerodynamic characteristics, which will be obtained upon completion of the development stage. Each aircraft has its own targets. Therefore, the characteristic for which the reserve is allocated depends on the specific type of aircraft. For example, the characteristics of a commercially available engine are known values. If the power plant is at the development stage, there is a risk of non-implementation of the TTT of the engine, which may affect the non-implementation of the TTT of the aircraft. The use of insufficiently mastered materials or technologies in the aircraft structure can lead to a significant deviation in the mass characteristics of the airframe units, etc. Therefore, the identification of the main elements that introduce uncertainty in the level of aircraft performance should be made in each case specifically.

The allocation of reserves for the characteristics of aircraft elements is carried out, as a rule, on the basis of experience gained from previous projects, and to a certain extent subjective. At the same time, it is not possible to quantify the risk of non-fulfillment of TTT in terms of performance characteristics. Knowing the contribution of the uncertainties of the input parameters (for the values of which a margin is allocated) to the uncertainty of the final values of the flight characteristics will allow the person who makes the decision to reasonably establish reserves in the design of the aircraft and take a more reasonable risk.

Approaches to the formation of the appearance of an aircraft with inaccurate determination of the initial data are reflected in many scientific works, for example [1-10]. The relevance of this issue is also confirmed by scientific works [11-15].

To assess the technical solutions adopted in the design, it is necessary to associate particular changes in various parameters (weight, aerodynamic characteristics) with a change in the aircraft flight characteristics.

2. Materials and Methods

Let's introduce the concept of the influence coefficient (IC) – a dimensionless coefficient showing the ratio of the change in function (f) to the change in input parameters (x_i) in the relative form:

$$K_{x_i\_f} = \frac{\Delta \hat{f}}{\Delta \hat{x}_i} \quad (1)$$

where $\hat{f} = \frac{\Delta f}{f_1}$, $\Delta \hat{x}_i = \frac{\Delta x_i}{x_{i_1}}$, Δf and Δx_i – uncertainty of the function and the input parameters, f_1 and x_{i_1} - expected value of the function and the input parameters.

In the absence of a relationship between the initial parameters, ICs are the coefficients of linear decomposition of the inaccuracy of determining the function from the errors in determining the parameters (in the relative form):

$$\frac{\Delta f}{f_{expect}} = \sum_{i=1}^{n} K_{x_i\_f} \frac{\Delta x_i}{x_{i_{expect}}} \quad (2)$$

The numerical value of the IC is equal to the percentage change of flying qualities and performances (FQP) when the parameter changes to 1%.

This paper assesses the influence of such parameters as mass (m) and zero lift-drag coefficient (Cx_0) on:
- maximum flying speed - V_{max};
- service ceiling - H_{max};
- specific excess power (max) - V_y^*;
- sustained load factor - n_y;
- acceleration time - t_p.

For a comprehensive study of the considered IC, the research is carried out in two directions:
1. Analytically, by deriving IC from the formula dependencies of the FQP;
2. Numerical calculations of the IC by calculating the changes in FQP with varying mass and zero lift-drag coefficient.

The analytical method makes it possible to establish the dependence of IC on specific aircraft parameters. These dependencies are derived from the flight dynamics formulas and can be applied to absolutely any aircraft with a similar flight principle. However, to represent the dependencies in a relatively simple and easy-to-use form, it is necessary to introduce some assumptions that can be used to perform the required mathematical transformations. The formulas obtained require verification, since one and the same introduced assumption can give different errors in magnitude, depending on the type of aircraft and its purpose.

The verification of the derived formulas for IC can be carried out by the calculation method. The computational method, in contrast to the analytical one, allows one to obtain the exact value of IC (without assumptions), but does not allow one to establish the dependence of IC on the aircraft parameters. The IC is calculated using the FQP calculation program detailed in section 2.

Thus, comparison of the analytical findings with numerical calculations allows to prove the accuracy of the first.

3. Calculation of the influence coefficient by analytically method

For various parameters, the degree of their uncertainty can differ significantly depending on the level of elaboration of certain elements of the aircraft. The accuracy of calculating the mass of

aircraft at the stage of the preliminary project is 6-10%, outline design 3-5%, working draft 1-2% [5].

The influence coefficient, as described above, is indicated $K_{x_{i\_f}}$. According to the IC mass for sustained load factor is indicated $K_{m\_n_y}$, IC C_{x_0} to specific excess power - $K_{C_{x_0}\_V_y^*}$, etc. We use the following notation:

$\Delta\widehat{m} = \frac{m_2 - m_1}{m_1}$ $\Delta\widehat{Cx}_0 = \frac{Cx_{0_2} - Cx_{0_1}}{Cx_{0_1}}$ – relative changes in mass and zero lift-drag coefficient.

3.1. The influence of parameters on maximum flying speed

The maximum flying speed can be expressed from the equality of thrust and drag:

$$V_{max} = \sqrt{\frac{2P_{max}}{(C_{x_0} + C_{x_i})\rho S}} \quad (3)$$

where C_{x_i} – inductive reactance coefficient, P_{max} – thrust at which maximum speed is reached, S – wing area, ρ – air density.

The inductive component of resistance when flying at maximum speed is small. In this case $C_{xi} \approx 0$ and the maximum flight speed can be calculated by the formula:

$$V_{max} = \sqrt{\frac{2P_{max}}{C_{x_0}\rho S}} \quad (4)$$

Let's get the formula for $K_{Cx_0 - V_{max}}$:

$$K_{Cx_0 - V_{max}} = \frac{\frac{V_{max2}}{V_{max1}} - 1}{\frac{Cx_{0_2}}{Cx_{0_1}} - 1} = \frac{\sqrt{\frac{Cx_{0_1}}{Cx_{0_2}}} * \sqrt{\frac{P_2}{P_1}} - 1}{\frac{Cx_{0_2}}{Cx_{0_1}} - 1}$$

Let $\tilde{c} = \frac{Cx_{0_2}}{Cx_{0_1}}$, and the coefficient $\tilde{p} = \frac{P_2}{P_1}$ will show the characteristics' flow of the altitude-speed performance (ASP) (with a change in V_{max}, the engine thrust will change). Then the influence coefficient Cx_0 to V_{max} will be equal to:

$$K_{Cx_0 - V_{max}} = -\frac{1 + \sqrt{\frac{\tilde{p}}{\tilde{c}}}}{\tilde{c} - 1} \quad (5)$$

The coefficient $\tilde{p}$ corresponds to a specific aircraft because its value is influenced by both the engine's ASP and the intake losses ($\Delta\bar{p}_{вх}$) and nozzle ($\Delta\bar{p}_c$):

$$\tilde{p} = f(\Delta\bar{p}_{вх}, \Delta\bar{p}_c, \tilde{c}) \quad (6)$$

The maximum flying speed is not always determined by the required and available thrusts. Limitations may include temperature, structural strength, stability and control of the aircraft.

As described above, it may be concluded that the biding of the coefficient $K_{Cx_0 - V_{max}}$ to the parameters of the aircraft is impractical, because it will not give a true picture of understanding the impact of the change Cx_0 to V_{max}.

3.2. The influence of parameters on service ceiling

The effect changes on the ceiling can be estimated by the formula [16]:

$$\Delta H_{max} = -6.3\Delta m\ [km] \quad (7)$$

In this case, IC will be equal to:

$$K_{m\_H_{max}} = \frac{\frac{Hmax_2 - Hmax_1}{Hmax_1}}{\frac{m_2 - m_1}{m_1}} = \frac{\frac{-6.3\Delta\widehat{m}}{Hmax_1}}{\Delta\widehat{m}} = \frac{-6.3}{Hmax_1}$$

$$\boldsymbol{K_{m\_H_{max}} = -\frac{6.3}{H_{max}}} \quad (8)$$

The value of the coefficient is related to the initial ceiling of the aircraft and does not depend on the percentage change in mass (Table 1).

Table 1: The dependence of $K_{m\_H_{max}}$ on the service ceiling.

Initial ceiling H_{max}, km	$\boldsymbol{K_{m\_H_{max}}}$	$\Delta\widehat{m}$	ΔH_{max}, %	ΔH_{max}, m
20	**-0.315**	0.01 (1%)	0.315	63
		0.05 (5%)	1.575	315
18	**-0.35**	0.01 (1%)	0.315	63
		0.05 (5%)	1.575	315
16	**-0.394**	0.01 (1%)	0.315	63

In order not to become attached to the initial ceiling, it makes sense to speak not about a relative change in H_{max}, but about the absolute. Therefore, with an increase in mass by 5% the ceiling decreases by 315 meters, 10% - 630 meters. An increase in the mass of 1% leads to a decrease in the ceiling of about 63 meters.

3.3. The influence of parameters on specific express power

In order not to become attached to the initial ceiling, it makes sense to speak not about a relative change in H_{max}, but about the absolute. Therefore, with an increase in mass by 5% the ceiling decreases by 315 meters, 10% - 630 meters. An increase in the mass of 1% leads to a decrease in the ceiling of about 63 meters. We write the specific excess power as follows [16]:

$$V_y^* = \frac{P - X}{G} V \quad (9)$$

where P – engine thrust, X – aircraft drag, G – aircraft weight, V – flight speed.

We assume that the flying speed at which the specific excess power is calculated is constant ($V_1 \approx V_2$), then $K_{m\_V_y^*}$ will be equal to:

$$K_{m\_V_y^*} = \frac{\frac{V_{y_2}^*}{V_{y_1}^*}-1}{\frac{m_2}{m_1}-1} = \frac{\frac{m_1}{m_2}\left(\frac{P-X_2}{P-X_1}\right)-1}{\frac{m_2}{m_1}-1} = \frac{\frac{m_1}{m_2}\left(\frac{P-X_1-\Delta X_i}{P-X_1}\right)-1}{\frac{m_2}{m_1}-1} = -\frac{1-\frac{m_1}{m_2}\left(1-\frac{\Delta X_i}{P-X_1}\right)}{\frac{m_2}{m_1}-1} = \left\langle t=\frac{m_2}{m_1}; K_{X_i}=1-\frac{\Delta X_i}{P-X_1}\right\rangle = -\frac{1-\frac{K_{X_i}}{t}}{t-1}$$

$$\boldsymbol{K_{m\_V_y^*}} = -\frac{\mathbf{1}+\frac{\boldsymbol{K_{X_i}}}{\boldsymbol{t}}}{\boldsymbol{t-1}} \quad (10)$$

where $t=\frac{m_2}{m_1}; K_{X_i}=1-\frac{\Delta X_i}{P-X_1}, \Delta X_i$ - increase in inductive component due to mass increase, $\Delta P = P - X$ – abundance of power.

Note that the higher ΔP, the smaller contribution to the value of the IC makes ΔX_i. With small changes in mass (in this case $\Delta X_i \to 0$) and a relativity high value of ΔP we can assume that $K_{m\_V_y^*} \approx -\frac{m_1}{m_2}$. Then at Δm=10% $K_{m\_V_y^*} \approx -0.91$, and the coefficient value without the assumptions described above will be even less.

We similarly take out $K_{C_{x_0}\_V_y^*}$:

$$K_{C_{x_0}\_V_y^*} = \frac{\frac{V_{y_2}^*}{V_{y_1}^*}-1}{\frac{Cx_{0_2}}{Cx_{0_1}}-1} = \frac{\left(\frac{P-X_2}{P-X_1}\right)-1}{\frac{Cx_{0_2}-Cx_{0_1}}{Cx_{0_1}}} = \frac{\left(\frac{P-X_1-\Delta X_0}{P-X_1}\right)-1}{\frac{Cx_{0_2}-Cx_{0_1}}{Cx_{0_1}}} = -\frac{1-\frac{\Delta X_0}{P-X_1}-1}{\frac{Cx_{0_2}-Cx_{0_1}}{Cx_{0_1}}} =$$

$$\left\langle \tilde{c}=\frac{Cx_{0_2}}{Cx_{0_1}}\right\rangle = -\frac{\Delta X_0\, Cx_{0_1}}{\Delta P\left(Cx_{0_2}-Cx_{0_1}\right)} = -\frac{\left(Cx_{0_2}-Cx_{0_1}\right)\frac{\rho V^2 S}{2} Cx_{0_1}}{\Delta P\left(Cx_{0_2}-Cx_{0_1}\right)} = -\frac{X_0}{\Delta P}$$

$$K_{C_{x_0}\_V_y^*} = -\frac{X_0}{\Delta P} \quad (11)$$

If we neglect the inductive component of the resistance, then the formula takes the form:

$$\boldsymbol{K_{C_{x_0}\_V_y^*}} = -\frac{\mathbf{1}}{\frac{\boldsymbol{P}}{\boldsymbol{X_0}}-\mathbf{1}} \quad (12)$$

The value of the coefficient does not depend on the percentage change C_{x_0} and is associated with a specific aircraft.

The max value of the specific excess power for fighters is achieved as a rule at M=0.8-0.9 near the ground. We calculate the IC value according to the formula (11) and (12) for several aircraft. At M=0.85 ram airflow is q=51250 H/m^2, the IC excluding and taking into account the inductance is presented in Table 2.

Table 2: The values $K_{C_{x_0}\_V_y^*}$ are calculated by the formula (4) and (5)

Air-craft	M	Cx_0	Cx_i	S, m^2	P, H	$\boldsymbol{K_{C_{x_0}\_V_y^*}}$ **excluding** $\boldsymbol{Cx_i}$	$\boldsymbol{K_{C_{x_0}\_V_y^*}}$ **taking into account** $\boldsymbol{Cx_i}$
Mig-29	0.9	0.025	0.0005	38	200000	**-0.398**	**-0.401**
Su-24 (69°)	0.8	0.0263	0.0029	51	230000	**-0.367**	**-0.383**
Mig-23 (72°)	0.9	0.0225	0.0026	34.2	150000	**-0.438**	**-0.460**

From Table 2 it is seen that the inductive component of the resistance makes an insignificant contribution to the value of the IC. When Cx_0 is changed by 10% the differences in the scatter of V_y^* excluding and taking into account Cx_i are about 0.2%, which can be considered a negligible value at the preliminary design stage. Therefore, to simplify, you can use the formula (5).

3.4. The influence of parameters on sustained load factor

The sustained load factor can be expressed through the formula [16]:

$$n_y = \sqrt{\frac{(P-X_0)qS}{A(mg)^2}} \quad (13)$$

where P – engine thrust, X_0 – non-inductive resistance, q – velocity head, A – drag-due-to-lift factor, g – acceleration of gravity, then:

$$K_{m\_n_{y\,уст}} = \frac{\frac{n_{y_2}}{n_{y_1}}-1}{\frac{m_2}{m_1}-1} = \frac{\frac{m_1}{m_2}-1}{\frac{m_2}{m_1}-1} = \left\langle t=\frac{m_2}{m_1}\right\rangle = \frac{\frac{1}{t}-1}{t-1} = \frac{1-t}{t(t-1)} = -\frac{1}{t}$$

$$\boldsymbol{K_{m\_n_y}} = -\frac{\boldsymbol{m_1}}{\boldsymbol{m_2}} \quad (14)$$

The value of the coefficient depends on the percentage change in mass (Table 3).

Table 3: Dependence of $K_{m\_n_y}$ on Δm

Mass change Δm, %	$\boldsymbol{K_{m\_n_y}}$	Change of the sustained load factor Δn_y, %
5	**-0.952**	*4.8*
10	**-0.909**	*9.1*
15	**-0.869**	*13*

Similarly, we get $K_{Cx_0\_n_y}$:

$$K_{Cx_0\_n_{y\,уст}} = \frac{\frac{n_{y_2}}{n_{y_1}}-1}{\frac{Cx_{0_2}}{Cx_{0_1}}-1} = \frac{\sqrt{\left(\frac{P-X_2}{P-X_1}\right)}-1}{\frac{Cx_{0_2}-Cx_{0_1}}{Cx_{0_1}}} = \frac{\sqrt{\left(\frac{P-X_1-\Delta X_0}{P-X_1}\right)}-1}{\frac{Cx_{0_2}-Cx_{0_1}}{Cx_{0_1}}} = \langle Cx_{0_2} - Cx_{0_1} > 0\rangle = \sqrt{\left(\frac{P-X_1-\Delta X_0}{P-X_1}\right)\frac{(Cx_{0_1})^2}{(Cx_{0_2}-Cx_{0_1})^2}} - \frac{1}{\Delta\widehat{Cx}_0} =$$

$$\sqrt{\left(1-\frac{\Delta X_0}{P-X_1}\right)\frac{{Cx_{0_1}}^2}{(Cx_{0_2}-Cx_{0_1})^2}} - \frac{1}{\Delta\widehat{Cx}_0} =$$

$$\sqrt{\frac{{Cx_{0_1}}^2}{(Cx_{0_2}-Cx_{0_1})^2} - \frac{qS(Cx_{0_2}-Cx_{0_1})}{qS\left(\frac{P}{Sq}-Cx_{0_1}\right)}\frac{{Cx_{0_1}}^2}{(Cx_{0_2}-Cx_{0_1})^2}} - \frac{1}{\Delta\widehat{Cx}_0} =$$

$$\sqrt{\frac{1}{(\Delta\widehat{Cx}_0)^2}-\frac{{Cx_{0_1}}^2}{(\frac{P}{Sq}-Cx_{0_1})(Cx_{0_2}-Cx_{0_1})}}-\frac{1}{\Delta\widehat{Cx}_0}=$$

$$\sqrt{\frac{1}{(\Delta\widehat{Cx}_0)^2}-\frac{Cx_{0_1}}{\left(\frac{P}{Sq}-Cx_{0_1}\right)(\Delta\widehat{Cx}_0)}}-\frac{1}{\Delta\widehat{Cx}_0}=$$

$$\sqrt{\frac{1}{(\Delta\widehat{Cx}_0)^2}-\frac{Cx_{0_1}\Delta\widehat{Cx}_0}{\left(\frac{P}{Sq}-Cx_{0_1}\right)}\frac{1}{(\Delta\widehat{Cx}_0)^2}}-\frac{1}{\Delta\widehat{Cx}_0}=\frac{1}{\Delta\widehat{Cx}_0}\sqrt{1-\frac{Cx_{0_1}\Delta\widehat{Cx}_0}{\frac{P}{Sq}-Cx_{0_1}}}-$$

$$\frac{1}{\Delta\widehat{Cx}_0}=\frac{1}{\Delta\widehat{Cx}_0}\left(\sqrt{1-\frac{Cx_{0_1}\Delta\widehat{Cx}_0}{\frac{P}{Sq}-Cx_{0_1}}}-1\right)$$

$$K_{Cx_{0\_}n_y}=\frac{1}{\Delta\widehat{Cx}_0}\left(\sqrt{1-\frac{Cx_{0_1}\Delta\widehat{Cx}_0}{\frac{P}{Sq}-Cx_{0_1}}}-1\right) \qquad (14)$$

The coefficient value slightly depends on the percentage change Cx_0 (differences – in thousandths). When Cx_0 changes by 10%, the formula takes the form:

$$\boldsymbol{K_{Cx_{0\_}n_y}=10\left(\sqrt{1-0.1\frac{Cx_{0_1}}{\frac{P}{Sq}-Cx_{0_1}}}-1\right)} \qquad (15)$$

For example, we calculate the IC for several AK. At M=0.8 at a higher H=1000m ram-air flow is equal q=41200 H/m^2, (Table 4).

Table 4: The values $K_{Cx_{0\_}n_y}$

Aircraft	Cx_0	S, m^2	P (H=1000, M=0.8), H	$\boldsymbol{K_{Cx_{0\_}n_y}}$
Mig-29	0.024	38	170000	**-0.16**
Su-24 (69°)	0.026	51	206000	**-0.21**
Mig-23 (72°)	0.022	34.16	130000	**-0.18**

3.5. The influence of parameters on acceleration time

Acceleration time in horizontal flight in within speed range is [17]:

$$t_p=\frac{1}{g}\int_{V_н}^{V_к}\frac{dV}{n_{xa}}\approx\frac{V_к-V_н}{g*(n_{xa})_{cp}} \qquad (16)$$

Where $V_н$ and $V_к$ – initial and final speed, $(n_{xa})_{cp}$ – averaged longitudinal g load for within speed range.

Let's get the formula for $K_{m\_t_p}$:

$$\Delta(n_{xa})_{cp}=(n_{xa})_{cp2}-(n_{xa})_{cp1}=\frac{(P_{cp}-X_{2_{cp}})}{G_2}-\frac{(P_{cp}-X_{1_{cp}})}{G_1}=$$

$$\frac{\frac{m_1}{m_2}P_{cp}-\frac{m_1}{m_2}X_{1_{cp}}-\frac{m_1}{m_2}\Delta X-P_{cp}+X_{1_{cp}}}{G_1}=-\frac{P_{cp}\left(1-\frac{m_1}{m_2}\right)-X_{1_{cp}}\left(1-\frac{m_1}{m_2}\right)+\Delta X\frac{m_1}{m_2}}{G_1}=$$

$$-\frac{(P_{cp}-X_{1_{cp}})\left(1-\frac{m_1}{m_2}\right)+\Delta X\frac{m_1}{m_2}}{G_1}\text{, then:}$$

$$K_{m\_t_p}=\frac{\frac{t_{p_2}}{t_{p_1}}-1}{\frac{m_2}{m_1}-1}=\left(\frac{t_{p_2}}{t_{p_1}}-1\right)\frac{1}{\Delta\hat{m}}=\left(\frac{(n_{xa})_{cp1}-(n_{xa})_{cp2}}{(n_{xa})_{cp2}}\right)\frac{1}{\Delta\hat{m}}=$$

$$\left(\frac{1}{\frac{(n_{xa})_{cp2}}{-\Delta(n_{xa})_{cp}}}\right)\frac{1}{\Delta\hat{m}}=\left(\frac{1}{-1-\frac{1}{\frac{\Delta(n_{xa})_{cp}}{(n_{xa})_{cp1}}}}\right)\frac{1}{\Delta\hat{m}}=$$

$$\left(\frac{1}{\frac{1}{\frac{(P_{cp}-X_{1_{cp}})\left(1-\frac{m_1}{m_2}\right)G_1}{(P_{cp}-X_1)G_1}+\frac{\Delta X*m_1*G_1}{m_2(P_{cp}-X_1)G_1}}-1}\right)\frac{1}{\Delta\hat{m}}=$$

$$\left(\frac{1}{\frac{1}{1-\frac{m_1}{m_2}+\frac{m_1}{m_2}\frac{\Delta X}{(P_{cp}-X_{1_{cp}})}}-1}\right)\frac{1}{\Delta\hat{m}}=\left(\frac{1}{\frac{1}{1-\frac{m_1}{m_2}\left(1-\frac{\Delta X}{(P_{cp}-X_{1_{cp}})}\right)}-1}\right)\frac{1}{\Delta\hat{m}}$$

$$\boldsymbol{K_{m\_t_p}=\left(\frac{1}{\frac{1}{1-\frac{m_1}{m_2}\left(1-\frac{\Delta X_{i_{cp}}}{\Delta P_{cp}}\right)}-1}\right)\frac{1}{\Delta\hat{m}}} \qquad (17)$$

where ΔP_{cp} and $\Delta X_{i_{cp}}$ – the average engine thrust margin and the average increase of the induced drag for the speed range acceleration.

The value of the coefficient slightly depends on the percentage change in the mass. Assuming $\Delta X_{i_{cp}}/\Delta P_{cp}=0$ (ΔX_i is significantly less than the average available thrust ΔP_{cp}) $K_{m\_t_p}$=1, and taking into account the influence of increased inductance (due to the increase in mass) the coefficient value will be slightly larger. For example, for Mig-29 during acceleration 600 – 1100 km/hour near the ground $K_{m\_t_p}=1{,}12$ (Table 5).

Table 5: The values $K_{m\_t_p}$ for Mig-29 (600 – 1100 km/hour on the deck)

m_1, kg	Δm, %	ΔC_y	ΔCx_i	P_{cp}, kgf	X_{cp}, kgf	ΔX_i, kgf	$\boldsymbol{K_{m\_t_p}}$
15000	10	0.011	0.001	16000	4000	130	**1.12**

If accept $\Delta Cx_{0_{cp}}\approx\Delta Cx_0$ and $Cx_{0_{1cp}}\approx Cx_{0_1}$, $K_{Cx_{0\_}t_p}$ takes the following form:

$$K_{Cx_{0\_}t_p}=\frac{\frac{t_{p_2}}{t_{p_1}}-1}{\frac{Cx_{0_2}}{Cx_{0_1}}-1}=\left(\frac{t_{p_2}}{t_{p_1}}-1\right)\frac{1}{\Delta\widehat{Cx}_0}=\left(\frac{1}{-1-\frac{(n_{xa})_{cp1}}{\Delta(n_{xa})_{cp}}}\right)\frac{1}{\Delta\widehat{Cx}_0}=$$

$$\left\langle\Delta(n_{xa})_{cp}=\frac{-\Delta X_0}{G_1}\right\rangle=\left(\frac{1}{\frac{P_{cp}-X_{1_{cp}}}{\Delta X_{0_{cp}}}-1}\right)\frac{1}{\Delta\widehat{Cx}_0}=$$

$$\left(\frac{1}{\frac{P_{cp}-X_{1_{cp}}-\Delta X_{0_{cp}}}{\Delta X_{0_{cp}}}}\right)\frac{1}{\Delta\widehat{Cx}_0}=\left(\frac{\Delta Cx_{0_{cp}}qS}{qS\left(\frac{P_{cp}}{qS}-Cx_{0_{1cp}}-\Delta Cx_{0_{cp}}\right)}\right)\frac{Cx_{0_1}}{\Delta Cx_0}=$$

$$\left\langle\begin{matrix}\Delta Cx_{0_{cp}}\approx\Delta Cx_0\\ Cx_{0_{1cp}}\approx Cx_{0_1}\end{matrix}\right\rangle=\frac{Cx_{0_1}}{\frac{P_{cp}}{qS}-Cx_{0_{1cp}}-\Delta Cx_{0_{cp}}}=\frac{1}{\frac{P_{cp}}{X_{0_{1cp}}}-1-\Delta\widehat{Cx}_0}$$

$$K_{Cx_0\_t_p} = \frac{1}{\frac{P_{cp}}{X_{0_{1cp}}} - 1 - \Delta\widehat{Cx_0}} \quad (18)$$

The coefficient value slightly depends on the percentage change Cx_0 and is associated with a specific aircraft. For Mig-29 $K_{Cx_0\_t_p}$ = 0.27 (Table 6).

Table 6: The values $K_{Cx_0\_t_p}$ for Mig-29 (600 – 1100 km/hour near the ground)

Cx0	Δ Cx0, %	Pcp, kgf	$X_{0_{1cp}}$, kgf	$K_{Cx_0\_t_p}$
0.025	10	16000	3300	**0.27**

4. Calculation of the influent coefficients by numerical method

To confirm the derived dependencies (8), (10), (12), (14), (15), (17), (18), we calculate the IC using the "Program for preparing the initial data and calculating the FQP". This software product was developed by Sukhoi Design Bureau and verified by many years of the application practice. The program uses weight, aerodynamic characteristics (Cx_0(M), Cx_i (C_y, M), Cy_{max}(M), C_y^a) and propulsive characteristics P (H, M, Dr) Qc(H, M, Dr), as well as some limitations (q_{max}, ny_{max}) and constants. The graphs Cx_0(M), Cx_i (C_y, M), Cy_{max}(M) C_y^a, P (H, M), Qc(H, M) are digitized from books of practical aerodynamics [18-21]. ICs are calculated according to the algorithm shown in the Figure 1.

We calculate IC for the following aircraft of the class of operational - tactical aviation: fighter 4th generation Su-27; frontline bomber Su-24; strike-fighter Su-25; fighter 4th generation Mig-29; fighter 3rd generation Mig-23; interceptor jet 3rd generation Mig-25.

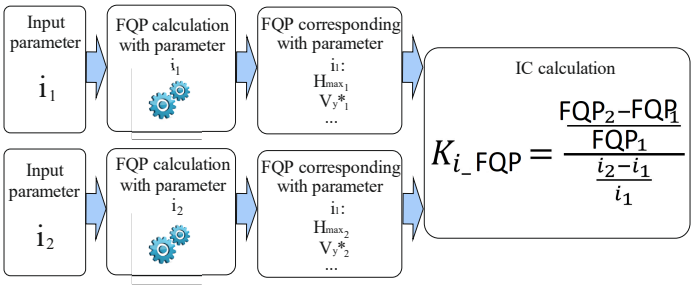


Figure 1: IC calculation scheme

Table 9 and Figures 2-11 show the results of calculating the IC programmatically. ICs were obtained when the mass changed by 10%, Cx_0 changed for M<1 by 10% and for M>1 by 18%. Such differences in the spread of Cx_0 are taken on the assumption that the inaccuracy of determining this parameter for the supersonic domain of flight is higher. Aircraft characteristics are taken in the calculations are presented in Table 7.

The effect on the service ceiling is shown in Figures 2 and 3, on the specific excess power - in Figures 4 and 5, on the sustained load factor - in Figures 6 and 7, for the acceleration time - in Figures 8 and 9. Figures 2-9 also show the values obtained from the above formulas.

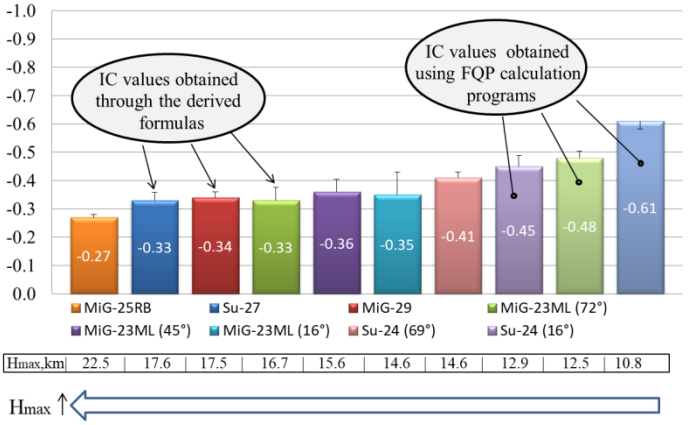


Figure 2: $K_{m\_H_{max}}$ values

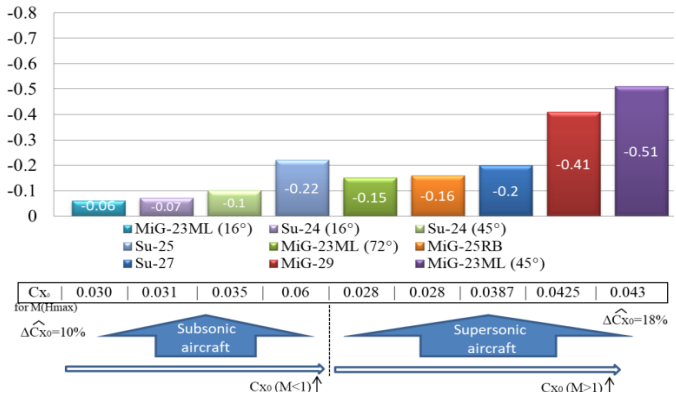
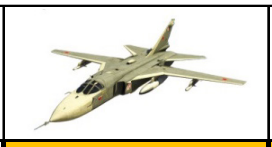


Figure 3: $K_{Cx_0\_H_{max}}$ values

Table 7: Aircraft characteristics, used in the calculations

Aircraft	MiG-29	MiG-25RB	MiG-23ML (72°)	MiG-23ML (45°)	MiG-23ML (16°)	Su-27	Su-25	Su-24 (69°)	Su-24 (45°)	Su-24 (16°)
Weight										
Mass in calculations, kg	15000	24000	18000	18000	18000	21200	13000	25000	25000	25000
Wing parameters										
Wing area, m^2	38	61.5	34.16	35.3	37.27	62	30.1	51	53	55.18
Wing extension	3.39	2.94	1.77	3.43	5.26	3.5	6	2.11	3.9	5.64
Leading edge a sweep angle, deg.	42	41	74	47	18	42	20	69	45	16
Power plant specifications										
Jet engine	2xRD-33	2xR-15-300	R-35	R-35	R-35	2xA1-31F	2xR-95 SH	2xA1-21 F-3	2xA1-21 F-3	2xA1-21 F-3
Thrust (H=0, M=0), kgf	16000	22400	12500	12500	12500	25000	8200	22400	22400	22400
Aircraft design parameters										
Starting thrust-to weight ratio	1.07	0.93	0.69	0.69	0.69	1.18	0.63	0.90	0.90	0.90
Wing load, kg/m^2	395	390	527	510	483	342	432	490	472	453

Table 8: IC for the studied aircrafts

	Su-27		**MiG-29**		**MiG-23ML (72°)**		**MiG-23ML (45°)**		**MiG-23ML (16°)**	
FQ&P	**m**	**Cx0**	**m**	**Cx0**	**m**	**Cx0**	**m**	**Cx0**	**m**	**Cx0**
Vmax	-0.07	-0.81	-0.09	-0.71	-	-	-0.14	-0.85	-0.01	-0.17
Hmax	-0.33	-0.2	-0.34	-0.41	-0.33	-0.15	-0.36	-0.51	-0.35	-0.06
Vy*	-0.91	-0.37	-0.91	-0.47	-1.02	-0.47	-0.98	-0.53	-0.91	-0.45
ny	-0.91	-0.09	-0.91	-0.09	-0.91	-0.11	-0.91	-0.1	-0.91	-0.09
tp	1.09	0.26	1.08	0.27	1.22	0.24	1.11	0.28	1.02	0.35
										
	MiG-25RB		**Su-25**		**Su-24 (69°)**		**Su-24 (45°)**		**Su-24 (16°)**	
FQ&P	**m**	**Cx0**	**m**	**Cx0**	**m**	**Cx0**	**m**	**Cx0**	**m**	**Cx0**
Vmax	-	-	-0.02	-0.24	-	-	-0.02	-0.18	-0.02	-0.14
Hmax	-0.27	-0.16	-0.61	-0.22	-0.41	-	-0.48	-0.1	-0.45	-0.07
Vy*	-0.96	-0.43	-0.94	-0.46	-0.98	-0.39	-0.96	-0.45	-0.94	-0.41
ny	-0.91	-0.21	-0.91	-0.26	-0.91	-0.19	-0.91	-0.14	-0.91	-0.18
tp	1.06	0.35	1.05	0.35	1.19	0.31	1.11	0.22	1.06	0.30

- subsonic aircraft

Sustained load factor on high 1000 m M=0.8 (for Su-25 M=0.6)

Acceleration time on the deck 600-1100 km/hour (for Su-25 500-800 km/hour)

Aircrafts images courtesy of Google images

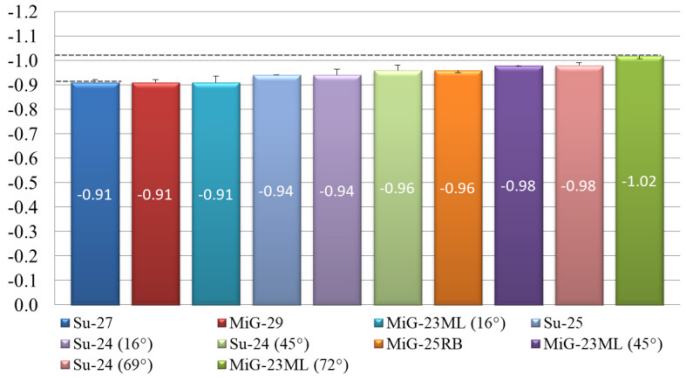


Figure 4: $K_{m\_V_y^*}$ values

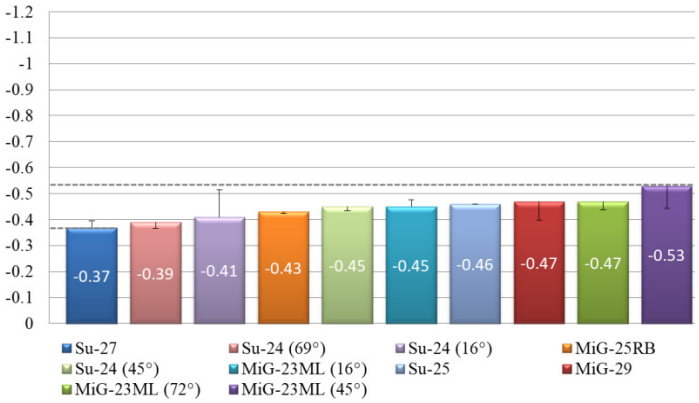


Figure 5: $K_{C_{x_0}\_V_y^*}$ values

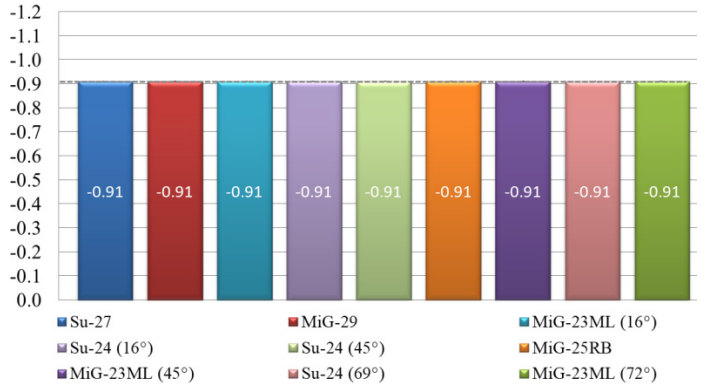


Figure 6: $K_{m\_n_y}$ values (H=1000m, M=0.8)

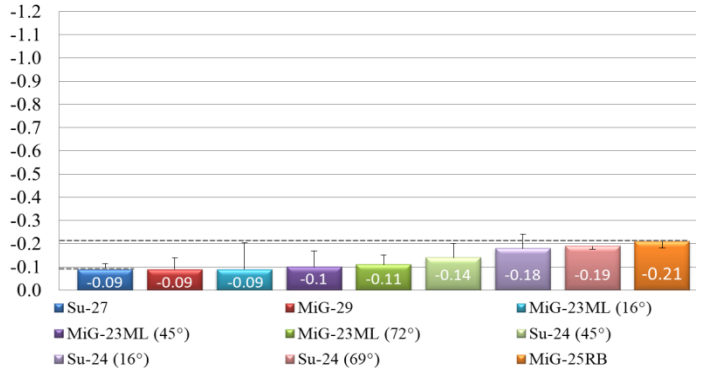


Figure 7: $K_{C_{x_0}\_n_y}$ values (H=1000m, M=0.8)

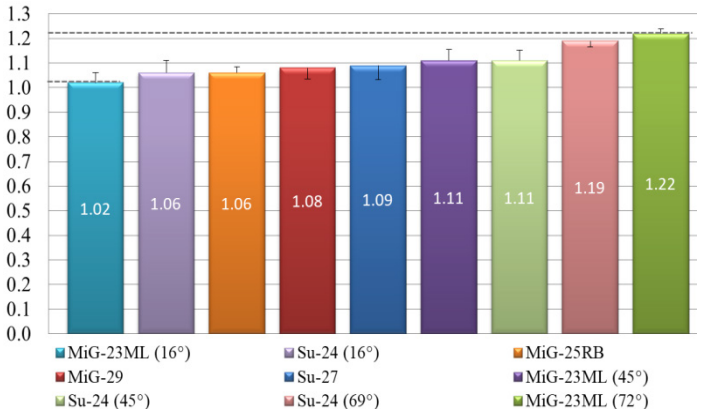


Figure 8: $K_{m\_t_p}$ values

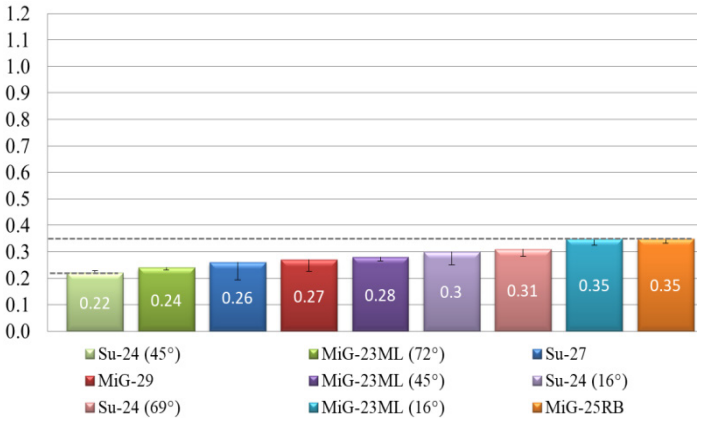


Figure 9: $K_{C_{x_0}\_t_p}$ values

5. Results and Discussions

As can be seen from Figures 2-10, the analytically obtained ICs are confirmed by numerical calculations. The absolute error when using the derived formulas is less than 0.1 of the IC value, which in terms of FQP values gives an error of less than 10%. At the initial stage of aircraft design (when the error of other constituent elements is comparable in magnitude), this is an acceptable indicator.

Summarize the calculated IC in Table 9, the percentage change in FQP is presented in Table 10.

Table 9: IC ranges for the considered OTA aircrafts

FQP	K_{m_FQP}	$K_{Cx_0\_FQP}$	
		$M_{max}<1$	$M_{max}>1$
V_{max}	-0.1	-0.1…-0.25	-0.7…-0.85
H_{max}	0.25…0.6	-0.05…-0.2	-0.15…-0.5
$V_y^* max$	-0.9….-1.05	-0.35…-0.55	-
n_y (H=1000m, M=0.8)	-0.91	-0.05…-0.2	-
t_p (600-1100 km/h, H=200m)	1…1.25	-0.2…-0.35	-

Table 10: Percent change in the FQP

FQP	With an increase in mass by 10%, the FQP deteriorates by:	With an increase in Cx_0 in the subsonic area by 10% and by 18% in supersonic the FQP deteriorate by:	
		$M_{max}<1$	$M_{max}>1$
V_{max}	1%	1…2.5%	12…15.5%
H_{max}	decreases by 630 m	0.5…2%	3…8%
$V_y^* max$	9….10.5%	3.5…5.5%	-
n_y (H=1000m, M=0.8)	9.1%	0.5…2%	-
t_p (600-1100 km/h, H=200m)	10…12.5%	2…3.5%	-

From Tables 9 and 10 we can conclude that the considered FQP, such as the maximum specific excess power, sustained load factor, and acceleration time, are most sensitive to a change in mass. The inaccuracy in determining the mass in 1% gives a spread of these FQP in about 1% (more accurate values are given earlier). The IC of the mass on the practical ceiling depends on the absolute value of the ceiling, so it makes sense to talk about its absolute change. With an inaccuracy of mass of 1%, the ceiling is determined with an accuracy of about 60-70 meters.

To inaccuracy C_{x_0} is most sensitive to the maximum flying speed (for M>1). The numerical value of the IC on V_{max} is determined by the curve crossing nature of the required and available thrust (Figure 11). Graphs Cx_0 (M) for the studied aircrafts are shown in Figure 10.

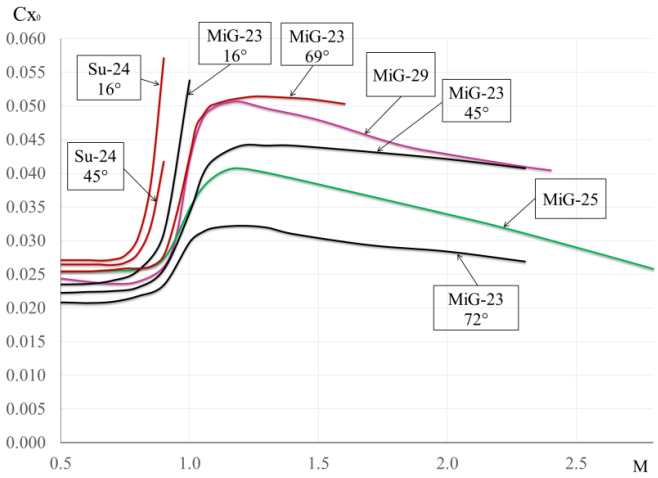


Figure 10: Graphs Cx0 for the studied aircrafts.

Figure 10 shows that in the area of maximum flight speed, the slope of the curves C_{x0} (M) is different for different aircraft.

The power-available curve depends on the engine unit's altitude speed performances, the required one on the drag run, therefore, the IC depends strongly on the specific aircraft (on the flow pattern of Cx_0 and altitude-speed performances), therefore,

linking this coefficient to the aircraft parameters is impractical. However, as the research showed, the IC value for subsonic aircraft is approximately 0.1…0.2 (Table 9), for supersonic: 0.7…0.9.

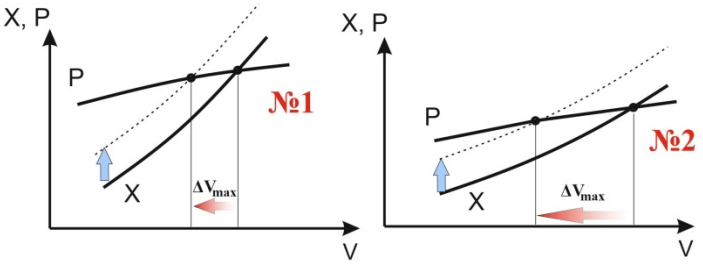


Figure 11. The different nature of Zhukovsky curve crossing. P – engine thrust curve, X – airplane drag curve.

6. Conclusions

The results of this work allow us to estimate the contribution of mass uncertainties and drag coefficient at zero lifting force to the range of variation of the FQP values.

The formulas obtained make it possible to determine the quantitative influence of the parameter on the value of the FQP. Verification of the formulas for some aircraft from the class of operational-tactical aviation showed that the error in determining the spread of flight characteristics according to the formulas obtained is within 10%.

The research showed that the mass influence coefficients Cx_0 on maximum specific excess power, sustained load factor and acceleration time practically do not depend on the parameters of the considered aircrafts of the same class. The IC of mass and Cx_0 on the maximum speed and practical ceiling associated with specific aircraft parameters. The sensitivity of the maximum speed to the zero lift-drag coefficient is determined by the curve crossing nature of the required and available thrust.

The inaccuracy in determining the mass in 1% gives a spread of maximum specific excess power, sustained load factor and acceleration time in about 0.9…1.3%. For the studied aircrafts the inaccuracy in determining the zero lift-drag coefficient of 1% gives a spread in the maximum speed within 0.7 ... 0.9% for supersonic aircraft and 0.1 ... 0.2% for subsonic aircraft.

References

[1] V.E. Denisov, V.K. Isaev, A.M. Ryabov, L.M. Shkadov, "Statisticheskaya ocenka harakteristik proektiruemogo samoleta s pomoschyu metoda Monte_Karlo" [Statistical evaluation of the characteristics of the designed aircraft using the Monte Carlo method], Uchenie zapiski CAGI [Scientific notes of TsAGI], **4**(2), 137–142, 1973.

[2] O.K. Yugov, O.D. Selivanov, Osnovi integracii samoleta i dvigatelya [Fundamentals of the integration of the aircraft and engine], Mashinostroenie [Mechanical Engineering], 1989.

[3] G.M. Skopec, "Vneshnee proektirovanie aviacionnih kompleksov: Metodologicheskie aspekti" [External design of aviation complexes: Methodological aspects], LENAND, 2017.

[4] M. Ravachol, Y. Deremeaux, Q.V. Dinh, E. Herbin, "Uncertainties at the conceptual stage: multilevel multidisciplinary design and optimization approach", ICAS 2008 26TH International Congress of the Aeronautical Sciences, Anchorage, Alaska; September 14-19, 2008.

[5] L.V. Myshkin, Prognozirovanie razvitiya aviacionnoi tehniki [Forecasting the development of aviation technology], Izdatelskii Dom Nauka [Publishing House Science], 2017.

[6] T. Pfeiffer, E. Moerland, D. Böhnke, B. Nagel, V. Gollnick, «Aircraft configuration analysis using low-fidelity, physics based aerospace framework under uncertainty considerations», ICAS 2014 29TH Congress of the International Council of the Aeronautical Sciences, St. Peterburg, Russia; September 7-12, 2014.

[7] D. Neufeld , J. Chung, K. Behdinan, "Aircraft Conceptual Design Optimization with Uncertain Contributing Analyses", AIAA Modeling and Simulation Technologies Conference, Chicago, Illinois; August 10–13, 2009. doi: 10.2514/6.2009-6237

[8] M.R. Kirby, D.N. Mavris, "Forecasting Technology Uncertainty in Preliminary Aircraft Design", AIAA, World Aviation Conference, San Francisco, CA; October 19-21, 1999.

[9] L.L. Green, H.Z. Lin, M.R. Khalessi, "Probabilistic methods for uncertainty propagation applied to aircraft design", AIAA, PredictionProbe, Inc. Newport Beach, CA 92660, 2002. doi:10.2514/6.2002-3140

[10] M. Hosseini, M. Nosratollahi, H. Sadati, "Multidisciplinary Design Optimization of UAV under Uncertainty", J. Aerosp. Technol. Manag., São José dos Campos, **9**(2), 169–178, 2017. doi: 10.5028/jatm.v9i2.725

[11] D.N. Mavris, D.A. DeLaurentis , D.S. Soban, "Probabilistic Assessment of Handling Qualities Characteristics in Preliminary Aircraft Design", AIAA 98-0492, 36th Aerospace Sciences Meeting & Exhibit Copyright, 1998.

[12] D.N. Mavris, D.A. DeLaurentis, D.P.Schrage, "System Synthesis in Preliminary Aircraft Design Using Statistical Methods", the 20th International Council of the Aeronautical Sciences (ICAS) Congress, Sorrento, Italy, September 8-13, 1996.

[13] G.S. Veresnikova, L.A. Pankova, V.A. Pronina, O.V. Ogorodnicov, I.I. Ikryanov, "Determining maneuverable aircraft parameters in preliminary design under conditions of uncertainty", International Conference on Knowledge Based and Intelligent Information and Engineering Systems, Marseille, France; September 6–8, 2017. doi: 10.1016/j.procs.2017.08.143

[14] D.A. Laushin, V.A. Klyagin, "K voprosu o vliyanii faktorov neopredelennosti na oblik samoleta" [To the question of the influence of uncertainty factors on the appearance of the aircraft], v sbornike tezisov XLIV Mejdunarodnoi molodejnoi nauchnoi konferencii Gagarinskie chteniya [In the collection of theses of the XLIV International Youth Scientific Conference Gagarin Readings], Moscow Aviation Institute, 2018.

[15] D.A. Laushin, V.A. Klyagin, "Uchet riskov, svyazannih s dostijeniem zadannih letno_tehnicheskih harakteristik samoleta" [Accounting for the risks associated with achieving the specified flight performance of the aircraft], Obscherossiiskii nauchno_tehnicheskii jurnal POLET [All-Russian scientific and technical journal POLET], 1, 28–32, 2019.

[16] A.M. Tarasenkov, V.G. Braga, V.T. Taranenko, Dinamika poleta i boevogo manevrirovaniya letatelnih apparatov [Dynamics of flight and combat maneuvering of aircraft, VVIA im. prof. N.E. Jukovskogo [Zhukovsky VVIA], 1984.

[17] G.S. Busgens, Aerodinamika, ustoichivost i upravlyaemost sverhzvukovih samoletov [Aerodynamics, stability and controllability of supersonic aircraft], Fizmatlit, 1998.

[18] "Prakticheskaya aerodinamika samoleta MiG-29" [Practical aerodynamics of the MiG-29], Moskva, slujebnoe izdanie VVS [Moscow, official publication of the Air Force], 1987. Available Online at: http://bookre.org/reader?file=543564

[19] "Prakticheskaya aerodinamika samoleta MiG-23ML i MiG-23UB" [Practical aerodynamics of the MiG-23ML and MiG-23UB], VVIA im. Jukovskogo [Zhukovsky VVIA], 1978. Available Online at: http://bookre.org/reader?file=543565

[20] "Prakticheskaya aerodinamika samoleta MiG-25RB" [Practical aerodynamics of the MiG-25RB], Voennoe izdatelstvo Ministerstva oboroni SSSR [Military Publishing House of the Ministry of Defense of the USSR], 1978. Available Online at: http://bookre.org/reader?file=543563

[21] "Prakticheskaya aerodinamika Su-24" [Practical aerodynamics of the Su-24]. Available Online at: http://bookre.org/reader?file=540492

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 127-132 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Infrared Uplink Implementation for Software Defined Visible Light Communication Systems

Oswaldo René Banda-Sayco*

Universidad Nacional de San Agustín de Arequipa, 04001, Peru

ARTICLE INFO

Article history:
Received: 30 July, 2020
Accepted: 01 October, 2020
Online: 08 November, 2020

Keywords:
Infrared Communication
Analog front-end
Visible Light Communication
Software Defined Radio
USRP
GNU Radio

ABSTRACT

The main advantage of software-defined visible light communication (VLC) systems is that they allow us to reuse modulation, coding, synchronization and medium access techniques used in conventional radio systems. This article focuses on the implementation of an uplink alternative using an infrared (IR) light communication system. For this goal, analog front-ends are designed and implemented. Therefore, they convert a radio frequency signal into an optical signal. In addition, a Universal Software Radio Peripheral (USRP) is used as a software-defined radio platform. The obtained bit error rate is 5.33×10^{-5} *using QPSK modulation with a bit rate of 4 Mbps at a distance of 160 cm between transmitter and receiver.*

1. Introduction

This paper is an extension of work originally presented in *2019 IEEE Colombian Conference on Communications and Computing (COLCOM)* [1]. This article aims to focus into the design process followed to implement analog front ends for the infrared (IR) communication uplink. On the other hand, an explanation is given about the blocks that make up the program to measure the bit error rate (BER) of the IR system. Finally, the signal-to-noise ratio (SNR) measurement is made and compared with the bit error rate (BER) obtained at different distances between transmitter and receiver.

Visible Light Communication (VLC) is currently being developed as a solution that enables the use of standard illumination light-emitting diodes (LEDs) to transmit data wirelessly. This type of communication is an alternative to mitigate the effects of saturation of the radioelectric spectrum in densely populated areas. Furthermore, it is known that the number of wireless devices connected to the internet is growing at an annual rate of 47% until 2021 [2]. This is mainly due to the greater access of terminal devices such as smartphones, tablets, laptops and recently to the deployment of the Internet of Things (IoT). For this reason, VLC is a technology that can serve as a complement of the currently deployed communication systems and provide access to the network. An important feature of VLC is that it can be used as a ubiquitous communication medium because LED lighting lamps are widely used in existing lighting infrastructure. On the other hand, VLC operates in the visible light band, which is not licensed and there is no electromagnetic interference generated by other communication equipment [3].

Information is sent and received through one or more LEDs on the transmitting side and a photodiode on the receiving side using a technique known as intensity modulation and direct detection (IM-DD) [4]. In such a way that the information that is sent produces variations in the level of illumination of the LED stage, these variations are imperceptible to the human eye. After that, the receiving photodiode converts the optical signal into an electrical signal for further processing.

In [5], the authors focused on the design and implementation of Analog Front-Ends, which would allow easy replication of the system through the use of low-cost components. The achieved bit rate was 20 kbps with a BER of 4.89×10^{-10} using OOK modulation and Manchester encoding. In [6], a similar paper was done, but it did not specify the components of the designed interfaces, a BER of 9.77×10^{-5} was measured at a bit rate of 2.5 Mbps and at a distance of 1.6 meters using BPSK modulation technique. In [7], an uplink based on infrared communication is proposed, the simulation result shows a value of 3.57 Gb / s using a transmitter with beam stearing and an angle diversity receiver. However, the experimental implementation stage of the system was not reached. In [8], it was possible to implement an infrared uplink by using an off-the-shelf receiver module (TSOP 1730) and the PPM modulation technique, however, the data transfer rate of

*Corresponding Author: Oswaldo René Banda Sayco, Arequipa, Peru, obanda@unsa.edu.pe

the infrared link reaches up to 1200 bps. In [9], a surveillance camera is used as a receiver for an uplink based on infrared light. In [10], an IR link was implemented using a silicon integrated beam steering device, the achieved bit rate was 12.5 Gb/s and the bit error rate (BER) was less than 10^{-9}.

This article shows the design and implementation process of Analog Front-Ends used on the transmitter and receiver side, achieving a bit rate of 4 Mbps and a BER of 5.33×10^{-5} using QPSK modulation at a distance of 1.6 m. These interfaces have the advantage of being coupled to Software Defined Radio (SDR) systems, in such a way that the modulation, coding and synchronization of the data is carried out within a computer connected to a Universal Software Radio Peripheral (USRP). This allows for rapid prototyping and measurement of the IR communication system.

Section two presents the methodology used to carry out this work, section three delves into the design and implementation of Front-Ends, section 4 shows the results of the BER and SNR. Finally, section 5 shows the conclusions of this research work.

2. Methodology

Figure 1 shows the methodology followed to design and implement the Analog Front-Ends. First, an analysis of the communication system to be implemented is made. This analysis allows us to know all the elements that are required to make the VLC system work correctly. In addition, it is necessary to define the design parameters that allow the development of Analog Front-Ends. After that, the second procedure is Development of the Front-Ends. To carry out this task, it has been divided into two threads: design and implementation.

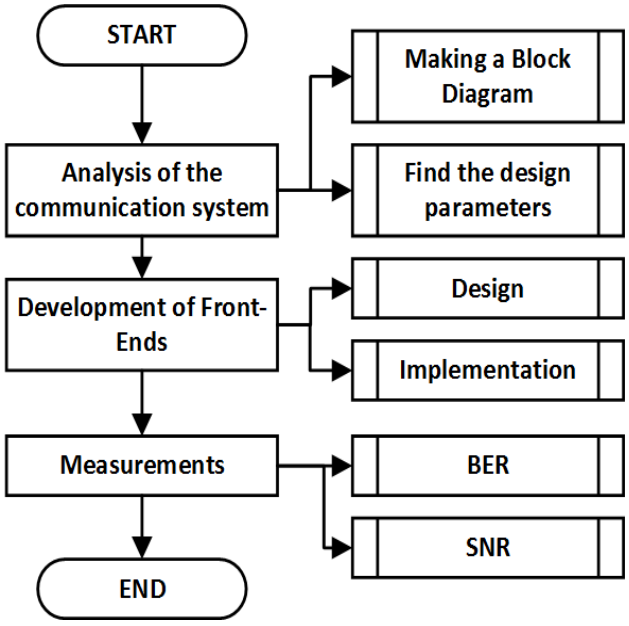


Figure 1: Methodology diagram

Finally, the last step is to carry out the corresponding measurements of the implemented communication system. To do this, a BER measurement is made as a function of the separation distance between transmitter and receiver. On the other hand, the measurement of the signal-to-noise ratio (SNR) is also carried out and a comparison is made with the previously measured BER.

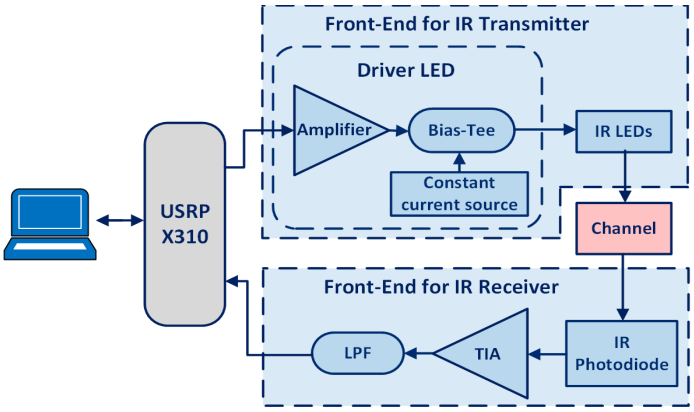


Figure 2: Block diagram of the IR communication system for BER measurement

3. Analysis of the Communication System

3.1. Block Diagram of the System

Figure 2 shows the block diagram of the IR communication system implemented for the purpose of making the BER measurement. This diagram is based on the research article previously made in [1] . It is observed that a laptop has a software stage where the modulation and synchronization of the information is carried out. After that, a USRP X300 is connected, which is a software-defined radio platform that allows to generate a modulated signal in the RF band. In this case, a central frequency of 5MHz has been used to carry out this test. Then, the signal goes to the transmitter front-end. This front-end is composed by an amplifier, a bias-tee, a constant current source and the IR LEDs. After that, the signal travels towards the receiver through the channel. In this case, the channel is the air in the indoor environment. On the other hand, the receiver front-end includes an IR photodiode, a transimpedance amplifier (TIA) and a low pass filter (LPF).

Table 1: Design Parameters for the IR link

Parameter	Description
Application	Indoor
Coverage range	1.6 m
Configuration	Direct LOS link
Bitrate	4 Mbps
Modulation	QPSK / BPSK
Transmitter	LED Specification (each one) - Radiant Intensity: 550 mW/sr - View angle: 10° - Wavelenght: 800 nm – 875 nm
Receiver	Photodiode specifications: - Code: SFH 203 FA - Spectral response: 800 nm – 1100 nm - Sensitive area: 1 mm^2
OP-AMPs	LT6268-10 and LT6268 with dual supply ±2.5 V_{DC}

3.2. Design Parameters

The design parameters of the electronic circuits for the transmitter and receiver analog interfaces are shown in Table 1. It is observed that the configuration used in this paper requires a line of sight (LOS), this is due to the fact that the angle of view of the LEDs arrangement is 10°. Which allows maximum energy transfer in only one direction. In this way, the use of an optical lens that amplifies the optical signal at the receiver is avoided. In addition to this, the other parameters taken into account to implement this communication system can be read.

4. Development of Front-Ends

Front-ends are developed based on the design parameters mentioned in the previous section. First, the design of the front-ends is done. After that, the physical implementation of them is made.

4.1. Transmitter Design

The transmitter front-end is composed by four blocks. The first block is an amplifier that must take the signal from the USRP and feed a LED array. The second block is a bias-tee that superimposes the RF signal coming from the amplifier on a constant current level coming from the third block called "Constant current source". The three previous blocks make up the LED driver. Finally, the last block is an array of LEDs, it is composed of eight high-power LEDs placed in series, which produce an optical signal that will travel through free space until it reaches the receiver front-end.

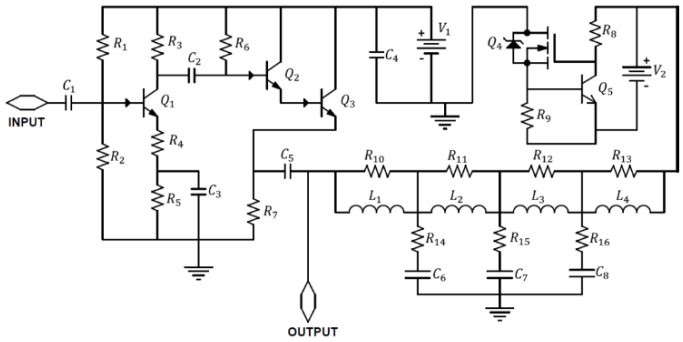


Figure 3: Schematic circuit of the Transmitter Front-end

Figure 3 shows the schematic circuit of the transmitter front-end, it can be seen that it includes a common emitter amplifier, that includes the transistor Q_1. After that, the circuit is followed by a Darlington amplifier, which is made of the transistors Q_2 and Q_3. The common emitter amplifier was designed following the procedure described in [11]. The design equations are as follows:

$$I_b = \frac{I_c}{\beta} \tag{1}$$

$$V_{ce} = \frac{V_1}{3} \tag{2}$$

$$R_3 = R_4 + R_5 \tag{3}$$

$$I_c(R_3 + R_4 + R_5) + V_{ce} = V_1 \tag{4}$$

$$R_3 = R_4 + R_5 = \frac{V_1 - V_{ce}}{2I_c} \tag{5}$$

$$R_{bb} = R_1 // R_2 \tag{6}$$

$$R_{bb} \ll \beta(R_4 + R_5) \tag{7}$$

$$Z_{in} = R_{bb} // \beta(r_e + R_4) \tag{8}$$

$$Z_{out} = R_3 \tag{9}$$

From (1) the base current is obtained from the value of the collector current, which in this case is 10 mA, in addition β has a value of 95. In (2) and (4), V_1 has a value of 12 V. The values of the resistors R_3, R_4 and R_5 are calculated from (4) and (5). On the other hand, (6) and (7) allow the calculation of R_1 and R_2. Next, (8) and (9) allow us to know the input and output impedances. Finally, a Darlington amplifier is inserted to reduce the output impedance and increase the output power.

In Figure 3, Q_4 is a MOSFET that provides a constant current source whose value is defined by (10). In this case, 140 mA was chosen as the I_D value because it allows the LEDs to keep turned on.

$$I_D = \frac{V_{BE}}{R_2} \tag{10}$$

I_D: drain current of MOSFET Q_4

V_{BE}: Base-Emitter Saturation Voltage of Transistor Q_5

LED's must be turned on to modulate their light intensity without turning them off. For this purpose, it is necessary to superimpose the AC signal coming from the amplifier with the DC level coming from the constant current source. The bias-tee is in charge of this task. Finally, this LED driver is connected to an array of eight LEDs in series. In this way, the Front-end transmitter has been designed.

4.2. Receiver Design

Figure 4 shows the schematic circuit of the receiver front-end, which is composed of a photodiode, a transimpedance amplifier and a low-pass filter. The photodiode makes it possible to transform an optical signal into an electrical one. In this project, the SFH 203 FA photodiode is used, which has a switching time of 5 ns, it is inversely polarized with a voltage of 12 V. Furthermore, R_3 and C_5 constitute a low-pass RC filter in order to eliminate noise from source V_1.

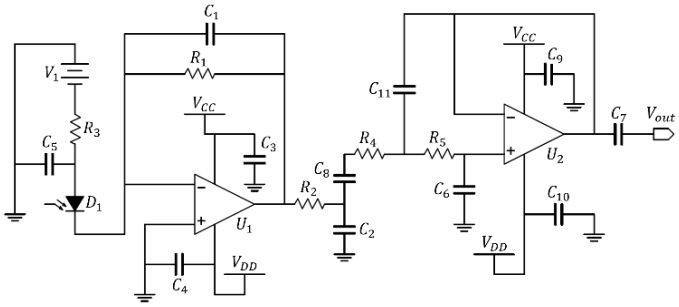


Figure 4: Schematic circuit of the Receiver Front-end

The current at the output of the photodiode enters a transimpedance amplifier. The gain of the transimpedance amplifier is equal to the feedback resistance R_1. After a voltage is obtained at the output of the transimpedance amplifier (TIA), the signal passes through a two-stage low-pass filter. The first stage is

made up of an RC filter composed of R_2 and C_2 and whose cutoff frequency can be calculated using (11). After that, the coupling capacitor C_8 is connected, this capacitor allows only the alternating component of the signal to pass. In this way, the voltage produced by the ambient light is minimized. Finally, a Chebyshev low-pass filter is responsible for eliminating high-frequency noise and adapting the output impedance for connection to the USRP equipment. The design of this filter was done following the procedure described in [12], the calculations were made using (12), (13), (14) and (15).

$$f_{c1} = \frac{1}{2\pi(R_2)(C_2)} \tag{11}$$

$$R_4 = R_5 = \frac{a_1}{2w_c C_6} \tag{12}$$

$$C_{11} = \frac{b_1}{{w_c}^2 R_4 R_5 C_6} \tag{13}$$

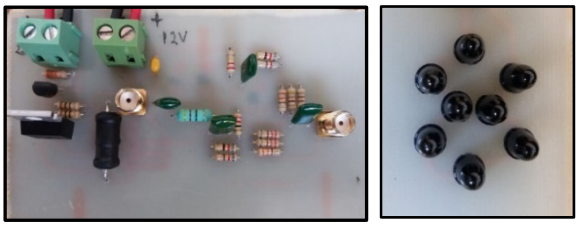
$$a_1 = 1.1813 \tag{14}$$

$$b_1 = 1.7775 \tag{15}$$

4.3. Implementation of Front-ends

The transmitter front-end was implemented as shown in Figure 5. The physical implementation of this circuit can be shown on two single printed circuit boards (PCB). Figure 5.a shows the LED driver. On the other hand, Figure 5.b is the IR-LED array. The reason why the LED array is not included in the same PCB is because it allows to do further tests with different LED configurations. In this way, it is avoided to build new front-ends in order to speed up the tests and improve the transmitting analog interface. Both circuits are connected by a coaxial cable in the experimental setup.

Figure 6 shows the physical circuit of the receiver front-end for the IR uplink. It is observed that the output connection is made through a female SMA connector, to which an RG58 coaxial cable is connected.



a) b)

Figure 5: Physical implementation of the transmitter Front-end.

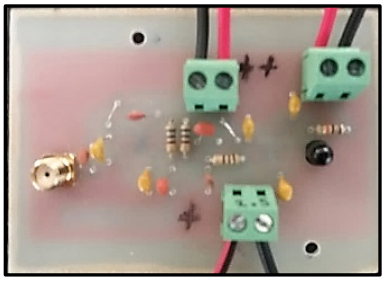


Figure 6: Physical implementation of the receiver Front-end

5. Measurements

Once the Front-ends have been implemented, the Bit Error Rate (BER) measurement is performed. To do this, it is necessary to establish the experimental setup shown in Figure 7. This setup is based on the block diagram described in the third section of this paper. The connection between the laptop and the USRP is made through a gigabit ethernet cable. The connections between the USRP and the front-ends are made through RG58 coaxial cable.

To carry out the BER measurement, the program shown in Figure 8 is used. This program was developed in GNU Radio from the information available in [6], with the difference that now QPSK modulation and the rate are used. sampling rate is changed to 4 MS/s.

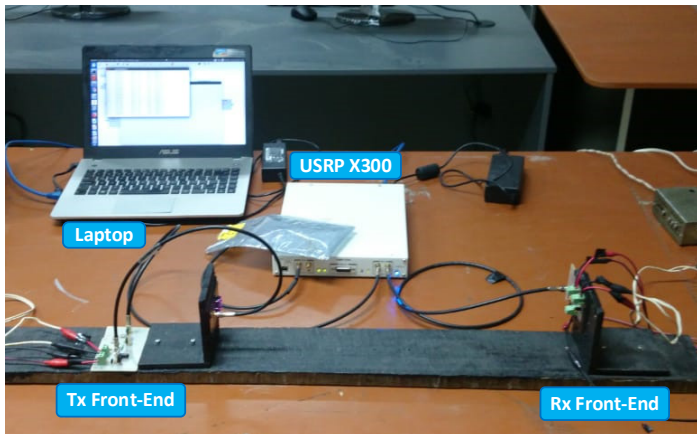


Figure 7: Experimental setup for BER measurement

The explanation of the blocks that make up this program is as follows:

- The "PRBS Source B" block is responsible for generating a random sequence of bits, which is repeated indefinitely.
- “Preamble Insert” is a block that inserts a bit preamble, which is used to indicate the moment at which the BER measurement starts.
- "Root Raiced Cosine Filter" is a block that is used in transmission and reception to minimize intersymbolic interference.
- “Costas Loop” is responsible for the frequency recovery of the carrier signal.
- Subsequently, the "Decode using Preamble Synch" block takes care of the demodulation and detection of the preamble to start the BER measurement.
- The “Binary Slicer” block transforms the demodulated signal into bits. These bits can have two values 0 or 1.
- Finally, the "PRBS Sink B" block is in charge of comparing the received bits with the original sequence and showing the bit error rate (BER).

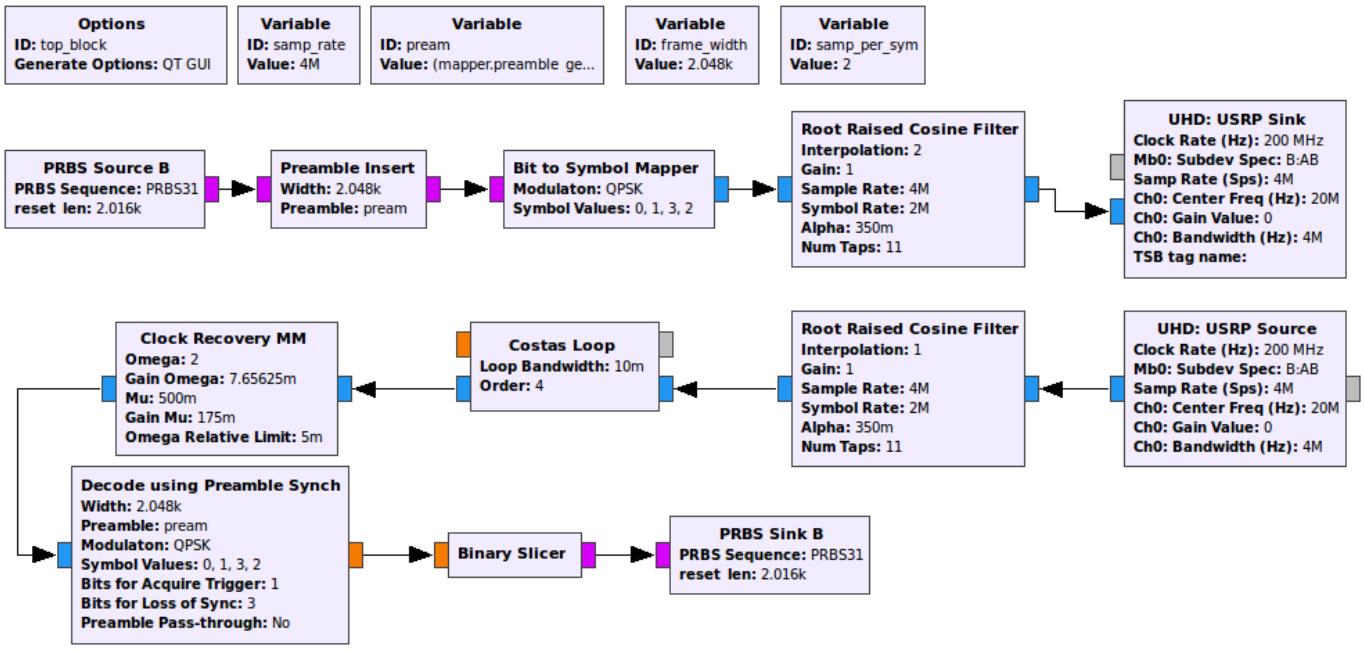


Figure 8: Program in GNU Radio for BER measurement using QPSK modulation

Figure 9 shows the results obtained from the BER measurement as a function of the separation distance between transmitter and receiver. The red graph indicates the BER measured using QPSK modulation at a bit rate of 4 Mbps. A BER of 5.33×10^{-5} was obtained at a distance of 160 cm between transmitter and receiver. The blue graph shows the BER obtained using BPSK modulation, which has a value of 1.35×10^{-5} at a distance of 160 cm and the bit rate is reduced to 2 Mbps.

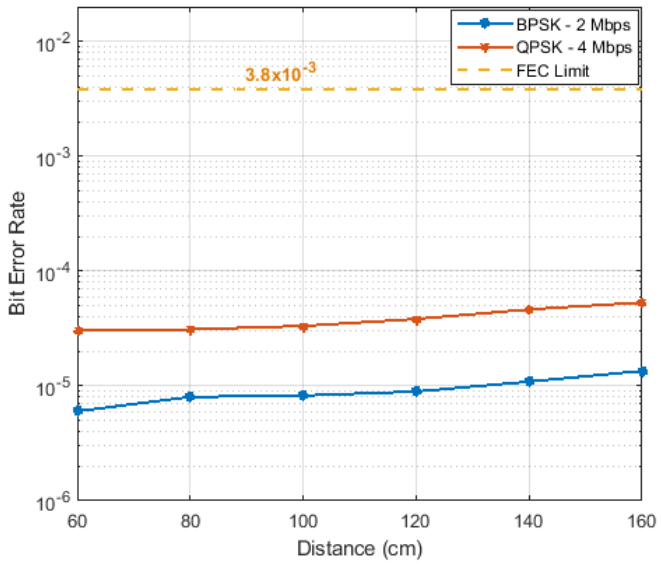


Figure 9: BER with respect to distance

Finally, signal-to-noise ratio (SNR) is measured at different distances between the transmitter and receiver using BPSK modulation. From the obtained data, it has been possible to make the graph of Figure 10, in which the BER is observed as a function of the SNR. With a SNR of 36.2 dB, a BER of 1.35×10^{-5} was measured. In the other hand, with a SNR of 46.4 dB, a BER of 8.032×10^{-6} was obtained. It can be seen that, if the SNR increases, a lower BER is measured. In this way, having an SNR greater than 45 dB, the BER seems to converge, however this is due to the fact that a finite number of packets were sent. Therefore, it is convenient to have the least amount of noise in the system or, failing that, to increase the power of the transmitted signal.

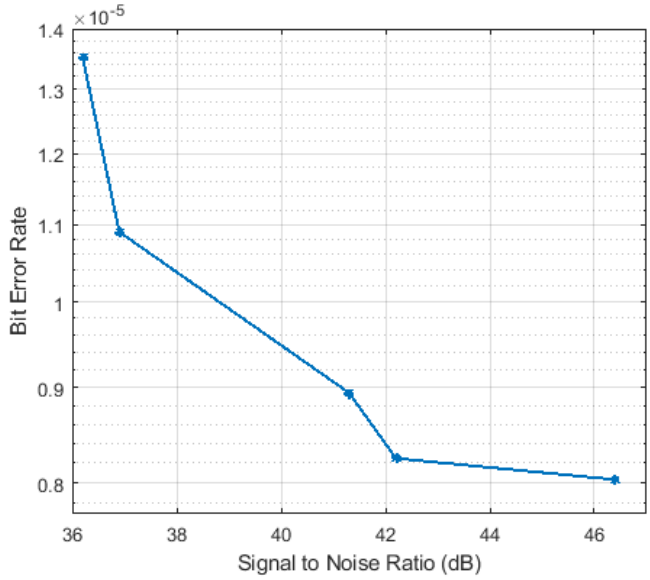


Figure 10: Measured BER vs SNR

6. Conclusions

- A bit error rate of 5.33×10^{-5} has been measured when transmitting at a bit rate of 4 Mbps using QPSK modulation at a distance of 160 cm between transmitter and receiver for an infrared light communication system.

- Signal to noise ratio decreases when the distance between transmitter and receiver is greater. In turn, the bit error rate increases as there is a lower signal-to-noise ratio.
- The transmitter front-end provides a constant current level of 140 mA, this keeps the LED stage turned on, while at the same it is modulated the intensity of the optical signal.
- The receiver front-end is composed by a transimpedance amplifier and a two-stage low-pass filter. The output voltage does not exceed 2 Vpp, which is the input limit for a USRP X300.
- A software program was developed in GNU Radio to measure the bit error rate of the infrared communication link. The used sampling rate is 4 MS/s and the central operating frequency is 5 MHz.
- Front-ends have been designed and implemented for infrared light communication that can be used in Software Defined Radio systems as an alternative for VLC uplink. Taking into account the requirements of this type of links.
- The main advantage of this system is that, as it is based on software defined radio, the modulation, coding and synchronization stages are implemented within a personal computer. Therefore, to implement it, it is necessary to use analog front-ends that allow the conversion of RF signals to optical signals.
- The main disadvantage is that the bit rate is limited by the capacity of the computer processor where the transmission and reception programs are executed, it was necessary to carry out tests with different personal computers and the one with the best performance was chosen.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgment

The author thanks to the National University of San Agustin of Arequipa for providing access to the laboratories to carry out the tests of this paper. On the other hand, the author also thanks to professors Alex Cartagena Gordillo, Pablo Lizardo Pari Pinto and Ruben Matheos Herrera.

References

[1] O.R. Banda Sayco, A.C. Gordillo, "Development of an IR system as an alternative for VLC uplink," in 2019 IEEE Colombian Conference on Communications and Computing (COLCOM), IEEE: 1–5, 2019, doi:10.1109/ColComCon.2019.8809131.

[2] M. Uysal, "Visible Light Communications: From Theory to Industrial Standardization," in Optical Fiber Communication Conference (OFC) 2019, OSA, Washington, D.C.: Th3I.4, 2019, doi:10.1364/OFC.2019.Th3I.4.

[3] Z. Wang, Q. Wang, W. Huang, Z. Xu, Visible Light Communications Modulation and Signal Processing, 2017.

[4] P. Kamsula, Desing and implementation of a bi-directional visible light communication testbed, University of Oulu, 2015.

[5] M. Basha, M.J.N. Sibley, P.J. Mather, A. Makama, "Design and Implementation of a Tuned Analog Front-End for Extending VLC Transmission Range," in 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), IEEE: 173–177, 2018, doi:10.1109/iCCECOME.2018.8658815.

[6] O.R.B. Sayco, A.C. Gordillo, "Design and Implementation for a USRP – Based Visible Light Communications Transceiver," in 2019 UNSA International Symposium on Communications (UNSA ISCOMM), IEEE: 1–5, 2019, doi:10.1109/UNSAISC.2019.8712826.

[7] O.Z. Alsulami, M.T. Alresheedi, J.M.H. Elmirghani, "Infrared Uplink Design for Visible Light Communication (VLC) Systems with Beam Steering," in 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), IEEE: 57–60, 2019, doi:10.1109/CSE/EUC.2019.00020.

[8] T. Adiono, A. Pradana, R.V.W. Putra, W.A. Cahyadi, Y.H. Chung, "Physical layer design with analog front end for bidirectional DCO-OFDM visible light communications," Optik, **138**, 103–118, 2017, doi:10.1016/j.ijleo.2017.03.046.

[9] T.C. Bui, S. Kiravittaya, "Demonstration of using camera communication based infrared LED for uplink in indoor visible light communication," in 2016 IEEE 6th International Conference on Communications and Electronics, IEEE ICCE 2016, 2016, doi:10.1109/CCE.2016.7562615.

[10] K. Wang, Z. Yuan, E. Wong, K. Alameh, H. Li, K. Sithamparanathan, E. Skafidas, "Experimental demonstration of indoor infrared optical wireless communications with a silicon photonic integrated circuit," Journal of Lightwave Technology, 2019, doi:10.1109/JLT.2018.2889252.

[11] Hunter College - City University of New York, 15. Transistor Amplifier Design and Measurement, Http://Www.Hunter.Cuny.Edu/Physics/Courses/Physics222/Repository/Files/Pdf/ElectronicsLab15.Pdf, 1–13.

[12] T. Kugelstadt, Active Filter Design Techniques, Elsevier: 365–438, 2009, doi:10.1016/B978-1-85617-505-0.00020-X.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 133-142 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Experimental Investigation of RC Footings Resting on Sand Strengthened with Concrete Jacketing

Mohamed Attia Fouda[*,1], Mahmoud Elkateb[1], Tamer Elkateb[2], Ayman Khalil[1]

[1]*Structural Engineering Department, Ain Shams University, 11311, Cairo, Egypt*

[2]*Geotechnical Engineering Department, Ain Shams University, 11311, Cairo, Egypt*

ARTICLE INFO

Article history:
Received: 25 September, 2020
Accepted: 29 October, 2020
Online: 08 November, 2020

Keywords:
Reinforced concrete footings
Stress distribution
Stress redistribution
Concrete jacketing
Bond
Dowels and Strengthening
Cohesionless Soil

ABSTRACT

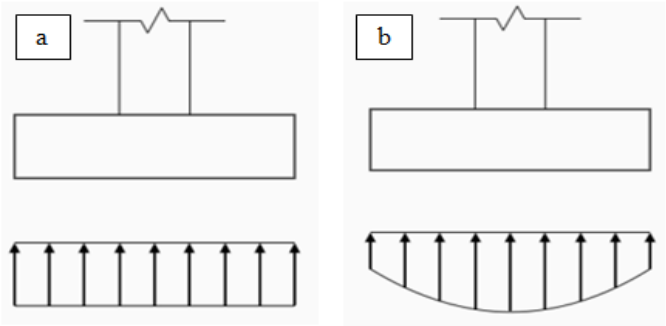
An experimental study regarding structural strengthening of existing reinforced concrete footing was conducted on six full scale reinforced concrete footings, strengthened with concrete jacketing under concentric and eccentric loads. The footings had constant dimensions and flexural steel reinforcement which was designed to prevent brittle failure due to crushing of concrete before steel yielding. Concrete jacketing was installed on all sides and top of existing concrete for footings subjected to concentric loading, while concrete jacketing was installed on three sides and top of existing concrete for footings under eccentric loading. The bonding between the existing and the new concrete was done either by applying epoxy bonding with dowels or by using an epoxy bonding agent on all sides only. This was to investigate the effect of each strengthening technique on the stress distribution beneath reinforced concrete footings rested on cohesionless soil. The dowels were used and designed to prevent any premature shear failure. The experimental results showed that the stress distribution beneath the center of footings subjected to concentric loading was higher than that beneath the control specimen by approximately 10%. The results also showed that the contact stress distribution beneath control and strengthened footings were not consistent with the available theoretical solutions. The effect of using dowels, in connecting old and new concrete, on stress distribution under strengthened footings and load-carrying are presented.

1. Introduction

Strengthening of existing Reinforced Concrete (RC) footings is often conducted when there is a need to increase the load-carrying capacity of such footings. Failure of RC footings is governed by bending, punching shear and one-way shear. The traditional way for strengthening, which is to increase the footing dimensions by adding a new layer of concrete on the top and sides of the existing footing, is an effective method for strengthening. The effect of the bonding technique, between existing and new concrete, on the stress distribution beneath strengthened footing is not known, and it is still based mostly on engineering judgment. This is due to the lack of sufficient research related to estimating stress distribution beneath strengthened footings. Bonding between existing and new concrete is considered as a main factor for load transfer, and for the concrete jacket to act as one part with the old concrete. The contact stress distribution beneath RC Footings resting on soil depends on the type of soil, as well as the stiffness of the footing [1]. Previous research regarding strengthening of existing RC Footing has focused on improving punching and shear behavior [2-5]. In 2017, Gia Toai Truong [2] investigated the effects of inclined shear reinforcement, high-strength concrete, and additional concrete cast with amorphous metallic fibers (AMFs) on the punching capacity of existing footings, the authors stated that the effect of AMFs and high strength concrete on punching resistance was higher than the effect of using inclined shear reinforcement. In 2007, J. Hegger [3] investigated the punching behavior for rigid isolated RC footings resting on sand for different footing depths. The authors also developed a design model for assessing punching strength using the soil-structure interaction, and it was found that the contact stress distribution beneath the rigid footings rested on sand was higher at the center of the footing and lower at edges. In 2013, Bonic and Folic [4] studied the punching behavior on RC footings, subjected to concentric loading, resting on sand. This study showed that the actual punching failure load was higher than the expected failure load, calculated according to the international

[*]Corresponding Author: Mohamed Attia, attia87_2009@hotmail.com

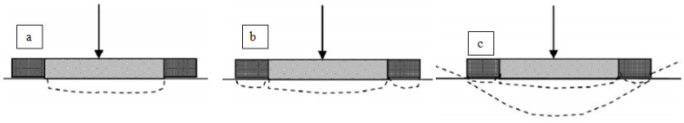
codes [5-7]. The authors stated that the theoretical assumption of uniform distributed stress under footing was conservative. In 2019, Wang-Xi Zhang [8] investigated the punching behavior of RC footings, subjected to eccentric loading, resting on rubber block. The results showed that the stress distribution was linear at low load levels, while it tends to be highly concentrated under the point of load application with load increase. Furthermore, the nonlinearly distributed stress influenced and increased the punching shear strength as well as the punching shear force of the footing. In 1990, Mosley [9] demonstrated the contact stresses distribution beneath a typical rigid footing resting on sand as shown in Figure 1.



In 2015, Todor Vacev [10] investigated the punching behavior of rigid isolated RC column footings, subjected to concentric loading, resting on sandy soil, and found that the stress distribution beneath the footing was different from the theoretical assumption. However, the stresses were concentrated under center of the footing. In 1996, Muller, G. U. [11] analyzed the contact stress distribution under rigid square footings, resting on cohesionless soil, and subjected to eccentric and concentric loading. It was found that the stress distribution was concentrated under center of the footing for eccentric loading, instead of being uniform linear triangular as in the theoretical analysis.

Various codes world-wide [5-7] do not quantify the resistance load of elements repaired with each Strengthened method. Moreover, in all published studies, the variable overloading punching effect is usually examined. Therefore, it is required to measure and examine the influence of the interface surface on the redistribution of the contact stresses.

Dowels and epoxy bonding agent were selected to connect the interface of the jacketed elements based on the outcomes from previous research [12], [13]. In 2013, M.N. EL Siragy [14] assessed the settlement profile, shown in Figure 2, and load-carrying capacity for rigid strip steel footings, resting on sand, after being strengthened with two steel plates from the sides. The authors found that the increase in the footing area led to reduction both the deformation and plastic flow of soil particles under the loaded footing.



2. Research Significance

The current codes and design guidelines [5-7] do not account for the effect of strengthening on the load-carrying capacity and stress distribution beneath strengthened footings, so, in this study, full-scale footings were tested to study the real load-displacement behavior and stress distribution, as the behavior of a small footing will be different than that of a large footing when rested on sand medium under the same gravity load [15]. The current study also investigates the effect of using concrete jacketing for existing RC Footing, using both dowels and epoxy bonding agent or applying epoxy bonding agent only at the interface between the old and new concrete.

3. Experimental Program

3.1. Geometry and detailing of tested samples

Six full scale square reinforced concrete footing specimens were tested under concentric and eccentric static loading till failure. These footings were divided into two groups; each group consists of three footings, a control specimen and two footings strengthened by concrete jackets (with/without) dowels. Table (1) summarizes the test groups dimensions and illustrates the strengthening methodology. Concrete jackets were used to study the effect of dowels and the bonding agent, on the whole bonding strength under eccentric and concentric loads.

Table 1: Test Matrix

Group	Symbols	Dimensions before strengthening (mm)	Dimensions after Strengthening (mm)	Loading Type	Strengthening Methodology
1	C1(control)	1200x1200x400	1200x1200x400	Concentric	---
	G1F1	800x800x250			Concrete jacket with shear dowels on top and sides of existing concrete.
	G1F2	800x800x250			Concrete jacket with shear dowels on top of existing concrete only.
2	C2(control)	1200x1200x400	1200x1200x400	Eccentric (50mm)	---
	G2F1	900x1000x250			Concrete jacket with shear dowels on top and sides of existing concrete.
	G2F2	900x1000x250			Concrete jacket with shear dowels on top of existing concrete only.

All reinforced concrete footings were tested on compacted dense sand inside a square wooden box with side length, and thickness of 2400 and 800 mm respectively. The box was strengthened with steel pipes from all sides. Figure 3-5 show the reinforcement details in all control and strengthened footings, Flexural reinforcement arrangement in the strengthened footings were constant and similar to the control specimens. Figure 6 shows the test setup for all footings. Each tested specimen was subjected to concentrated gravity load induced a vertical hydraulic jack, using load control mode, at which the load was increased by incremental values of 5 kN until failure load level was attained.

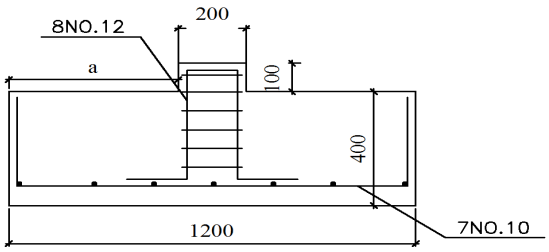


Figure 3: Reinforcement details for (a) a=500 mm for Control-1. (b) a =450 mm for Control-2.

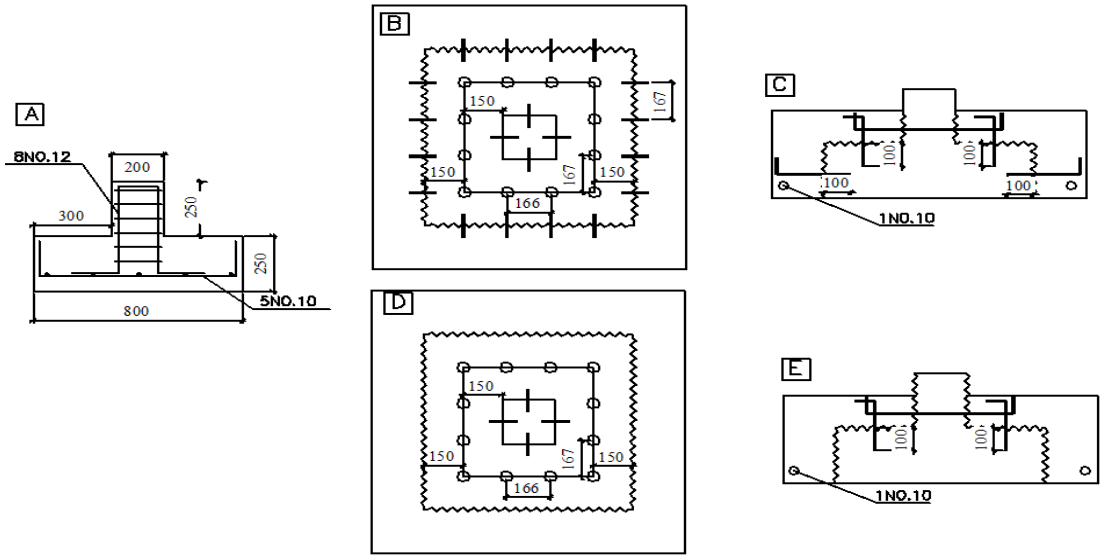


Figure 4: Reinforcement and dowel details for (a) G1F1and G1F2 before strengthening. (b) G1F1 after strengthening plan view. (c) G1F1 after strengthening elevation view. (d) G1F2 after strengthening plan view. (e) G1F2 after strengthening elevation view.

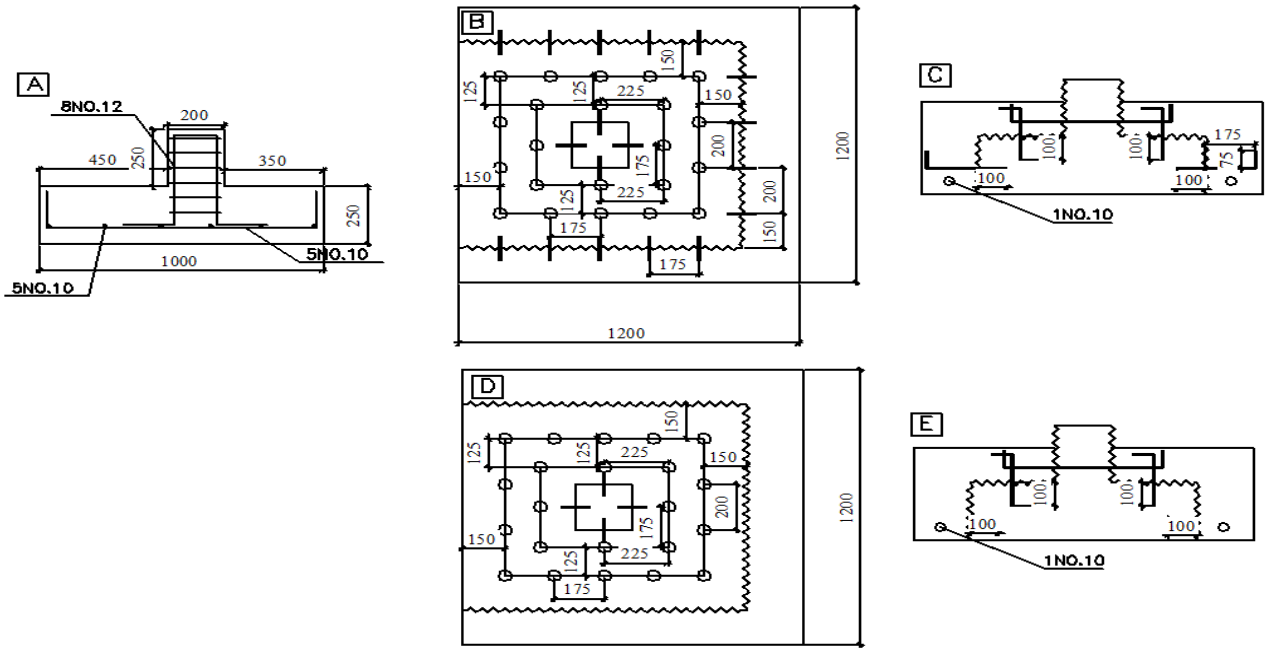


Figure 5: Reinforcement and dowel details for (a) G2F1and G2F2 before strengthening. (b) G2F1 after strengthening plan view. (c) G2F1 after strengthening elevation view. (d) G2F2 after strengthening plan view. (e) G2F2 after strengthening elevation view.

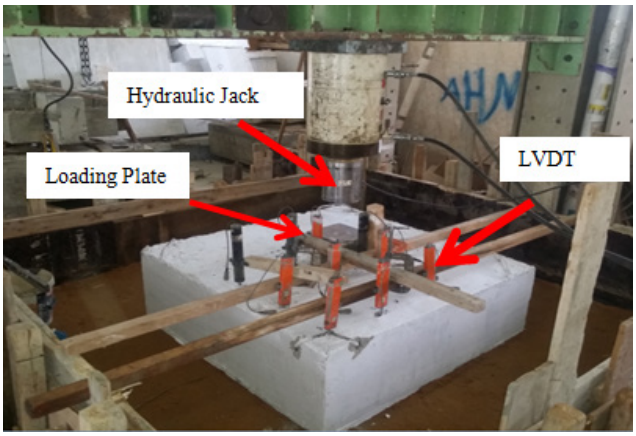


Figure 6: Experimental test setup.

3.2. Materials

3.2.1. Properties of Sand

Clean dense sand was used in the laboratory testing. The particle size distribution was determined through sieve analysis (D60=1, D30=0.4 and D10=0.2). The soil can be classified as well graded sand according to the Unified Soil Classification System (USCS) (ASTM Standard D2487, 2011) [16]. The density of the sand placed in the models was 20 kN/m^3, while the friction angle was 43.9°, the friction was relatively high as the roughness coefficient between soil and concrete footing is responsible for increasing the bearing capacity of soil [17]. The soil was prepared and compacted up in layers of 250 mm thickness with vibrator machine as shown in Figure 7, using 10.5% optimum water content.

Figure 7: Compaction process

3.2.2. Concrete

The strengthened footings were cast in two stages. Initially, the concrete ingredients were mixed uniformly with proper proportions of 350 Kg cement, 630 Kg sand, 1260 Kg crushed stone, and 175 Kg water. The platform was cleaned, so that no foreign particle could be mixed with concrete. Precautions were taken during the placing and transportation of concrete so that consistency remained in the mixes, and to prevent segregation. Concrete was thoroughly compacted around reinforcement using electric vibrator. After 40 days, when the initial casting of concrete reached its ultimate strength, the concrete jacket was applied on existing footings using sikadur-32 epoxy bonding agent. During the initial casting and concrete jacketing operations, concrete cylinders and cubes were cast from each type of concrete mix. The actual compressive and tensile strength of concrete of individual footings were determined by testing of the corresponding cylinders and cubes on the day of casting the concrete jackets and on the day of testing. The measured characteristic compressive strengths (f_{cu}) of original concrete, after 28 days, was 24.8 MPa, while it was 33 MPa for the column stub. The compressive strength of the concrete jacketing was 25.2 MPa, with a coefficient of variation (COV of about 8.2%) based on 12 cubes tested according to the British Standards [18].

3.2.3. Reinforcement

10 mm diameter deformed bars were placed at 50 mm cover as main reinforcement for all the footings and 12 mm diameter deformed bars were placed in column stub. The measured yield tensile strength (f_y) of bars was 407 MPa and the corresponding yield strain is 0.00205, while the ultimate tensile strength was 611 MPa based on tensile strength test according to ASTM A615 / A615M [19].

3.3. Surface preparation and installing dowel bars

The preparation of the surface was done when the original concrete footing specimens were 35 days old. The surfaces of the RC footings were roughed by hand-chiseling from all sides and top surface. Dowel bars of 10 mm diameter and 250 mm length rod were installed in original footings between the interfaces of the layered construction. These bars were embedded into the initial cast concrete by drilling after 40 days of casting with 100 mm embedded length and were installed using sikadur-31CF (Thixotropic epoxy resin adhesive). Figure 8 shows surface perpetration process, dowels drilling and placement in original footing and casting concrete jacketing.

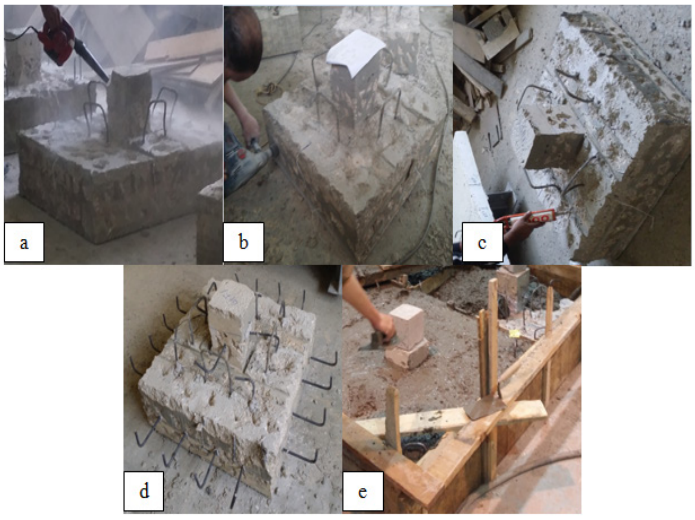


Figure 8: Preparation process (a) Roughening the concrete surface. (b) Drilling. (c) Installing epoxy. (d) Installing and placement dowels. (e) Casting new jacket.

3.4. Instrumentation acquisition system

The data acquisition system included two data acquisition modules for all the strains, displacements, and load cells. Ten

LVDTs were used to record the vertical displacement. Figure 9 shows the distribution of LVDTs on group-1 and group-2 respectively.

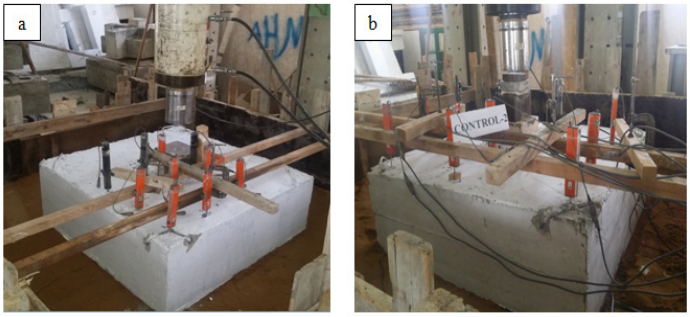


Figure 9: LVDTs distribution for (a) Group-1. (b) Group-2.

Concrete strain gauges were installed on the upper reinforced concrete footing on one direction only for group-1; strains were placed on to orthogonal direction for group-2 as shown in Figure 10).

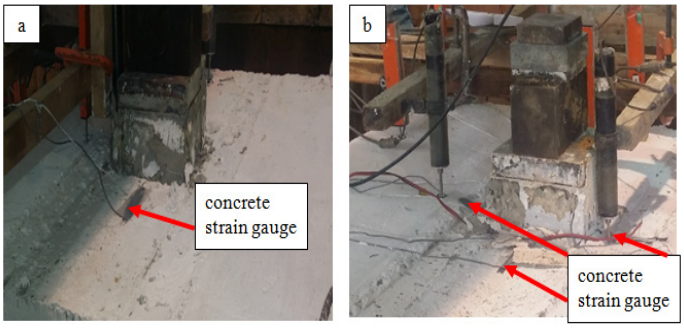


Figure 10: Concrete strain gauge distribution on (a) Group-1. (b) Group-2.

Reinforcement Strain gauges (6mm-length) were attached to steel rebar's in one direction for group-1 and in two orthogonal directions for group-2 as shown in Figures 11 (a and b) respectively. Figures 11 (c and d) show the distribution of additional reinforcement strain on the newly placed steel bars for strengthened footings G1F1, G1F2, G2F1 and G2F2.

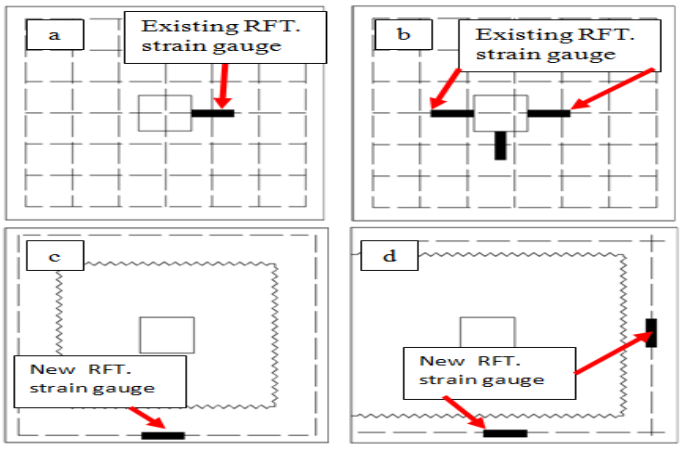


Figure 11: Reinforcement strain gauge distribution (a) Group-1. (b) Group-2. (c) Strengthening footing group-1. (d) Strengthening footing group-2.

Shear dowels, at typical locations, had strain gauges at their mid-length in longitudinal direction in specimen G1F1, while strain gauges were placed on shear dowels in orthogonal directions in specimen G2F1. Figure 12 shows a typical layout for strain gauge placed on shear dowels.

Figure 12: Typical reinforcement strain gauge placed on shear dowels.

Figure 13(a) shows a TY-350 pressure cell placed between the reinforced concrete footing and the soil surface to determine the contact stress distribution below the footing. Thus, it was decided to cover a pressure cell with thin steel plate to protect the pressure cell from damage. Four soil pressure cells were distributed below the reinforced concrete footing for groups (1&2) are shown in Figures (13) (b & c) respectively.

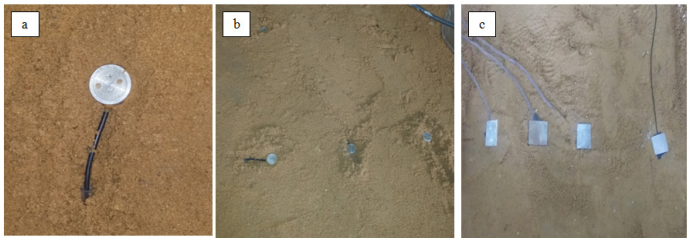


Figure 13: Distribution of load cell pressures beneath concrete footing (a) Typical TY-350 pressure cell. (b) (Group-1). (c) (Group-2).

4. Experimental Results

The application of concrete jacketing increased the ultimate load-carrying capacity of strengthened footings compared with those without strengthening. Moreover, the load-carrying capacity of G1F1 and G2F1 were higher than that of G1F2 and G2F2. However, the difference in the ultimate capacity was not significant. The increases in loading capacity are due to the increase in the concentration of stress at the center of the footing for footings in groups (1&2). The actual moment is calculated using compatibility of ultimate strains, where the location of the neutral axis is determined, and the moment is calculated by multiplying the tension force times the moment arm.

4.1. Failure modes and crack patterns

All footings failed in flexure due to steel reinforcement yielding, no brittle failure and debonding for concrete jacketing was observed. All side cracks were observed during test, while cracking on bottom surface of footings were observed at the end of the tests. The stiffness of the footings decreased with propagation of cracks at increased load levels. Figure 14 shows the typical crack patterns at failure for each footing. The behavior of the strengthened Footings G1F1, G1F2, G2F1, and G2F2 was close to the control specimens in terms of the crack initiation, crack pattern and failure modes whilst having different failure

loads. Increased strains in the flexural reinforcement were measured before failure. Experimental failure loads for different groups are listed in Table (2). Side cracks first appeared at bending moment level of about 50 kN.m in all footings and the corresponding load level were in the range of 700 to 800 kN. The theoretical moment at failure load, using the uniform distribution theory beneath footing rested on soil (Force/Area), was calculated as $WL^2/2$. From Table (2), it is noted that the actual moment at failure is equal to 82% of the theoretical moment which means the stress distribution under concrete footing resting on cohesionless soil was not consistent with the uniform distribution stress theory. Furthermore, the load-carrying capacity for all strengthened footing was higher than control samples, and the stress distributions under strengthened footing were the main reason for increasing the footing capacities.

4.2. Reinforcement and Concrete Strains

For all footings, compressive strains at the top side of the concrete were measured, in addition to the tensile strains of the lower main reinforcement. Figure 15 shows the applied load vs. the compressive strains of the upper side of the footing beside the column stub, showing a similar behavior for all footings under concentric loads. Regarding eccentric footings, the behaviour of the strains in the eccentric direction was different than that in the perpendicular direction, due to the concentration of stresses in eccentricity direction. Concrete crushing did not occur in all specimens and all compression strains did not exceed the plastic limit of 0.003. It was also noticed that the load-strain relationship was almost identical in the linear elastic stage.

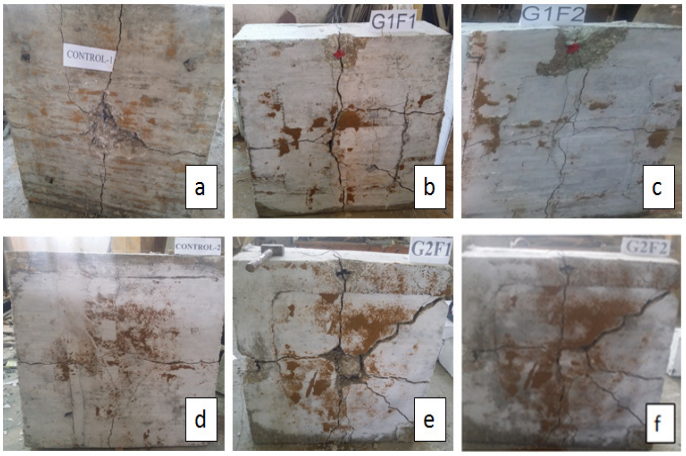


Figure 14: Typical cracking patterns at Failure for footings (a) C1. (b) G1F1. (c) G1F2. (d) C2. (e) G2F1. (f) G2F2.

Figure 16 shows the flexural reinforcement strains measured for concentric and eccentric footing. The strain gauges were installed within the column face region. All maximum measured strain values for existing reinforcements were above the yield limit which indicated that the failure occurred due to flexure, while all recorded strain values for dowels were below the yield strain limit, which indicates that the flexural debonding did not occurred during loading process.

Table 2: Test Results

Group	Label	Failure Load (kN)	% Increase in load-carrying Capacity	Theoretical moment at failure using uniform stress distribution (kN.m)	Actual moment at failure (kN.m)	Max vertical settlement at the center of the footing at control peak load (mm)
1	C1(control)	921	---	79.86	66.13	16.8
	G1F1	994.5	7	79.86	66.51	17.78
	G1F2	980	6	79.86	66.51	18
2	C2(control)	1000	---	79.1	66.5	17.4
	G2F1	1051	5.1	79.1	66.13	17.95
	G2F2	1020	2	79.1	68.24	18.1

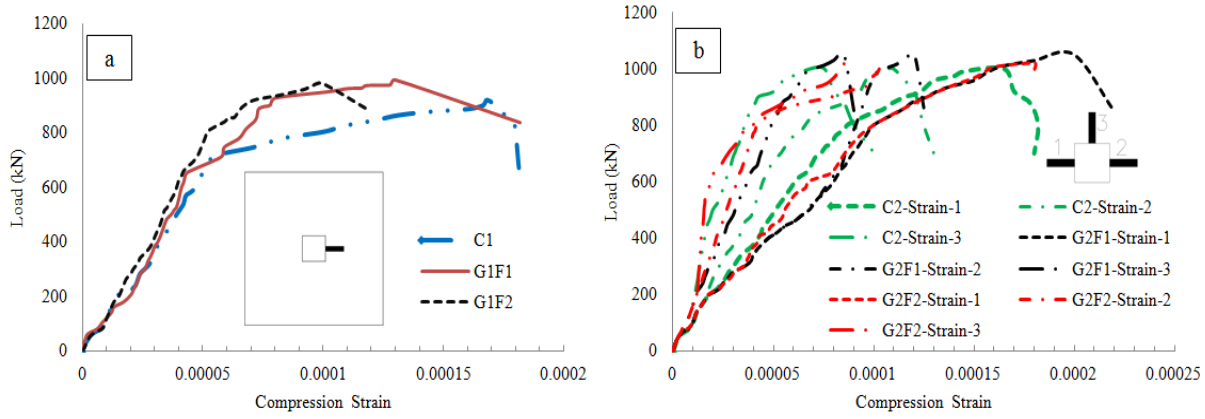


Figure 15: Applied load versus compressive strain (a) Concentric loading (b) Eccentric loading.

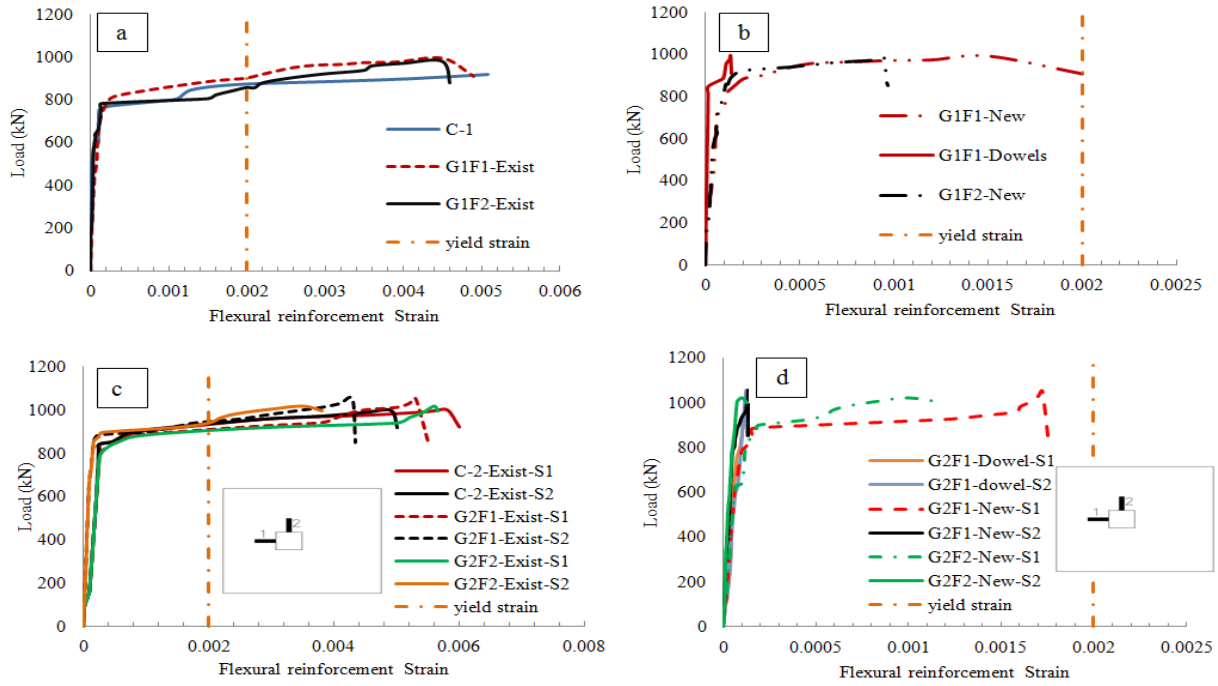


Figure 16: Applied load versus Flexural reinforcement strain (a) Existing reinforcement under concentric loading. (b) New reinforcement under concentric loading. (c) Existing reinforcement under eccentric loading. (d) New reinforcement eccentric loading.

4.3. Settlement under concrete footing

The settlement of the control and strengthened reinforcement concrete footing was determined at the same points along the width of the footing, as indicated in Figures 9 (a& b) for concentric and eccentric footings. The normalized average settlement versus the applied load of the reinforced concrete footing is illustrated in Figure 17. The behaviors of the concentric and eccentric footing were in good agreement until a load level of about 400 kN. After that, due to the eccentricity effect, a clear deviation in measured average settlement between group-1 and group-2 was noticed. Before cracking of the concrete, the settlement of all points along the width of the footing was similar. The rate in increase in average settlement after cracking was higher than that before cracking.

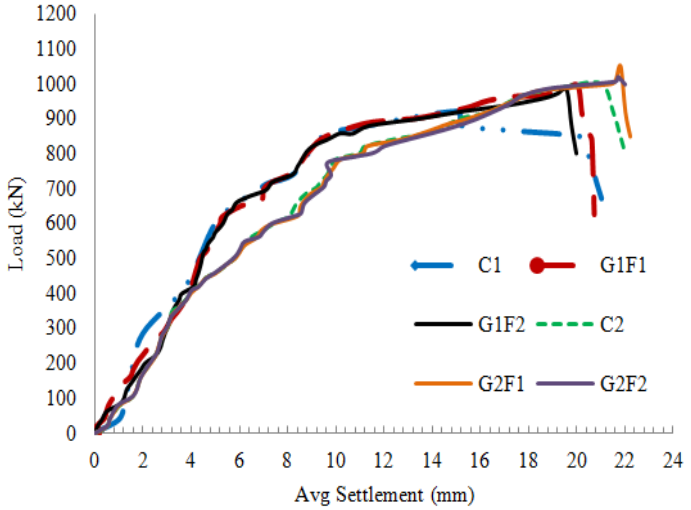


Figure 17: Applied load versus Average Settlement.

For the elastic stage, the behavior of all strengthened and control footings were similar and the elastic stiffness was almost the same according to the linear load-displacement relationship shown in Figure 17. After the concrete cracked, the behavior changed as the stiffness of the reinforced concrete footing decreased and the settlement significantly increased. Average settlements were not affected by using dowels compared with the footings strengthened with epoxy bonding agent only. The maximum settlement at the center of the footings increased with the increase in the contact stresses below the footings as shown in Table 2. The maximum settlement under strengthened footing G1F1, G1F2 were 5% and 7% higher than that of C1 respectively, while the settlement under strengthened footing G2F1, G2F2 were higher than that of C2 by 3% and 4% respectively.

4.4. Contact Pressure Distribution

The contact Stress distribution below control and strengthened concrete footing using concrete jacket with/without dowels were measured using pressure gauges as shown in Figure 13. Figure 18 shows the shape of contact stress distribution at center line of the footing for group-1 at 50% and 100% of the ultimate load of the control specimen C1.

Figure 18 shows the behavior of contact stress pre and post cracking. There was no crack initiation observed at 50% of the failure load of C1. The recorded stress at center line of footing C1 was close to the theoretical uniform stress distribution, but the stress at the edge was less than the theoretical assumption. The stress under center of strengthened footing G1F1 is higher than C1 by 9% and 12% in cases (a) & (b) respectively. Moreover, The stress under edge of strengthened footing G1F1 is lower than C1 by 16% in case (a) and 22% in case of (b) which means the contact surface between existing and new concrete play an important role in stress distribution under strengthened footing.

The stress under center of strengthened footing G1F2 is higher than that of C1 by 18% in case (a) and 13.5% in case (b). While, the stress under edge of strengthened footing G1F2 is lower than that of C1 by 23% and 25% in cases (a) and (b) respectively. The ratio between the center to the edge stress increased with increasing the applied load, which is due to that the effect of cracking. The flexibility of the footing increased with increasing the flexural cracks leading the concentration of the stress to be shifted towards the center of the footing. It is noticed that for the control footings in group-1, the shape of contact stress before cracking was saddle-shaped, but after cracking the shape of contact stress takes place as a parabolic shape for control footing, while the contact stress was a bell shape beneath strengthened footings. Figure 19 shows the shape of contact stress distribution at edge line of the footing for group-1 at 50% of C1 failure load and 100% of C1 failure load.

From Figure 19, it is noticed that the recorded stress at center line of C1 seems close to the theoretical uniform stress distribution assumption. The stress under center of strengthened footing G1F1 is less than that of C1 by 16% in case (a) and 22% in case of (b). Additionally, the stress under edge of strengthened footing G1F1of above figure is lower than that of C1 by 5% in case (a) and 22% in case of (b). The stress under center of strengthened footing G1F2 of the above figure is lower than C1 by 23% in case (a) and 25% in case of (b). Furthermore, the stresses under the edge of strengthened footing G1F2 are lower than that of C1 by 9% and 15% in cases (a) and (b) respectively.

Figure 20 shows the shape of contact stress distribution at the center line of the eccentric footing for group-2 at 50% and 100% of the failure load of the control specimen C2. The stress distribution under eccentric loading is concentrated under the column and completely different than the theoretical assumption (MY/I), these results were in agreement with the experimental results for non strengthened footings by G. U. Muller et al. [11]. The stress concentration under center line of column stub for G2F1 and G2F2 is higher than that of C2 by 1% and 5% respectively for case (a) and 2% and 5.6% respectively for case (b). Moreover, the stress concentration under footing edge at maximum compression zone G2F1 and G2F2 is higher than that of C2 by 1.5% for case (a), while it is lower than that of C2 by 2% and 6% respectively for case (b). The stress concentration under footing edge at minimum compression zone G2F1 and G2F2 is lower than C2 by 16% and 23% respectively for case (a) and lower than C2 by 18% and 25% respectively for case (b).

Figure 21 shows 3-Dimensional Contact stress distribution beneath control footings C1 and C2 at 50% and 100% of their failure loads using Surfer maps version 13.

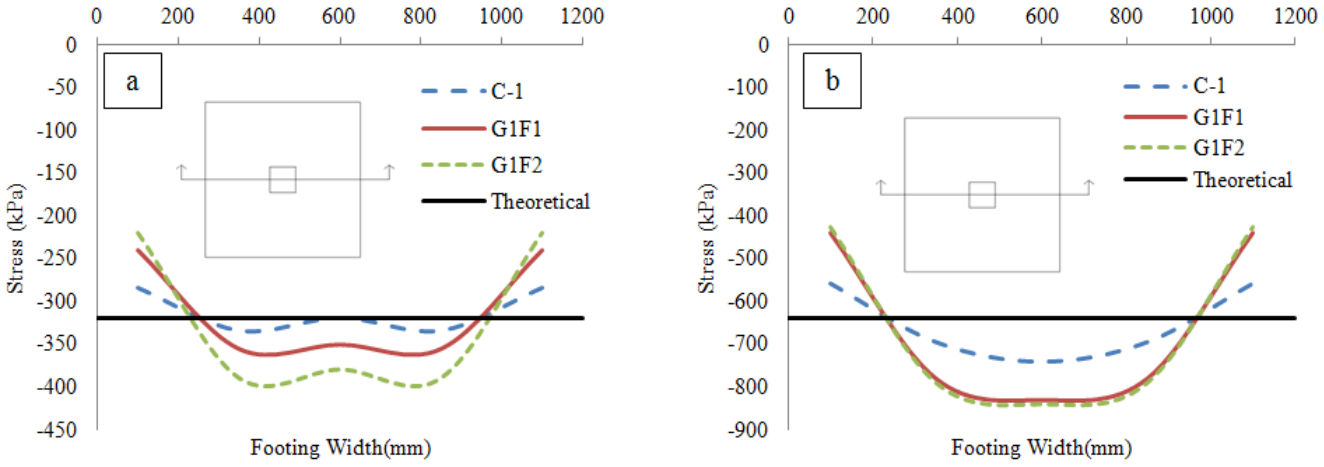


Figure 18: Contact stress distribution beneath footing center line (group-1) at (a) 50% of C1 failure load (b) 100% of C1 failure load.

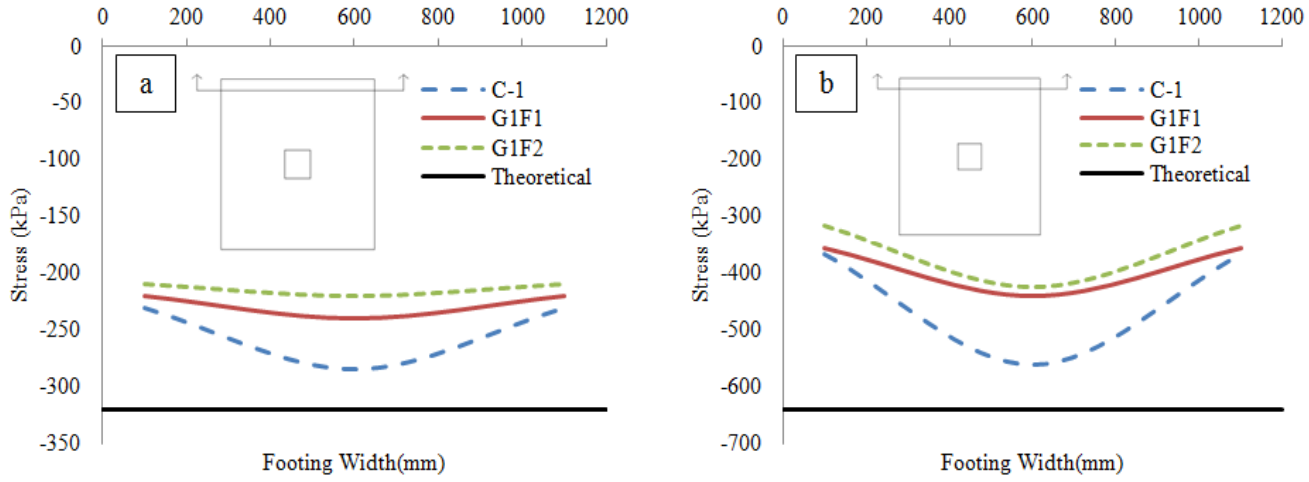


Figure 19: Contact stress distribution beneath footing edge line (group-1) at (a) 50% of C1 failure load (b) 100% of C1 failure load.

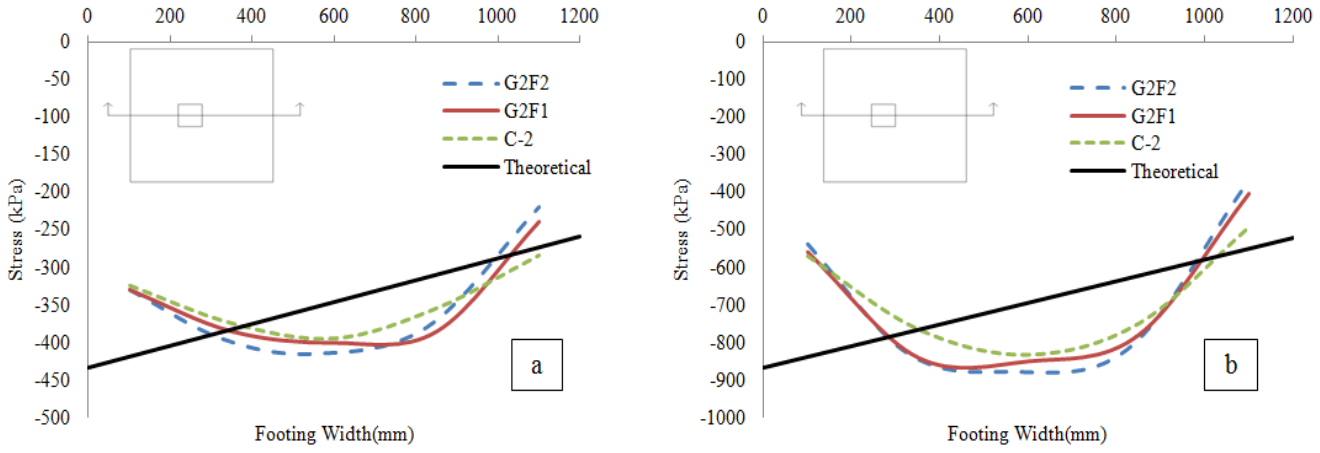


Figure 20: Contact stress distribution beneath footing center line (group-2) at (a) 50% of C2 failure load (b) 100% of C2 failure load.

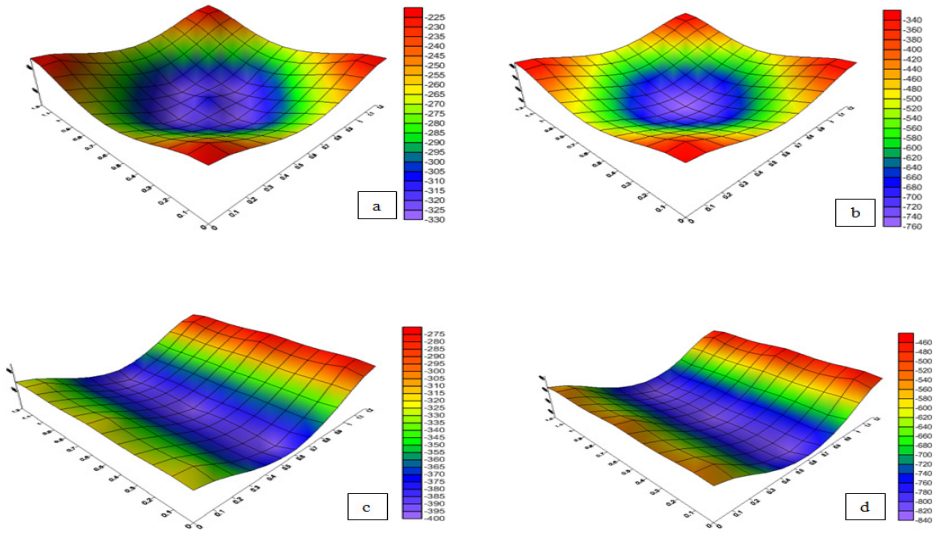


Figure 21: 3-D demonstration the contact stress distribution beneath control footings C1 and C2 (a) 50% of C1 failure load. (b) 100% of C1 failure load (c) 50% of C2 failure load. (d) 100% of C2 failure load.

5. Summary and Conclusions

In the current study, six isolated concrete footing were tested to evaluate the structural performance after strengthening for concentric and eccentric footing. From the test results, the increase in load-carrying capacities was calculated, the stresses and settlement distribution beneath control and strengthened footings were measured and the effect of using dowels in addition to epoxy bonding agent was evaluated an compared to the effect of using epoxy bonding agent only. The following conclusions are summarized:

- In both concentric and eccentric loading, the load-carrying capacities of strengthened footings were close to that of non-strengthened ones with same dimensions.
- Using dowels to connect the existing footing to the concrete jacket did not increase the load-carrying capacity significantly compared to using bonding agent only.
- The contact stress distributions beneath control and strengthened footings were not consistent with the theoretical solutions, the assumption of uniformity solution gives expected higher moment than the actual results and leads to a conservative structural design.
- The stress concentration under column center was increased by average 10% and 15% for concentric and eccentric jacketing respectively compared to the control specimens.
- The increase in settlement after jacketing was approximately 6% and 3.5% for concentric and eccentric jacketing respectively compared to the control specimens.
- The contact stresses seem to be saddle-shaped before cracking and the concentration of stresses tend to be shifted towards the center of footing after cracking during concentric loading.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] H.M. Algin, "Practical formula for dimensioning a rectangular footing," Engineering Structures, **29**(6), 1128–1134, 2007, doi:10.1016/j.engstruct.2006.08.009.

[2] G.T. Truong, K.K. Choi, H.S. Kim, "Punching-shear behaviors of RC-column footings with various reinforcement and strengthening details," Engineering Structures, 151, 282–296, 2017, doi:10.1016/j.engstruct.2017.08.037.
[3] J. Hegger, M. Ricker, B. Ulke, M. Ziegler, "Investigations on the punching behaviour of reinforced concrete footings," Engineering Structures, 29(9), 2233–2241, 2007, doi:10.1016/j.engstruct.2006.11.012.
[4] Z. Bonić, R. Folić, "Punching of column footings - Comparison of experimental and calculation results," Gradjevinar, **65**(10), 887–899, 2013, doi:10.14256/jce.916.2013.
[5] P.J. Gusia, "A new European Standard on products and systems for Protection and Repair of concrete structures – EN 1504," IABSE Symposium Report, **93**(22), 9–14, 2011, doi: 10.2749/222137807796120067.
[6] A.C.I. Committee, 318-19 Building Code Requirements for Structural Concrete and Commentary, 2019, doi: 10.14359/51716937.
[7] CSA/A23.3, Canadian Standard Association, Code for the design of concrete structures for buildings, CSA/A23.3-14, Toronto, Ontario, Canada, 2014.
[8] W.X. Zhang, B. Li, H.J. Hwang, J.Y. Zhang, L.J. Xiao, W. jian Yi, H.G. Park, "Punching shear strength of reinforced concrete column footings under eccentric compression: Experiment and analysis," Engineering Structures, **198**(August), 2019, doi:10.1016/j.engstruct.2019.109509.
[9] W.H. Mosley, J.H. Bungey, Design of Reinforced Concrete Design, Macmillan Education UK, London: 154–191, 1990, doi:10.1007/978-1-349-13058-0_7.
[10] T. Vacev, Z. Bonić, V. Prolović, N. Davidović, D. Lukić, "Testing and finite element analysis of reinforced concrete column footings failing by punching shear," Engineering Structures, **92**, 1–14, 2015, doi:10.1016/j.engstruct.2015.02.027.
[11] Muller, G.U. "Photoelastic Determination of Contact Stresses of Foundations", Journal of Geotechnical Engineering, **122**(8), 692–696, 1996, doi: 10.1061/(ASCE)0733-9410(1996)122:8(692)
[12] E.N.B.S. Júlio, D. Dias-da-Costa, F.A.B. Branco, J.M.V. Alfaiate, "Accuracy of design code expressions for estimating longitudinal shear strength of strengthening concrete overlays," Engineering Structures, **32**(8), 2387–2393, 2010, doi:10.1016/j.engstruct.2010.04.013.
[13] E.N.B.S. Júlio, F.A.B. Branco, V.D. Silva, "Concrete-to-concrete bond strength: Influence of an epoxy-based bonding agent on a roughened substrate surface," Magazine of Concrete Research, **57**(8), 463–468, 2005, doi:10.1680/macr.2005.57.8.463.
[14] M.N. EL Siragy "Improving the ultimate capacity of loaded strip footing using additional contact area under excessive loads", Life Sci J, **16**(12), 147-153, 2019, doi:10.7537/marslsj161219.19
[15] Aiban, S. A., & Znidarčić, D. "Centrifugal Modeling of Bearing Capacity of Shallow Foundations on Sands", Journal of Geotechnical Engineering, **121**(10), 704–712, 1995, doi: 10.1061/ (ASCE) 0733-9410(1995)121:10(704)
[16] ASTM Standard D2487. "Standard Practice for Classification of Soils for Engineering Purposes", Unified Soil Classification System, 2011, doi: 10.1520/D2487-11.
[17] T. Gnananandarao, R.K. Dutta, V.N. Khatri, "Model studies of plus and double box shaped skirted footings resting on sand," International Journal of Geo-Engineering, **11**(1), 2020, doi: 10.1186/s40703-020-00109-0.
[18] British Standard Institute, "BS EN 12390-3:2001 Testing hardened concrete - Part 3: Compressive strength of test specimens," BSI Standards Publication, 4–10, 2001.
[19] ASTM International, "Standard Specification for Deformed and Plain Carbon-Steel Bars for Concrete Reinforcement," Aashto, M31, 6, 2016, doi: 10.1520/A0615.

Notations

- f_{cu}: concrete Compressive Strength.
- f_y: Steel Yield Stress.
- RC: Reinforced Concrete.
- W: uniform distribution under footing (Force/Area).
- L: Distance between edge of columns and edge of footing.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 143-147 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Electro-tactile Stimulation for Augmenting Finger Motoric Learning

Daniel Sutopo Pamungkas[*,1], Arjon Turnip[2]

[1]*Mechatronics Department, Politeknik Negeri Batam, 29432, Indonesia*

[2]*Department of Electrical Engineering, Universitas Padjadjaran, 45311, Indonesia*

ARTICLE INFO

Article history:
Received: 28 July, 2020
Accepted: 30 October, 2020
Online: 08 November, 2020

Keywords:
Finger training
Haptic feedback
Electro-tactile feedback
Piano learning

ABSTRACT

Finger skills are achieved through the training process. This training process will be more straightforward if a training system is used. A novel finger motoric training system is introduced in this paper. This system is to facilitate a beginner piano player to learn the basic piano skill. The system is comprised of a tracking finger sensor, a personal computer, and multimedia feedback. Subjects train with passive and active mode. The passive mode, the subject's fingers are doing nothing when the information is given, whereas the subjects are playing a virtual piano under the sensor. While the information which fingers have to tap are provided by visual, audio, and enhanced with electrical haptic sensation. The experimental results demonstrate that the system equipped with electro-tactile feedback makes the subjects more responsive to both modes. Moreover, this system enabled the user to accelerate learning a new skill.

1. Introduction

Motoric skill learning can be trained with repetition, also utilizing stimulants from the surrounding environment [1]. Such motoric skills, for example, are typing, writing, and playing a musical instrument. To acquire dexterous motoric skills, practice with repetitive movements is needed. Moreover, these practices need persistence from the instructors and the learner. On the other hand, a stimulus can be used to augment the learning process.

Visual and audio information are commonly used as stimulants. Learners are showed how to do the skill accurately. In this process, they are played, or listened to, the appropriate sound. A haptic sensation can be applied to the learner using mechanical interaction. Some researchers utilize these stimuli to enhance the method to train the motoric skills, for example, playing the piano [1, 2], writing Japanese characters [3], writing Arabic and Japanese characters [4], and typing the Braille alphabet [5]. However, these systems are cumbersome, and the user finds it difficult to move their fingers.

To overcome these shortcomings, electrical stimulus is used. This information is given to the skin of the user. The electro-tactile feedback system is proven to become feedback used for controlling mobile [6], arm robot [7] also fingers [8], information for haptic sensors [9], enhanced virtual game [10]. To examine the potential usage of electrical signal, a multimedia learning system for a motoric finger is established. This paper is an extension of work initially presented in the 2019 International Conference on Sustainable Engineering and Creative Computing (ICSECC) [11]

For the experiment, subjects are divided into three groups. They are trained using different methods. Information in the form of audio and visual means are given to the first group to train the finger movements of each subject. In the other groups, the subjects utilized audiovisual means, enhanced with electrical sensations to practice. These groups consist of subjects which do the experiments with active mode and subjects with passive mode. The results show that the group using an electro-tactile feedback system is capable of playing the music appropriately with both active and passive modes. Moreover, they learn more rapidly than the other group. Besides that, the proposed system is simple and easy to use. The haptic sensation also does not cause pain to the subject.

This paper is organized as follows: Section II delivers a brief overview of related works within this field of study. Section III presents the details of system implementation. Section IV describes the experiment, which demonstrates the effectiveness of

[*]Corresponding Author: Daniel S Pamungkas, Politeknik Negeri Batam, daniel@polibatam.ac.id

the system for finger motoric learning. Finally, Section V delivers the concluding remarks.

2. Background

The process or action to achieve an ability to do, or to gain knowledge, is known as a learning process. The actions done in this process require stimulation from the surroundings. Moreover, repetition is needed in the learning process [12]. In this process, one thing can be connected to others, or known as associate learning. Stimulus or information is given repeatedly, resulting in a reaction, until the reaction can be an automatic response and become habitual. This stimulus can be associated with a reaction. Pavlov shows this condition with his experiments, in which the subject dogs have a new behavior. The dogs are able to associate the sound of a metronome with food [13]. Some physical training can be involved in associate learning, such as finger movements.

The skilled fingers motions are involved for a specific action, such as typing a computer keyboard or playing the piano. These finger skills are challenging to learn using only verbal instructions. Such training will be effective and efficient if the learner does a lot of exercises. Some software have been developed to assist people to learn to type in a keyboard or to play the piano. These software are intended to support the user in typing or playing the piano with all fingers. Visual stimulation is used by these software, for example, colors or lights. Other programs utilize audio information, e.g., the sound of the notes. Moreover, some of the applications apply both audio and visual information. However, another sense has excellent potential to transmit data, namely the sense of touch or tactile taste, which uses the skin to deliver stimulation.

Human skin is sensitive to several sensations, for example, vibration, heat, force, and electric signal (voltage or current). The skin has some sensory receptors, namely mechanoreceptors, thermoreceptors, nociceptors, and proprioceptors. These receptors are to accommodate all the sensations. Moreover, these receptors also enable humans to feel texture, paint, etc. [a]. Tactile sense causes humans to have the ability to detect and classify objects through their skin. This sense is sometimes called haptic sensations.

The term *haptic* comes from the word 'hapto' from Greek. This word means ‘to touch and handle objects’ [14]. Several researchers introduce haptic information utilizing vibration [15], heat [16], force [17], and electric sensations [6]. These information aims to enrich the user to interpret or interact with the application more effectively.

Haptic feedback has been used widely in various applications. A simulator to learn how to perform a surgery [17], remotely controlling a robot’s actions [18], a tactile sensor to sense the roughness of a surface [9] and gaming experience using VR [10] are the examples of the application of this feedback. Also, there are applications for training fingers skills using haptic feedback. For instance, haptic feedback is used for learning to play the piano [1,2], and using a Braille keyboard [5]. Moreover, this feedback is combined with other feedback, namely visual and audio, for writing characters [3] [4].

One of the rehabilitation procedures for a person with spinal cord injury (SCI) is to learn how to play the piano. For this training, subjects are equipped with a vibrating glove. This glove is equipped with a vibration motor on every finger. The vibrations help the subject to know which finger has to strike the keys of the piano [1]. The subjects are required to move their fingers while training. Such a method is called an active mode. Markow et al. report that this system enables the subject to play an easy song.

Others report that a passive mode training using vibration cues has positive results. The passive mode method is the method in which the subject trains without actually doing the activity about which they learn. The learning process is given using their senses. In [2] the subjects are trained to play a song with a piano using a vibration glove. The audio and haptic feedbacks are given to the subjects while they are doing the activity. Subjects are listening to the song, while the vibrations in their fingers give information which finger has to move related to the song. These experiments show that this information, including haptic feedback, enables the user to play the piano reflexively.

In [5], active and passive modes are used to train blind subjects to type in a Braille keyboard. This system has a keyboard that has the capability to train the user to use audio and vibration. The vibration information is given to the user's fingers. The subjects are able to use both hands to learn using this system. In these studies, the experiments are done with active and passive modes. The authors argue that both methods using audio and haptic information can be used to learn to type.

Another haptic information is force feedback. One of the devices able to give kinesthetic feedback to the user is called The Phantom. This device is used to train the learner to write Japanese [3] and Arabic [4] characters. This device helps the hand of the user to write the characters correctly. Both studies show that the force feedback is able to assist the subjects in training to write several newly-learned characters efficiently.

Mechanical feedback, such as vibration and force feedback, is able to transmit information for finger skill training. Haptic feedback has a great potential to be a training method, however it is cumbersome and costly. These devices hinder the movement of users in typing on a keyboard or playing the piano.

One type of haptic feedback is electro-tactile feedback. This applies an electric current to forward information to the skin. This feedback gives sensation to the skin and directly to the nerves. Electrotactile feedback system is simple and comfortable to use. Data from electric sensations can be informed through differences in amplitude or frequency. Thus, electrotactile feedback has various ranges of information that can be transmitted to the user [6]. In the next section, detailed methods to demonstrate the potential of electrotactile feedback are presented.

3. System

3.1. Overview

The proposed system for finger training is a computer with a screen and sound system, in addition to a gesture sensor, also with an electro-tactile feedback system. The electro-tactile feedback has different modes to vary the frequency, amplitude and the width of the electrical pulse. These variations have to be adjust to the subjects wearing this device. This system is comfortable to wear and does not hinder the user’s movements [6-11].

The frequency, amplitude and pulse width have a wide range of options, the frequency was varied between 10Hz to 120Hz. For this experiment stimulus pulses are set at 20Hz frequency (Figure 7). The amplitude of the pulses is adjusted by the user between 40-80V to achieve a comfortable condition. The intensity is controlled by the pulse width. The width of signal is varied, particularly between 10 to 100μs.

This system can be seen in Figure 1. Figure 2 shows the diagram block of the system. The screen of the computer displays the character of the note and which finger(s) to bend, while the corresponding frequency of the note is sounded using the computer's sound system. The electrodes in the fingers of the subject are transmitting the electric sensation. These sensations are given to the fingers appropriate to the note. The gesture sensor is used for tracking the fingers of the subjects. This sensor is placed near the top of the screen facing down to detect the movement of the subject fingers.

The learning is done by a passive and active method. The audio, visual, and haptic information are given to the subject while learning in the passive method, whereas the hand tracking device with the feedback is used when active learning is performed. In the test period, only a hand gesture sensor is used with the computer.

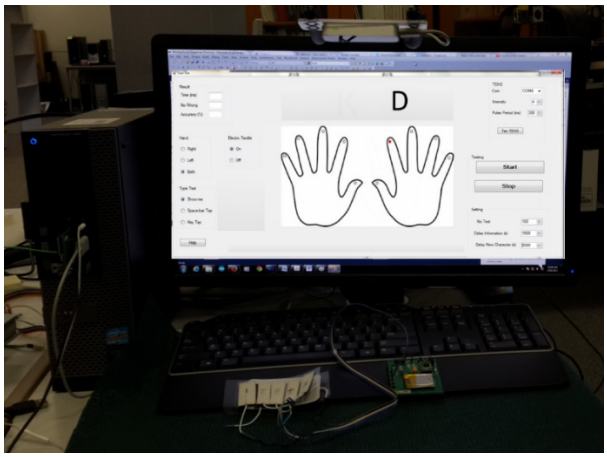


Figure 1: Arrangement of the system

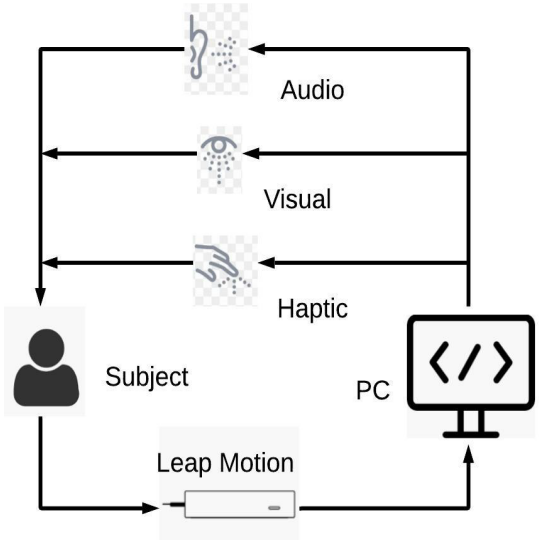


Figure 2: The Diagram block of the system

3.2. Hand tracking and software

One of the main components of the proposed system is the hand and fingers tracking sensor. For these tasks, a leap motion sensor is used. Leap motion is able to detect fingers and hands accurately from around one meter in three dimensions. The accuracy of this sensor is around 0.1 mm. This device is equipped with two infra-red cameras and three infrared LEDs [19]. This sensor has two modes. The first mode faces upward; the hands are above the leap motion sensor, known as a desktop mode. In the second mode, the HMD, the leap motion faces downward. HMD mode is used for this system. A custom bracket is utilized for placing the sensors. In this position, it is possible to have noise from reflective surfaces, so a dark mat is placed to cover the desk. The mat is shown in Figure 1.

The finger training multi senses system has several modes to train the subjects. The first mode is training using three senses; visual, audio, and haptic. The second mode uses visual and audio information, and the third uses audio and haptic information. The first and the second modes will be used to train the subject with the active method, while the third one is used to train with the passive method. The active mode requires subjects to tap the fingers below the gesture sensor.

The software is able to detect the correct or the wrong fingers. The function of this feature is to test the result of the training. The GUI of the system can be seen in figure 3. Moreover, the system is able to train the subject to play piano with only the right hand/notes only, or with both hands (notes and simple chord).

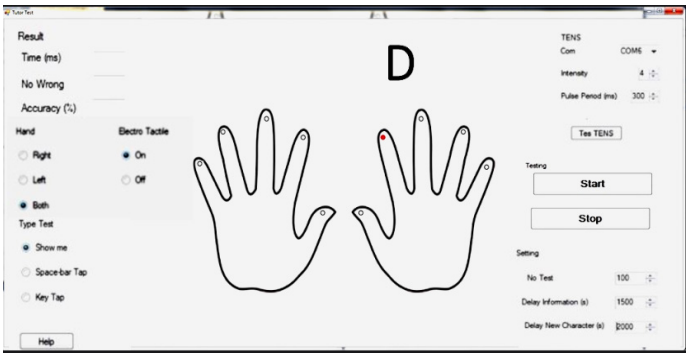


Figure 3: GUI of the system

The software is developed to display the characters of the note. Also, this software shows the appropriate fingers that have to bend. Figure 4 shows the fingers as appropriate to the note in a piano for the left hand (Figure 4.a) and right hand (Figure.4.b). The left-hand fingers of the human are placed in the a lower octave than the right hand. The left hand is trained to press the fundamental note of the chord. For example, if we want to play the C chord (C-E-G) the subjects are trained to press C, and so on such as in Figure 4.a. The left hand and the corresponding notes of the chord are shown in Figure 4.b. The right hand of the subject is trained to play the notes from C to G. Using this arrangement, this software is able to instruct the user to perform simple music. The user can play with only the right hand (melody only) or with both hands (melody with simple chord).

This software also has the capability to produce sound of note. This voice corresponds with the finger as defined in Figure 3. The

output of this information is transmitted to the computer sound system. Moreover, the haptic feedback augments the information for the user. Electric signal is transmitted to the nerve using electrodes patch. Fig 5 shows the illustration of the back of human palms. All the human fingers of both hands are placed the electrode. The software has facility to send electrical sensations appropriate to the visual and audio information. The next section will provide the detailed explanation of the electro-tactile feedback system.

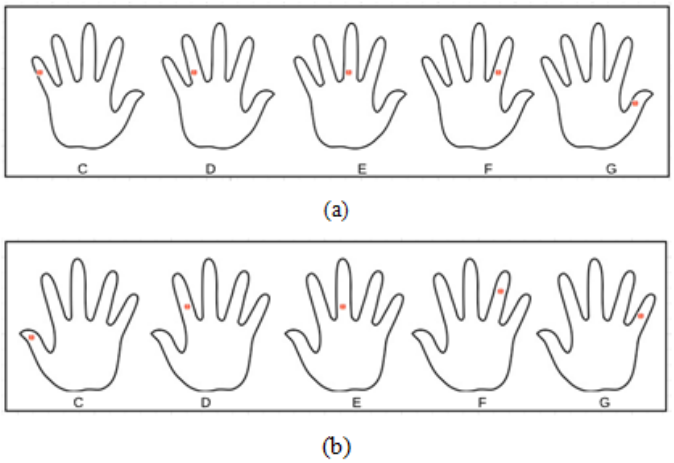


Figure 4: The corresponding fingers with the notes (a) left hand (b) right hand.

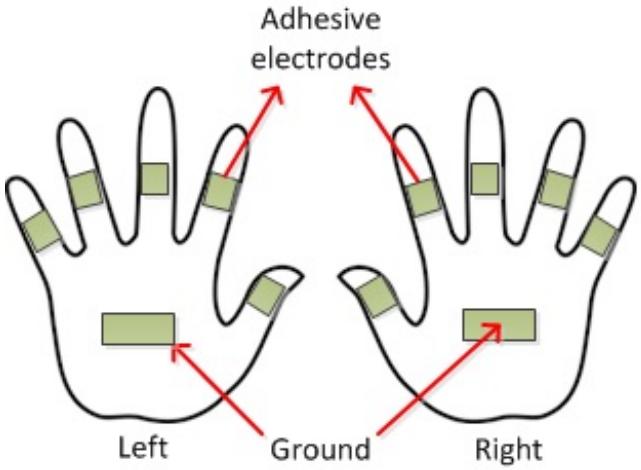


Figure 5: The electrodes in hands of the subject

3.3. Haptic system

The electro-tactile feedback system, which we call TENS (Transcutaneous Electrical Nerve Stimulation) system, comprises of three main components, namely transmitter, receiver and electrodes. The first component is the transmitter; the function of which is to transmit the signal from the computer to the receiver wirelessly. The main element of this component is a Bluetooth transmitter. The receiver receives the information from the transmitter and transformed it into electrical pulses. The five channels receivers with one ground are used. For the experiments, two transmitters are used for both hands. This structure enables the subject to receive ten signal frequencies. The last component is the electrode, placed in the knuckles, and a ground is placed in the middle of each hand. This arrangement allows the nerves of the subjects to obtain the information. The electric signal has frequencies that vary in the range of 10Hz to 120Hz. The width of the pulse varied between 10μs to 100μs. Moreover, the width of the pulse is a factor in controlling the intensity of the signal. The signal has an amplitude range between 40V until 80V. Those features are able to be adjusted. This adjustable function is to attain the subjects' comfort. Every subject has their own preferences. In the following section, the experiment with the results will be shown.

4. Experiment Methods and Results

This learning system is to train the subjects with passive and active mode at improving the basic skills of playing the piano. These trials were directed mainly to confirm the potential of the proposed system using audio and visual information and augmented with electrotactile feedback system. This objective can be done by comparing the experiment with a training system using only audio and visual feedback.

To do the experiment, nine subjects are divided into three groups. All the subjects are never trained to play the piano. They also have never used the electrotactile feedback. The subject's ages vary between 15 until 21 years old. The subjects are comprised of five males and four females. The first group trains using the active mode with audio, visual and haptic information. The second group practices with the active mode, only using audio and visual feedback. The passive mode, with audio and haptic information, is given to the third group. The subjects doing the active mode have to tap their fingers as appropriate as the given information.

Before the trials, all the subjects are given basic music theory, focusing on the use of appropriate fingers and hands for specific note or chords. For the first five minutes, they are trained to use the system. The aim of this training is to enable the subjects play with the virtual piano. Subjects were trained to tap their fingers according to the notes that appear on the screen until they become used to use their fingers under the sensors. This is followed with an initial test. In this state all the subjects tried to play a simple song. The system only shows the letter of the notes that they have to tap, and the errors are counted. Errors include the use of wrong fingers or missed notes.

After the initial test, for the groups where haptic feedback is used, the intensity of the electric signal is adjusted for each subject. This is for the comfort of the subject, also to adapt with the signal. This is then followed with five minutes training of the subjects with the method in their groups. In this time all the subjects are given information about which fingers must be bent in accordance with the letters of the notes shown in the screen. Then, in the last five minutes, all the subjects tried to play music assisted with the feedback and the letters of the notes in the screen. After a short break, the final test is conducted to prove the effectiveness of the haptic feedback. This task is the same with the initial task. The errors are calculated and the percentage of improvement of each subject can be counted.

Based on the experiment, on average, the first group, who learned to bend the fingers utilizing visual and audio information, and enriched with electrical sensation, on average achieved an improvement of around 40%. On the other hand, the second group, who learned through the active method using visual and audio

feedback, but without haptic feedback, achieved 20% improvement. While the last group, which used the passive method, had about 35% improvements. Table 1 shows the number of errors for each subject before and after being trained in the method they use.

Analysis of the outcome of the experiment shows that all the training methods offers positive results. All the methods are able to improve the musical skills of the subjects. However, the methods which employ the haptic feedback show a more significant performance improvement compared to the method not involving haptic feedback. This can be concluded that the systems that involve haptic information are more effective. Moreover, subjects who use electric stimulation feel that the feedback is not painful and easy to use.

Table 1: Error before and after training

Active Electro Tactile Feedback			**Active Without Electro Tactile Feedback**			**Passive With Electrotactile feedback**		
Subject No	**No Errors**		**Subject No**	**No Errors**		**Subject No**	**No Errors**	
	Before	After		Before	After		Before	After
1	25	15	4	28	22	7	30	19
2	27	16	5	27	21	8	23	15
3	22	13	6	26	20	9	24	15

5. Conclusion

This paper introduces a novel system to train fingers to play the piano using multimedia information. The audio, video and haptic feedback is used to teach. A computer's speaker is used to provide audio information. The screen of the computer is used to show visual feedback, while haptic information is transmitted utilizing electrical sensation. This proposed system employs a gesture sensor to become a virtual piano, and a custom software is built to accommodate all the inputs and outputs of the system.

The results of the tests indicate that the system involving electro-tactile feedback using active and passive training methods have positive results. The electro-tactile feedback system enables to train the fingers of the users to move in association with the information reflexively. Moreover, this system augments to train users the new finger skill better than the method only using audio and visual. The improvement of the subjects who use haptic feedback is superior to the subjects who utilize the conventional one. Electro-tactile feedback is inexpensive, convenient to use, simple to wear, and does not hinder the user from moving their fingers.

This system has shown its potential to enhance finger training devices. It is desirable to conduct further research with more subjects, as well as a comparison with existing training systems to measure the performance of this proposed system.

Acknowledgment

This work is financially supported by Penelitian Terapan Unggulan Perguruan tinggi research fund at Indonesian Ministry of Education. Also acknowledgment to Politeknik Negeri Batam for the laboratory facilities.

References

[1] T. Markow, N. Ramakrishnan, K. Huang, T. Starner, M. Eicholtz, S. Garrett, H. Profita, A. Scarlata, C. Schooler, A. Tarun, D. Backus, "Mobile music touch: Vibration stimulus in hand rehabilitation," in 2010 4th International Conference on Pervasive Computing Technologies for Healthcare, Pervasive Health 2010, IEEE Computer Society, 2010, doi:10.4108/pervasivehealth.2010.8.

[2] D. Kohlsdorf, T. Starner, "Mobile music touch: The effect of primary tasks on passively learning piano sequences," in Proceedings - International Symposium on Wearable Computers, ISWC, 2010, doi:10.1109/ISWC.2010.5665877.

[3] M.A. Eid, M. Mansour, A.H. El Saddik, R. Iglesias, "A haptic multimedia handwriting learning system," in Proceedings of the ACM International Multimedia Conference and Exhibition, ACM Press, New York, New York, USA: 103–108, 2007, doi:10.1145/1290144.1290161.

[4] J. Bluteau, S. Coquillart, Y. Payan, E. Gentaz, "Haptic Guidance Improves the Visuo-Manual Tracking of Trajectories," PLoS ONE, **3**(3), e1775, 2008, doi:10.1371/journal.pone.0001775.

[5] C. Seim, D. Quigley, T. Starner, "Passive haptic learning of typing skills facilitated by wearable computers," in Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, New York, New York, USA: 2203–2208, 2014, doi:10.1145/2559206.2581329.

[6] D.S. Pamungkas, K. Ward, "Electro-tactile feedback system for achieving embodiment in a tele-operated robot," in 2014 13th International Conference on Control Automation Robotics and Vision, ICARCV 2014, 2014, doi:10.1109/ICARCV.2014.7064573.

[7] D.S. Pamungkas, K. Ward, "Tele-operation of a robot arm with electro tactile feedback," in 2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics: Mechatronics for Human Wellbeing, AIM 2013, 2013, doi:10.1109/AIM.2013.6584175.

[8] D. Pamungkas, K. Ward, "Electro-tactile feedback system for a prosthetic hand," in 22nd Annual International Conference on Mechatronics and Machine Vision in Practice, M2VIP 2015, 2015.

[9] D. Pamungkas, K. Ward, "Tactile sensing system using electro-tactile feedback," in ICARA 2015 - Proceedings of the 2015 6th International Conference on Automation, Robotics and Applications, 2015, doi:10.1109/ICARA.2015.7081163.

[10] D.S. Pamungkas, K. Ward, "Electro-Tactile Feedback System to Enhance Virtual Reality Experience," International Journal of Computer Theory and Engineering, **8**(6), 465–470, 2016, doi:10.7763/ijcte.2016.v8.1090.

[11] D.S. Pamungkas, A. Turnip, "Electro-Tactile Cues for a Haptic Multimedia Finger Motoric Learning System," in ICSECC 2019 - International Conference on Sustainable Engineering and Creative Computing: New Idea, New Innovation, Proceedings, Institute of Electrical and Electronics Engineers Inc.: 127–132, 2019, doi:10.1109/ICSECC.2019.8906989.

[12] Principles of Neural Science, Fifth Edition (Principles of Neural Science (Kandel)): 9780071390118: Medicine & Health Science Books @ Amazon.com.

[13] I.P. Pavlov, "Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex," Annals of Neurosciences, **17**(3), 2010, doi:10.5214/ans.0972-7531.1017309.

[14] M. Mihelj, D. Novak, S. Beguš, Introduction to Virtual Reality, Kluwer Academic Publishers, 2014, doi:10.1007/978-94-007-6910-6_1.

[15] Y. Yoon, D. Moon, S. Chin, "Fine Tactile Representation of Materials for Virtual Reality," Journal of Sensors, **2020**, 7296204, 2020, doi:10.1155/2020/7296204.

[16] R.L. Peiris, Y.L. Feng, L. Chan, K. Minamizawa, "Thermalbracelet: Exploring thermal haptic feedback around the wrist," in Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, New York, New York, USA: 1–11, 2019, doi:10.1145/3290605.3300400.

[17] K. Rangarajan, H. Davis, P.H. Pucher, "Systematic Review of Virtual Haptics in Surgical Simulation: A Valid Educational Tool?," Journal of Surgical Education, **77**(2), 337–347, 2020, doi:10.1016/j.jsurg.2019.09.006.

[18] A. Abiri, J. Pensa, A. Tao, J. Ma, Y.Y. Juo, S.J. Askari, J. Bisley, J. Rosen, E.P. Dutson, W.S. Grundfest, "Multi-Modal Haptic Feedback for Grip Force Reduction in Robotic Surgery," Scientific Reports, **9**(1), 1–10, 2019, doi:10.1038/s41598-019-40821-1.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 148-154 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Aviation MRO: Impact of Physical Environment Factors on Job Performance in Aircraft Maintenance Organization

Kamal Jaiswal[1,*], Serdar Dalkilic[1], Evangelos Papageorgiou[1], Balgopal Singh[2]

[1]*Higher Colleges of Technology, Muroor Raod, Abu Dhabi, 25035, United Arab Emirates*

[2]*Banasthali Vidyapith, Banathali, Distt.- Tonk, Rajasthank, 304022, India*

ARTICLE INFO

Article history:
Received: 27 August, 2020
Accepted: 11 October, 2020
Online: 08 November, 2020

Keywords:
Aircraft Maintenance
Physical environments
Job performance
Safety
Work Efficiency

ABSTRACT

A happy employee with high morale is the best asset of an organization. The morale of an individual working in an organization can be impacted by both positive and negative ways by physical environment factors of the workplace. The physical environment of an organization impacts a lot on the work performance of an individual. It becomes more significant in case of aviation maintenance organization formally called Maintenance Repair and Overhaul (MRO) organization as the work performed by an individual requires full accuracy and directly linked with the safety of the aircraft and its occupants. This can only be achieved by providing a safe and positive working atmosphere to individuals working in MROs. Considering the same, this study targets to deliver a brief status report of physical environment factors disturbing work performance in Maintenance Repair Overhaul (MRO) agencies. This study initially explains the impact of different physical environment factors. The study also discusses the impact of unsafe work due to improper working environment in aircraft maintenance organizations. The individual's working in MROs located in the United Arab Emirates were considered for testing the physical environment factors for this case study. This case study aims to provide the status of different physical environment factor of organization causing problems at work for individuals working in aircraft maintenance organizations. The result of the study indicated that MROs in UAE are providing a conducive working atmosphere to employees, however, the lack of workplace is still an area for improvement.

1. Introduction

The human being is manipulating the environment for his survival and comfort contrary to the environment or nature is human's best friend and nearest surrounding. Unsafe and dangerous conditions may arise if we manipulate nature incorrectly and can make the life of human being very difficult. [1].

The efficiency of individuals and the work morale of individual at the workplace might get affected due to the hazardous environment resulting from the poorly administered surrounding. To determine the performance, morale and efficiency of an individual at a workplace, the workplace environment is a key factor. The workplace environment includes the area and condition in which an employee is executing the job however a positive and effective work environment is the condition when the output of the individual working for an organization is same as expected as per job objective [2]. The physical environment where the job is being carried out by an individual significantly affect the performance of an individual.

The bodily atmosphere as a characteristic of the workplace atmosphere has an uninterrupted influence on human intelligence and can slowly variate interpersonal communications and therefore, efficiency. It happens because the features of space/room or an area have significance on the efficiency and happiness of individuals [3]. Several types of research have concluded that the features of a working area affect the satisfaction of the users. These features of a work area are preferred by the user and extremely important to the user for productivity and work satisfaction. These features include proper lighting, good ventilation, soundproofing, temperature and workspace etc. [4].

*Corresponding Author: Kamal Jaiswal, kamaljswl@gmail.com

The factors like proper illumination of the workplace and suitable work resources such as furniture can have an encouraging effect on the health of an individual and subsequently on the efficiency. Other environment elements of a workplace such as controlled temperature, soundproofing, good ventilation etc. play a significant role in the productivity of an organization [5]. These elements ultimately influence the overall working atmosphere and decide the productivity and satisfaction.

The employees who are bodily and mentally motivated to do their job, the performance is expected to be enhanced, and safety is improved. A well-working environment can significantly reduce the absenteeism and conclusively enhance work efficiency [6]. The aircraft maintenance organization work under a complex system and the productivity of the individual decide the success of the organization. The efficiency of individual working in aircraft maintenance organization largely rely on working condition or physical working environment and it is the most important factor for safety as well [7]. The safety management system used in the aviation sector also address this issue and emphasize on providing a good physical environment to the individuals to perform the job safely. However, it is quite challenging for the management to provide a user-friendly environment to an individual in aircraft maintenance organization as the certain task to be performed on aircrafts requires an individual to work under adverse condition [8].

The output of an organization is directly linked with the satisfaction of employees with the physical environment of the workplace. The administration should regularly monitor and ensure that the physical working environment of the organization is conducive to work for the individual performing the job. The commitment of the administration to deal with work environment influences the motivation of the individual to perform properly and it should be assessed regularly and management should make necessary changes whenever required by employees to improve the efficiency and safety [9]. The changes may include, such as a change in the layout of the workplace or the addition of more resources.

Today's administration has to alter the way of working in order to get real modifications in these facets. The organization has to devote more time on governing these issues of the work situations rather than micromanage. This can be accomplished by enlightening on the physical workplace and amenities, assigning tasks, cumulative liability, and inspiring teamwork. This will encourage faith and devotion between the workers and boosts improved cooperation amongst them.

The capability of an employee to successfully perform a job in a tuff work environment is limited. The studies have revealed that a workspace feature changes the approach of employees and intensify their output. It also states that hazardous and insalubrious workplace atmosphere in terms of poor ventilation, incorrect illumination, disproportionate sound etc. disturb employees' efficiency and healthiness

2. Physical Working Environment

Aircraft maintenance engineers work inside a "complex system". There are plentiful characteristics inside this structure that encroach on the aircraft maintenance engineer, outspreading from his knowledge, skills and capabilities, the atmosphere in which he/she works, to the values of the organization for which he/she works. Even beyond the actual company employee works for, the governing requirements laid down for his/her job evidently impact on his/her comportment. All features of this structure may promote headed for inaccuracies that the engineer might make [10].

To well apprehend efficiency on jobs, it is vital to study the relationship between job performance, individuals, and circumstances elements. Job performance is a very substantial factor manipulating the cost-effectiveness of any organization. Performance is a key element for organizations as employees' performance directs to business success. Similarly, performance is vital for individuals, as accomplishing responsibilities can be a basis of happiness [2].

All the activity and behaviours that are executed by an individual, leading to attain the objectives of an organization can be termed as Job performance. The outcome of a task done by an individual or a group of individual in an organization on a specific time directly indicating that how well the individual or a group of the individual did that job while considering the organization's goal can be described as Performance. The physical environment such as operating conditions, procedures, resources and equipment to perform the job can majority impact the job performance of an employee [11].

In aviation, specifically in MROs, the job performance of an individual is very significant as the job performed by them is safety-critical. High level of job performance from an individual is required in the aviation industry to have a safe operation. The physical work environment and its impact have been enormously considered since the conditions can hamper, interfere with, or set restrictions on the variety of work behaviours that are exhibited which, in turn, possibly influences task performance [3].

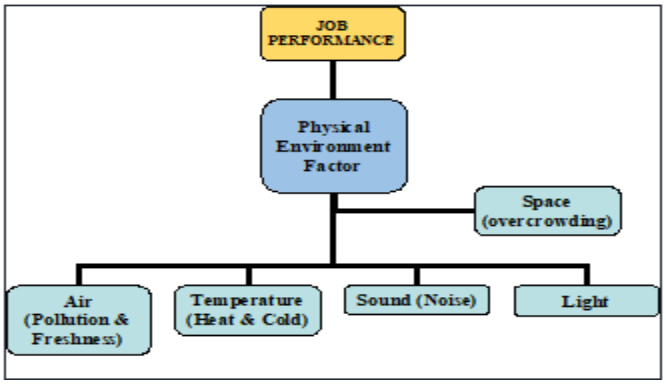


Figure 1: Physical Environment Factor

3. Physical Environment Factor that Influences the Performance

A vigorous interface is discovered between job performance and physical working environment, by the description of what is envisioned by the work environment and literature study findings [12]. The employees' performance, satisfaction, social relations and health is directly linked with physical environment of workplace. Generally, it is presumed that the bodily design of workstations and the ecological condition of workstation are

substantial elements in organizational performance. The study of Badayai (2012) has established earlier studies and categorized five key aspects of an unpleasant working environment as shown in figure 1.

3.1. Air Factor

The behaviour of individuals working in aircraft maintenance organization can significantly alter the job performance due to poor air quality resulting from lack of ventilation or any other factor. The quality of air significantly affect the health of individuals and describe the comfort level of the individual. Sometimes the quality of air can be very poor, specifically individual working with chemicals and in the confined area of aircraft in aircraft maintenance organization [5].

Utmost conceivably unsafe indoor air toxins are radon, asbestos, inorganic material, environmental tobacco smoke, organics, biological and non- ionizing radiation. Supplementary polluting agents such as odours and dust can be the source of serious uneasiness and feelings of reluctance that may lead to a fall in efficiency and job performance. A good ventilation system can overcome this issue and provide a sense of comfort with a combination of controlled temperature. Apart from this, some protective apparatus maybe use to avoid inhalation of pollutants in certain working conditions [10].

3.2. Temperature

The discomfort caused by temperature can be very painful and can significantly reduce job performance. Several types of research have established that the temperature of the workplace has a lot to do with how we interact with others in an organization. The behavior of individuals can significantly change when the temperature of the workplace is not conducive such as too hot or too cold [10]. Working under a condition either hot or cold can cause the lack of attention specifically in those tasks which require a great amount of focus such as crown inspection of rivets on aircraft or Non-destructive testing in workshops.

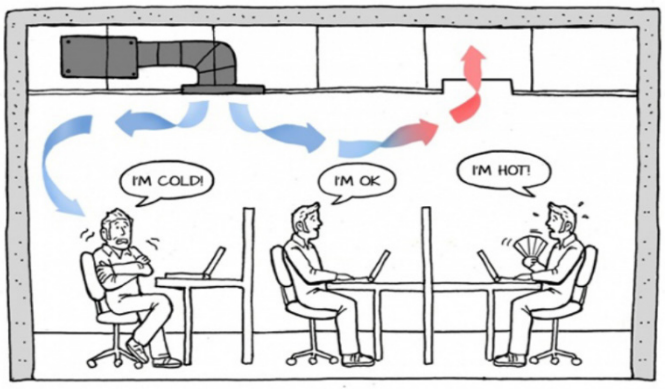


Figure 2: Effect of Temperature

A workplace, which is too hot or too cold, can be the source of stress and can affect the health of individuals. For example, Heat stress can make the body to drop electrolytes and H2O quicker, triggering low psychological act and reduced motor skills. Your musculoskeletal system will be making efforts to save energy, which will absolutely slow your minds, making finishing tasks and escaping errors a hard accomplishment [4].

3.3. Sound (Noise)

Sound is amongst the substantial environmental factors for people's well-being, and it has a vital part in both physical and mental injuries, and it also disturbs individuals' performance and efficiency [6]. The aircraft maintenance engineering is frequently exposed to high noise level such as noise coming from Turbine Engine of the aircraft or from pneumatic equipment used by them. These high level of noise can significantly reduce the ability of an individual to perform efficiently and also have a serious impact on an individual's health both physical and mental [9].

Figure 3: Noise

3.4. Illumination

Poor illumination often gets unnoticed in the workplace, as we talk about mental health and well-being, and the focus is firmly on creating happier and healthier workplaces. But bad lighting is related to a range of ill-health effects, both physical and mental, such as eye strain, headaches, fatigue and also stress and anxiety in more high-pressured work environments [1]. This becomes more significant in terms of aircraft maintenance, as several tasks carried out by aircraft mechanics in aircraft maintenance organization require adequate lighting. The improper lighting can cause serious safety risk as some defects or faults on aircraft may get unnoticed due to lack of lights or bright lights [6].

3.5. Workspace

The individual performing particular task requires an adequate workspace to perform the job properly, however, sometimes it is a bit impossible to have sufficient workspace while performing aircraft maintenance such as C checks performed as a part of major maintenance require several teams of aircraft mechanics to work together [8]. Having a confined space to work, affect the ability of an individual to perform the job and the situation may become worse when it comes with other adverse condition such as noise, time pressure etc.Captions must be written in sentence case (e.g., Macroscopic appearance of the samples.).

Figure 4: Workspace in MRO

4. Research Methodology

The objective of this research paper is to analyze the different physical environment factors disturbing work performance of individuals employed in the Maintenance Repair Overhaul (MRO) organization in the United Arab Emirates (UAE). The profile of intended respondent for this research is individual working in MRO inside the UAE and doing in maintenance activities such as aircraft maintenance technicians, aircraft maintenance engineers and other secondary staff. UAE is a key centre of aviation and patently has the finest airlines in the world. UAE also has the top capability for maintenance of aircraft in the region and unquestionably is the hub of premium MROs in the world. The MROs are functioning under a very multifaceted structure and use innovative technologies for daily operations. The agenda of the research was discussed with the intended audience to have an impartial examination of research elements.

Qualitative and quantitative approaches were engaged in measuring the physical environment factors. The research data was collected through paper-based surveys after a comprehensive conversation with individuals working in aviation MROs. The intended audience considered for this research were the employees of aircraft maintenance organization employed at a diverse level. This research was conducted as a part of the student's project and as project students have to go out and meet industry professionals in different aviation organization. The student involved in the research were studying a four-year Bachelor's Degree program in Aircraft Maintenance Engineering Major.

The structured questionnaire was used and the selection of questions was based on different elements of physical environment factors causing the problem at the workplace. The questionnaire includes the physical environment elements, which can possibly cause problems at the workplace, were related to:

- Inadequate lighting
- Noise
- Excessive Heat
- Excessive Cold
- Overcrowding
- Ventilation
- Fumes
- Maintenance equipment

Cronbach's alpha is used to test the reliability and validity. The Cronbach's alpha value is more than 0.70. The Cronbach's alpha of the questionnaire found to be 0.76. The statistical method used to investigate the data is the frequency and percentage analysis. The data was collected using online survey platform survey monkey. The MS excel software was used to carry out the statistical analysis. The outcomes of the study were arranged after carrying out an in-depth examination of the data collected concerning physical environment elements affecting job performance. Table I shows the statistics of the research survey.

Table 1: Sample Table

STATISTICS OF THE RESEARCH SURVEY		
#	Research Material	
	Type of Material	Quantity
1	Total questions included in questionnaire	22
2	Total questionnaires distributed	200
3	Percentage of responses received	85%

5. Investigation Results

The research outcomes are indicated in the form of the illustration, column and pie tables for the comfort of explanation. The percentage of replies is 85%; it is good and imitates the individual's mindfulness and enthusiasm in relation to the physical environment of the organization.

Figure5 shows that a total of 53% of respondents NEVER had problems due to poor or inadequate lightings at the workplace. Total 38% of respondents choose that SOMETIMES they have problems due to improper lighting. Total 9% respondent OFTEN have problems caused by inadequate lighting at the workplace. In aviation, the adequate lighting is quite critical especially when individuals are working in complex and small aircraft parts during the night and in a dark atmosphere.

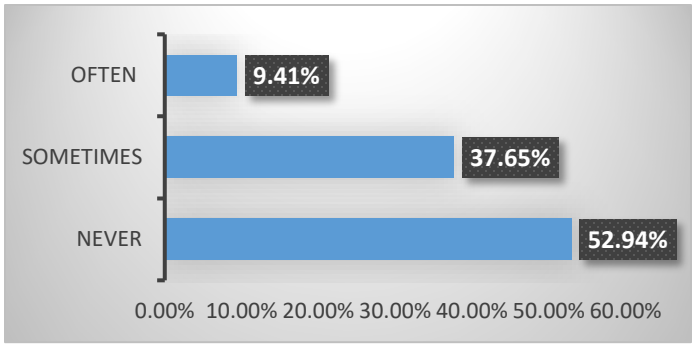


Figure 5: Poor/inadequate Lighting

Figure 6 shows frequency responses to the problem caused by noise at the workplace. As shown, 41% of respondents NEVER have a problem due to noise at the workplace, and 43% of them SOMETIMES have problems due to noise. However, 16% of respondent OFTEN have problems due to noise. Noise can be very annoying and can have a serious impact on health, which can lead to degradation of work performance

Figure 7 shows responses to the effectiveness of individuals exposed to excessive heat. As indicated in the figure, 49%, respondent NEVER have excessive heat working condition, 39% SOMETIMES have excessive heat working condition, and 11%

respondent OFTEN have excessive heat working condition. This is a clear indication that working under excessive heat condition cannot be avoided, but it affects your performance and individuals get less effective when adverse conditions arise.

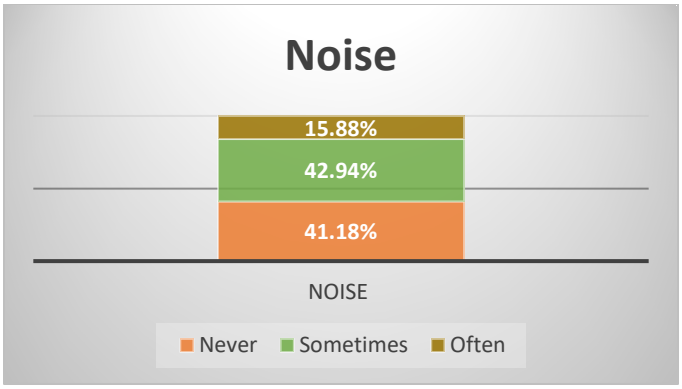


Figure 6: Noise

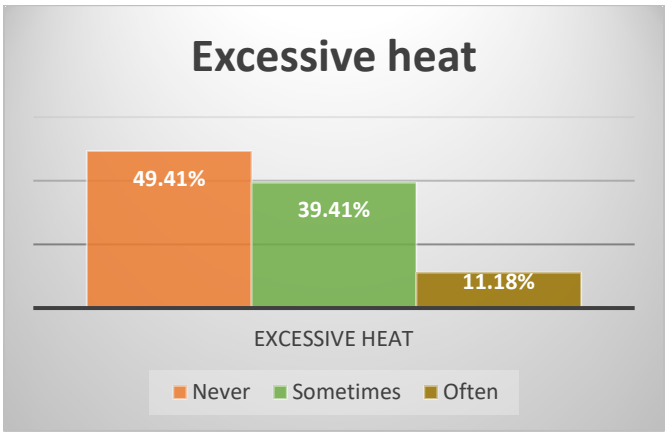


Figure 7: Excessive Heat

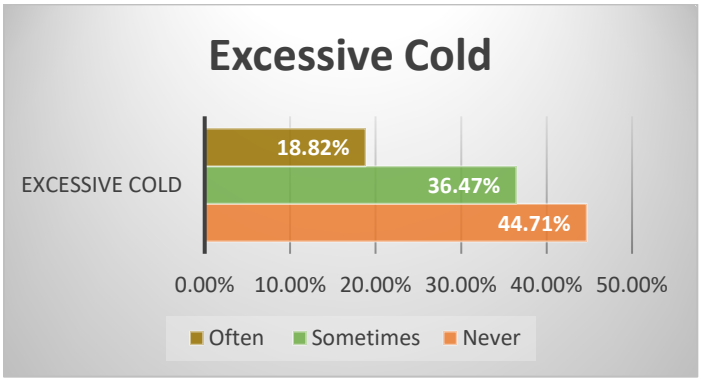


Figure 8: Excessive Cold

Figure 8, shows that 45% of respondents have acknowledged that they NEVER a condition of excessive cold at the workplace while 36% of respondents SOMETIMES have the excessive cold condition at the workplace and almost 19% respondent OFTEN have the excessive cold condition at the workplace. These percentages seem to be quite significant and require action by the organization to control the physical environment condition of the workplace to make it more conducive.

Figure 9 shows responses to the overcrowding of the workplace leading to lack of workspace. About 45%, 40%, and 15% have responded to NEVER, SOMETIMES and OFTEN respectively. This result reflects that the majority of the people have an issue related to overcrowding of the workplace.

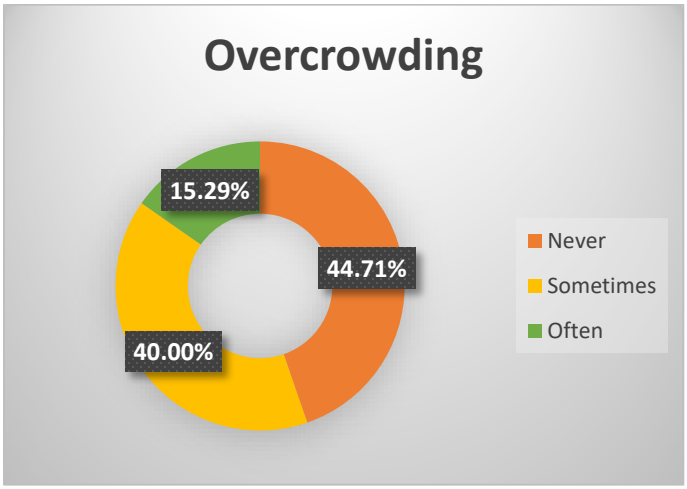


Figure 9: Overcrowding

Figure 10 shows responses to poor ventilation condition at the workplace. About 51%, 38% and 11% have responded to NEVER, SOMETIMES and OFTEN respectively. These results reflect that ventilation condition needs some improvement and it becomes more significant when dealing with hazardous chemicals.

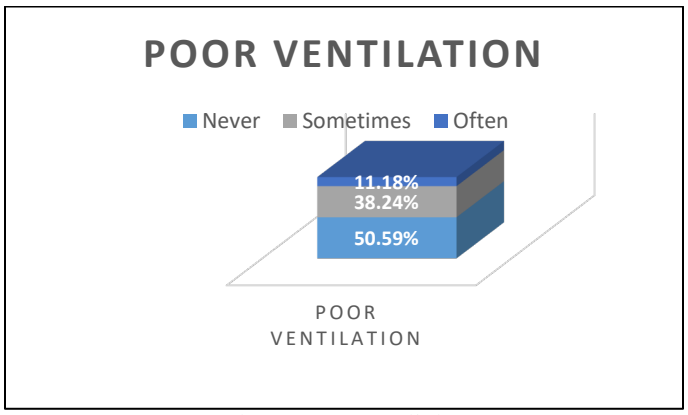


Figure 10: Poor Ventilation

Figure 11, shows responses to the question related to the presence of Dust and Fumes at the workplace. About 54% of respondents NEVER have problems related to dust/fumes, 29% respondent SOMETIMES have problems with dust/fumes and 16% respondents OFTEN have problems with dust/fumes. It indicates that only small no. of the individual having a problem with dust or fumes and thus interfering with their work.

Figure 12, shows the results of poor maintenance condition of equipment used by individuals causing the problem at work. It was found that 52% of respondents NEVER have the problem of poor maintenance of equipment they use, 35% respondent SOMETIMES have the problem of poor maintenance condition of the equipment they use and only 12% respondent OFTEN have the problem of poor maintenance of equipment they use. This

matter is directly related the work performance as equipment used by an individual in the organization provides the basis of good maintenance of aircraft.

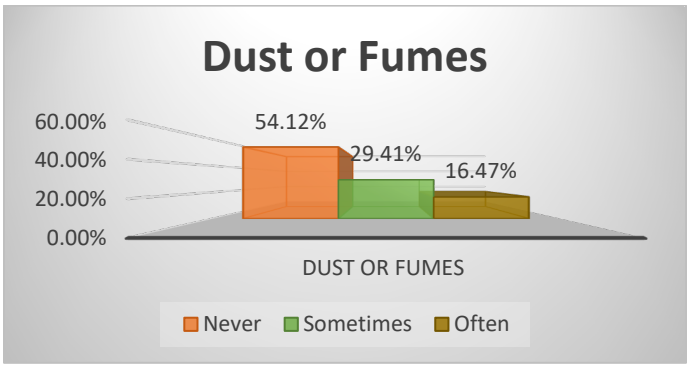


Figure 11: Dust or Fumes

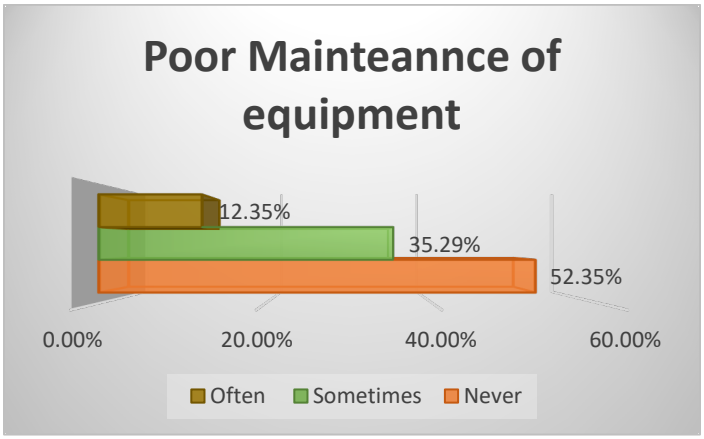


Figure 12: Poor Maintenance of Equipment

6. Discussion and Recommendation

This research specified that (reference to Figure 5) people working in MROs in the UAE typically does not have an issue related to poor illumination, and it does not affect their work. It is a optimistic signal that organizations are aware of this and have the necessary resources in place. However, a substantial number of individuals do have a problem with poor lighting sometimes which further need to be investigated why.

The study presented (a reference to Figure 6) that individuals working in aircraft maintenance organization frequently exposed to Noise, which should be the matter of concern for MROs as noise can seriously degrade the performance of individuals. It is also found (a reference to Figure 9) that a large number of individuals have problems with overcrowding leading to lack of workspace to perform the job specifically for the individuals working in base maintenance.

The analysis from Figure 10, evidently specifies that a great number of persons never have an issue due to poor ventilation which can reduce the air quality of the workplace. However, some respondent reported that they often have an issue with poor ventilation and this need to fix as poor ventilation lead to poor air quality and can result in a poor health condition of individual.

Figure 7 and Figure 8, clearly indicates that the majority the individuals does not have issues related to excessive heat or excessive cold condition. It reflects that temperature control of the workplace is properly maintained to provide a conducive atmosphere to individual working in MRO.

Exploration of figure 12 specifies that the most of the individuals working in MRO's at UAE does not have an issue related to poor maintenance of equipment's they use. However, a significant number of respondents have indicated that they sometimes or often have problems due to poor maintenance of equipment's.

The underneath mention fact specifies the zones of development as per the results of this study, and the succeeding recommendations are identified:

- Increasing attentiveness of the physical work environment and its optimistic effects on the organization and the individuals.
- Management taking proactive action to resolve the problem related to poor working conditions.
- Additional efforts are required to improve working conditions to guarantee enhanced work performance.
- Gathering more often with staff to improve their comfortability to report and to obtain opinion to lift the whole work performance.

Further, a study with hefty no. of the participant and a comprehensive questionnaire is suggested. The study can be very valuable if it is linked to statistics of MRO's worldwide.

7. Discussion and Recommendation

In this exploration, efficiency and significance of physical environment factors affecting work performance in MROs were examined. The MROs in UAE have delivered great backing throughout the course the research and research data was formed with support from MROs. It is concluded that MROs generally have very good physical environment. The MROs are taking all necessary action to provide a conducive atmosphere for employees. However, there are areas for improvement for a better physical environment including the improvement in terms of overcrowding leading to lack of workspace and the introduction of a robust system for proactive actions related to a physical environment.

Acknowledgements

The study was led by endless and unceasing backing from Higher Colleges of Technology (U.A.E.) and different MROs within the United Arab Emirates. The delicate and specialized support by Each Subject Matter Expert has educated me the finest research guidance, and I am appreciative for the generous support delivered by them.

References

[1] A.A. Shikdar, "Identification of ergonomic issues that affect workers in oilrigs in desert environment." International Journal of Occupational Safety and Ergonomics, **10**: 169-177, 2002, doi.org/10.1080/10803548.2004.11076605

[2] F.O. Becker, "Workspace creating environments in organization." New York, N.Y. : Praeger, x, 225 p. : ill, 1981.

[3] F. Becker "Improving organizational performance by exploiting workplace flexibility." Journal of Faculty Management, 1: 154-162, 2002.
[4] C.C. Berger, HH Ehrsson, "Mental Imagery Changes Multisensory Perception." Current Biology, **23**(14), 1367-1372, 2013, doi.org/10.1016/j.cub.2013.06.012.
[5] P.E. Mohr, J.J. Feldman, J.L. Dunbar, et al., "The societal costs of severe to profound hearing loss in the United States." Internationl Journal of Technology Assess. 2000.
[6] R.G. Geen, "Preferred stimulation levels in introverts and extroverts: Effects on arousal and performance." Journal Personality Social Psycholog, 46**,** 1303–1312, 1984.
[7] C.H. Friend, Aircraft maintenance management, Harlow: Longman scientific and technical, 1992.
[8] GCAA-U.A.E., Civil Aviation Requirement-145, 05 revision-01, GCAA-U.A.E., February 2016.
[9] A. Kumar, P.K. Sinha, "Human Error Control in Railways." Jordan Journal of Mechanical and Industrial Engineering, **2**, 183–190, 2008.
[10] S.A. Stansfeld, M.P. Matheson, "Noise pollution: non-auditory effects on health." British Medical Bullettin., 68:243–257, 2003. doi.org/10.1093/bmb/ldg033
[11] K. Persson-Waye, R. Rylander, S. Benton, H.G. Leventhall, "Effects on performance and work quality due to low frequency ventilation noise." Journal of Sound Viberation.;205:467–474, 1997 doi.org/10.1006/jsvi.1997.1013
[12] W.W. Clark, B.A. Bohn, "Effect of noise in hearing." JAMA the Journal of the American Medical Association; **281**(17),1658-1659, 1999. DOI: 10.1001/jama.281.17.1658

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 155-159 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Inferring Topics within Social Networking Big Data, Towards an Alternative for Socio-Political Measurement

Khalid Ait Hadi*,[1], Rafik Lasri[1], Abdellatif El Abderrahmani[2]

[1]*Department of Computer Sciences, Polydisciplinary Faculty of Larache, Abdelmalek Essaadi University, 93000, Morocco*

[2]*Department of Computer Sciences, Dhar El Mahraz Sciences Faculty, Sidi Mohamed Ben Abdellah University, 30050, Morocco*

ARTICLE INFO

Article history:
Received: 03 September, 2020
Accepted: 28 October, 2020
Online: 08 November, 2020

Keywords:
Language processing
Data Mining
Topics detection

ABSTRACT

This research sought to measure some socio-political indicators using millions of opinionated messages from social network sourced big data. Thus, and using an enhanced mixed method for sentiment analysis and a fusion model algorithm to infer topics from short text, this study attempted to demonstrate the value of computational approaches in measuring some phenomena in the real social world and quantifying public opinion fluctuations in response to certain socio-political issues. The validity of the experimental results was examined by comparing them with data obtained from representative surveys, thus providing a better understanding of the relationships between online and offline opinion dynamics. This contribution is intended to be multidisciplinary, both useful for policymakers and opinion analysts to explore public trends and to inquire into socio-political issues.

1. Introduction

Nowadays, in the communication fields, market research and public opinion analysis, as in many other disciplinary domains, social media are mobilized as new societal trends observatories [1, 2]. The evolution of these networking platforms in the first decade of the 21st century has enabled Internet users without specific technical skills to easily publish and share content on their concerns [3]. The participation architecture on which these services are based has facilitated the information co-production, and has offered spaces for socio-political engagement [4].

In this context, millions of people interact daily on social networks, producing several hundreds of millions of messages of all kinds and of any content. These messages represent valuable clues concerning practices, representations and Internet users' opinions. As such, they form a particularly interesting material to investigate when looking at citizen public behavior studies. Furthermore, due to the development of technical and computational tools that have made it possible to collect, archive and analyze huge data volumes, namely big data, these numerical phenomena can be objectified and quantified, on a large scale, for the investigation's needs. Social media analytics open up a promising way for the quantitative study of certain substantial subjects of declarative surveys, such as quantifying a social object, capturing behavioral representations and uncovering political intentions.

In this work, we seek to exploit data extracted from Facebook and use computational methods to measure some socio-political indicators. This article does not claim to provide affirmative answers about the effective correlations between the phenomena observed on social media and those perceived in the real world. On the contrary, it aims to help support the observation indicating that social networking data constitute a particularly interesting material to investigate when trying to study citizens' behavior, public opinion's fluctuations and social trends, particularly where declarative surveys and others conventional approaches turn out to be often costly, time-consuming and labor-intensive [5].

This paper is an extension of work originally presented in 2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS) [6]. An initial work, which was interested in measuring of notoriety capital on Facebook of a political figure. Here, we will not be limited to measuring political popularity, but we will improve this research to be able to evaluate other socio-political indicators. Likewise, we will not settle for the use and study of messages formulated only in one language, as was the French case in [6], but we will extend the prospecting to include also social networking data expressed in classical Arabic and dialect as well.

*Corresponding Author: Khalid Ait Hadi, khalid.aithadi@gmail.com

Another improvement of the present work is linked this time to the observation relating to the fact that topic models, like Latent Dirichlet Allocation (LDA) and its derivatives, are not efficient in the analysis of short texts due to data sparsity [7]. Alas, it is clear that the overwhelming majority of data extracted from social networks consists of short texts. Here, we will overcome this inconvenience by opting for an improved Biterm Topic Model (BTM) model based on the heat matrix [8].

The rest part of the paper is structured as follows. Section 2 presents an overview of related work. Section 3 is dedicated to the data collection process description and the suggested approach implementation. An experimental evaluation of the method is also presented in the same section. Finally, discussions, conclusions and suggestions for future research openings are made in the last section.

2. Related Work

Studies seeking to examine predictive capacity of social media data and to establish existence of possible correlations between the facts perceived on social media and those observed in the effective world have been subject to a rich and growing literature since the beginning of the 21st century [2].

Many recent studies have used social media data as a new sensor of societal trends and a new predictor of economic or political phenomena, ranging from stock market volatility to box office performance or, in a certain perspective that interests us more directly, public opinion fluctuations and social dynamics [9, 10]. Research in this area has been driven by a positivist approach that is certainly strengthened owing to the unprecedented possibilities offered by these new social platforms, in the sense that they allow unheard-of access to the data shared daily by people as well as to their networks, regardless of any temporal or geographic consideration.

In this context, a review article published in 2014 [11] has listed more than a hundred studies devoted solely to political messages published on the social network Twitter. The penchant of researchers for Twitter can be explained by easier access to data, where, unlike Facebook, the vast majority of accounts are “accessible” and by the brevity of posts on Twitter, limited to 280 characters.

Always on Twitter, a meta-analysis developed by Gayo-Avello [12] examined the predictive usefulness of data published on this social network, leading to the conclusion that such data have relatively predictive power and provide some information referring to electoral consultations results.

It should be noted on the socio-political aspect that the most commonly studied subject is that concerning electoral outcomes prediction using social media data [13]. In the opposite, studies relating to the political popularity evaluation, for example, or to the measurement of social indicators are not numerous, and here too Twitter has often been called upon to make forecasts. In addition, and almost globally, these studies have not examined the validity of the predictions by making comparisons to conventionally more admissible results types, such as those from censuses and surveys [14].

It should also be emphasized, according to the existing scientific literature [1], that most results in the field of forecasting with social media data claim to have produced predictive results to some extent, registering a positive statistical correlation between offline and online data. However, several scientists have criticized these results, which draw up a fundamental limitation to these works that is linked to the lack of sociological representativeness of the populations recorded on social networks [15]. Though, dismissing social media data on this ground would fail to capture the opinion forming dynamics. Some studies have even argued that representativeness search is no longer consubstantial with the need to sample populations in the big data era. The arguments put forward in this wake point to fact that debates led by certain politically active groups prevail over those who develop in society at large [16], and that individuals active on social networks are thought leaders, more politicized than average, influential in their entourage and acquaintances and whose opinions count more than those of "ordinary" individuals. To poll these opinion leaders would indirectly mean polling their entourage and, ultimately, the entire population. In addition, discussions on certain social platforms would above all reflect the concerns and themes put on the agenda by the mainstream media [17]. In many ways, the social network appears to be an echo chamber for the media field, and the associated media agenda would indirectly influence citizens' concerns [18].

Regarding the diversity of methods used in measuring public opinion, whether sentiment-based or social network-based approaches, it should be underlined that there is no unanimity about the most efficient methods in terms of prediction, and several researchers have reported some performance with different approaches and implementations.

Thus, and particularly in the category of sentiment-based studies, where opinions tone is used as a behavior’s indicator, two lines of research stand out in this context, one is based on pre-established lexicas [19], while the other relies on new sentiment models specifically for political messages [20].

Although concerning the dictionary-based sentiment classifiers, some studies have criticized the reliability of these approaches for predictions [21, 22], advancing in this regard some deficiencies such as the incorrect classification of the word in the lexicon, the lack of words disambiguation, and the neglect of contextual inference. Imperfections that are more accentuated in the socio-political context, with the emergence of difficulties especially linked to the lack of subtleties of socio-political language [23].

Through this literature review, it is seen that the majority of research has focused on the electoral and public opinion forecasts, without identifying and analyzing the reasons behind these opinions development, problematic that the present work is concerned with. In addition, there are two more obstacles that we will overcome in this paper, firstly by using a data source other than Twitter, Facebook in this case, and secondly by comparing our results with those of a declarative survey [24].

3. Methodology and Experimental Results

As previously announced, this research aims precisely to support the predictive interest of data extracted from Facebook. This involves, on the one hand, developing algorithm to analyze a very large amount of messages with socio-political connotation in

order to identify their polarity and target and, on the other hand, to compare these information to opinion poll statistics so as to better grasp similarity degrees between online and offline opinion dynamics. For this, we will consider the same period covered by the public opinion survey 'The Arab Barometer' [24] conducted in Morocco from May 7 – June 11, 2016.

Data collection has been made using '*Rfacebook*' (a package that provides a series of functions that allow R users to access Facebook's API to collect public status updates that mention specific tags). Therefore, and in order to cover a very large sample of socio-politically concerned Moroccan Facebook users, the choice has been made so as to extract both messages and comments written in French or Arabic, even dialect. On this register, we admit that debate around socio-politically charged issues is also, if not largely, done in the mother tongue and that disregarded educational level, people remain always sensitive to socio-political themes and use informal language to take part in the debate.

Therefore, and to test our approach, we have looked at two socio-political issues: (I1) 'opinions about politics' and (I2) 'perceptions of freedoms'. For the first theme, we have collected data (posts and comments) corresponding to the following queries: 'politique', 'gouvernement', 'politique gouvernementale', 'سياسة', 'حكومة', and 'السياسة الحكومية'. The second theme was examined using three tags: 'libertés', 'حرية', and 'الحريات'.

3.1. Sentiment analysis

It would be necessary before approaching the opinion mining phase, and in the ultimate objective to recover pure personal opinion, to operate some text preprocessing and cleaning steps: stemming, removal of URLs, punctuations, stopwords, screen names, special caracters and duplicate comments.

For sentiment analysis, two approaches will be used, one reserved for data written in French and the other for those formulated in Arabic or dialect language. For the first category, the choice focused on the use of Carousel greedy algorithm with Cat Swarm Optimization based Functional Link Artificial Neural Networks (CSO-FLANN), a technique for sentiment mining, enhanced in term of accuracy and computational effort [25]. For the other category of data, we used 'SentiArabic', a sentiment lexicon package for standard Arabic [26] coupled with the 'MADAR' corpus, a collection of parallel sentences covering the dialects of 25 cities from the Arab World [27], and for our case we choose the package relating to the Moroccan city 'Fez'. Using these tools, we found 1.15 million positive, 1.48 million negative, and 3.37 million neutral posts and comments for the first issue (I1) 'opinions about politics', and 2.58 million positive, 6.02 million negative, and 4.4 million neutral posts and comments for the second issue (I2) 'perceptions of freedoms'.

3.2. Inferring Topics

Our approach focuses on detecting the main reasons behind a positive or negative impression expressed by Facebook users on a given theme.

Unlike the Latent Dirichlet Allocation (LDA)-based models, which are relatively inefficient in short text processing, where they generate a high dimensional and sparse data problem [7], we opt here for a model called HMBTM (Heat Matrix based Biterm Topic Model), based on Biterm Topic Model (BTM) and improved by introducing the heat matrix. This is also merged with VSM (Vector Space Model) for counting similarities, thus inducing accuracy improvement and dimensionality reduction [8]. Applying an implementation of VSM- HMBTM to reveal the topics from the set of negative posts and comments for the first issue (I1), and from the set of positive posts and comments for the second issue (I2), and using thereafter Maximum Likelihood Estimation (MLE) [28], 906 topics were selected as the optimum number of topics for the first issue (I1), and 798 for the second issue (I2). We admitted here that people express their agreement with positive sentiments as they express their critical positions with negative sentiments.

By examining the most used terms, and despite overlaps and similarities between several themes, we try to assign a unifying label in order to bring together the most relevant subjects. For example, we have assigned for the second issue (I2) 'perceptions of freedoms', the label "freedom of association" to topics containing: joint associations, civil associations, civil organizations… and the label "political participation" to a set including the terms: political parties, free election, partisan activities…

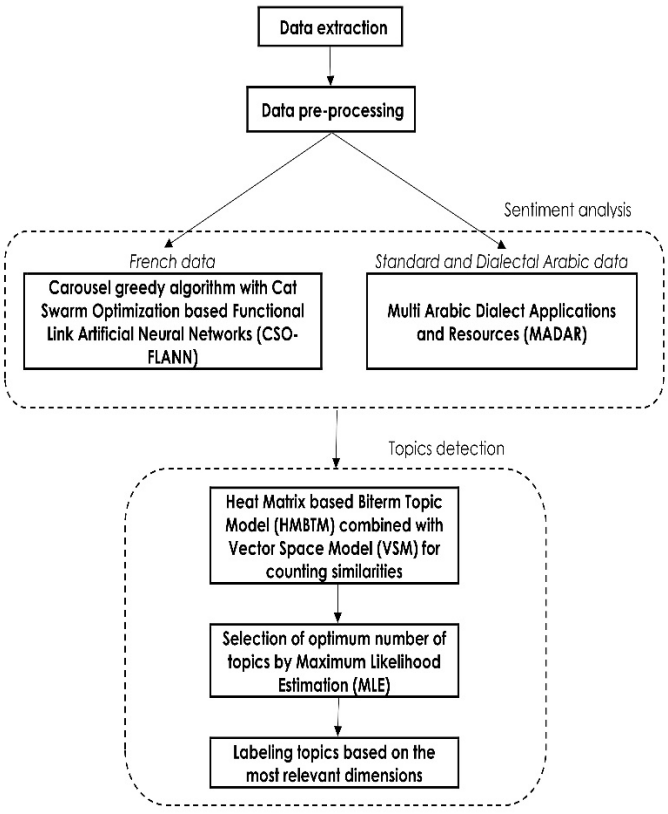


Figure 1: The construction process of our approach

We adopt the same process for the first issue (I1) 'opinions about politics. By examining the main related words such as those in Table 1, we agreed to assign a label based on the most relevant dimensions. For example, we reserved the label "complexity of politics" to topics containing: blurry political landscape, multiplicity of political concepts, number of political parties...

Table 1: A Sample of First Issue's Topics

Complexity of politics	Government not concerned with citizens
multiplicité innombrable ضبابية مبهم	opportunisme détachement لامبالاة ابتعاد
Freedom to criticize the government	Obligation to support the government
reproches critiquable انتقاد خوف	soutien supporter تأييد مساندة

The construction process of our approach is illustrated in Figure 1.

Subsequently, we explore the distribution of labels to get an idea of the topics' distribution and weight for Facebook users. Therefore, Figure 2 represents the rates corresponding to the four relevant subjects that emerge from the analysis, compared to those established by of a declarative survey [24]. Obviously, the values have been scaled given the difference in measurements and scales adopted in each situation.

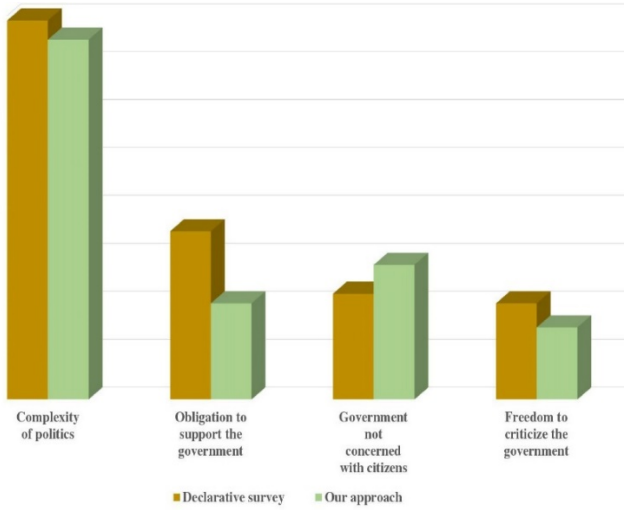


Figure 2: Distribution of First Issue's Topics

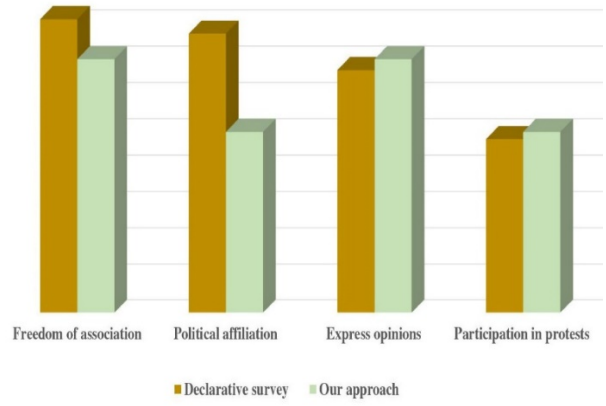


Figure 3: Distribution of Second Issue's Topics

For the second issue (I2) 'perceptions of freedoms', four topics were the most salient: freedom of association, political affiliation, express opinions and participation in protests. Figure 3 compares the weight of these topics to that resulting from the declarative survey [24].

4. Discussion and Conclusions

We have presented a research approach whose objective is to compare the results obtained via two techniques for measuring opinions: the declarative survey, on the one hand, and the analysis of messages published on Facebook, on the other hand. The work sought to situate the data produced by computational methods in relation to the results obtained by surveys. Although the current results do not release certainties, but it appears that, our approach opens up promising horizons, in the sense that it provides very indicative results about the configuration of future declarative surveys results, particularly when the latter are costly in time and even materially. The results presented through this research demonstrate the potential of data extracted from social networks to provide basic and essential information about public opinion dynamics.

However, the results as they stand confirm the observation so much raised by several researchers, indicating that the analysis of social networks would not replace, at least now, public opinion studies based on conventional polls [14]. Nevertheless, it offers indicator data and informative tools to refine our assimilation of public opinion and socio-political behavior. It also opens up perspectives for better capturing indicators relating to engagement and trends measurement, and assessing by the way, the feasibility of many promises in terms of renewing mobilization practices in socio-political field.

One of the possible openings of the present work is to diversify the platforms of data extraction, something that could play, in our point of view, a crucial role to improve the analyses precision. Broadening the base of opinionated data would undoubtedly amount to increasing the representativeness index, so much criticized in this area. One of the concerns that are shaping up recently is the future unavailability of social networks data when the confidentiality of private life is increasingly questioned. End-to-end encrypted messaging applications are attracting new users every day. Social platforms further restrict access to user data to ensure confidentiality of privacy or simply for purely commercial reasons.

One of the possible directions of research is also the one linked to the following question: does not reasoning in terms of positive or negative tones lead to neglecting an essential dimension of the analysis of socio-political messages on social networks, namely the strong majority of messages considered neutral (something that can be seen from the results above)? This leads us to question the silence or neutrality of users and to apprehend its socio-political significance, which would constitute a major issue for future promising research.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. F. Schober, J. Pasek, L. Guggenheim, C. Lampe, F. G. Conrad, "Research Synthesis: Social Media Analyses for Social Measurement" Public Opinion Quarterly, **80**(1), 180–211, 2016. https://doi.org/10.1093/poq/nfv048

[2] H. Schoen, D. Gayo-Avello, P. Takis Metaxas, E. Mustafaraj, M. Strohmaier, P. Gloor, "The Power of Prediction with Social Media" Internet Research, **23**(5), 528–543, 2013. https://doi.org/10.1108/IntR-06-2013-0115
[3] D. Cardon, V. Jeanne-Perrier, F. Le Cam, N. Pélissier, "Présentation" Réseaux, **24**(137), 9–25, 2006.
[4] N. A. Jackson, D. G. Lilleker, "Building an Architecture of Participation? Political Parties and Web 2.0 in Britain" Journal of Information Technology & Politics, **6**(3-4), 232–250, 2009.
[5] W. Wang, D. Rothschild, S. Goel, A. Gelman, "Forecasting Elections with Non-representative Polls" International Journal of Forecasting, **31**(3), 980–991, 2015. https://doi.org/10.1016/j.ijforecast.2014.06.001
[6] K. Ait Hadi, R. Lasri, A. El Abderrahmani, "Social Media Reputation and Political Popularity: Study of a Moroccan Case" in 2019 IEEE International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS), Taza, Morocco, 2019. https://doi.org/10.1109/ISACS48493.2019.9068920
[7] L. Jiang, H. Lu, M. Xu, C. Wang, "Biterm Pseudo Document Topic Model for Short Text" in 2016 IEEE International Conference on Tools with Artificial Intelligence (ICTAI), San Jose, CA, USA, 2016. https://doi.org/10.1109/ICTAI.2016.0134
[8] Q. Liqing, J. Wei, L. Haiyan, F. Xin, "Microblog Hot Topics Detection Based on VSM and HMBTM Model Fusion" IEEE Access, **7**, 120273–120281, 2019. https://doi.org/10.1109/ACCESS.2019.2932458
[9] E. Kalampokis, E. Tambouris, K. Tarabanis, "Understanding the Predictive Power of Social Media" Internet Research, **23**(5), 544–559, 2013.
[10] D. Rousidis, P. Koukaras, C. Tjortjis, "Social Media Prediction: a Literature Review" Multimed Tools Appl, **79**, 6279–6311, 2020. https://doi.org/10.1007/s11042-019-08291-9
[11] A. Jungherr, "Twitter in Politics: A Comprehensive Literature Review" Social Science Research Network, 2014. https://doi.org/10.2139/ssrn.2402443
[12] D. Gayo-Avello, "A Meta-analysis of State-of-the-art Electoral Prediction from Twitter Data" Social Science Computer Review, **31**(6), 649–679, 2013. https://doi.org/10.1177/0894439313493979
[13] S. C. McGregor, R. R. Mourão, L. Molyneux, "Twitter as a Tool for and Object of Political and Electoral Activity: Considering electoral context and variance among actors" Journal of Information Technology & Politics, **14**, 154–167, 2017. https://doi.org/10.1080/19331681.2017.1308289
[14] M. M. Skoric, J. Liu, K. Jaidka, "Electoral and Public Opinion Forecasts with Social Media Data: A Meta-Analysis" Information, **11**(4), 187, 2020. https://doi.org/10.3390/info11040187
[15] T. H. McCormick, H. Lee, N. Cesare, A. Shojaie, E. S. Spiro, "Using Twitter for Demographic and Social Science Research: Tools for Data Collection and Processing" Sociol Methods Res. **46**(3), 390–421, 2017. https://doi.org/10.1177/0049124115605339
[16] A. Dür, "How Interest Groups Influence Public Opinion: Arguments Matter more than the Sources" European Journal of Political Research, **58**(2), 514–535, 2019. https://doi.org/10.1111/1475-6765.12298
[17] J. T. Feezell, "Agenda Setting through Social Media: The Importance of Incidental News Exposure and Social Filtering in the Digital Era" Political Research Quarterly, **71**(2), 1–13, 2018. https://doi.org/10.1177/1065912917744895
[18] Y. Tsfati, J. Cohen, Perceptions of Media and Media Effects, The International Encyclopedia of Media Studies: Media Effects/Media Psychology, 1st Edition, Blackwell Publishing Ltd, 2013.
[19] M. Ibrahim, O. Abdillah, A. F. Wicaksono, M. Adriani, "Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in a Twitter Nation" in 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, USA, 2015. https://doi.org/10.1109/ICDMW.2015.113
[20] D. Contractor, T. A. Faruquie, "Understanding Election Candidate Approval Ratings using Social Media Data" in 2013 International Conference on World Wide Web (www '13 Companion), Rio de Janeiro, Brazil, 2013. https://doi.org/10.1145/2487788.2487883
[21] J. E. Chung, E. Mustafaraj, "Can Collective Sentiment Expressed on Twitter Predict Political Elections?" in 2011 AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 2011.
[22] B. O'Connor, R. Balasubramanyan, B. R. Routledge, N. A. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series" in 2010 AAAI Conference on Weblogs and Social Media, Washington, DC, USA, 2010.
[23] D. Gayo-Avello, "Don't Turn Social Media into Another 'Literary Digest' Poll" Communications of the ACM, **54**, 121–128, 2011. https://doi.org/10.1145/2001269.2001297
[24] The Arab Barometer, "Morocco Five Years after the Arab Uprisings", 2017. https://www.arabbarometer.org/wp-content/uploads/Morocco_Public_Opinion_Survey_2016.pdf
[25] K. Ait Hadi, R. Lasri, A. El Abderrahmani, "An efficient Approach for Sentiment Analysis in a Big Data Environment" International Journal of Engineering and Advanced Technology (IJEAT), **8**(4), 263–266, 2019.
[26] R. Eskander, "SentiArabic: A Sentiment Analyzer for Standard Arabic" in 2018 Language Resources and Evaluation Conference (LREC'18), Miyazaki, Japan, 2018.
[27] H. Bouamor, et al., "The MADAR Arabic Dialect Corpus and Lexicon" in 2018 Language Resources and Evaluation Conference (LREC'18), Miyazaki, Japan, 2018.
[28] W. H. Greene, Econometric Analysis, 7th Edition, Pearson Education, Prentice Hall, 2012.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 160-168 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Research of the Effect of Rotation and Low-Frequency Vibration on the Robotic Assembly Process

Mikhail Vladimirovich Vartanov, Trung Ta Tran*, Van Dung Nguyen

Faculty of Mechanical Engineering, Moscow Polytech University, Moscow, 107023, Russia

ARTICLEINFO

Article history:
Received: 28 August, 2020
Accepted: 23 October, 2020
Online: 08 November, 2020

Keywords:
Robotic Assembly
A Force/Torque Sensor
The Effect of Rotation
Vibrational Device

ABSTRACT

The prospects for the introduction of assembly automation in the industry are largely determined by the reliability of the assembly process. One of the reasons for reducing its reliability is parts jamming. In order to increase the technological reliability of the assembly process, various physical and technical effects are applied. The article presents the results of an experimental study using the rotational movement effect and low-frequency vibrations of the base part in order to reduce assembly efforts and the probability of part jamming.

1. Introduction

Assembly tasks using industrial robots are constantly becoming more complicated over the years. Their main problem is jamming even with slight mismatching. In the case the assembly gap is comparable to the form error, assembly based on rigid basing becomes impossible. The assembly becomes especially problematic when the assembly gap in the joints is comparable to the repeatability of an industrial robot.

The case of the parallel-plane movement of the assembly parts has been considered in known dynamic models of the robotic assembly of cylindrical joints, using passive adaptation [1, 2], as well as control algorithms. In the case of the plane-parallel movement of the installed part, significant sliding friction forces can occur at the contact points of the assembly parts, in which there is a small gap. It leads to jam. Thus, reducing sliding friction forces is one of the ways to increase the reliability of the assembly process. For that purpose, the rotation and the low-frequency vibration can due to provide this effect.

The efficiency of rotational movement for ensuring assembly conditions during the robotic assembly process was considered in works [3-8] and others.

In the work [6] the assembly using the rotational effect with the airflow was experimentally studied. However, an analytical model was not built. In [7, 8], the influence on the assembly process of oscillations along the conjugation axis and along the inclined axis was studied. In work [3], the assembly of parts was studied when the installed part rotates with respect to the non-inertial reference frame and the base part rotates with respect to the inertial reference frame. However, the simultaneous superposition of the effect of rotation and low-frequency vibrations has not previously been considered.

The use of low-frequency vibration remains one of the most promising areas in the field of automatic assembly. This is confirmed by the results of theoretical and experimental works [9-11].

The modern approach to the theory of experiment assumes the need to achieve the required research quality. For this purpose, it is necessary to build a model of the experiment [12].

The effectiveness of technological operation modeling in robotic assembly and the process as a whole based on statistical experimental design allow studying the influence of many process characteristics. It is necessary to establish a connection among the n independent non -random variables $x_1, x_2 \dots x_n$ and the random variable y that depends on them. When processing the results of a multi-factor experiment, the mathematical model of the process, namely the function $\mathrm{M(y)} = f\,(\mathrm{x_1}, \mathrm{x_2} \dots \mathrm{x_n})$ is usually represented as a polynomial of a specific degree. The expression $\mathrm{M(y)} = f\,(\mathrm{x_1}, \mathrm{x_2} \dots \mathrm{x_n})$ is called the regression equation of the mathematical expectation of a random variable y for non- random variables $x_1, x_2 \dots x_n$, and the type of function can be linear or curved [13].

In this article, the authors study the rotation effect and low-frequency vibration to increase reliability in a robotic assembly process. A mathematical dynamics model of the robotic assembly process was presented. A multifactor experiment of this assembly process was carried out, the results and discussion were presented.

*Corresponding Author: Trung Ta, Tran, Rizhskij proezd, 15k2, Moscow, 129626, Russian Federation, +7 905 531 29 68, trungta82@gmail.com

2. Mathematical Model

2.1. Choosing a kinematic assembly scheme

The shaft *3* is held by the gripper *2* of the robot *1*. The working movements of the robot manipulator are the progressive movement of the gripper vertically downwards to the specified position. The vibrational device *5* is a three-link manipulator, each of its links is driven by a separate drive. The base bushing *4* is kept in the center of the vibrational disk, attached to the last link of the vibrational device (Figure 1).

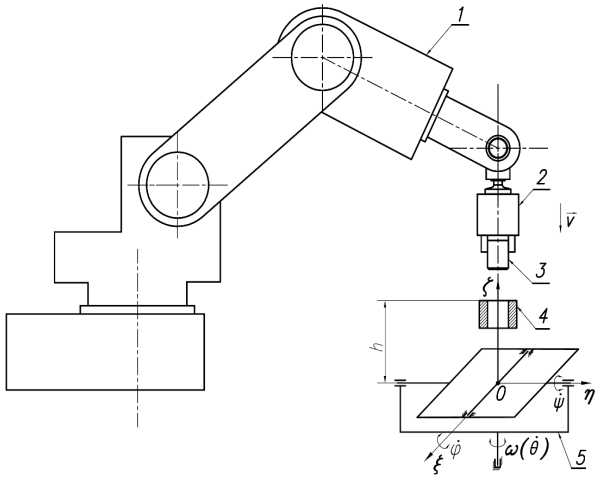


Figure 1: Experimental installation (principle scheme): 1. Industrial robot; 2. The gripper with a force/torque sensor; 3. The shaft, 4. The bushing; 5. The vibrational device;

Four reference frames were defined with a common origin at point O: $Oxyz$ is a fixed reference frame; $Ox_1y_1z_1$ and $Ox_2y_2z_2$ are reference frames, attached to the first and second links of the vibrational device respectively, $O\xi\eta\zeta$ is the reference frame, attached to the third link of the vibrational device. The reference frame $Cx_3y_3z_3$ is attached to the shaft, its origin coincides with the shaft mass center (Figure 2).

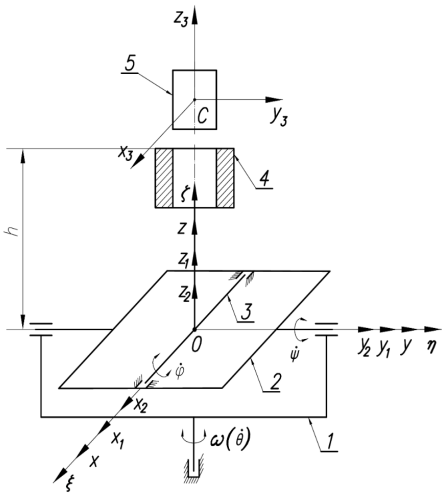


Figure 2: The scheme for determining coordinate systems: 1 - the first link of the vibrational device; 2 - the second link of the vibrational device; 3 - the third link of the vibrational device; 4 - the bushing; 5 - the shaft.

The first link performs rotational motion with a constant angular velocity ω around the fixed axis Oz. In this case, the reference frame $Ox_1y_1z_1$ attached to the first link, at the time t, will rotate with respect to the fixed reference frame $Oxyz$ by an angle θ:

$$\theta = \omega t \tag{1}$$

The second and third links oscillate around mutually perpendicular axes with a phase shift equal to the angle $\pi/2$ as follows:

$$\psi = A\sin kt\,;\, \varphi = A\cos kt \tag{2}$$

where A is the amplitude; k – the angular frequency.

2.2. The mathematical model of the motion dynamics of the shaft mass center with respect to the non-inertial reference frame attached to the base part

Previously, the authors published a mathematical model of the assembly process of a cylindrical joint using rotation and low-frequency vibrations [14]. Therefore, this article presents only a brief mathematical statement.

Formulation of the mathematical model.

On the basis of the theorem on the mass center movement of the system, the movement of the shaft in Figure 3 can be described by a relation of the form:

$$m\vec{\mathbf{a}}_c = m\vec{\mathbf{g}} + \vec{\mathbf{N}} + \vec{\mathbf{F}}_f + \vec{\mathbf{P}} \tag{3}$$

where: m is the mass of the part;

$\vec{\mathbf{a}}_c$ – the absolute acceleration of the cylindrical part mass center;

$\vec{\mathbf{g}}$ – acceleration of gravity;

$\vec{\mathbf{N}}$ – the normal reaction of the vibrational disk plane;

$\vec{\mathbf{F}}_f$ – the sliding friction force;

$\vec{\mathbf{P}}$ – the assembly force.

The relative movement of the mass center of the part with respect to the non-inertial reference frame attached to the bushing can be represented by the equation:

$$m\vec{\mathbf{a}}_{\mathbf{cr}} = m\vec{\mathbf{g}} + \vec{\mathbf{N}} + \vec{\mathbf{F}}_{\mathbf{f}} + \vec{\mathbf{F}}_{\mathbf{ce}}^{\mathbf{fict}} + \vec{\mathbf{F}}_{\mathbf{cC}}^{\mathbf{fict}} + \vec{\mathbf{P}} \tag{4}$$

where $\vec{\mathbf{a}}_{\mathbf{cr}}$ – the relative acceleration of the shaft mass center;

$\vec{\mathbf{F}}_{\mathbf{ce}}^{\mathbf{fict}}$, $\vec{\mathbf{F}}_{\mathbf{cC}}^{\mathbf{fict}}$ – the force of moving space and Coriolis inertial forces of the shaft mass center;

Equation (4) in projections on the axis of the non-inertial reference frame $O\xi\eta\zeta$ has the form:

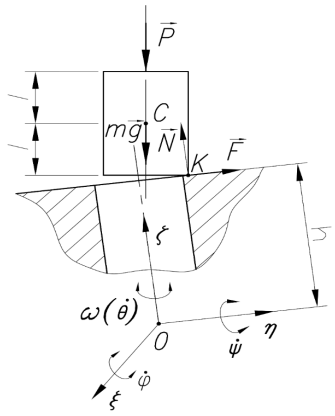


Figure 3: Scheme of forces at single-point contact of the assembled parts

$$\begin{aligned} m\ddot{\xi}_c &= (mg)_\xi + F_{f\xi} + F_{ce\xi}^{fict} + F_{cC\xi}^{fict} + P_\xi, \\ m\ddot{\eta}_c &= (mg)_\eta + F_{f\eta} + F_{ce\eta}^{fict} + F_{cC\eta}^{fict} + P_\eta, \\ m\ddot{\zeta}_c &= (mg)_\zeta + F_{ce\zeta}^{fict} + F_{cC\zeta}^{fict} + N + P_\zeta. \end{aligned} \tag{5}$$

where ξ_c, η_c, ζ_c are the coordinates of shaft mass center (point C) in the reference frame $O\xi\eta\zeta$.

To determine the projections of all forces on the axis of the non-inertial reference frame $O\xi\eta\zeta$, the apparatus of homogeneous coordinate transformation matrices was used [5].

Differential equations of motion of the shaft mass center with respect to the non-inertial reference frame attached to the base part

In [14] the components of (5) were determined, a mathematical model of the dynamics of the relative motion of the shaft mass center in its one-point contact with the plane of the bushing can be as follows:

$$\begin{aligned} m\ddot{\xi}_c &= p\sin\psi - fN\frac{\dot{\xi}_k}{\sqrt{\dot{\xi}_k^2+\dot{\eta}_k^2}} + m(\xi_c\dot{\psi}^2 - \eta_c\ddot{\psi}\sin\varphi \\ &\quad - 2\eta_c\dot{\varphi}\dot{\psi}\cos\varphi - 2\eta_c\omega\dot{\varphi}\sin\varphi - \zeta_c\ddot{\psi}\cos\varphi \\ &\quad + 2\zeta_c\dot{\varphi}\dot{\psi}\sin\varphi - 2\zeta_c\omega\dot{\varphi}\cos\varphi\cos\psi \\ &\quad + \xi_c\omega^2\cos^2\varphi + \eta_c\omega^2\sin\psi\sin\varphi\cos\psi \\ &\quad + \zeta_c\omega^2\sin\psi\cos\varphi\cos\psi) - 2m\dot{\zeta}_c(\dot{\psi}\cos\varphi + \omega\cos\psi\sin\varphi) \\ &\quad + 2m\dot{\eta}_c(-\dot{\psi}\sin\varphi + \omega\cos\psi\cos\varphi); \\ m\ddot{\eta}_c &= -p\cos\psi\sin\varphi - fN\frac{\dot{\eta}_k}{\sqrt{\dot{\xi}_k^2+\dot{\eta}_k^2}} \\ &\quad + m(\xi_c\ddot{\psi}\sin\varphi + \eta_c\dot{\psi}^2\sin^2\varphi + \eta_c\varphi^2 - 2\eta_c\omega\dot{\varphi}\sin\psi \\ &\quad + \zeta_c\dot{\psi}^2\cos\varphi\sin\varphi + \xi_c\omega^2\sin\psi\cos\psi\sin\varphi \\ &\quad + \zeta_c\dot{\varphi} + \eta_c\omega^2\sin^2\dot{\varphi}\sin^2\psi + \zeta_c\omega^2\sin^2\psi\sin\dot{\varphi}\cos\varphi \\ &\quad + 2\xi_c\omega\dot{\psi}\sin\psi\cos\varphi - 2\eta_c\omega\dot{\psi}\cos\psi\sin\varphi\cos\varphi \\ &\quad - 2\zeta_c\omega\dot{\psi}\cos\psi\cos^2\varphi + \eta_c\omega^2\cos^2\varphi - \zeta_c\omega^2\sin\varphi\cos\varphi) \\ &\quad - 2m\dot{\xi}_c(-\dot{\psi}\sin\varphi + \omega\cos\psi\cos\varphi) + 2m\dot{\zeta}_c(\dot{\varphi} - \omega\sin\psi); \end{aligned} \tag{6}$$

$$\begin{aligned} N &= m\ddot{\zeta}_c - m(\xi_c\omega^2\cos\psi\sin\psi\cos\varphi \\ &\quad + \eta_c\omega^2\sin^2\psi\sin\varphi\cos\varphi - \eta_c\omega^2\sin\varphi\cos\varphi \\ &\quad + \zeta_c\omega^2\sin^2\psi\cos^2\varphi + \zeta_c\omega^2\sin^2\varphi - 2\zeta_c\omega\dot{\varphi}\sin\psi \\ &\quad + \xi_c\ddot{\psi}\cos\varphi + \eta_c\dot{\psi}^2\sin\varphi\cos\varphi - \eta_c\ddot{\varphi} + \zeta_c\dot{\psi}^2\cos^2\varphi \\ &\quad + \zeta_c\dot{\varphi}^2 - 2\xi_c\omega\dot{\psi}\sin\psi\sin\varphi + 2\eta_c\omega\dot{\psi}\cos\psi\sin^2\varphi \\ &\quad + 2\zeta_c\omega\dot{\psi}\cos\psi\sin\varphi\cos\varphi) + 2m\dot{\eta}_c(\dot{\varphi} - \omega\sin\psi) \\ &\quad - 2m\dot{\xi}_c(\dot{\psi}\cos\varphi + \omega\cos\psi\sin\varphi) + p\cos\psi\cos\varphi. \end{aligned}$$

where N – the normal reaction of the plane of the bushing. f – the term coefficient of friction; $\dot{\xi}_k$ and $\dot{\eta}_k$ – the projections of the velocity of the contact point $"K"$ on the axis $O\xi, O\eta$ of the reference frame $O\xi\eta\zeta$.

Derivatives $\dot{\xi}_k$ and $\dot{\eta}_k$ are determined as follows [14]:

$$\begin{aligned} \dot{\xi}_k &= r\omega\cos\theta\cos\psi - r\sin\theta\sin\psi\cdot\dot{\psi} \\ &\quad + l\cos\psi\cdot\dot{\psi} + \dot{\xi}_c; \\ \dot{\eta}_k &= r(\omega\cos\theta\sin\psi\sin\varphi + \sin\theta\cos\psi\cdot\dot{\psi}\sin\varphi \\ &\quad + \sin\theta\sin\psi\cos\varphi\cdot\dot{\varphi} - \omega\sin\theta\cos\dot{\varphi} - \cos\theta\sin\varphi\cdot\dot{\varphi}) \\ &\quad - l(-\sin\psi\cdot\dot{\psi}\sin\varphi + \cos\psi\cos\varphi\cdot\dot{\varphi}) + \dot{\eta}_c \end{aligned} \tag{7}$$

and the coordinate ζ_c is determined as follows [14]:

$$\begin{aligned} \zeta_c &= h + r\left(\sin\psi\cos\varphi\sin\theta + \sin\varphi\cos\theta\right) \\ &\quad + l\cos\psi\cos\varphi \end{aligned} \tag{8}$$

The obtained mathematical models (6) and (7), (8) established the relationship between the components of the forces with factors of rotation movement and low-frequency vibrations.

3. Experimental Installation

3.1. Determining the Factors

A multifactor experiment of the process was carried out as follows. Upon reaching the operating mode of vibration, the shaft was fed into the assembly zone in the vertical direction with the contact of the faces of the assembled parts (Figure 1). Under the influence of vibrational and rotational movements of the experimental installation, the shaft was combined with the bushing and the assembly process was carried out. The contact moment of the assembled part faces was taken as the starting point. The completion of the assembly was recorded when the shaft was inserted to a depth of 50 mm. The resulting value of the assembly force was considered as the output optimized parameter (response) of the system when studying the process.

The following parameters were taken as input parameters (factors) in the study of the system: the angular frequency (rad/s); the linear amplitude of the last link of the vibrational device (mm); the linear velocity of the last link of the robot (mm/s); the angular velocity of the first link of the vibrational device (rad/s) and the gap in the joint (mm) (Table 1). All of them meet the appropriate requirements of controllability and integrity used to factors, as well

as the requirements of compatibility and lack of linear correlation applied to a set of factors.

Table 1: Input Factors

Factors	Symbol
The angular frequency of vibrational device (rad/s);	X_1
The linear amplitude of the vibrational device (mm);	X_2
The linear velocity of the last link of the robot (mm/s);	X_3
The angular velocity of the first link of the vibrational device (rad/s)	X_4
The gap in the joint (mm)	X_5

3.2. Selecting the Type of Model

A mathematical description of the assembly force components can be obtained using statistical experimental design [16].

The magnitude of the components of force and torque in the contact zone is measured by a force/torque sensor installed on the last link of the robot. Since the experimental sample was a shaft, then in the experiment 4 components of force and torque were investigated: axial force – F_z, torque around the z-axis – T_z, axial force – F_y and torque around the x-axis – T_x. These force and torque components are dependent on the input process factors. The mathematical model of the research object was understood as the form of the response function connecting the optimization parameter with the process factors. In general, this function can be written as follows:

$$F = f(X_i), \ (i = 1..5) \tag{9}$$

Selecting a type of model means choosing the type of function and its mathematical form. Suppose that the dependencies can be approximated with sufficient accuracy by equations of power type. So, for example, the relationship (9) can be expressed by the equation:

$$F = cX_1^{\alpha} X_2^{\beta} X_3^{\gamma} X_4^{\theta} X_5^{\mu} \tag{10}$$

where $X_i, i = 1..5$, process factors; $c, \alpha, \beta, \gamma, \theta, \mu$ – constant values.

After logarithm, equation (10) is linearized:

$$\ln F = \ln c + \alpha \ln X_1 + \beta \ln X_2 + \gamma \ln X_3 + \theta \ln X_4 + \mu \ln X_5 \tag{11}$$

The possibility of approximating dependences (10) by equations in the form (11) is established by verifying the hypothesis of the linear model adequacy when expressing the results of the experiment with a polynomial in the form:

$$y = b_0 + \sum_{1 \le i \le 5} b_i x_i + \sum_{1 \le i < j \le 5} b_{ij} x_i x_j \tag{12}$$

where $y = ln\, F$, $x_i -$ coded value of X_i.

Coded value x_i of factor i:

$$x_i = \frac{2(\ln X_i - \ln X_{i\max})}{\ln X_{i\max} - \ln X_{i\min}} + 1, \ (i = 1..5) \tag{13}$$

where X_i – natural value of factor i; $X_{i_{max}}$, $X_{i_{min}}$ – natural values of the upper and lower levels, respectively.

3.3. Design of Multifactor Experiments in the Search for Optimal Technological Conditions and Design Parameters of Equipment.

According to the verified experiment results, it is assumed that on the considered small interval, the response function can be described by a polynomial of the first degree. Due to the linearity of the model, the factors varied at two levels. At the same time, the zero levels were taken on the basis of previous studies [2] (Table 2).

Table 2: Levels of Factors and Variable Intervals

Factors	Levels			Range of variation	Dimension
	-1	*0*	*+1*		
X_1	14	16	18	2	rad/s
X_2	1,8	2,0	2,2	0,2	mm
X_3	10	20	30	10	mm/s
X_4	0	0,25	0,5	0,25	rad/s
X_5	0,01	0,03	0,05	0,02	mm

When conducting a full factorial experiment with varying factors at two levels, one replicate of it would involve $2^5 = 32$ experiments. It would make the experiment cumbersome enough if at least three replicates were performed. Therefore, the fractional factorial design of the experiment $N = 2^{5-1}$ was chosen. It reduces the number of experiments of one replicate to 16 due to the system of mixing linear effects of factor x_5 with the least significant interaction effects, but at the same time the number of replicates is increased to three. All experiments, according to the principle of randomization, were carried out in random order. Three replicates for all selected conditions were implemented (Table 3).

Table 3: Experiment Planning Matrix

№	x_0	x_1	x_2	x_3	x_4	x_5	x_1x_2	x_1x_3	x_1x_4	x_1x_5	x_2x_3	x_2x_4	x_2x_5	x_3x_4	x_3x_5	x_4x_5
1	+	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
2	+	+	-	-	-	+	-	-	-	+	+	+	-	+	-	-
3	+	-	+	-	-	+	-	+	+	-	-	-	+	+	-	-
4	+	+	+	-	-	-	+	-	-	-	-	-	-	+	+	+
5	+	-	-	+	-	+	+	-	+	-	-	+	-	-	+	-
6	+	+	-	+	-	-	-	+	-	-	-	+	+	-	-	+
7	+	-	+	+	-	-	-	-	+	+	+	-	-	-	-	+
8	+	+	+	+	-	+	+	+	-	+	+	-	+	-	+	-
9	+	-	-	-	+	+	+	+	-	-	+	-	-	-	-	+
10	+	+	-	-	+	-	-	-	+	-	+	-	+	-	+	-
11	+	-	+	-	+	-	-	+	-	+	-	+	-	-	+	-
12	+	+	+	-	+	+	+	-	+	+	-	+	+	-	-	+
13	+	-	-	+	+	-	+	-	-	+	-	-	+	+	-	-
14	+	+	-	+	+	+	-	+	+	+	-	-	-	+	+	+
15	+	-	+	+	+	+	-	-	-	-	+	+	+	+	+	+
16	+	+	+	+	+	-	+	+	+	-	+	+	-	+	-	-

3.4. Experimental Installation

Experimental Installation has been established (Figure 4). A force-torque sensor *2* (Schunk FT AXIA 80 Ethernet) was installed on the last link of the robot *1*. The base bushing *6* is kept in the vibrational device *7*, described in [6] and [17].

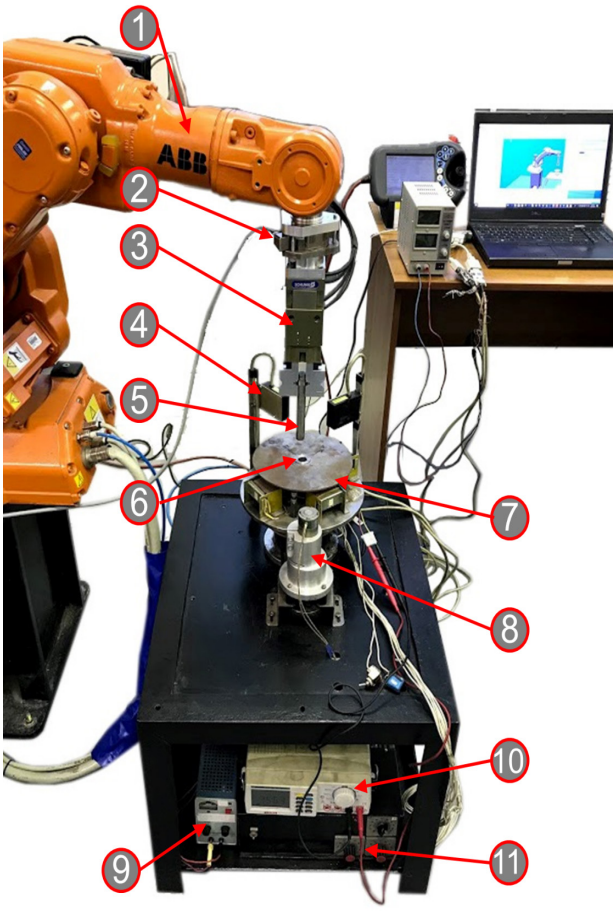


Figure 4: Experimental installation: 1. Industrial robot ABB IRB 140; 2. The force/torque sensor Schunk FT AXIA 80 Ethernet; 3. Gripper; 4. Triangulation Laser Sensors Riftek RF603.2; 5. The shaft; 6. The bushing; 7. The vibrational device; 8. DC motor; 9. Controlled converter; 10. Electronic frequency meter (Mastech M9803R); 11. The generator of low-frequency oscillations.

The experimental procedure was as follows. The shaft *5* was held by gripper *3* of the robot *1*. Vibrations were provided by the original generator of low-frequency oscillations *11*. The generator of low-frequency oscillations was controlled by the electronic frequency meter *10*, and the actual vibration amplitude of the vibrational device was monitored by contactless laser sensors *4*. The first link of the vibrational device performed a rotational motion with a constant angular velocity ω around a vertical fixed axis by a DC motor *8*. The DC motor was controlled by a controlled converter *9*. All sensor signals were collected and processed by the computer. Thus, during the experiments, continuous monitoring of the current values of technological factors was ensured.

The choice of force/torque sensor Schunk FT AXIA 80 Ethernet was determined by the required process sensitivity. The default value of sensitivity is 0.005 (Where: 0 = no damping - robot maintains speed; 1 = full damping - robot decelerates quickly) and its technical data is presented in Table 4.

Table 4: Technical Data of the Sensor Schunk FT AXIA 80 Ethernet

Storage and Operating Conditions			
Storage Temperature [°C]		-20 to +85	
Operating Temperature [°C]		0 to +65	
Electrical Specifications			
Power Source		***DC Power***	
Voltage [V] min. max.		 12 30	
Maximum Power Consumption [W]		1.5	
Measurement Ranges			
Parameter	***F_{xy} [N]***	***F_z [N]***	***T_{xyz} [Nm]***
Measurement Range 0	500	900	20
Measurement Range 1	200	360	8

4. Results and Discussion

4.1. Screening of Coarse Observations in Parallel Experiments

For each row of the experiment planning matrix, according to the results of *n* parallel experiments, $\bar{y}_j$ – arithmetic mean value of optimization parameter was found:

$$\bar{y}_j = \frac{1}{n}\sum_{u=1}^{n} y_{ju} \tag{14}$$

Where u is the number of parallel experiment; y_{ju} is the value of the response in the u parallel experiment of the j row of the experiment planning matrix.

In order to estimate the deviations of the response from its mean value for each row of the experiment planning matrix, the dispersion s_j^2 of the experiment was calculated according to n parallel experiments (Table 5). Statistical dispersion was estimated by an expression of the form:

$$s_j^2 = \frac{1}{n-1}\sum_{u=1}^{n}(y_{ju} - \bar{y}_j)^2 \tag{15}$$

Table 5: Experimental Results and Dispersion of Forces F_y, F_z and Torques T_x, T_z

№	F_y		F_z		T_x		T_z	
	$\bar{y}$	s_j^2	$\bar{y}$	s_j^2	$\bar{y}$	s_j^2	$\bar{y}$	s_j^2
1	2.442	0.015	1.260	0.007	-1.715	0.00759	4.262	0.002
2	2.297	0.016	1.196	0.009	-1.556	0.00983	4.252	0.002
3	2.722	0.006	1.496	0.004	-1.368	0.00656	4.330	0.006
4	2.735	0.037	1.524	0.004	-1.135	0.00272	4.380	0.001
5	2.400	0.062	1.291	0.002	-1.504	0.01142	4.176	0.004
6	2.373	0.012	1.236	0.012	-1.736	0.00990	4.459	0.005
7	2.745	0.020	1.542	0.019	-1.363	0.01318	4.557	0.003
8	2.913	0.017	1.578	0.017	-1.241	0.00276	4.415	0.003

№	F_y		F_z		T_x		T_z	
	$\bar{y}$	s_j^2	$\bar{y}$	s_j^2	$\bar{y}$	s_j^2	$\bar{y}$	s_j^2
9	3.341	0.012	1.858	0.002	-0.742	0.00113	3.942	0.004
10	3.075	0.029	1.814	0.001	-0.613	0.01398	4.306	0.002
11	3.204	0.002	1.873	0.003	-0.597	0.00232	4.368	0.001
12	3.557	0.012	1.919	0.002	-0.589	0.00109	4.180	0.001
13	3.321	0.017	1.930	0.001	-0.892	0.00393	4.169	0.006
14	3.240	0.006	1.854	0.004	-0.771	0.00157	4.214	0.003
15	3.684	0.008	2.050	0.009	-0.688	0.00201	4.297	0.011
16	3.543	0.004	2.067	0.009	-0.621	0.00306	4.448	0.001

4.2. Verification of the Uniformity of Dispersions

The dispersion homogeneity was tested by the Fisher F criterion, which is the ratio of the largest to the smallest variance:

$$F = \frac{s_1^2}{s_2^2} \tag{16}$$

where s_1^2 и s_2^2 - respectively, the largest and smallest values of the response variance in different experiments.

If the observed value is $F < F_t$ (F_t is the table value of the criterion F for the corresponding freedom degrees $f_1 = n_{\max} - 1 = 2$, $f_2 = n_{min} - 1 = 2$ and the accepted significance level of 0.05 [16]) then the variances are uniform (Table 6).

Table 6: Testing the Homogeneity of Dispersion by the Fisher Criterion

$F < F_t = 19.2$							
F_y		F_z		T_x		T_z	
14.62876	Yes	13.23305	Yes	18.30958	Yes	12.82161	Yes

With uniform duplication of experiments, the homogeneity of a number of dispersions was checked using the Cochrane G criterion, which is the ratio of the maximum dispersion to the sum of all dispersions:

$$G = \frac{s_{\max}^2}{s_1^2 + s_2^2 + \ldots + s_N^2} = \frac{s_{\max}^2}{\sum_{j=1}^{N} s_j^2} \tag{17}$$

The dispersions are homogeneous if $G < G_t$ (G_t is the table value of the criterion G for the conditions: N is the number of compared variances $N = 16$, and n is the number of parallel experiments $n = 3$ [16]). If $G > G_t$, then the variances are inhomogeneous, and this indicates that the investigated value y does not obey the normal law (Table 7).

Table 7: Testing the Homogeneity of Dispersion by the Cochrane Criterion

$G < G_t = 0.3346$							
F_y		F_z		T_x		T_z	
0.222947	Yes	0.211286	Yes	0.182417	Yes	0.150301	Yes

4.3. Determination of Regression Coefficients

According to the results of the experiment, the coefficients of the model were calculated. The regression coefficients are determined by the formulas:

$$b_0 = \frac{1}{N}\sum_{j=1}^{N} \bar{y}_j; \; b_j = \frac{1}{N}\sum_{j=1}^{N} x_{ij}\bar{y}_j; \qquad b_{jl} = \frac{1}{N}\sum_{j=1}^{N} x_{ij}x_{lj}\bar{y}_j. \tag{18}$$

where i, l – numbers of factors; x_{ij}, x_{lj} – coded values of factors $i$and l in the j-th experiment. These formulas are obtained using the least squares method.

The significance of each coefficient of the regression equation was checked by using Student t-criterion, comparing the obtained values with the table values in the corresponding degrees of freedom and significance level:

$$t_j = \frac{|b_j|}{s_{\{b\}}} \tag{19}$$

where $s_{\{b\}}$ – the dispersion of the regression coefficients, determined by formula:

$$s_{\{b\}}^2 = \frac{s_y^2}{N \cdot n} \tag{20}$$

If the dispersion s_j^2 of the experiments are homogeneous, the dispersion s_y^2 of the reproducibility of the experiment is calculated by the expression:

$$s_y^2 = \frac{1}{N}\sum_{j=1}^{N} s_j^2 \tag{21}$$

Table 8: Significance of each Coefficient of the Regression Equation

b_i	F_y	F_z	T_x	T_z
b_0	2.975	4.297	1.655	-1.071
b_1	–	0.035	–	0.038
b_2	0.163	0.075	0.101	0.12
b_3	0.053	0.045	0.038	-0.031
b_4	0.396	-0.057	0.265	0.382
b_5	0.045	-0.071	–	–
b_{12}	0.057	-0.051	–	–
b_{13}	–	–	–	-0.028
b_{14}	–	–	–	–
b_{15}	–	–	–	–
b_{23}	–	–	–	–
b_{24}	–	–	-0.044	-0.055
b_{25}	–	–	–	-0.034
b_{34}	–	–	–	–
b_{35}	–	–	–	0.038
b_{45}	0.04	–	–	–

Student criterion t is calculated for each regression coefficient. Statistically insignificant coefficients can be excluded from (16). The verification results are summarized in the Table 8. The regression equations for the considered response functions $y_{F_y}, y_{F_z}, y_{T_x}, y_{T_z}$ can be written based on the data given in Table 8.

4.4. Verification of Model Adequacy

After calculating the coefficients of the model and checking their significance, the dispersion of adequacy was determined. The residual dispersion or the adequacy dispersion characterizes the scattering of empirical values of y with respect to the calculated y determined by the found regression equation. The adequacy dispersion was determined by the formula:

$$s_{ad}^2 = \frac{n\sum_{j=1}^{N}(\bar{y}_j - \hat{y}_j)^2}{f} = \frac{n\sum_{j=1}^{N}(\bar{y}_j - \hat{y}_j)^2}{N-(k+1)} \quad (22)$$

where $\bar{y}_j$ – arithmetic mean value of the optimization parameter in the j-th experiment; $\widehat{y}_j$ – value of the optimization parameter calculated by the model for the conditions of the j-th experiment; f – the number of freedom degrees; k – the number of factors.

The final step in processing the experimental results was to verify the hypothesis of the adequacy of the found model. Hypothesis verification was carry out according to the Fisher criterion:

$$F = \frac{s_{ad}^2}{s_y^2} \quad (23)$$

If the value $F < F_t$ (F_t – table value of the criterion, $F_t = 2.1$ for the accepted significance level of 0.05 and the corresponding numbers of freedom degrees $f_1 = N - (k + 1) = 16 - (5 + 1) = 10, f_2 = N(n - 1) = 32$ [16]) then the model is considered adequate. For $F > F_t$, the adequacy of hypothesis is rejected (Table 9).

Table 9: Result of Model Adequacy Checking

	F_y	F_z	T_x	T_z
s_{ad}^2	0.021959	0.003026	0.005974	0.008727
s_y^2	0.017306	0.003367	0.00647	0.005815
F	1.269	0.899	0.923	1.501
$F < F_t$	Yes	Yes	Yes	Yes

In this case, we can conclude that the resulting model satisfies the adequacy condition.

For the transition from the coded values to natural values of the factors in the regression equation, values x_i factors on expressions (12) were substituted by (13) and regression coefficients values b_i were referenced in Table 8, obtained:

$$\begin{aligned}\ln F_y = {} & 11.111 - 3.111\ln X_1 - 10.875\ln X_2 \\ & +0.096\ln X_3 + 0.188\ln X_4 + 0.117\ln X_5 \\ & +4.521\ln X_1 \ln X_2 + 0.016\ln X_4 \ln X_5\end{aligned} \quad (24)$$

$$\begin{aligned}\ln F_z = {} & -5.322 + 3.062\ln X_1 + 11.931\ln X_2 \\ & +0.082\ln X_3 - 0.188\ln X_4 \\ & -0.088\,\mathrm{n}\, X_5 - 4.405\ln X_1 \ln X_2\end{aligned} \quad (25)$$

$$\begin{aligned}\ln T_x = {} & 1.458 + 0.470\ln X_2 + 0.069\ln X_3 \\ & +0.182\ln X_4 - 0.141\ln X_2 \ln X_4\end{aligned} \quad (26)$$

$$\begin{aligned}\ln T_z = {} & -4.670 + 1.459\ln X_1 - 1.075\ln X_2 \\ & +1.392\ln X_3 + 0.244\ln X_4 - 0.406\ln X_1 \ln X_3 \\ & -0.176\ln X_2 \ln X_4 - 0.421\ln X_2 \ln X_5 \\ & +0.045\ln X_5 + 0.086\ln X_3 \ln X_5\end{aligned} \quad (27)$$

Potentiating expressions (24 – 27), the dependence of forces and torques on the studied factors of the robotic assembly process was found:

$$F_y = \frac{66893.374 X_3^{0.096} X_4^{(0.188+0.016\ln X_5)} X_5^{0.117} X_1^{4.521\ln X_2}}{X_1^{3.111} X_2^{10.875}} \quad (28)$$

$$F_z = \frac{0.005 X_1^{(3.062-4.045\ln X_2)} X_2^{11.931} X_3^{0.082}}{X_4^{0.018} X_5^{0.088}} \quad (29)$$

$$T_x = 4.298 X_2^{(0.470-0.141\ln X_4)} X_3^{0.069} X_4^{0.182} \quad (30)$$

$$\begin{aligned}T_z = {} & 0.009 X_1^{(1.459-0.406\ln X_3)} X_2^{(-1.075-0.176\ln X_4-0.421\ln X_5)} \\ & X_3^{(1.392+0.086\ln X_5)} X_4^{0.244} X_5^{0.045}\end{aligned} \quad (31)$$

According to (28 - 31), nomograms can be constructed, which allow, for practical purposes, to determine the components of the force and torque acting on the installed part when determining factors for the robotic assembly process.

A suitable response for the above regression model F_z (29) was constructed and nomograms (Figure 5) were constructed with paired combinations of these factors while the other factors were at the zero level.

Figure 5 illustrates the dependence of the force F_z on variable factors in the assembly process. It is clear that the angular velocity and the gap in the joint are inversely proportional to the force F_z, while the vibration frequency, the vibration amplitude and the linear velocity are proportional to the force F_z. Also, there is an interaction effect of two factors of the angular frequency and the vibration amplitude. The maximum value of the force F_z was 87.17 N (Figure 5.f), the minimum value of F_z was 57.86 N (Figure 5. g).

Force F_z decreases significantly with increasing the angular velocity (Figure 5. c, f, h, j). Force F_z decreases rapidly at the initial stage, and then more slowly decreases in the investigated values range of the angular velocity. The maximum value of the force F_z in the absence of the rotational movement was 87.17 N (Figure 5. f), the minimum value of F_z with the rotational movement was 65 N (Figure 5. h). Also, F_z decreases with an increase in the gap in the joint. With the small gap, the maximum value of the force F_z was 84.7 N (Figure 5. j). With the large gap, the minimum value of the force F_z was 57.86 N (Figure 5. g).

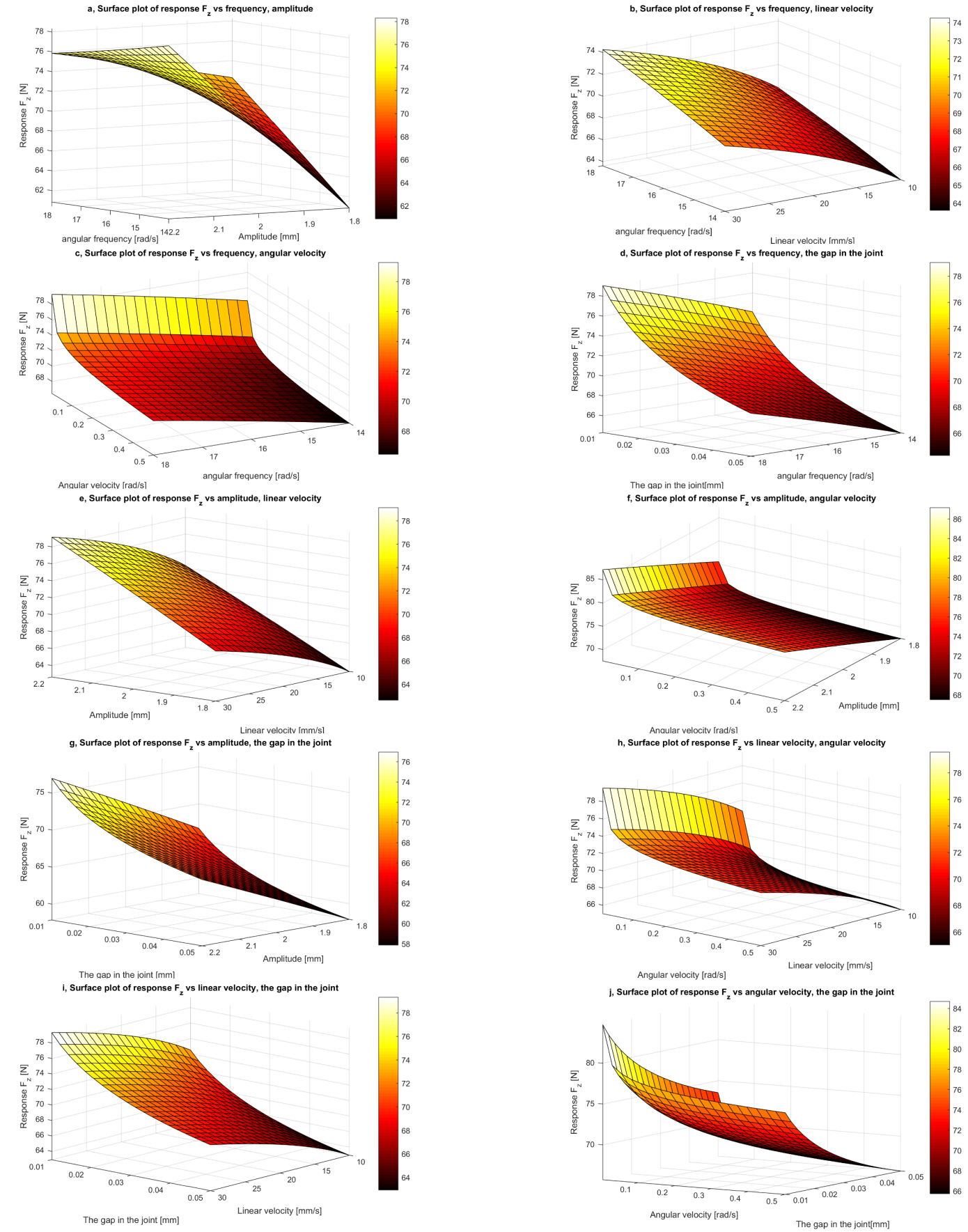


Figure 5: Three-dimensional nomograms showing the behavior of force F_z affected by factors of the robotic assembly process.

The force F_z is proportional to the angular frequency and vibration amplitude. The force F_z decreases with a simultaneous increase or decrease of both these factors (Figure 5. a – Figure 5. g). The angular frequency and vibration amplitude interact with each other, affecting the magnitude of the force F_z. When both factors changed in the investigated values range of them, the change in the value of the force F_z was small (60.8 – 78.3 N) (Figure 5. a). The effect of linear velocity on force F_z is negative (Figure 5.b, e, h, i), however, decreasing the linear velocity the assembly time will increase.

5. Conclusion

The conducted experimental studies revealed the influence of variable factors on the components of the assembly force and torque. Mathematical models that establish the relationship between the components of the friction forces and torques with disturbing factors are formed. The adequacy of the mathematical model of the process is proved. The effects of the interaction of two factors are identified.

The effect of increasing the assembly speed is negative, which correlates with the theory of friction. At the same time, the greatest effect of reducing the assembly force was achieved in the areas corresponding to the lower boundaries of the variation intervals. In various experiments, the decrease in assembly force ranged from 14 to 33%.

The conducted experiments proved the presence of the effect of vibrations and rotation on increasing the technological reliability of the assembly process. Further research will be associated with a change in the zero level of variation of the input factors. It is supposed to check the adequacy of the mathematical model and the results of physical experiments.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors express their gratitude to the Department “Technologies and Equipment of Mechanical Engineering”; Ph.D., Assoc. Prof., Petukhov Sergey Leonidovich, Senior lecturer Mishin Vladimir Nikolaevich and Ph.D., Assoc. Prof., Arkhipov Maksim Viktorovich, Faculty of Mechanical Engineering, Moscow Polytechnic University for the opportunity given to them to work in the laboratory, for significant support and useful suggestions for improving the research.

References

[1] M.V. Vartanov, L.V. Bozhkova, Z.K. Bakena Mbua, “Robot assembly of complex shafts on the basis of passive adaptation and low-frequency vibration”, Russian Engineering Research, **30**, 1092–1094, 2010, https://doi.org/10.3103/S1068798X10110055

[2] A.L. Simakov, Обоснование методов и средств адаптации соединяемых деталей на базе принципов автоматического управления и выявленных взаимосвязей при автоматизированной сборке, Sc.D Thesis, Kovrov State Technological Academy, 2003.

[3] M.G. Kristal, Производительность и надежность сборочных автоматов: монография – Volgograd: VolSTU, 2011.

[4] L.B. Chernyakhovskaya, “Influence of rotational movement of a peg on the automatic peg-in-hole assembly of cylindrical details,” Assembly in mechanical engineering, instrument making, **6**, 7-13, 2016 (article in Russian with an abstract in English).

[5] L.B. Chernyakhovskaya, Кинематический и динамический анализы автоматической сборки цилиндрических деталей: Монография - Samara: Samara State Technical University, 2011.

[6] D.M. Levchuk, Исследование и разработка методов относительного ориентирования сборочных единиц соединения во вращающемся потоке газов при автоматической сборке, Ph. D Thesis, Moscow Automotive Institute, 1974.

[7] B. Bakšys, J. Baskutienė, A. Chadarovičius, “Simulation of vibratory alignment of the parts to be assembled under passive compliance,” Mechanika, **19**(1), 33-39, 2013, doi: https://doi.org/10.5755/j01.mech.19.1.3625.

[8] B. Baksys, J. Baskutiene, “The directional motion of the compliant body under vibratory excitation,” International Journal of Non-Linear Mechanics, **47**(3), 129–136, 2012, doi: https://doi.org/10.1016/j.ijnonlinmec.2012.01.008.

[9] A.A. Ivanov, “Vibration assembly systems," Assembly in mechanical engineering, instrument making, **5**, 7-10, 2013 (article in Russian with an abstract in English).

[10] M.G. Kristal, I.A. Chuvilin, “Research of dynamics of internal member lower flank cradle vibrating meshing,” Assembly in mechanical engineering, instrument making, **4**, 13-17, 2008 (article in Russian with an abstract in English).

[11] V.G. Shuvaev, V.A. Papshev, “Compound oscillatory motion forming device for space orientation of automatically assembling details,” Assembly in mechanical engineering, instrument making, **11**, 15-18, 2009 (article in Russian with an abstract in English).

[12] A.I. Khuri and S. Mukhopadhyay, “Response surface methodology,” Wiley Interdiscip. Rev. Comput. Stat., **2**(2), 128–149, 2010, doi: http://dx.doi.org/10.1002/wics.73

[13] M. A. Bezerra, R. E. Santelli, E. P. Oliveira, L. S. Villar, and L. A. Escaleira, “Response surface methodology (RSM) as a tool for optimization in analytical chemistry,” Talanta, **76**(5), 965–977, 2008, doi: 10.1016/j.talanta.2008.05.019.

[14] T. T. Tran, M. V. Vartanov and M. V. Arkhipov, "Application of the Part Rotation Effect for Reliability of the Robotic Assembly Process," in 2019 International Conference on System Science and Engineering (ICSSE), 25-30, 2019, doi: 10.1109/ICSSE.2019.8823286.

[15] Paul, Richard. Modelling, Trajectory Calculation and Servoing of a Computer Controlled Arm [Text] / R. Paul; Translation from English. A. F. Vereshchagin, V. L. Generozov / Ed. E.P. Popova. - Moscow: Nauka, 1976.

[16] M.G. Kristal, Обработка результатов планирования экстремального эксперимента, Volgograd, 2019

[17] L.V. Bozhkova, M.V. Vartanov, E.I. Kolchugin “Experimental setup for robotic assembly on the basis of passive adaptation and low-frequency oscillations,” Assembly in mechanical engineering, instrument making, 1, 5-7, 2009 (article in Russian with an abstract in English).

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 169-174 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

A Novel Approach to Augment the Opto-Electronic Properties of Stannic Oxide (SnO_2) Thin Films by Vanadium Doping

A. Victor Babu[*,1], S. Murugan[1], D.C. Bernice Victoria[2]

[1]*A.V.C. College Autonomous, (Affiliated to Bharathidasan University, Tiruchirappalli), Mannampandal, 609305, India*

[2]*Department of Physics, TELC School of Higher Education, Pudukkottai, 622003, India*

ARTICLE INFO

Article history:
Received: 28 August, 2020
Accepted: 23 October, 2020
Online: 08 November, 2020

Keywords:
Doping Concentration
Optical Properties
SnO_2 Thin films
Spray Pyrolysis

ABSTRACT

The Tin Oxide: Vanadium (SnO_2: V) thin films have been deposited on glass substrates with various Vanadium concentration at a substrate temperature;500°C using microprocessor-controlled spray pyrolysis technique. The structural, compositional, morphological and optical analyses of the samples are characterized by XRD, EDAX, SEM and UV-Visible spectroscopy. The XRD analysis shows tetragonal rutile structure formation in lattice configuration of virgin and Vanadium doped SnO_2 thin films. The uniform crystallinity is appeared on the surface of crystal and it is proved by extreme diffraction peak for 3at% Vanadium doped SnO_2 thin films. The optical transmittance increases from virgin to 3at% with increase in doping concentration. The transmittance for 4at% is minimum compared with 3at% of vanadium doping. The FTIR analysis illustrates the bond formation between Sn-O and V-O respectively. This also validates the strong force constant for the consistent crystal structure.

1. Introduction

Thin film technologies paved the way of preparing the most promising material [Transparent conducting oxides] in optical energy controlled electronic devices; solar cells, displays, opto-electronic interfaces devices and electronic-circuitries, owing to its properties such as electrically conductive nature with high transmittance of light [1].The accomplishment of such properties is performed by successive doping with wide-band gap optical semiconductors (E_g > 3.1 eV) and are pivoted to advanced technologies[2-5]. In_2O_3 is considered to be the leading TCO in concerned with their properties as well as its commercial success. However, ITO is playing a major role in electronics market; it is of high cost and limited availability [6-9]. SnO_2 is economically cheap compared to indium oxide. It is an ideal candidate for many applications. Its high chemical stability is considered to be an inherent advantage of SnO_2 [10] and it makes SnO_2 appreciably adoptable for applications in challenging atmosphere, for preparing transparent and strong-conductive electrode on photo enabled electrodes or as a selectively solar energy penetration coating [11, 12]. Its applications include surface coatings, environmental and human health [13] gas sensors [14] organic pollutants and toxic water [15] and heterojunction catalysts [16], Tin dioxide (SnO_2), having lattice parameters a = b = 4.737 Å and c = 3.826 Å with tetragonal Rutile structure is an important n-type semiconductor. Each tin atom at the corners of a regular octahedron is bounded to six oxygen atoms and three tin atoms surround each oxygen atom [17]. High ability to let incoming radiation in the UV region, high reflectance in the IR region, high electrical conductivity, high melting temperature, harmless nature, easy doping and plenty in nature. These are the remarkable characteristics of SnO_2 [18]. For optoelectronic devices and solar cell applications, SnO_2 thin films is a very suitable candidate as it is having a good electrical and optical properties [19].

Many research investigations have been performed on heavily doped and rather doped tin oxide thin films using various available techniques [20-22]. Although there are several techniques for deposition, spray pyrolysis technique is the most preferred technique due to its simplicity, low cost and for large area coating [23]. Doping with certain p block and d block elements augments the light sensitive properties and the ability of transmitting electricity of SnO_2 thin films [24]. As the ionic radius of Vanadium ($V4^+/Sn4^+$ = 0.58/0.71Å) is smaller than the host $Sn4^+$, it can be easily doped into SnO_2 [25]. After screening the literatures, it is revealed that, few researches have analyzed the

*Corresponding Author: A. Victor Babu, A.V.C. College (Autonomous), Mannampandal, Email: victornicy@gmail.com

properties of Vanadium fused SnO_2 thin films [26-28], it is of particular interest to deposit Vanadium doped SnO_2 thin films by micro controller-assembled spray pyrolysis technique. In the present work, the Vanadium doped SnO_2 thin films have been characterized and they are investigated according to the operated properties. As a first step SnO_2: V thin films are synthesized. Then structural and light sensitive properties of the customized thin-films are keenly observed and the data is validated. This work intends to improve the characteristics of SnO_2 thin films by doping Vanadium at different concentration.

2. Experimental methods

2.1. Synthesis of vanadium doped SnO_2 thin films

In the blended mixture of doubly de-ionized water and intensive solvent; methanol in the ratio 1:1, the precursor material $SnCl_4.5H_2O$ (Tin (IV) Chloride Pentahydrate) (35.06g) is dissolved to obtain 0.1M starting solution and few drops of hydrochloric acid is added to get a clear solution. To the precursor, well grained powder V_2O_5(Vanadium Pentoxide) is added in the ratio of 0,1,2,3&4at% (0g, 0.509g, 1.018g, 1.527g, 2.036g). A reservoir filled with precursor solution is mounted on stand. The area of the substrate is 2.5x7.5cm^2 (optically flat glass slides). Formerly cleaning of the substrates are achieved using hydrochloric acid, non-polar solvent; acetone and sterilized water by suitable ultrasonic cleaner for 35 minutes to terminate organic substance and another contaminant from the surface.

The micro controller-operated spray pyrolysis set up is displayed in Figure1. It consists of a syringe needle having the diameter of 0.6mm connected to the reservoir and its height can be altered to get 0.002mL/Min at the needle tip. Distance of 8cm is kept between the needle and the surface of the substrate. Substrate attracts the precursor solution as high voltage of about (9.65KV) is applied between needle and hot-plate. To optimize the thickness (500nm) different volume of precursor solution is sprayed and the dimensions between the two surfaces of all the sprayed samples are verified using standard method. A timer is connected to high voltage power supply, relay and hot plate. The timer is set in such a way that when hot plate is on, high voltage will be in off condition and vice-versa. The charge presents in substrate holder and syringe needle, substrate temperature, heat inside the chamber are controlled by means of relay, thermocouple arrangement and exhaust fan. The distance of the syringe needle is adjusted up and down to obtain a continuous thin film. In the experiment, five sequential sets of undoped and Vanadium doped SnO_2 thin films are fabricated by controlling the substrate temperature at 500^0C.The optimized time allotment for deposition for each sample is determined as 44minutes by counting the spray intervals of time. The alternating spray process is employed for this study at a spray for 5seconds and an interval of 10seconds for each cycle.

2.2. Characterization Technique

In our present study the surface morphology of the prepared samples is characterized by employing scanning electron microscope (CARLZEISS) shown in Figure 2. The compositional details are analyzed by (EDAX) technique. The lattice parameters are examined by X-ray diffractometer (SHIMADZU-XRD 6000) (Cu–Ka) of wavelength 1.520Å. The profilometer is used to find the dimension between two surfaces of the films by making suitable stylus technique at four different areas on the outside of film and the average dimension is found to be around 500 nm. Weight gain method is also used to find the thickness. The optical absorption and translucence of the films are measured in the wavelength range of 300–1100nm using a Unico 4802 UV–Vis double beam spectrophotometer. FTIR spectra are recorded by means of SHIMADZU1800-UV Fourier transform infrared spectrometer.

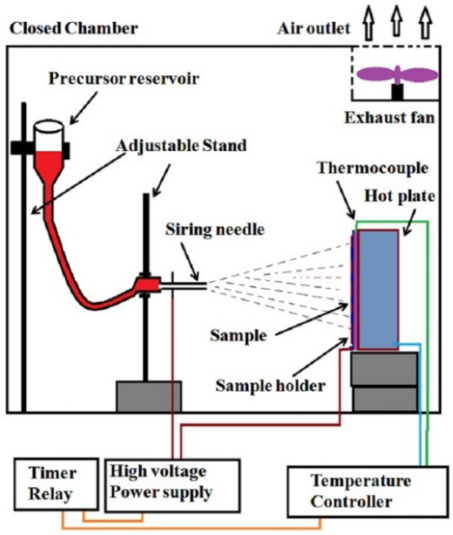


Figure 1: Schematic diagram of the Home-made spray pyrolysis setup.

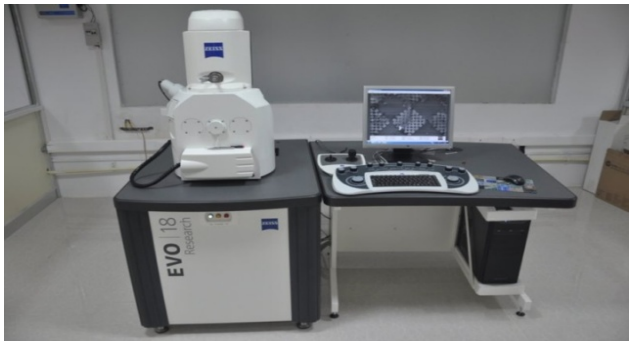


Figure 2: Carl Zeiss scanning electron microscope.

3. Results and Discussion

3.1. Structural properties

Figure 3 shows the XRD spectra of virgin and Vanadium doped SnO_2 thin films. The highest diffraction peaks are (110), (101), (200), (211), (220). The coincidence of diffracted peaks with tetragonal structure of SnO_2 (JCPDS card number: 41-1445) is observed for all the recorded peaks of Vanadium doped SnO_2 thin films [29]. No phase consistent to the dopant or other compound is observed. Though the doping level of Vanadium is increased from 1at% to 3at%, Vanadium oxide phases are not

dominated in Vanadium operated SnO_2 films. The intensity of the peak corresponding to the predominant plane (110) in the primary phase is increased on Vanadium doping until 3at% and further addition of 1at% doping concentration, the intensity decreases due to the exchange of $Sn4^+$ ions with Vanadium ions in the lattice of SnO_2 film, whereas the intensity of the plane (211) in the secondary phase rapidly increases with increase in the concentration of Vanadium. This shows better atomic arrangement and lower scattering in these planes. The structural analysis is done for the predominant peak and the results are given in Table 1. The optimized crystallite size of virgin SnO_2 and doped SnO_2 films for all the dopant levels are calculated using Debye-Scherrer formula [30]

$$D= K\lambda/\beta\cos\theta \quad (1)$$

where D is the mean crystallite size, K is a particle size dependent constant whose value is 0.9, λ is the diffraction wavelength of CuKa (λ = 1.5406 Å). β is the full width at half maximum (FWHM) and θ is the diffracted angle. The lattice parameters and crystallite size for the predominant peak (110) of the prepared samples are given in Table1. It is confirmed that the increase in the concentration of Vanadium from 0 to 4at% results in the range of crystallite size from 55.72nm to 41.79nm. The decrease in crystallite size may be due to the density enrichment in the nucleation centers of the doped samples and also may be due to lattice distortion. For further addition of 1at% Vanadium, dopant dominates the crystalline size.

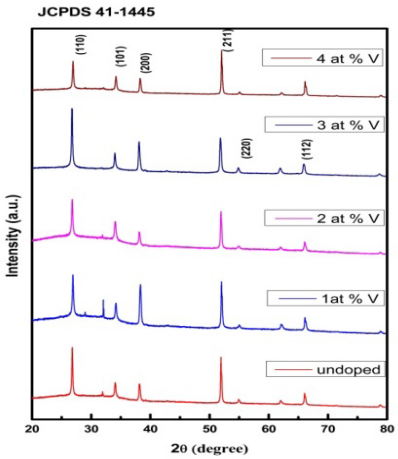


Figure 3: XRD Pattern of undoped and V (1at%, 2at%, 3at%, 4at%) doped SnO_2 thin films.

Table 1: Doping concentration ratio (1at%, 2at%, 3at%, 4at%) for SnO_2 thin films.

Predominant peak	Angle2θ (degree)	d spacing (Å)	FWHM	Crystallite size (nm)
Undoped (110)	26.806	3.3232	0.153	55.72
1at%V (110)	26.908	3.3107	0.180	47.37
2at%V (110)	26.932	3.3079	0.175	48.73
3at%V (110)	26.793	3.3248	0.204	41.79
4at%V (110)	26.748	3.3302	0.172	49.76

3.2. Morphological Analysis

The surface morphology of virgin and Vanadium doped SnO_2 thin films deposited at a temperature of 500°C with different Vanadium content concentration are shown in Figure 4. The SEM image of virgin SnO_2 exhibits smaller grains. It is evident from the images that the surface morphology reveals significant changes on doping [31]. For 1at% of Vanadium doped SnO_2 thin film, molecules of different morphology and sizes are formatted due to inadequate Vanadium doping. Grain size decreases when doping concentration of Vanadium increases [32]. This is observed in the SEM image for 2at% Vanadium doped SnO_2 thin film. A film having superior quality with uniform grains as proved by structural analysis is observed for 3at% Vanadium doped SnO_2 thin film. Above 3at% of the dopant concentration, grains are found to increase in its size and appear with small sized grains in between large sized grains.

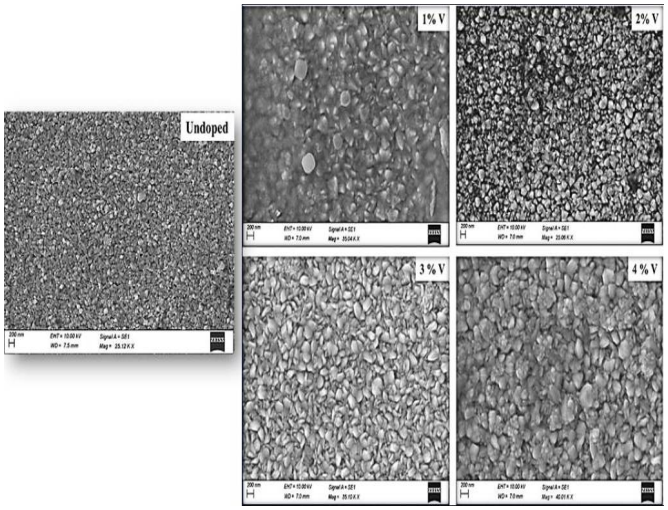


Figure 4: SEM images of undoped and V (1at%, 2at%, 3at%, 4at%) doped SnO_2 thin films.

3.3. Elemental analysis

From the EDAX analysis recorded in Figure 5. The consistent peaks Sn, V and O are observed in the EDAX pattern of all the Vanadium doped samples. For all V-doped SnO_2 samples the existence of Vanadium is also detected. Both vanadium extent and relative intensity are directly proportional to each other.

3.4. Optical properties

Optical transmittance spectra of virgin and Vanadium doped SnO_2 thin films are shown in Figure 6. In the visible region (400–700 nm), the observed translucence is about ~60–80%. For virgin SnO_2 thin film transmittance is minimum compared to the vanadium doped SnO_2 thin film. Transmittance increases with dopant concentration. Increase in transmittance may be due to the

improved crystallinity of the film. In our results, transmittance increases from virgin to 3at% with increase in dopant concentration. Transmittance decreases with further 1at% of Vanadium doping. The decrement of optical transmittance is made due to the increment of particle size and it is evidenced from the SEM image and also may be due to the crystal defects created by doping.

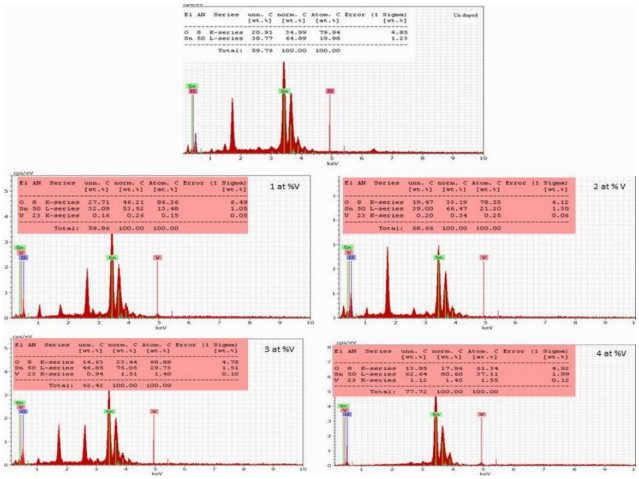


Figure 5: EDAX analysis spectra of undoped and V (1at%, 2at%, 3at%, 4at %) doped SnO_2 thin films.

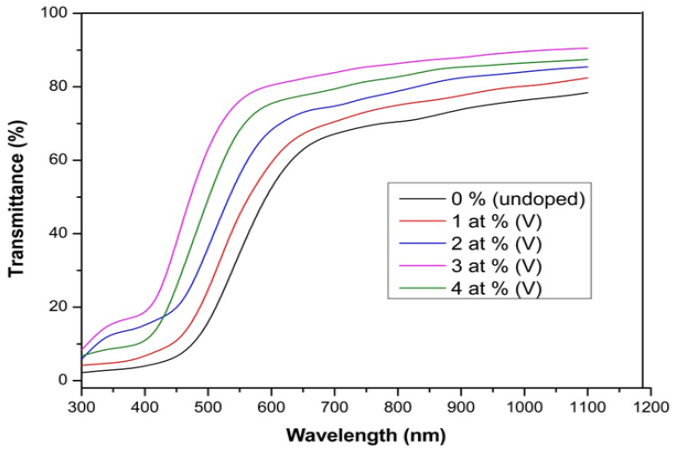


Figure 6: Transmittance spectra of undoped and V (1at%, 2at%, 3at%, 4at %) doped SnO_2 thin films.

The observed absorbance modulated spectra of virgin SnO_2 and Vanadium doped SnO_2 thin films at different concentrations are depicted in Figure 7. Gradually increasing absorption from NIR to the visible region is observed. The absorption edge is prolonged to the visible region due to the fusion of Vanadium ions into SnO_2 ions. A substantial red-shift is absolutely observed on the increment in doping attentiveness of vanadium species [33]. Absorption band edge of the films is observed by the calculated first derivative of the optical transmittance. The band gap is an important property which determines the quality of the deposited thin films. In Figure 8, the curves of dT/dλ vs hν have been drawn and showed the stress of doping on the direct optical band gap (Eg).The band gap values are observed to be ranged from 2.970 eV to 3.130 eV. There is fluctuation in band gap energies with doping concentrations of Vanadium. High value of band gap energy is a result of lower-quality films. In the present case good quality film is obtained in 3at% of Vanadium doping concentration.

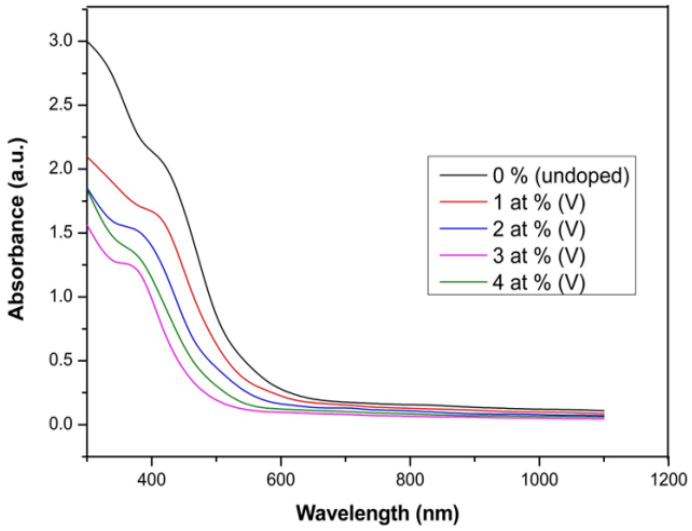


Figure 7: Absorption spectra of undoped and V (1at%, 2at%, 3at%, 4at%) doped SnO_2 thin films.

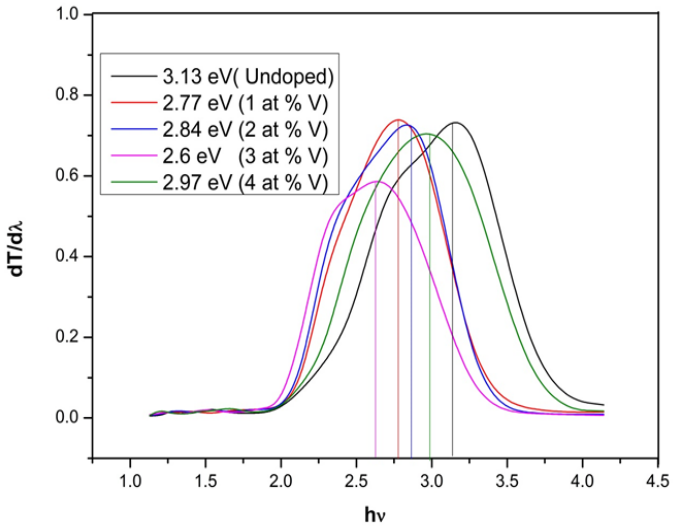


Figure 8: Plot of dT/dλ vs. (hυ) for undoped and V (1at%, 2at%, 3at%, 4at%) doped SnO_2 thin films.

3.5. FTIR analysis

Figure 9 displays the FT-IR spectra of virgin and Vanadium doped SnO_2 thin films fabricated by microprocessor-controlled spray pyrolysis method. As the present case is inorganic compound, all the IR peaks are concentrated in the region of 200-1200cm^{-1}. Particularly, the stretching, in plane and out of plane bending modes for Sn-O bond are observed at 750, 740 and 710cm^{-1}, the in plane bending modes are found at 400, 380 and 360cm^{-1} and the out of plane bending modes are occupied at 280, 260 and 250cm^{-1}. All the observed bands are validated by the literatures [34]. The V-O bond stretching, bending modes are determined at 610, 320 and 210cm^{-1}. All the spectroscopical impressions illustrate the strong coupling of V on SnO_2 configuration.

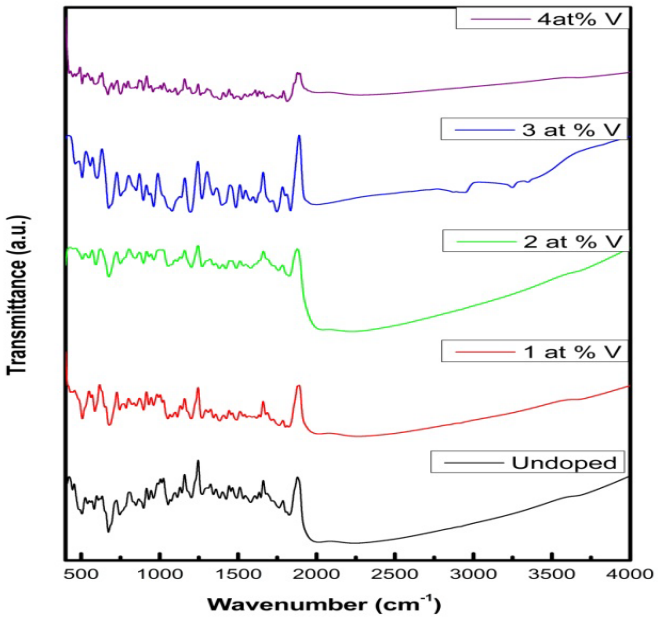


Figure 9: FTIR spectra of undoped and V (1at%, 2at%, 3at%, 4at%) doped SnO_2 thin films.

4. Conclusions

In this work, the popular heavy metal species Vanadium is doped on SnO_2crystal material and it is formed as thin films with the doping concentration of 1at%, 2at%, 3at% and 4at% respectively. The compositional, morphological, structural and optical analysis is made on samples and the results are discussed. From the observed data, the optimized state of doping concentration is found to be 3at%.The vibrational study is confirmed the active presence of strong bonds such as Sn-O and V-O. The presence of Sn-O-V bonds and angles are also validated from the complex form of vibrational bands.The film at doping of 3at % is having superior quality with uniform grains as proved by structural. From the optical studies, it is revealed that, the optical intensive transmittance increases from virgin to doped SnO_2 at 3at% which leads to the decrement of band gap energy. The acquired results from structural, morphological, vibrational and optical studies demonstrated the effectiveness of doped thin film.

Acknowledgment

The authors gratefully acknowledge to Alagappa University and Kalasalingam University for providing the facilities to characterize the samples.

References

[1] A. Stadler, "Transparent Conducting Oxides-An Up-To-Date Overview," Materials, **5**(4), 661-683, 2012, doi: 10.3390/ma5040661.

[2] S.C. Dixon, D.O.Scanlon, C.J.Carmalt, I.P.Parkin, "N-Type doped Transparent conducting binary oxides: an overview," Journal of Chemical, **4**(29), 6946−6961, 2016, doi: 10.1039/C6TC01881E.

[3] P. Ivan Parkin, D. Troy Manning, "Intelligent thermo chromic windows," Journal of Chemical Education, **83**(3), 393−400, 2006. doi: 10.1021/ed083p393.

[4] S. Das, V. Jayaraman, "SnO_2: A comprehensive review on structures and gas sensors," Progress Materials Science, **66**(23), 112−255, 2014, doi: 10.1016/j.pmatsci.2014.10.001.

[5] Hosono, "Recent progress in transparent oxide semiconductors: Materials and device application," Thin Solid Films, **515**(15), 6000−6014, 2007, doi: 10.1016/j.tsf.2006.12.125.

[6] M. Aliyu, S. Hossain, J. Husna, N. Dhar, M.Q. Huda, K. Sopian, N. Amin, "High quality indium tin oxide (ITO) film growth by controlling pressure in RF magnetron sputtering," 38th IEEE Photovoltaic Specialists Conference; IEEE, 002009−002013, 2012, doi: 10.1109/PVSC.2012.6317992.

[7] R. Gupta, K. Ghosh, S. Mishra, P. Kahol, "Structural, optical and electrical characterization of highly conducting Mo-doped In_2O_3 thin films," Applied Surface Science, **254**(13), 4018−4023, 2008, doi: 10.1016/j.apsusc.2007.12.035.

[8] T. Minami, "New n-Type Transparent Conducting Oxides, MRS Bulletin, **25**(8), 38−44, 2000, doi: 10.1557/mrs2000.149.

[9] A. Benjamin, D. Williamson, Thomas J. Featherstone,Sanjayan S. Sathasivam, Jack E. N. Swallow, Huw Shiel, Leanne A. H. Jones, Matthew J. Smiles, Anna Regoutz, Tien-Lin Lee, Xueming Xia, Christopher Blackman, Pardeep K. Thakur, Claire J. Carmalt, Ivan P. Parkin, Tim D. Veal and David O. Scanlon, "Resonant Ta Doping for Enhanced Mobility in Transparent Conducting SnO_2," Chem. Mater. **32**(5), 1964−1973, 2020, doi: 10.1021/acs.chemmater.9b04845.

[10] J.L. Vossen, Kern.W, (Eds.) Thin Film Processes II; Academic Press: Cambridge, MA, USA, 1991.

[11] F. Lungwitz, Ramon Escobar-Galindo, Daniel Janke, Erik Schumann, Robert Wenisch, Sibylle Gemming, Matthias Krause, "Transparent conductive tantalum doped tin oxide as selectively solar transmitting coating for high temperature solar thermal applications," Solar Energy Materials Solar Cells, **196**(1), 84–93, 2019, doi: 10.1016/j.solmat.2019.03.012.

[12] L. Kavan, L. Steier, Michael Grätzel, "Ultrathin Buer Layers of SnO_2 by Atomic Layer Deposition: Perfect Blocking Function and Thermal Stability," The Journal of Physical Chemistry C, **121**(1), 342–350. 2017, doi: 10.1021/acs.jpcc.6b09965.

[13] T. Chang, X. Cao, Shan-Hu Bao, Shi-Dong Ji, Hong-Jie Luo, Ping Jin,"Review on thermochromic vanadium dioxide based smart coatings: from lab to commercial application," Advances Manufacturing, **6**(1), 1-19, 2018, doi:10.1007/s40436-017-0209-2.

[14] C. Wang, Miao-Ting Chen, D. Lai, "Vanadium tin oxide nanoparticles with gas sensing and catalytic activity," Journal of the American Ceramic Society, **94**(12), 4471-4477, 2011, doi.10.1111/j.1551-2916.2011.04733.x.

[15] M. Wei, X. Peng, Qi-Sheng Liu, Fang Li, Ming-Ming Yao, "Nano crystalline TiO_2 composite films for the photo degradation of formaldehyde and oxytetracycline under visible light irradiation," Molecules, **22**(6), 1-13, 2017, doi:10.3390/molecules22060950.

[16] K.H. Abass, M. Mohammed, "Fabrication of ZnO: Al/Si Solar Cell and Enhancement its Efficiency Via Al-Doping," Nano Biomedicine and Engineering, **11**(2), 170- 177, 2019, doi.10.5101/nbe.v11i2.p170-177.

[17] W. Cai-hong, Xiangfeng Chu, M.Wu, "Highly sensitive gas sensors based on hollow SnO_2 spheres prepared by carbon sphere template method," Sensors and Actuators B Chemical, **120**(2), 508-513, 2007, doi: 10.1016/j.snb.2006.03.004.

[18] D.S. Ginley, Clark Bright, "Transparent Conducting Oxides," MRS Bulletin, **25**(8), 15 -18, 2000, doi:10.1557/mrs2000.256.

[19] A. Abdelkrim, S. Rahmane, "Optoelectronic properties of SnO_2 thin films sprayed at different deposition times," Chinese Physics B, **25**(4), 2016. doi:10.1088/1674-1056/25/4/046801.

[20] G. Sanon, R. Rup, A. Mansingh, "Band-gap narrowing and band structure in degenerate tin oxide (SnO_2) films," Physical Review B, **44**(11), 5672–5680, 1991, doi: 10.1103/physrevb.44.5672.

[21] K. S. Shamala, L.C.S. Murthy, K. Narasimha Rao "Studies on tin oxide films prepared by electron beam evaporation and spray pyrolysis methods," Bulletin of Materials Science, **27**(3), 295–301, 2004, doi: 10.1007/BF02708520.

[22] S. Hamzaoui, M. Adnane, "Effects of temperature and r.f. power sputtering on electrical and optical properties of SnO_2," Applied Energy, **65**(1-4), 19-28, 2000, doi: 10.1016/S0306-2619(99)00045-8.

[23] S. Chandra, K. Ravichandran, G. George, T. Arun, P. Rajkumar, "Influence of Fe and Fe+F doping on the properties of sprayed SnO_2 thin films," Journal of Materials Science Materials in Electronics,**27**(9),9558–9564, 2016, doi: 10.1007/s10854-016-5008-1.

[24] B. Zhang, Y. Tian, J. X. Zhang and W. Cai, "The FTIR studies of SnO_2: Sb (ATO) films deposited by spray pyrolysis," Materials Letters. **65**(8), 1204-1206, 2011, doi: 10.1016/j.matlet.2011.01.052.

[25] A. Popa, D. Toloman, O. Raita, M. Stan, O. Pana, T.D. Silipas, L.M. Giurgiu, "Ferromagnetic behavior of vanadium doped SnO_2 nanoparticles annealed at different temperatures," Journal of Alloys and Compounds, **591**, 2014, doi:10.1016/j.jallcom.2013.12.189.

[26] W.B. Soltan, M.S. Lassoued, S. Ammar, T. Toupance**, "**Vanadium doped SnO_2 nano particles for photo catalytic degradation of methylene blue," Journal of Materials Science Materials in Electronics, **28**(4),15826–15834, 2017, doi: 10.1007/s10854-017-7477-2.

[27] J. Mazloom, F.E. Ghods, H. Golmojdeh,"Synthesis and characterization of vanadium doped SnO_2 diluted magnetic semiconductor nano particles with enhanced photo catalytic activities", Journal of alloy compounds,**639**(1),393-399,2015, doi: 10.1016/j.jallcom.2015.03.184.

[28] R. Nasiraei, M. R. Fadavieslam, and H. Azimi-juybari, "Structural, photoconductive, thermoelectric and activation energy measurements of V-doped transparent conductive SnO_2 films fabricated by spray pyrolysis technique," Pramana Journal of Physics, **87**(2), 1-10, 2016, doi: 10.1007/s12043-016-1237-3.

[29] Y. Fu, H. Ma, Zhenl¨u Wang, Wanchun Zhu, Tonghao Wu, Guo-jia Wang. "Characterization and reactivity of SnO_2-doped V_2O_5/γ-Al_2O_3 catalysts in dehydrogenation of isobutane to isobutene," Journal of Molecular Catalysis A: Chemical, **221**(1), 163–168, 2004, doi:10.1016/j.molcata.2004.07.016.

[30] S. Chacko, N. Sajeeth Philip , K.G.Gopchandran, Peter Koshy, V.K.Vaidyan, "Nanostructural and surface morphological evolution of chemically sprayed SnO_2 thin films," Applied Surface Science, **254**(7), 2179-2186, 2008, doi: 10.1016/j.apsusc.2007.09.027.

[31] J.F. Chang, H.H. Kuo, I.C. Leu, M.H. Hon, "The effects of thickness and operation temperature on ZnO: Al thin film CO gas sensor," Sensors and Actuators B: Chemical, **84**(2), 258–264, 2002, doi: 10.1016/S0925-4005(02)00034-5.

[32] C. Venkata Reddy, B. Babu, S.V. Prabhakar Vattikuti, R.V.S.S.N. Ravikumar, Jaesool Shim "Structure and optical properties of vanadium doped SnO_2 nanoparticles with high photo catalytic activities," Journal of Luminescence, **179**, 26-34, 2016, doi:10.1016/j.jlumin.2016.06.036.

[33] J. Liu, R. Yang, S. Li, "Synthesis and Photocatalytic Activity of TiO_2/V_2O_5 Composite Catalyst Doped with Rare Earth Ions," Journal of rare Earths, **25**(2), 173–178, 2007, doi: 10.1016/S1002-0721(07)60068-4.

[34] A. R. Khataee, M. B. Kasiri, "Photocatalytic degradation of organic dyes in the presence of nanostructured titanium dioxide: influence of the chemical structure of dyes," Journal of Molecular Catalysis A: Chemical, **328**(1-2), 8–26, 2010, doi:10.1016/j.molcata.2010.05.023.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 175-184 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Applications of the Heuristic Optimization Approach for Determining a Maximum Flow Problem Based on the Graphs' Theory

Simona Kirilova Filipova-Petrakieva*

Department Theory of Electrical Engineering, Faculty of Automation, Technical University of Sofia, Sofia, 1797, Bulgaria

ARTICLE INFO

Article history:
Received: 17 July, 2020
Accepted: 09 October, 2020
Online: 08 November, 2020

Keywords:
Maximum flow optimization
Graphs' theory
Power systems
Information systems
Communication systems
Transport systems
Manufacturing systems

ABSTRACT

In the present paper, a universal approach for determining the maximum flow problem in directed graph for solving different problems is applied. It can be considered as a heuristic and it can be used as a denomination for analyzing an arbitrary system with a mathematical description as a directed graph. The physical nature of the flows passing through the arcs of the system considered can be energy, information, transportation, or material. Then the studied system will be power, communication, transport, or manufacturing, respectively. In the present article five types of maximum flow problems are considered. Each of them is solved analytically by the universal heuristic method. The common between these problems is expressed in the fact that its behavior can be presented with the same mathematical model in the form of a directed graph including one initial (pending) vertex and one final (blocked) vertex, respectively. These vertexes are called either real (if they exist) or fictive (if they are introduced additionally) source and receiver depending on the topology of the associated directed graph's model. The final solution obtained with this approach for Problem 1 is compared with the similar one found through Ford-Fulkerson's, Edmonds-Karp's and Dinic's algorithms and it has been shown to be better than those determined by them.

1. Introduction

The maximum flows problem in industrial systems (energy, information or material) can be modeled by graphs' theory which adequately describe the topology of the studied system [1].

Note: For simplicity of the present analysis, it will be used only the term flow, regardless of type of the studied system - power [2, 3], information [4], communication [5, 6] or transportation [7].

Energy flows are optimized in power systems. In this sense, in [8] a review of the stochastic optimization methods, solving the optimal power flow problem. An application of the graphs' theory for the energy flows optimization in electric networks is made in [9]. The transformation of the linear model (e. g. linear goal function and linear constraints as unequations) to the directed graph's model is made there. Thus, the optimization procedure is expressed analyzing the resulting directed graph.

The definition of the flows problem in information systems is made in [10]. In [11] a review of approaches modelling information flow for organizations is presented. A new graph theoretical concept for analyzing the information flows passing through the directed graph's model, using guessing numbers is proposed in [12].

In [13], the problem of optimization the information flows passing through the uncertain networks with a given constrained capacities of the arcs is considered. The solution bases on the two NP-hard subproblems. First, it calculates the expected information flow associated to a source vertex and next it makes an optimal choice of arcs.

The flows optimization passing through the manifacturing systems is widely discussed issue. One possible solution to this problem is given by the graph's theory. For example, in [14] using the shortest path method for flows optimization in cellular manufacturing systems. In [15] an optimization for single-part flow-line configurations of reconfigurable manufacturing systems (RMS) is suggested. In [16] the other optimization approach of the flows in RMS systems is proposed. This approach also uses the graphs' theory for modeling the considered system. In [17] third model for optimizing demand period cost (fixed plus operating) of RMS flow-line configurations, based also on graphs' theory, is suggested.

A wide range of problems arise in real-world networks such as Internet, local information networks, energy systems, etc., which

*Corresponding Author: Simona Kirilova Filipova-Petrakieva, petrakievas-te@tu-sofia.bg

are modeled by directed graphs. In [18] a review of a range of theoretical primitives for generating a simple graph based on the given degree sequence is proposed. This is based on the graphs' theory and the computer networks. Several problems for which efficient algorithms for determining the theoretical solutions are proposed. Since in practice the graphs' models consists of tens of thousands of vertexes, approximation algorithms are particularly important.

The present paper is organized as follows. In a Section 2 the directed graph's model with one source – one recipient is defined, as well as the three necessary and sufficient conditions for performing its topological analysis. Five basic tasks arising in real systems, modeled with directed graphs, are also defined. In Section 3 an universal algorithm for solving all these five problems is briefly described. Solutions of these five type of problems applying the universal algorithm is proposed in Section 4. The applicability of this algorithm is visualized in Section 5 on an illustrative example describing problem 1. In the same section, a comparison is made with the obtained solutions to the same illustrative example using traditional methods based on minimum-cut the theorem, such as: Ford Fulkerson, Edmonds Carp and Dinic algorithms. The paper finishes with the conclusion remarks on the universality and advantages of the suggested algorithm.

The novelty of the proposed method consists in its universal application in determining the maximum flow passed through a arbitrary directed graph, whose model contains one initial and one final nodes. It is applicable to analysis of different systems - information, power, manufacturing, etc. The algorithm can be used for solving the five types of problems, defined in the next section

2. Problem statement

The maximum flow problem passing through the arcs of an directed graph is defined in [1].

In order to refine the considered graph's model, some basic concepts and initial hypotheses will be briefly introduced.

Let's consider a directed graph $G = < V, L >$, with $V = \{V_s\}$, $s = \overline{1,n}$ vertexes and $L = \{L_{ij}\} = \{L^k\}$, $k = \overline{1,m}$ arcs. The latter is associated to the system's channels. Let's $f_{ij} = f(V_i \to V_j) = f(V_i, V_j)$ is the flow passing through the arc L_{ij}. The capacity of this arc is q_{ij} [4].

A necessary and sufficient condition for determining the maximum flows passing through the arcs of the studied directed graph satisfies the following three assumptions.

Assumption 1

Balance condition: The sum of the incoming flows in a vertex V_i is equal to the respective sum of the outgoing ones, i.e.

$$\begin{aligned} &\sum_{V_j} f(V_i, V_j) - \sum_{V_s} f(V_s, V_i) = 0 \\ &V_j \in M^-(V_i);\ V_s \in M^+(V_i) \end{aligned} \quad . \tag{1a}$$

where: V_i is intermediate vertex for the considered directed graph;

$M^+(V_i)$ and $M^-(V_i)$ are the sets of input and output vertexes for the vertex V_i, respectively.

Assumption 2

Sequencing condition: The flow f_{ij} passing through the arc L_{ij} must not exceed its capacity q_{ij}., i.e.

$$f(V_i, V_j) \le q_{ij};\ \forall\ L_{ij} \in L. \tag{1b}$$

Assumption 3

Invariability condition: In absolute value, the flow passing through the directed graph from the source vertex to the recepient vertex must be exactly equal to that in the opposite direction.

$$\begin{aligned} &\sum_{V_j} f(V_{src}, V_j) - \sum_{V_i} f(V_i, V_s) = f_\Sigma \\ &V_j \in M^+(V_{src});\ V_i \in M^-(V_{src}) \\ &\sum_{V_j} f(V_r, V_j) - \sum_{V_i} f(V_i, V_{rec}) = -f_\Sigma \\ &V_j \in M^+(V_{rec});\ V_i \in M^-(V_{rec}) \end{aligned} \quad , \tag{1c}$$

where f_Σ is the flow passing through the directed graph.

$$\begin{aligned} &f_\Sigma = \sum_{V_j} f(V_{src}, V_j) = \sum_{V_i} f(V_i, V_{rec}) = 0 \\ &V_j \in M^+(V_{src});\ V_i \in M^-(V_{rec}) \end{aligned} \quad . \tag{2}$$

The first condition can be violated if part of the flow is dissipated in a vertex of the graph. This can be observed in the energy transfer, when each vertex is a specific physical device that consumes electricity and at the same time can dissipate it in the form of heat.

The second condition cannot be violated, since real systems are considered where it is physically impossible for a flow to pass through an arc that is greater than its capacity.

In real energy systems there are transfer losses, conversion losses, etc. In this case moving in the both directions of the graph (from the source to the recipient and vice versa) these losses cannot be equal, but they are very close.

Note: Without a loss of generality it is assumed, that the directed graph's model of the system has 1 source and 1 recipient vertexes. In general case, it can have multiple sources (N_s) and multiple recipients (N_r). Then the graph is transformed to a directed graph of the first type by adding fictive source V_{src} and recipient V_{rec} vertexes, respectively (Figure 1).

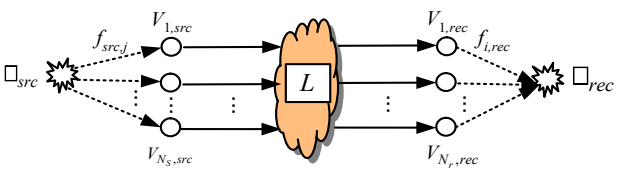


Figure 1: Graph's model with multiple sources (N_s) - multiple recipients (N_r)

In real systems depending on their parameters the following 5 types of maximum flow problems can be arised. In the first four of them the flow through the arcs f_{ij} is constant, and in the last one – it varies.

Problem 1: The arcs' capacities q_{ij} are known and the capacities of all graph's vertexes are infinity, i.e. $P_s \to \infty$.

Problem 2: The arcs' capacities q_{ij} are known and limited in a given range $q_{ij}^{\min} \le q_{ij} \le q_{ij}^{\max}$.

Problem 3: The arcs' capacities q_{ij} and the capacities P_s of the graph's vertexes $V_s \in V$ are known.

The Problems 1, 2 and 3 discuss the directed graph's model with one source and one recipient.

Problem 4: It consideres the directed graph with multiple sources $\{V_{src,l}\}$, $l = \overline{1, N_s}$ and multiple recipients $\{V_{l,rec}\}$, $l = \overline{1, N_r}$ (Figure 1). The arcs' capacities q_{ij} are known.

Problem 5: The arc's capacities q_{ij} of all the graph's arcs $L_{ij} \in L$ are known, but the balance condition (1a) is violated.

$$f_{in}\left(V_i, V_j\right) > f_{out}\left(V_i, V_j\right) \tag{3a}$$

or

$$f_{in}\left(V_i, V_j\right) < f_{out}\left(V_i, V_j\right), \tag{3b}$$

where $f_{in}\left(V_i, V_j\right)$ and $f_{out}\left(V_i, V_j\right)$ are the input and the output flows, associated to the arc $L_{ij} = \left\langle V_i, \quad V_j \right\rangle$, respectively.

For each of the above defined problem it obtains the maximum flow passing through the directed graph from the source to the recipient vertexes.

The basic aim and the contribution of the author of this paper consists of the suggesting the universal general algorithm for obtaining the maximum flow started from the source vertex V_{src} and reached to the final graph's vertex – either real V_{rec} or fictive $\mathbf{V}_{rec}$.

There are always losses when transmitting information or energy through the graph's model from the source to the recipient. They decrease the flow $f_{in}\left(V_i, V_j\right)$, i.e. the condition (3a). The condition (3b) is satisfied in some types of manufacturing systems.

For all 5 maximum flow problems defined above, *the condition* (1b).

The input flow for the system is generated by the source vertex V_{src} and it is named as $f_{src} = f(V_{src})$. The output flow $f_{rec} = f(V_{rec})$ is the flow, consumed by the recipient V_{rec}. It s assumed that the flow passing through the intermediate graph's vertexes is zero.

The different types of flows are defined in [1]. For the future analysis, we will describe them again in briefly.

Definition 1: The **allowed** flow is:

$$f^{allowed} = \sum_{M(V_i)} f_{out}\left(V_i, \quad V_j\right) - \sum_{M(V_j)} f_{in}\left(V_i, \quad V_j\right) = \begin{cases} f(V_{src}), & V_i = V_{src} \\ -f(V_{src}), & V_i = V_{rec} \\ 0, & \forall V \ne V_{src}, V_{rec} \end{cases}. \tag{4}$$

Definition 2: The **maximum flow** is:

$$f^{\max} = \max_{i=1, n_{\max}} \left[f_i^{allowed}\left(V_{src}, \quad V_{rec}\right) \right] \tag{5}$$

where $f_i^{allowed} = \left(V_{src}, \quad V_{rec}\right)$ are all allowed flows with number i passing through the directed graph from the source to the recipient vertexes.

Definition 3: The **conventionally optimal flow** $f^{opt} = \left(f^{opt}\left(V_{src}\right), \quad f^{opt}\left(V_{rec}\right)\right)$ is, if arbitrary allowed flow $f_i^{allowed} = \left(f_{src}, \quad f_{rec}\right)$ satisfies at least one of the conditions (6)

$$\begin{aligned} &f^{+}\left(V_{src}\right) \ge f^{-}\left(V_{src}\right), \ if \ \ f^{+}\left(V_{rec}\right) \ge f^{-}\left(V_{rec}\right) \\ &f^{-}\left(V_{rec}\right) \ge f^{+}\left(V_{rec}\right), \ if \ \ f^{+}\left(V_{src}\right) \ge f^{-}\left(V_{src}\right) \end{aligned}, \tag{6}$$

where: $f^{+}(V_{src})$ - the flow, generated from the vertex V_{src} passing through the graph's model from V_{src} to V_{rec};

$f^{-}(V_{src})$ - the flow, consumed of the vertex V_{src} passing through the graph's model from V_{rec} to V_{src};

$f^{+}(V_{rec})$ - the flow, consumed of the vertex V_{rec} passing through the graph's model from V_{src} to V_{rec};

$f^{-}(V_{rec})$ - the flow, generated from the vertex V_{rec} passing through the graph's model from V_{rec} to V_{src}.

Definition 4: The **maximum allowed flow** is $f^{*} = \left(f^{*}\left(V_{src}\right), \quad f^{*}\left(V_{rec}\right)\right)$, if it satisfies all the conditions (4), (5) and (6) simultaneously.

3. Algorithm of the universal approach for determining the maximum flow problems

Each problem defined in the previous paragraph can be reduced to the optimization maximum flow problem, where the the goal function is:

$$f_{\Sigma}^{\max} = \max \sum_{V_j} f\left(V_k, V_j\right) = \max \sum_{V_i} f\left(V_i, V_r\right) \tag{7}$$

and the constraints:

$$\begin{cases} \sum_{V_j} f(V_i, V_j) - \sum_{V_s} f(V_s, V_i) = 0 \ \ if \ \ V_i \neq V_{src}, \ V_i \neq V_{rec} \\ \sum_{V_j} f(V_i, V_j) - \sum_{V_s} f(V_s, V_i) = f_\Sigma; \ V_i = V_{src} \\ \sum_{V_j} f(V_i, V_j) - \sum_{V_s} f(V_s, V_i) = -f_\Sigma; \ V_i = V_{rec} \end{cases}, \quad (8)$$

where $f_\Sigma^{\max}$ is the maximum flow passing through the minimum intersection $L_{\min}^{C}$ of the considered directed graph.

The basic directed graph, which is divided into two separate ones G^A and G^B this function is:

$$f_\Sigma^{\max} = f_\Sigma^{\max}(L_{\min}^{C}) = \sum_{L_{\min}^{C}} q_{ij}, \quad (9)$$

where: $L_{\min}^{C} = \left[G^A \to G^B \right]$ $\quad V_{src} \in G^A$; $\quad V_{rec} \in G^B$;
$G^A \cap G^B = \varnothing$; $G^A \cup G^B = G$.

The approach, solving the optimization problem (7) with constraints (8) is proposed and described in details in [19]. Each iteration consists of two procedures – vertexes marking and flows correction. The iteration process stops when the minimal intersection $L_{\min}^{C}$ in the directed graph's model is determinated.

Before starting the algorithm, it is necessary to know the structure of the directed graph modeling the considered system as well as the capacities q_{ij} of each graph's arc. It is assumed that the initial values of the flows f_{ij} passing through the graph's arcs $L_{ij} \in L$ are zero.

In **Procedure 1** a vertexes marking is made. Each node is assigned the appropriate pair of marks, moving from source to the recipient vertexes. This marking indicates where the respective flow comes from and its maximum allowable value that can be passed through the respective arc. When all vertexes are marked then the algorithm goes to the **Procedure 2**. This procedure moves from the recipient to the source vertexes correcting the flows passing through each of the arcs in the respective sequences of arcs in the considered directed graph. If all the arcs entering the vertex V_{rec} are saturated, then the maximum flow $f_\Sigma^{\max}$ is determined and the procedure stops. Otherwise it returns to the beginning of Procedure 1.

Note: An arc is saturated if the flow that passes through it f_{ij} is exactly equal to its capacity q_{ij} .

4. Solutions to the defined 5 types of the problems

The maximum flow determination approach, proposed in [19], can be applied to solve each of the typical problems, defined in Section 2.

Problem 1

Before starting the algorithm the directed graph's model is determined. In this model the values of the arcs capacities $\{q_{ij}\}$ and vertexes infinity capacities $P_s \to \infty, \ s = \overline{1,n}$ are known.

To simplify the future exposition and without loss of the generality the maximum flow $(f_\Sigma^{\max})$ is obtained for the intersection of the directed graph's model shown in Figure 2. For simplicity the elements (vertexes and arcs) included in the beginning of the graph are aggregated in the set M^- .

In this problem the condition (1b) is satisfied, i.e. the flow f_{ij} passing through the arc L_{ij} , have to be smaller or equal its capacity q_{ij} .

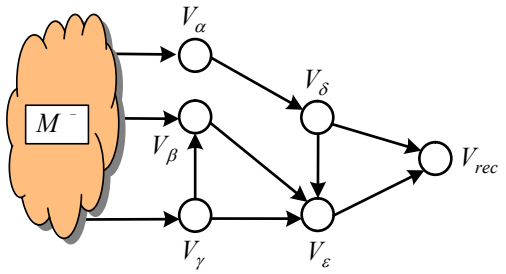


Figure 2: Idealized system described by the graph's model

Based on the minimum intersection $L_{\min}^{C}$ (see Figure 2), which satisfy the condition (4), the respective maximum flows are:

$$\begin{cases} f_{\Sigma_A}^{\max} = q_{\beta\varepsilon} + q_{\gamma\varepsilon} + q_{\delta\varepsilon} \\ f_{\Sigma_B}^{\max} = q_{\delta,rec} \end{cases}. \quad (10)$$

Conclusions:

- The maximum allowed flow $f_{\Sigma_A}^{\max}$ is equal to the sum of the capacities of the arcs $L_{\beta\varepsilon}$, $L_{\gamma\varepsilon}$, $L_{\delta\varepsilon}$, because they beyond to the minimum intersection $L_{\min}^{C}$ and they are saturated (their capacity is exhausted).
- The maximum allowed flow $f_{\Sigma_B}^{\max}$ for the arc $L_{\alpha\delta}$ is equal of the capacity of the arc $L_{\delta\varepsilon}$, because it is assumed that the condition $q_{\delta\varepsilon} < q_{\alpha\delta} + q_{\delta\varepsilon}$ is satisfied. Therefore, the arc $L_{\delta\varepsilon}$ is saturated and it is added to the minimum intersection $L_{\min}^{C}$.
- The final maximum flow, determined by Eq. (4) and Eq. (8) is $f_\Sigma^{\max} = f_\Sigma^{\max}(L_{\min}^{C}) = f_{\Sigma_A}^{\max} + f_{\Sigma_B}^{\max}$.

Problem 2

Before starting the algorithm the directed graph's model is determined. In this model the values of the arcs capacities $\{q_{ij}\}$ are known and each of them must belong to an initially defined range, i.e. $\leq q_{ij} \in \left[q_{ij}^{\min}, \ q_{ij}^{\max} \right]$.

The solution of the Problem 2 requires a transformation the basic directed graph's model to the new one, where:

- Each arc is replaced by two new arcs - L_{rg} and L_{gs}.
- The capacity of each of these new arcs determines as follows:

- about the arc $L_{rg} \to q_{rg} = q_{rs}^{\max} - q_{rs}^{\min}$;
- about the arc $L_{gs} \to q_{gs} = q_{rs}^{\min}$.
- It introduces the fictive arc L_{fict} with infinity capacity, i.e. $q_{fict} \to \infty$.

Then the Problem 2 has a solution only when the condition $f_{\Sigma}^{\max} = \sum_{q_{ij}^{\min} \neq 0} q_{ij}^{\min}$ is satisfied.

The transformed graph's model is the same as in Problem 1. Hence, the maximum flow $f_{\Sigma}^{\max}$ is determined in the same way.

Problem 3

Before starting the algorithm, the directed graph's model is determined. In this model the values of the arcs capacities $\{q_{ij}\}$ and vertexes capacities P_s, $s = \overline{1,n}$ are known.

The specificity of this problem requires that the additional condition (8) to be included:

$$f[L^{\oplus}(V_i)] \le P_s; \ \forall V_s \in V, \tag{11}$$

where: $f[L^{\oplus}(V_i)]$ is the flow entering to the vertex $V_s \in V$;

$L^{\oplus}(V_i)$ is the set of arcs, which are directly connected with the vertex $V_s \in V$ only by one arc.

To get the solution of this problem it is necessary to transform the basic directed graph's model to the new one with $(n+m)$ vertexes where *n* and *m* are the number of vertexes and arcs in the initial directed graph, respectively.

- Each vertex $V_s \in V$ is replaced by two vertexes V_{s_1} and V_{s_s}, which are connected by the arc $L_{s_1 \to s_2} = < V_{s_1}, V_{s_2} >$;
- The capacity of the arc $L_{s_1 \to s_2} = < V_{s_1}, V_{s_2} >$ is obtained as follows:

$$q_{s_1 \to s_2} = P_s; \ \forall V_s \in V. \tag{12}$$

The transformed graph's model is the same as in Problem 1. Hence, the maximum flow $f_{\Sigma}^{\max}$ is determined in the same way.

Problem 4

The initial conditions for this problem are the same as these in Problem 3 with the following additions. Fictive source $\mathbf{V}_{src} = \{V_{l,src}\}$, $l = \overline{1, N_s}$ and receiver $\mathbf{V}_{rec} = \{V_{l,rec}\}$, $l = \overline{1, N_r}$ are introduced, where each of them is associated with more than one real one (Figure 1). It is assumed that the graph's vertexes have infinite capacities, i.e. $q_{ij} \to \infty$.

The graph's model from the Problem 4 has $(n + N_s + N_r)$ vertexes and $(m+2)$ arcs compared to the models in Problems 1, 2 and 3, where *n* and *m* are the number of vertexes and arcs in the initial directed graph from Problem 1, respectively.

The resulting graph's model is the same as in Problem 1. Hence, the maximum flow studied $f_{\Sigma}^{\max}$ is determined in the same way.

Problem 5

Before starting the algorithm the directed graph's model is determined. In this model the values of the arcs capacities $\{q_{ij}\}$ and the values of the coefficients $\{\Psi_{ij}\}$, correcting the flow, passed through all of the graph's arcs are known. The latter are always positive and they reflect to the increasing ($\delta_{ij} > 0$), resp. decreasing ($\delta_{ij} > 0$) of the flow in the arc respective L_{ij}.

The next three steps describe the application of the considered heuristic algorithm [19] for solving Problems 1, 2 and 3, resp. 4 and 5, which are reduced to Problem 1.

1. Flows' ccorrection passing through the graph's arcs

The flows' correction is performed in the following way:

$$\Psi_{ij} . f_{in}(V_i, \ V_j) = f_{out}(V_i, \ V_j). \tag{13}$$

Note: If $\forall \ \Psi_{ij} = 1$ then the condition (1a) is satisfied and the directed graph's model is associated to the Problems 1, 2 or 3.

2. Determining the graph's paths from the source vertex V_{src} to the recipient vertex V_{rec}, where the resulting flow increases

If the conditions

$$\begin{aligned} &f_{in}(V_\alpha, \ V_\omega) < q_{\alpha\omega} \\ &\forall \text{"stright" arc } L_{\alpha\omega} = \langle V_\alpha, \ V_\omega \rangle \end{aligned}; \tag{14a}$$

and

$$\begin{aligned} &f_{in}(V_\alpha, \ V_\omega) > q_{\alpha\omega}, \\ &\forall \text{"opposite" arc } L_{\omega\alpha} = \langle V_\omega, \ V_\alpha \rangle \end{aligned}. \tag{14b}$$

are satisfied, then there is a path from the vertex $V_\alpha \in V$ to the vertex $V_\omega \in V$ in the analyzed directed graph, where the resulting flow increases.

The example of this path including two arcs with straight direction and two ones with opposite direction and their capacities q_{ij} and correcting coefficients Ψ_{ij}, is shown in Figure 3.

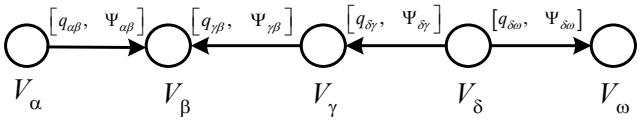


Figure 3: Path with initial vertex V_α and final vertex V_ω, including two straight and two opposite arcs

3. Obtaining the active cycles in directed graph's model

In directed graph's model $G = < V, L >$ there is an active cycle $C(G)$, connected by the recipient vertex V_{rec}, if the following conditions are satisfied:

- The value of the summary correcting cycle's coefficient is bigger than 1, i.e. $\Psi\left[C(G)\right] > 1$.
- The cycle's capacity is not zero, i.e. $q\left[C(G)\right] \neq 0$.
- The cycle includes a vertex V_λ, which is directly connected to the recipient vertex V_{rec}.

The example of this type cycle is shown in Figure 4.

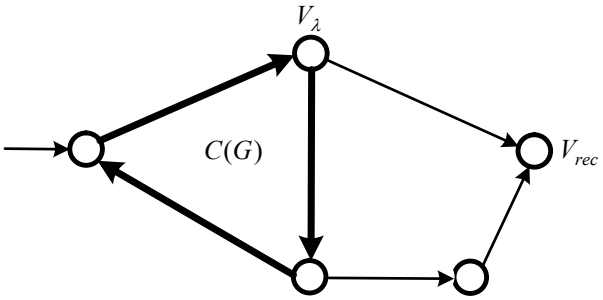


Figure 4: A cycle in directed graph's model

Note: It is allowed in directed graph's model to have cycles, connected to the source vertex V_{src}.

5. Application of the maximum flow suggested approach to an illustrative example about Problem 1

The algorithm for finding the maximum flow [19] will be illustrated for solving the Problem 1. Let's the following example of the directed graph's model is analyzed (Figure 5a). On the directed graph $G = < V, L >$ the following quantities are indicated:

1 The capacities q_{ij} of the respective arc $L_{ij} = \langle V_i, \ V_j \rangle \in L$ are shown below it.

2 The values of the flows $f_{ij} = f(V_i, V_j)$ passing through the arc L_{ij} are underlined and are written above to the arc. It is assumed that the initial values of these flows for all arcs are zeroes.

3 The vertexes marks $\left[M_1(V_i), \ M_2(V_i)\right]$ are noted in the following way:

- $M_1(V_i) = \pm V_j$, with respect to an increasing (+) or decreasing (–) flow passing through the arc L_{ij};
- $M_2(V_i) = f_{add}(V_i)$ - the maximum value of the additional flow $f_{add}(V_i)$ from the source vertex V_{src} to the vertex V_i passing through the sequence of the arcs, where the flow increases. This flow is determined as follows:

$$\begin{aligned} & f_{add}(V_i) = \min\left[f_{add}(V_j), \ q_{ij} - f\left(V_j, \ V_i\right) \right], \\ & \text{if } \ f\left(V_j, \ V_i\right) < q_{ij} \ \wedge \ V_i \in M^{+}(V_j) \\ & f_{add}(V_i) = \min\left[f_{add}(V_j), \ f\left(V_i, \ V_j\right) \right], \\ & \text{if } \ f\left(V_j, \ V_i\right) > 0 \ \wedge \ V_i \in M^{-}(V_j) \end{aligned} \qquad (15)$$

It is assumed that the initial marks of all vertexes are zeroes. And the source vertex V_{src} gets the mark $M(V_{src}) = \left[V_{src}, \ \infty\right]$, i.e. the generated flow from the source vertex is infinite.

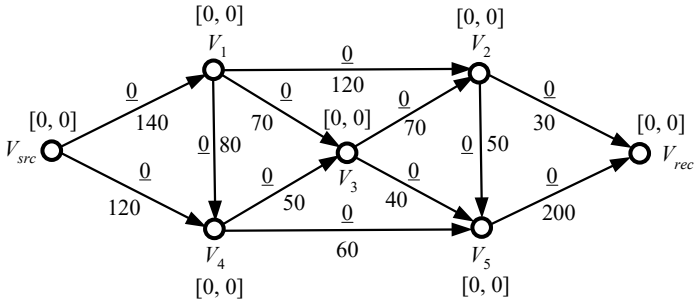


Figure 5a: Illustrative example for solving a maximum flow problem

When applying the heuristic algorithm [19], each iteration consists of 2 steps: first the vertices in the considered directed graph are marked and checked, and second the flows passed through the arcs in the graph's model are corrected. When marking the vertices, the algorithm moves from the beginning to the end of the graph, and when correcting the flows – vice versa, taking into account the capacity of each arc.

The required maximum flow is determined when all incoming arcs in the end graph's vertex are saturated. For the present example, this happens to the third iteration. The iterations are following.

Iteration 1

Vertexes marking (Figure 5b)

The source vertex V_{src} is marked with:

$$M\left(V_{src}\right) = \left[+V_{src}, \ \infty\right].$$

The vertexes V_1 and V_4, that are directly connected with V_{src} get the following marks:

$$\left| \begin{array}{l} M\left(V_1\right) = \left[+V_{src}, \ \ 140\right], \ M_2 = \min\left[\infty, \ 140 - 0\right] = 140 \\ M\left(V_4\right) = \left[+V_{src}, \ \ 120\right], \ M_2 = \min\left[\infty, \ 120 - 0\right] = 120 \end{array} \right. .$$

Therefore, the vertex V_{src} is marked and checked.

Let's move to the next vertex - V_1, which is marked but not checked. The latter is connected to the vertexes V_2 and V_3. So their marks are determined as follows:

$$\left| \begin{array}{l} M\left(V_2\right) = \left[+V_1, \ \ 120\right], \ M_2 = \min\left[140, \ 120 - 0\right] = 120 \\ M\left(V_3\right) = \max\left\{ \begin{array}{l} \left[+V_1, \ \ 70\right], \ M_2 = \min\left[140, \ 70 - 0\right] = 70 \\ \left[+V_4, \ \ 50\right], \ M_2 = \min\left[120, \ 50 - 0\right] = 50 \end{array} \right\} = \left[+V_1, \ \ 70\right] \end{array} \right.$$

Therefore, the vertex V_1 is marked and checked.

The marks of the other vertexes are determined by analogy as follows:

$$\left| \begin{array}{l} M\left(V_5\right) = \max\left\{ \begin{array}{l} \left[+V_2, \ \ 50\right], \ M_2 = \min\left[120, \ 50 - 0\right] = 50 \\ \left[+V_3, \ \ 40\right], \ M_2 = \min\left[70, \ 40 - 0\right] = 40 \\ \left[+V_4, \ \ 60\right], \ M_2 = \min\left[120, \ 60 - 0\right] = 60 \end{array} \right\} = \left[+V_4, \ \ 60\right] \\ M\left(V_{rec}\right) = \left[+V_2, \ \ 120\right], \ M_2 = \min\left[120, \ 30 - 0\right] = 30 \end{array} \right.$$

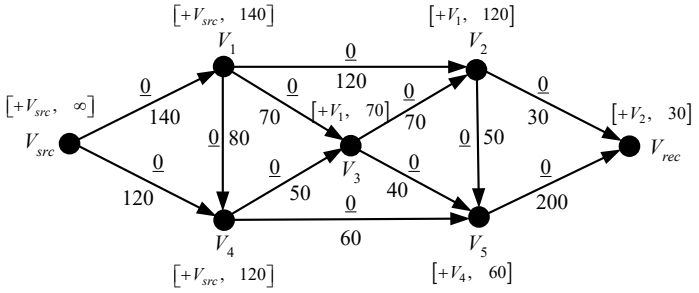


Figure 5b: Results from vertexes marking in an Iteration 1

So, all vertexes in the directed graph are marked and checked.

Flows correction (Figure 5c)

The vertex V_{rec} is chozen as a start vertex and we move from the end V_{rec} to the beginning V_{src} of the directed graph. Using Eq. (11) it determines the flow

$$f_{add}\left(V_{rec}\right) = M_2\left(V_{rec}\right) = 30\,.$$

Thus, the consequence of the arcs

$$\langle V_{rec},\ \ V_2\rangle \rightarrow \langle V_2,\ \ V_1\rangle \rightarrow \langle V_1,\ \ V_{src}\rangle$$

is formed basing on the first mark coefficients $M_1\left(V_i\right)$ of the respective vertexes. Then, using Eq. (12) the corrected values of the flows only in this arcs are calculated as follows:

$$f(V_2,V_{rec}) = f(V_1,V_2) = f(V_{src},V_1) = 0 + f_{add}\left(V_{rec}\right) = 30\,.$$

Then the condition

$$f(V_2,V_{rec}) = q_{2rec} = 30$$

is satisfied for the arc $\langle V_2,\ \ V_{rec}\rangle$. Then the arc is saturated and it is marked with bold.

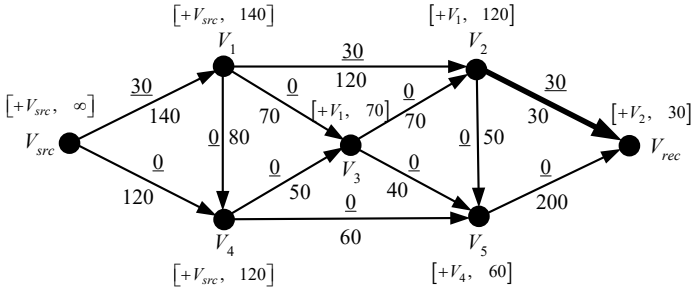


Figure 5c: Results from the flows corrections in the end of an Iteration 1

Since the arcs $\langle V_5,\ \ V_{rec}\rangle, \langle V_2,\ \ V_5\rangle, \langle V_3,\ \ V_5\rangle$ and $\langle V_4,\ \ V_5\rangle$ are not saturated the algorithm passes to next iteration – Iteration 2.

Moving from V_{rec} to the V_{src} as a result of this iteration a flow equal to 30 units is passed through each of the arcs $\langle V_{src},\ \ V_1\rangle, \langle V_1,\ \ V_2\rangle$ and $\langle V_2,\ \ V_{rec}\rangle$.

Iteration 2

Vertexes marking (Figure 5d)

Based on the information from Figure 5c, the new mark of the vertexes is set as follows:

$$\begin{aligned} M\left(V_{src}\right) &= \left[+V_{src},\ \ \infty\right] & M\left(V_2\right) &= \left[+V_1,\ \ 90\right] \\ M\left(V_1\right) &= \left[+V_{src},\ \ 110\right] & M\left(V_3\right) &= \left[+V_1,\ \ 70\right] \\ M\left(V_4\right) &= \left[+V_{src},\ \ 120\right] & M\left(V_5\right) &= \left[+V_4,\ \ 60\right] \\ & & M\left(V_{rec}\right) &= \left[+V_5,\ \ 60\right] \end{aligned}\,.$$

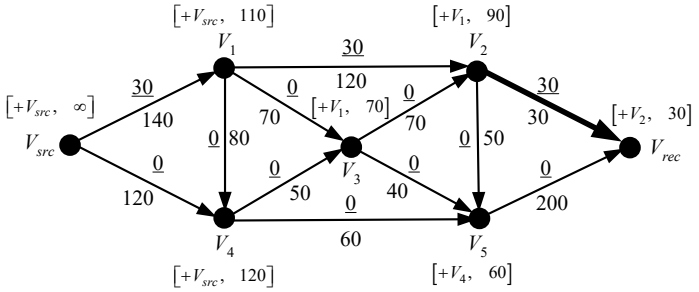


Figure 5d: Results from vertexes marking in an Iteration 2

Flows correction (Figure 5e)

Using Eq. (11) for the arc $\langle V_4,\ \ V_5\rangle$ the additional flow, is determined:

$$f_{add}\left(V_5\right) = M_2\left(V_5\right) = 60\,.$$

Thus, the consequence of the arcs

$$\langle V_{src},\ \ V_4\rangle \rightarrow \langle V_4,\ \ V_5\rangle$$

and

$$\langle V_{src},\ \ V_1\rangle \rightarrow \langle V_1,\ \ V_4\rangle \rightarrow \langle V_4,\ \ V_5\rangle$$

is formed basing on the first mark coefficients $M_1\left(V_i\right)$ of the respective vertexes. Using Eq. (12) the corrected values of the flows only in these arcs are corrected as follows:

$$\begin{aligned} f(V_{src},V_4) &= f(V_4,V_5) = 0 + f_{add}\left(V_5\right) = 60 \\ f(V_1,V_4) &= f(V_4,V_5) = 0 + f_{add}\left(V_5\right) = 60 \\ f(V_{src},V_1) &= 30 + f_{add}\left(V_5\right) = 30 + 60 = 90 \end{aligned}\,.$$

The arc $\langle V_4,\ \ V_5\rangle$ is saturated because the condition

$$f(V_4,V_5) = q_{45} = 60$$

is satisfied. Then the arc is marked bold.

Based on the additional flow $f_{add}\left(V_{rec}\right) = 60$, the corrected flows are:

$$f(V_5,V_{rec}) = f(V_4,V_5) = 0 + f_{add}\left(V_{rec}\right) = 60\,.$$

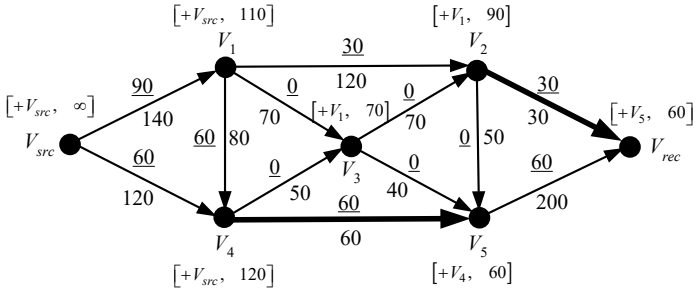


Figure 5e: Results from the flows corrections in the end of the Iteration 2

Since the arcs $\langle V_2, \ V_5 \rangle$ and $\langle V_3, \ V_5 \rangle$ are not saturated the algorithm passes to next iteration – Iteration 3.

Moving from V_{rec} to the V_{src} as a result of this iteration a flow equal to 60 units goes through each of the arcs $\langle V_{src}, \ V_4 \rangle, \langle V_4, \ V_5 \rangle$ and $\langle V_5, \ V_{rec} \rangle$. This additional flow of 60 units is then added to the current flows through the arcs $\langle V_1, \ V_4 \rangle$ and $\langle V_{src}, \ V_1 \rangle$, and finally, the flows of 60 and 90 units pass through the arcs, respectively.

Iteration 3

Vertexes marking (Figure 5f)

Based on the information from Figure 5e, the new mark of the vertex V_4 is set as follows (Figure 5f):

$$M(V_4) = [+V_{src}, \ \ 60].$$

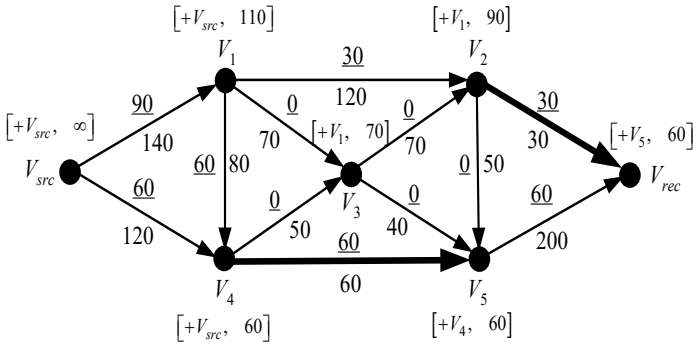


Figure 5f: Results from vertexes marking in an Iteration 3

3.2. Flows' correction (Figure 5g)

Using Eq. (11) the arcs $\langle V_2, \ V_5 \rangle$ and $\langle V_3, \ V_5 \rangle$ the following additional flows, are determined:

$$f_{add}^{V_2}(V_5) = \min\left[q_{25} = 50, \ M_2(V_5)\right] = \min[50, \ 60] = 50$$

and

$$f_{add}^{V_3}(V_5) = \min\left[q_{35} = 50, \ M_2(V_5)\right] = \min[40, \ 60] = 40 .$$

Thus, the other corrected flows are:

$$f(V_2, V_5) = f_{add}^{V_2}(V_5) = 50.$$
$$f(V_3, V_5) = f_{add}^{V_3}(V_5) = 40.$$

here

$$f(V_2, V_5) = 50, \ \text{and} \ \ q_{25} = 50 \Rightarrow \langle V_2, \ V_5 \rangle \ \text{is saturated.}$$
$$f(V_3, V_5) = 40, \ \text{and} \ \ q_{35} = 40 \Rightarrow \langle V_3, \ V_5 \rangle \ \text{is saturated.}$$

Therefore, both arcs $\langle V_2, \ V_5 \rangle$ and $\langle V_3, \ V_5 \rangle$ are marked bold.

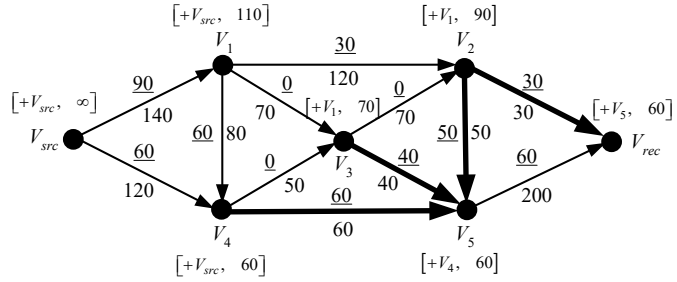


Figure 5g: Results from the flows corrections in the Iteration 3

The solution of the maximum flow problem is shown in Figure 5h, where the final vertexes are marked and flows corrections are made:

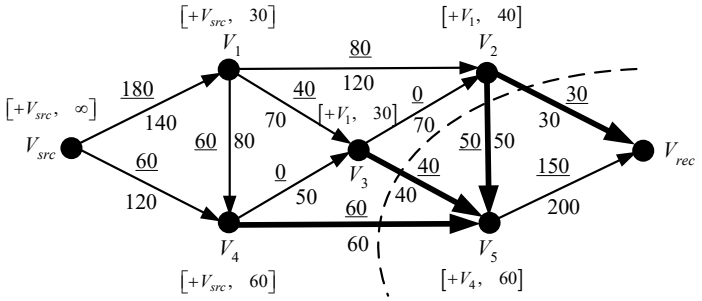


Figure 5h: Final maximum flow

Moving from V_{rec} to the V_{src}, as a result of this iteration, the following will happen. First, a flow equal to 40 units is passed through each of the arcs $\langle V_{src}, \ V_1 \rangle, \langle V_1, \ V_3 \rangle$ and $\langle V_3, \ V_5 \rangle$. This additional flow of 40 units is then added to the current flows through the arcs $\langle V_{src}, \ V_1 \rangle, \langle V_1, \ V_3 \rangle$ and $\langle V_3, \ V_5 \rangle$, and finally flows of 130, 40 and 40 units will pass through the arcs, respectively. Second, a flow equal to 50 units is passed through each of the arcs $\langle V_{src}, \ V_1 \rangle, \langle V_1, \ V_2 \rangle$ and $\langle V_2, \ V_5 \rangle$. This additional flow of 30 units is then added to the current flows through the arcs $\langle V_{src}, \ V_1 \rangle, \langle V_1, \ V_2 \rangle$ and $\langle V_2, \ V_5 \rangle$, and finally flows of 180, 50 and 50 units will pass through the arcs. Thus, the resulting flow through the arc will be 180 units.

After the last iteration, it is seen that all arcs, where a flow passing through, are saturated. Therefore, an end solution is reached and the final maximum flow is:

$$f_{\Sigma}^{\max} = f(V_2, \ V_{rec}) + f(V_2, \ V_5) + f(V_3, \ V_5) + f(V_4, \ V_5) =$$
$$= 30 + 50 + 40 + 60 = 180$$

In the analyzed illustrative example, the final maximum flow has been found after the third iteration, because all arcs of the cross section shown in Fig. 5h passing flow to the end vertex V_{rec} are saturated, i.e. the flow passing through them f_{ij} is exactly equal to the capacity of the corresponding arc q_{ij}.

A comparative study between the discussed heuristic algorithm and a standard min-cut algorithm for solving the maximum flow problem

The obtained solution with a heuristic approach of [19] for an illustrative problem shown in Figure 5a is compared with

analogous ones, determined by traditional methods based on min-cut theorem [20, 21, 22]. The latter is the following: "For any directed weight graph, the maximum flow passed through from the source to the recipient vertex is equal to the capacity of the minimum cut.". Then it is necessary to obtain the set of arcs with minimum weight that after removing from the graph there is no path from the source to the recipient vertexes. The following methods were used: Ford Fulkerson [23], Edmonds Karp [24] and Dinic algorithms [25, 26]. The calculations are made with free software Visualgo. The final the full solution of the considered problem about the maximum flow is obtained.

Note: In contrast to the method applied in this article, in the final solution Visualgo visualizes the flow that passes through each of the arcs and their remaining capacity, depending on the full arc's capacity, as shown in Figures 5a-5h.

The final solution obtained through the used methods are shown in Figures 6a, 6b and 6c.

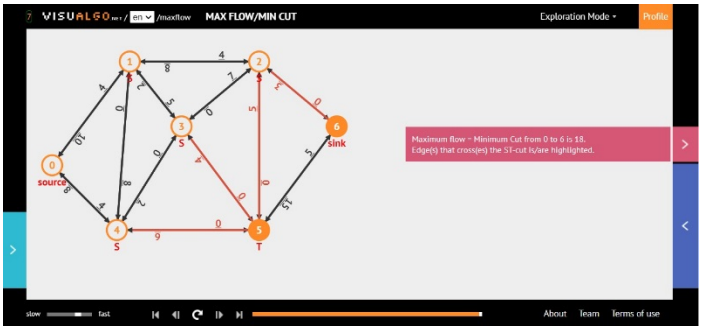


Figure 6a: Final maximum flow obtained by the Ford-Fulkerson's algorithm

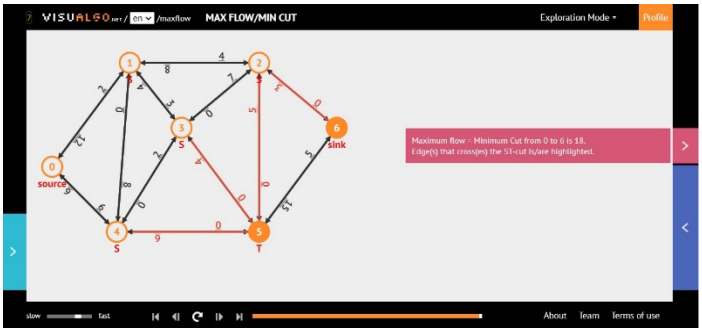


Figure 6b: Final maximum flow obtained by the Edmonds-Karp's algorithm

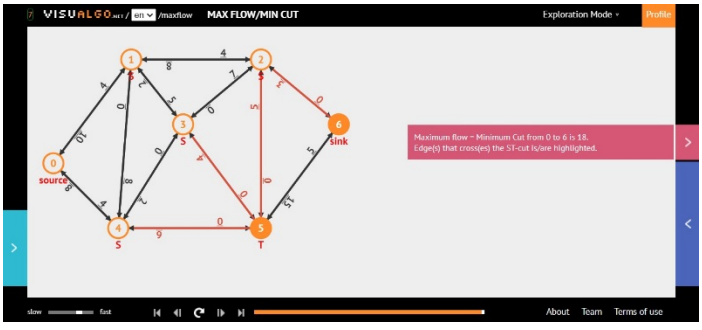


Figure 6c: Final maximum flow problem obtained by the Dinic's algorithm

The final maximum flow, determined by these three methods, based on min-cut theorem (Figures 6a, 6b and 6c), are the same as those of the heuristic approach applied in this article:

$$f_{\Sigma}^{\max} = f(V_2,\ V_{rec}) + f(V_2,\ V_5) + f(V_3,\ V_5) + f(V_4,\ V_5) = .$$
$$= 30 + 50 + 40 + 60 = 180$$

The advantage of the proposed approach is that the final solution is obtained on the 3rd iteration, and each iteration consists of two procedures.

In the method of Ford Fulkerson, the final solution is obtained on the 6th iteration. The Edmonds Karp algorithm consists of two built-in cycles. For the considered example the final solution is obtained on the 4th iteration. As the external cycle is executed 4 times and each time from its execution the internal is executed only once. The Dinic algorithm consists of two build-in cycles. For the considered example the final solution is determined after applying the external cycle 2 times. Initially, the internal cycle is rotated 2 times after that – it rotates 3 times. All this shows the better efficiency of the proposed algorithm

6. Conclusion

Determining the maximum flow and the distribution of the flows in the analyzed directed graph based on the interconnections in its topology is an important prerequisite for the efficient use of resources in the considered system. The solution of this optimization problem depends on the structure of the comsidered graph's model and the capacities of its arcs. The heuristic approach used in this article is universal not only in terms of the physical nature of the studied system, but also in terms of different types of problems considered - the input data of each of analyzed problem are different.

The solutions to the Problems 1, 2, 3 and 4 show that the flows passing through the arcs of the graph's model are constant. The solution to the Problem 5, however, shows that the flow is variable, so one can make a general conclusion when the new approach has been successfully applied in the case of violation of the balance condition.

In the present paper, the results of the solution for Problem 1 with heuristic algorithm are compared with these obtained from other analogous methods: Ford-Fulkerson's, Edmonds-Karp's and Dinic's algorithms. The time complexity of this algorithm is equal to that of the Ford-Fulkerson's and Edmonds-Karp's algorithms - $O\left[|V|.|E|^2\right]$, in contrast to that of the Dinic's algorithm, where it is $O\left[|V|^2.|E|\right]$. The illustrative example shows that it is faster to converge to the final solution comparing to the other considered algorithms.

The used universal algorithm is inapplicable to solving problems in which at least one of the three initial assumptions (defined in Problem statement) is not fulfilled, i.e. Balance condition, second condition for sequencing the flow passing through the arbitrary arc with its capacity and third condition for the invariability of the flow passing through the considered graph moving from the source to the recipient vertexes and vice versa.

One of the possible future development of the work is related to the program realization of the proposed approach for determining the maximum flow passing through a directed graph.

Other trend for the future development of the research on solving this problem is using an algorithm based on determining all possible paths in the considered directed graph, proposed in [27]. It uses two adjacent matrices corresponding to each vertex of the graph - these of the input and output arcs in the vertex, respectively. As a result of the algorithm, one will only be interested in the maximum flow that has passed the shortest path from the start to the end vertex in the directed graph.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgment

The publication of this manuscript was realized by financial support from the Scientific Research Sector of the Technical University of Sofia.

References

[1] H. A. Taha, Operations research: An introduction, Prentice Hall Price, 2017.

[2] J. A. Bondy, Murty, U. S. R., Graph Theory, Springer, 2008.

[3] N. Sayeekumar, Sh. Ahmed, S. Karthikeyan, S. Sahoo, I. Jacob, "Graph theory and its applications in power systems - a review", in 2015 IEEE International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 18-19 Dec. 2015, doi: 10.1109/ICCICCT.2015.7475267.

[4] J. Burch, Grudnitski, Information system. Theory and Practice, John Wiley & Sohn, 1997.

[5] S. L. Hakimi, "Application of graphs theory to problems in communication systems and networks", AD-A030 888. U. S. department of Commerce, National Technical Information Service, Northwestern Univ. Evenston III, Dept. of Electrical engineering, Air Force Office of Scientific Research, Boiling AFB, D C. 1976.

[6] G. Cherneva, H. Spiridonova, "Matching Filters Design for Information Transmit in Power Line Communication Systems", Mechanics, Transport, Communications Journal, **3/2**, XI-21- XI-27, 2018, (in Bulgarian).

[7] S. Guze, "Graph Theory Approach to Transportation Systems Design and Optimization", International Journal on Marine Navigation and Safety of Sea Transportation, **8**(3), Dec. 2014, doi: 10.12716/1001.08.04.12

[8] V. Radziukynas, I. Radziukyniene, Optimization Methods Application to Optimal Power Flow in Electric Power Systems, Optimization in the Energy Industry, Chapter 18, 409-436, Jan. 2009, doi: 10.1007/978-3-540-88965-6_18

[9] P. H. Nguyen, W. L. Kling, G. Georgiadis, M. Papatriantafilou, L. A. Tuan, L. Bertling, Application of the Graph Theory in Managing Power Flows in Future Electric Networks, New Frontiers in Graph Theory, Chapter 12, March 2012, doi: 10.5772/35527

[10] Schade, M. Begley. "Information flow diagram guidelines", 23 Sept. 2012, retrieved 20 March 2016.

[11] D. Christopher, T. Ashutosh, A. R. Jeffrey, "Modelling information flow for organisations: a review of approaches and future challenges", International Journal of Information Management, **33**(3), 597–610, June 2013, doi:10.1016/j.ijinfomgt.2013.01.009

[12] S. Riis, "Information flows, graphs and their guessing numbers", in 2006 IEEE 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 26 Feb. - 2 March 2006, doi: 10.1109/WIOPT.2006.1666516.

[13] Ch. Frey, A. Zufle, T. Emrich, M. Renz, "Efficient Information Flow Maximization in Probabilistic Graphs", IEEE Transactions on Knowledge & Data Engineering, **30**(5), 880-894, 2018, doi:10.1109/TKDE.2017.2780123

[14] K. A. Prasath, R. D. J. Johnson, "Fundamental Approach on Layout Optimization Using Graph Theory in Cellular Manufacturing Systems", Australian Journal of Basic and Applied Sciences, 9(33):180, October 2015.

[15] J. Dou, X.-Z. Dai, Z.-D. Meng, J. Li, "Optimization for single-part flow-line configurations of reconfigurable manufacturing system based on graph theory", Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems, CIMS, **16**(1), 81-89, 2010.

[16] J. Dou, X. Dai, Zh. Meng "Graph theory-based approach to optimize single-product flow-line configurations of RMS", International Journal of Advanced Manufacturing Technologies, **41**, 916–931, 2009, doi: 10.1007/s00170-008-1541-2

[17] J. Dou, X. Dai, Zh. Meng, "Optimization for Flow-Line Configurations of RMS Based on Graph Theory", in 2007 International Conference on Mechatronics and Automation, 1261-1266, doi: 10.1109/ICMA.2007.4303730

[18] M. Mihail, N. Vishnoi, "On Generating Graphs with Prescribed Vertex Degrees for Complex Network Modeling", College of Computing Georgia Institute of Technology, 2002.

[19] S. Filipova-Petrakieva, "A New Universal Optimization Approach for Maximum Flows in Power Systems", in 2019 IEEE XI International Conference BulEF'2019, 11-14.09.2019, doi: 10.1109/BulEF48056.2019.9030731

[20] G. B. Dantzig, D. R. Fulkerson, "On the Max-Flow MinCut Theorem of Networks", in Linear Inequalities, Ann. Math. Studies, no. 38, Princeton, New Jersey, 1956.

[21] G. B. Dantzig, D.R. Fulkerson, "On the max-flow min-cut theorem of networks", RAND corporation: 13, 9 September 1964, retrieved 10 January 2018.

[22] G. Eusden, "Max-Flow, Min-Cut - History and Concepts Behind the Max-Flow, Min-Cut Theorem in Graph Theory", 2013.

[23] L. R. Ford, D. R. Fulkerson, "Maximal flow through a network", Canadian Journal of Mathematics, **8**, 399–404, 1956, doi: 10.4153/CJM-1956-045-5.

[24] J. Edmonds, R. M. Karp, "Theoretical improvements in algorithmic efficiency for network flow problems", Journal of the ACM, **19**(2), 248–264, 1972, doi:10.1145/321694.321699.

[25] E. A. Dinic, "Algorithm for solution of a problem of maximum flow in a network with power estimation", Soviet Mathematics – Doklady, Doklady, **11**, 1277–1280, 1970.

[26] Y. Dinitz, "Dinitz' Algorithm: The Original Version and Even's Version", In Oded Goldreich; Arnold L. Rosenberg; Alan L. Selman (eds.). Theoretical Computer Science: Essays in Memory of Shimon Even, Springer, 218–240, 2006.

[27] V. M. Mladenov, Kr. N. Filipova, S. K. Petrakieva, B. Dimov, H. Uhlmann, "Analysis of Signal Competition in Asynchronous Ultra High-Speed Digital Circuits", Przeglad Elektrotechniczny (Electrical Review), **83**(11), 197-200, 2007.

ASTES

Advances in Science, *Technology and Engineering Systems Journal, Vol. 5, No. 6, 185-192 (2020)*

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

A Toolkit for the Automatic Analysis of Human Behavior in HCI Applications in the Wild

Andrea Generosi[1,*], Silvia Ceccacci[1], Samuele Faggiano[2], Luca Giraldi[2], Maura Mengoni[1]

[1]*Department of Industrial Engineering and Mathematical Sciences, Università Politecnica delle Marche, Ancona, 60131, Italy*

[2]*Emoj s.r.l., Ancona, 60131, Italy*

ARTICLE INFO

Article history:
Received: 03 August, 2020
Accepted: 07 September, 2020
Online: 08 November, 2020

Keywords:
User Experience
Deep Learning
Convolutional Neural Networks
Affective Computing
Gaze detection

ABSTRACT

Nowadays, smartphones and laptops equipped with cameras have become an integral part of our daily lives. The pervasive use of cameras enables the collection of an enormous amount of data, which can be easily extracted through video images processing. This opens up the possibility of using technologies that until now had been restricted to laboratories, such as eye-tracking and emotion analysis systems, to analyze users' behavior in the wild, during the interaction with websites. In this context, this paper introduces a toolkit that takes advantage of deep learning algorithms to monitor user's behavior and emotions, through the acquisition of facial expression and eye gaze from the video captured by the webcam of the device used to navigate the web, in compliance with the EU General data protection regulation (GDPR). Collected data are potentially useful to support user experience assessment of web-based applications in the wild and to improve the effectiveness of e-commerce recommendation systems.

1. Introduction

This paper is an extension of work originally presented in ISCT [1]. Understanding User eXperience (UX) in the wild is of paramount importance for online business (e.g., e-commerce) to gain information about users' opinion and behavior, in order to increase productivity, customize services and better drive strategic decisions. According to the definition provided by the international standard on ergonomics of human-system interaction, ISO 9241-210, UX is "a person's perceptions and responses resulting from the use or intended use of a product, system or service". It follows that the analysis of the sensations and emotions of the user with the product represents a central topic for UX research [2]. Over the last decade, increasingly accessible social media and tools offered by e-commerce sites themselves (e.g. Amazon.com), which encourages users to share their opinions, have made available a wealth of information potentially useful for understanding UX. Alongside, there has been a growing interest in sentiment analysis based on text analysis and user opinion mining. Several automatic approaches have been proposed to understand user emotion and extract UX information from online user online review [3-5]. However, though these tools are becoming increasingly accurate and effective, it is impossible to rely only on online reviews to understand how customers really feel. As reported by Jakob Nielsen in a study about user behavior in online communities [6], a significant number of customers (90%) are lurkers (i.e., they read and observe, but they do not contribute), while the 9% contribute intermittently within the only community and only the 1% account for most contributions. This phenomenon has been named "the 90-9-1 rule". Consequently, while systems able to get direct feedback from all the users (whether they are someone who will leave on online review or not) are still lacking, companies continue to spend millions of dollars on surveys and analysis of their customers to know what really works and what does not.

Moreover, recent years have shown an increasing interest in enhancing possibilities of human-computer interaction with e-commerce. Recommendation systems have started to attract interest in both business and research [7]. In the last few years there was significant improvement especially over the collaborative filtering approaches thanks to the advance in the field of machine learning and deep learning techniques [8]. However, the effectiveness of such a system remains generally low, as the algorithms used by most of the web e-commerce for providing shopping suggestions to customers use data collected from other customers, so that system difficulty meets the

*Corresponding Author: Andrea Generosi, a.generosi@pm.univpm.it

customer's expectation [7]. In this context, emotion detection represents one of the directions that can be taken to enhance the overall UX with these systems. In fact, it has been proved that recommender systems are more precise in providing relevant suggestions, when emotion variables are considered [9].

Nowadays, a variety of technologies exists to automatically recognize human emotions, spanning across facial expression analysis, acoustic speech processing and biological responses interpretation. The extensive deployment of smart devices equipped with sensors and connected to the Internet today opens the possibility of accessing a huge amount of data that until now would have been very difficult to collect. Smartphones and laptops equipped with cameras became an integral part of our daily lives. The use of cameras is so pervasive that we no longer worry about their presence, and we are not fully aware of how much information can be collected through them. For example, using appropriate systems, it is possible to process video images so that demographic (e.g., gender and age) and behavioral (e.g., emotions, eye movements) information can be easily extracted [10, 11]. This opens up the possibility of using technologies that until now had been restricted to laboratories to analyze UX in the wild.

In this context, this paper aims to extend and integrate the contribution of the work presented in ISCT, which introduced and evaluated a new system of emotion detection for mobile applications, in order to define a toolkit for automated collection of data related to users' emotions and behavior useful to:

- Support the collection of data related to the user behavior during the interaction with desktop and laptop web-based applications in the wild;
- Improve the effectiveness of recommendation systems.

Such a tool will take advantage of the system proposed in [1] to monitor user's emotions, through the acquisition of facial expression and implements new system based on deep learning algorithms to track the eye gazes from the video captured by the webcam of the device used to navigate the web, in compliance with the General data protection regulation (GDPR), the European Union regulation on the processing of personal data and privacy.

The remainder of this paper is organized as follows: Section 2 gives an overview of the state of art in the context of systems for the automatic detection of users' emotions and behavior and in particular of facial expression recognition and gaze tracking systems. Section 3 describes the overall system architecture with the Emotion recognition and Gaze detection modules. In Section 4 performance evaluation results of the proposed are reported. The last section summarizes the main paper contributions and highlights future outlook.

2. Research Background

Over the past years producers and the marketing/branding industry demonstrated an increasing interest in emotional aspects of user behavior, as understanding the emotional state of users is crucial for developing successful products and services [12]. Psychophysiological methods may offer data throughout the process of experience, which unfolds new possibilities for UX evaluation. This motivated the demand for automatic emotion recognition techniques as a tool for getting a larger quantity of more objective data. Several attempts have been made to explore the possibility to use sentiments analysis to extract UX information from online user reviews. However, online reviews have proved to be too generic and lacking contextual references to effectively support UX assessment [13]. Several studies, such as [14] and [15], proposed systems that allow us to correlate data from different typologies of data, e.g. eye-tracking fixations, sentiment analysis, body gestures, or facial expressions. However, these systems are designed to support only formal UX evaluation assessment, while are not suitable for studies in the wild. Based on the best of our knowledge, no studies so far have proposed the integrated use of gaze and emotion recognition systems, based on deep learning algorithms, to support the collection of relevant information useful for UX assessment of laptop and desktop web applications nor to power recommending systems. Only [11] proposed a system to support UX assessment based deep learning, but such a solution works only for mobile applications. While only [7] proposed to process data provided by a common pc webcam to enable users' gaze and emotion detection in order to manage an intelligent e-commerce recommendation system: it proposed respectively to use SVM algorithm to enable emotion recognition and a gradient-based method to perform gaze tracking. These algorithms, compared with most of Deep Learning CNNs, are more efficient for what concerns the computational performance, but less accurate [16].

2.1. Emotion recognition

In recent decades, research in the field of Human-Computer Interaction has shown an increasing interest in topics such as Affective computing, emotion analysis and study of human behaviors.

This has encouraged the development of numerous methods and tools for the recognition of human emotions, characterized by different levels of intrusiveness. Most of them relate to three main research areas: facial expression analysis, speech recognition analysis and biofeedback analysis. Instruments based on biofeedback currently available (e.g., ECG or EEG, biometric sensors) are still difficult to adopt for studies in the wild, although the use of smart watches and bracelets is increasingly widespread [17]. Instead, research efforts in the field of facial expression recognition systems have allowed the development of reliable and non-invasive systems. The theoretical model most widely used in the analysis of facial expressions to develop algorithms for the recognition of emotions is certainly represented by the Face Action Coding System (FACS) [18]. It allows the identification of six universal emotions (i.e., joy, surprise, sadness, anger, fear and disgust) by tracking the movements of the face muscles.

The vast majority of facial expression recognition systems are based on Convolutional Neural Networks (CNN): a particular mathematical model of Deep Learning that, unlike classical neural networks, present some layers, in the first part of the entire network, which apply a convolution instead of a general matrix multiplication to the images in input before passing the result to the next layers [10]. Several tools have been proposed in literature, such as [19-21]. Usually these systems refer to trained models using datasets built in controlled environments, where it

is possible to obtain the best accuracy scores, or, trained with data obtained by crawlers on the web, with low accuracy but mostly reflecting real contexts. To the best of our knowledge, approaches able to ensure a good accuracy using data obtained "in the wild" have not yet been evaluated. To ensure good accuracy in the recognition of human emotions in different contexts of use, the system of recognition of emotions proposed in this study adopts a hybrid approach. The resulting CNN, based on Keras and Tensorflow frameworks, has been trained merging three different public datasets.

2.2. Eye-gaze tracking

Tracking one's eyes movement is challenging yet important in HCI and computer vision fields. When it comes to products (as well as website) design, it is readily understandable how much it means to know where a user is looking at. In recent years research has aimed to achieve increasingly accurate results, using less and less invasive and off-the-shelf systems (especially webcams).

According to the results of a recent survey, gaze tracking techniques can be classified into two main categories: feature-based and appearance-based [22]. The first approaches require the acquisition of high-quality images to determine ocular characteristics, such as infrared corneal reflections [23], pupil center [24] and iris contour [25]. To determine the direction of the gaze independently to the head pose, these features are related to 3D eye models. Although such approaches allow high accuracy to be achieved, the accuracy of the geometric calculation is highly dependent on system calibrations that are often difficult or impossible for ordinary users to perform. In addition, as they need to extract small features of the eye, they require the use of special equipment (e.g. IR light sources, special glasses / contact lenses) to acquire special high-resolution infrared images. Such technologies are not normally available in uncontrolled environments. Conversely, appearance-based approaches do not need particular illumination conditions, so that they work well with natural lighting, and require only a single webcam [26]: they take image contents directly in input and try to implicitly extract some relevant features, thus establishing a mapping to screen coordinates. Applicability of this latter kind of methods is remarkably large, although the potential handling of low-resolution images makes their accuracy generally lower [26].

Recently, deep learning (DL) and convolutional neural networks (CNN) have been gaining interest for gaze estimation, so much so that many datasets [27] and network architectures [26, 28, 29] have been proposed over the last years. The gaze-tracking system described in this paper adopts an architecture similar to the one proposed by Krafka et al. [30], which represents one of the more solid CNN actually proposed for gaze tracking. It is based on the AlexNet model described in [31].

3. System Architecture

For the proposed toolkit a centralized architecture has been designed. This kind of architecture has been inherited from the system described in [11] and adapted to be used in synergy with web platforms instead of mobile applications. In Figure 1 are shown the main components: the Web Plugin and the Deep Learning Platform (DLP), respectively the frontend and the backend sides. The Web Plugin is a Javascript library with different functions usable by the developers (APIs) to get info about user's interactions with the web application, i.e. interactions timestamps, clicks and scroll coordinates and above all the webcam handling, so how and when to enable the webcam activation, shooting frequency etc. This feature is the most important because both the Deep Learning engines that compose the DLP make use of the user's face photos to return the information related to his behavior. Photos taken through the pc webcam are encoded in base64 from the Web Plugin and sent through a HTTPS POST to the REST interface exposed through a specific endpoint from the DLP and developed in Python. The received call is then decoded and parsed to get the original data and store it in the server volatile memory, to ensure the compliance of the system with GDPR about privacy. After that, the encoded photo is stored in three different Redis queues: in this way every DL Tracker module can process every photo asynchronously. These queues are used to store and manage the order of arrival of the data that will be given in input to the Gaze Tracker and Emo Tracker for the prediction, respectively, of the gaze coordinates on the screen and the universal Ekman's emotions. Both trackers are Convolutional Neural Networks (CNN) implemented with the Python frameworks Tensorflow and Keras. Whenever the photo processing is concluded for a CNN, the output is saved in a database.

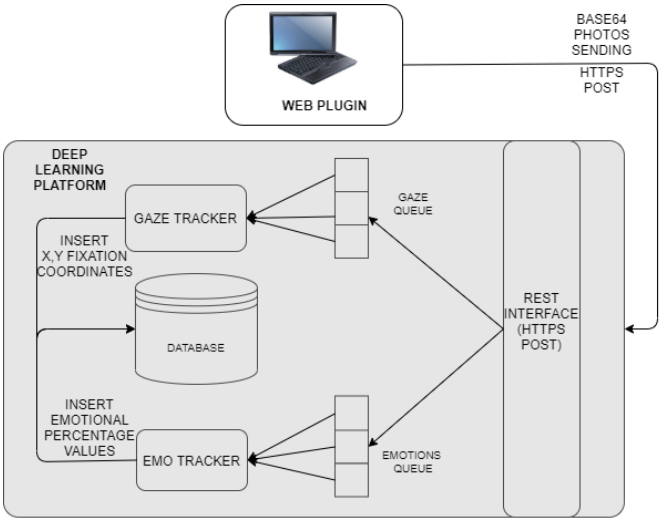


Figure 1: System architecture

3.1. Emo Tracker - the Emotion detection module

Based on [20], a state-of-the-art Deep Learning model that performs well in image recognition should also perform well in a facial expression recognition task: that's because basically the visual discrimination of human emotions is an image classification task. According to different review works, like [19], there exist different DCNNs (Deep Convolutional Neural Networks) that reach various accuracy levels, and those that reach the highest performances are always trained using datasets composed by images retrieved in controlled environments and labeled in labs. Instead, lower accuracy percentages are reached by those datasets with "in the wild" properties, especially retrieved by web crawlers, usually composed by blurred, decentralized or very low-resolution images with misclassified

labels [20]. In return, this kind of datasets are often significantly bigger than the other kind.

To achieve the best emotion recognition accuracy and performances, various tests were conducted. Firstly, by merging the public datasets CK+ (developed in laboratory) [32], FER+ (the FER dataset re-tagged version with crowdsourcing) and AffectNet (manually annotated), a big dataset is constructed. More in detail, FER+ dataset has a label accuracy about 90%, with just 35k photos [20], AffectNet instead has more than 1 million web crawled images, but it also provides more than 400k photos manually labeled by several experts. These datasets have been chosen because they are tagged with almost one of the universal Ekman's emotions labels and because they reach a labeling accuracy relatively high. Moreover, as mentioned, they belong to different dataset categories, so they provide the possibility to improve the resulting model reliability: the assumption is in fact that combining lab generated datasets with those with "in the wild" properties, it is possible to obtain a better accuracy for the in the wild benchmarks. To further improve resulting dataset quality, a Python script to filter all the images with none or more than one face has been developed. The resulting merged dataset at the end counts more than 250k photos, with data distribution shown in Figure 2.

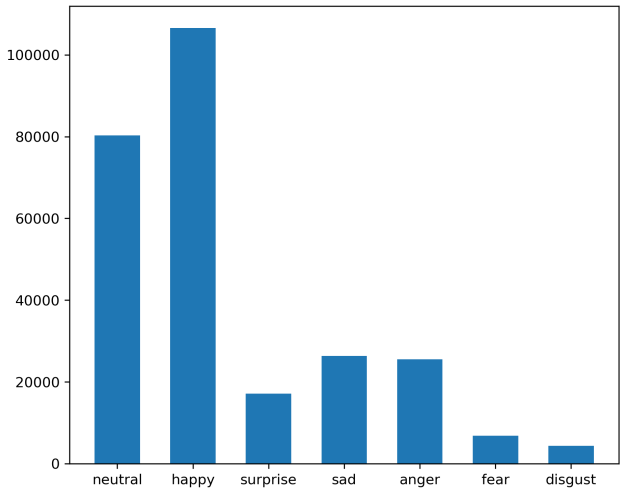


Figure 2: Data distribution

Every single dataset has then splitted into two different parts for training and validation, so to avoid distorting the results during validation phase: for this purpose it was chosen, empirically, to divide each dataset using the ratio 80% for training and 20% for validation, equally for each dataset. After constructing the dataset, preprocessing steps are performed such as facial alignment, centralization of the face respect to the image and face rotation, obtained using Dlib facial landmarks coordinates as reference. All the photos have then been scaled to 64x64 pixels, so to make them homogeneous in size.

For the implementation of the Emo Tracker module, different Python scripts that make use of Dlib, Tensorflow and Keras frameworks have been developed. In particular, Dlib is used to detect one or more human faces in a frame (Face Detection) and provide main face landmarks coordinates, while Keras and Tensorflow provide some of the most used APIs in the Deep Learning field. Trained networks models have been defined to take 64x64 pixels face images in grayscale by the input layer, and to return the Ekman's Emotions (joy, surprise, anger, disgust, sadness, fear and neutral) classification probability by the output layer. Different Keras model architectures such as Inception [33], VGG13, VGG16 and VGG19 [34], were also tested to compare the reached accuracy levels, as listed in Table 1.

Table 1: Performance of different models

Architectures	Accuracy (%)
VGG13	**75.48**
VGG16	74.48
VGG19	73.14
InceptionV2	75.26
InceptionV3	67.20

Network hyperparameters are initialized as it is stated on [20], then the other variations are also experimented such as validation split 0.1, 0.2, number of epochs 30, 50 and 100 and dynamic learning rate defined as

$$lr = lr \times (1 - \frac{epoch}{epoch})$$

Learning rate (lr) is initialized with 0.025 and updated on each epoch accordingly. As shown in table 1, VGG13 model reaches the best accuracy percentage in validation phase, accuracy values over time in relation to the epochs increasing are shown on Figure 3, as a reference of the carried-out experiment. Overfitting occurs during training using other deep neural network architectures, except for the VGG13, when epochs number is bigger than 30.

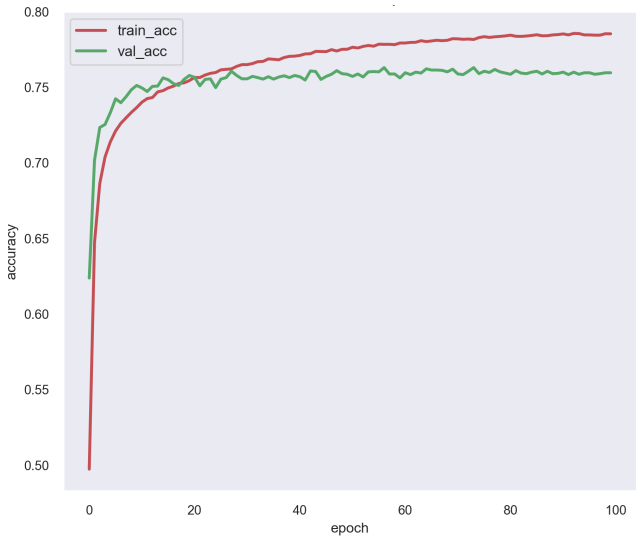


Figure 3: Training and validation accuracy of VGG13

3.2. Gaze Tracker - the Eye-Gaze Detection module

In 2016, Krafka et al. [30] published a paper with a solid CNN proposal for gaze tracking, together with a big dataset of crowdsourcing-collected faces; many researchers are referring to it since then. The Gaze Tracker module's network adopts an architecture similar to the one proposed in [30], which is based on the AlexNet model [31], but respect to that one, a dropout layer

was added to the right after each last convolutional layer. Also in this case, the output is the camera distance, in centimeters. The main challenge using this architecture has been to adapt it to an application context different from the one originally established: iTracker in fact has been trained to work with mobile devices, especially it can reach the best performances on iPhone devices, where it is possible to exactly know, dynamically, the phone camera position. Moreover, practically the whole dataset used to train the model, was built using photos taken by Apple devices. Our purpose, instead, is to propose a gaze detection module optimized to be used on laptop and desktop devices, while it has been decided to not consider a general model that involves also mobile displays in order to not undermine the performance of the model for the specific case taken into consideration by this research.

To achieve this objective, a new training dataset has been built from scratch, collecting pictures from informed volunteers who decided to contribute. For this purpose, a web application has been developed. Each participant was asked to stare at a red dot that was randomly displayed in 30 different screen positions, while the webcam took a picture for each of them; corresponding point coordinates were stored too, to ensure image-screen coordinates association. Concerning screen sizes, this application has been thought to be as general as possible (i.e., capable of dealing with different kinds of displays). However, due to infrastructural reasons, it is not possible to automatically retrieve physical dimensions of a screen from a web page application.

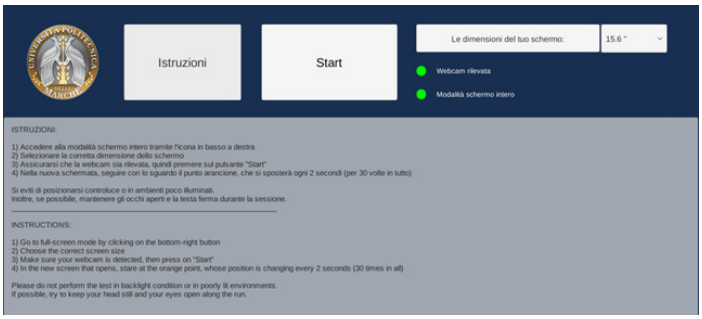


Figure 4: Main screen of the web application for collecting training pictures; by pressing on "Start", a new empty screen appears, in which the dots are shown in succession

For this reason, on the application interface, the user is asked to choose its own screen size, among the ones displayed in a dropdown list, ranging from 13 to 30 inches (i.e., the most common laptop and desktop sizes). By knowing the physical size of the screen and the display resolution (that can be automatically determined instead), it has been possible to express the dots coordinates in cm. At the end, 54 subjects agreed to participate, so the training dataset has been composed of 1620 photos.

The CNN for gaze tracking was implemented and trained using *python* programming language, running *Keras* API upon *TensorFlow* library support and with GPU acceleration.

Inputs of the CNN are the cropped face image (sized 224x224), the two cropped eyes images (sized 224x224) and a 1x4 face grid vector, expressing the portion of the entire image occupied by the face; in order to normalize the dataset, both the face and the eyes images are converted to grayscale, their contrast being increased by means of histogram normalization. The output is a two-dimensional vector containing x and y coordinates of the estimated display point (in centimeters). Faces and eyes are detected and cropped using the dlib [35] frontal face detector and 68 landmarks predictor, respectively. When a face box is found, the screen coordinates of its top left corner and its width and height are stored in the face grid vector; in this way implicit information can be retrieved about the user-screen relative pose. Adadelta [36] has been chosen among the optimizers implemented in *Keras*.

A python script was also realized for visualizing the scan path of the eyes over a displayed image; when launching it, a preset image is shown in full screen, while the webcam repeatedly takes pictures of the user and feeds the network with them. All the coordinates are stored and, when the image is closed, they are automatically processed in a manner that near points are clustered together, thus separating fixations from saccades [37]. Eventually, the colored clusters are depicted over the displayed image, together with straight lines connecting the subsequent ones, to give an idea on the path which the eyes followed overall.

4. Results

The proposed emotion and gaze detection systems were tested separately.

4.1. Evaluation of emotion detection performance

The test accuracy of each emotion category on the confusion matrix plotted as heatmap on Figure 5, shows some interesting results. The images with fear tag misclassified as surprise and disgust tag misclassified as anger has over 20% rate, these are results that are very often found in literature, mostly because these facial expressions are difficult to differentiate even for a human being, so they are often misinterpreted also in the dataset labeling phase. Moreover, as it can be seen, the emotion categories with more training samples resulting in higher accuracy compared to the ones with fewer training samples. Consequently, it is believed that a more evenly distributed dataset may improve the model accuracy further.

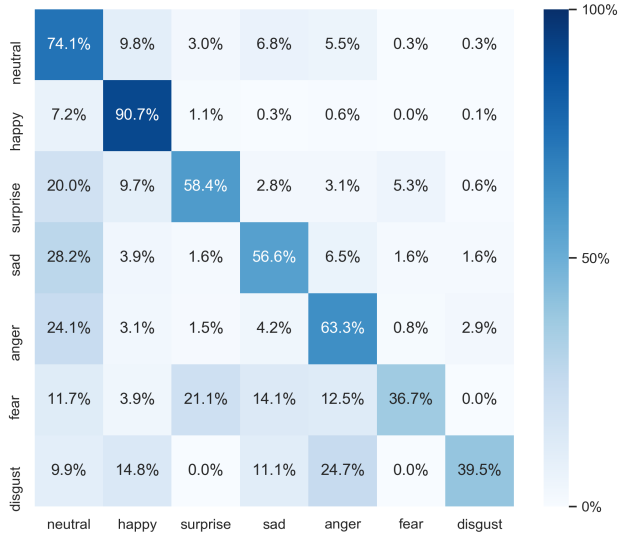


Figure 5: Accuracy of each emotion category

The model VGG13 we trained is also evaluated on EmotioNet [21] 2018 challenge dataset: the Ohio State University, on their website [38], has in fact made available their dataset to give anyone the opportunity to compare their results with those of the challenges of 2017 and 2018. The table shows the evaluation result.

Table 2: Emotionet evaluation results

Categories	Accuracy	F1
happy	**0.9770**	**0.9799**
anger	0.75	0.1198
disgust	0.0128	0.0099
sad	0.5955	0.2888
surprise	0.7059	0.3944

In this case only five emotions have been evaluated: this is because available photos related to Fear emotions were not enough to establish a meaningful comparison. In addition, no photos were made available for the comparison of the Neutral emotion.

4.2. Evaluation of eye-gaze detection performance

To evaluate the performance of the prediction the same software implemented to gather training samples has been used. The human eye detected using dlib [35] landmarks is the input of the python script that crops the region of interests and predicts the coordinates. They are expressed in centimeters and take the top-left screen corner as origin. A total of 20 subjects agreed to take part in the test and for each participant the predictions have been made for 30 different screen positions.

Every time the test is repeated the real and predicted coordinates have been saved in a database. To calculate the error between the predicted point and the real one the following formula has been used:

$$e_m = \frac{\sum_{i=1}^{n} | x_{i\,real} - x_{i\,pred} |}{n}$$

where $x_{i\,real}$ is the real coordinate where the subject *i* is looking at, $x_{i\,pred}$ is the predicted coordinate, *n* represents the total number of participants, and e_m the mean error (in centimeters) that has been calculated for any screen position. The same formula was used to evaluate the error on the y coordinate.

At this stage, 30 values of e_m have been calculated., The mean errors have been aggregated in 12 more meaningful values, as no significant variations have been found in comparison with the high number of screen positions. Each of them represents the error in a specific screen area (see Figure 6 and 7).

Both the heatmaps in Figure 6 and 7 show a higher error value in the top side of the screen (up to 11 cm), but interesting results have been achieved for the other screen positions. In this case the error goes between 0.02cm to 7.2cm. As the heatmap shows, the model still needs some improvement since the accuracy strongly depends on the area the subject stares at, but this experiment proves that it is possible to reach very accurate gaze predictions.

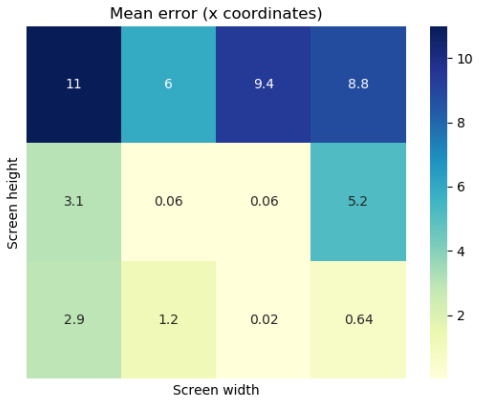


Figure 6: Mean error (cm) for each screen area for the x coordinate.

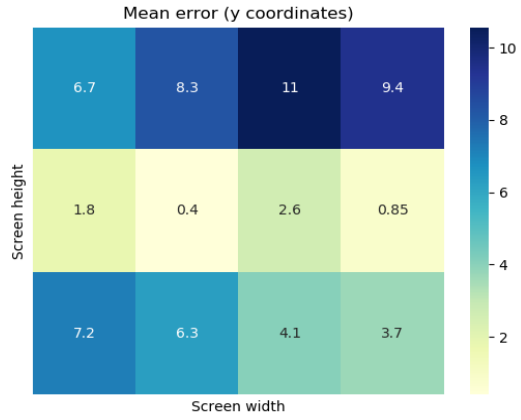


Figure 7: Mean error (cm) for each screen area for y coordinate.

5. Conclusion and discussion

This work proposed a system able to collect, data about user interactions/behavior with web applications potentially useful to support UX analysis. In particular, the proposed toolkit makes use of a deep learning-based platform to allow an "in the wild" data collection during user's interactions with a web application/platform just using a normal webcam. Future studies will be conducted to define an automatic system able to support the analysis of collected data, visualize UX outcomes through the provision of proper KPIs, and suggest useful design guidelines, based on the observed users' behavior, to improve UX design.

Regarding the emotion detection module, experimental results showed that the accuracies of facial expression categories such as fear and disgust are low mainly because of the small dataset of the corresponding category. The other facial expression categories however have reached relatively high classification accuracies. The evaluation results on EmotioNet dataset also supports the results of our experiment. Future work will be focused on the collection of more data on the small dataset categories to improve the classification accuracies of those corresponding categories

and on the exploration of generative adversarial networks to increase recognition accuracy.

About the gaze detection module, has been showed that the average error that can be reached is widely improvable and still does not reach the levels of accuracy of iTracker: this is due both to the small size of the dataset used for training and to the heterogeneity of the environments with which the user may be operating, in particular distance between user and screen, screen size, webcam position, etc. One of the future main works are undoubtedly the expansion of the dataset used to train the gaze network, unfortunately still very limited, adopting crowdsourcing tools.

Finally, data related to users' emotions and behavior, automatically collected through the proposed tool, can be used to improve recommendation systems based on collaborative filtering approaches. Current recommendation systems actually consider only data related to product features, customer preferences, customer demographic data (e.g., age, gender) historical data of purchases, and environmental factors (e.g., time, location). The availability of a huge amount of data related to what users observe the most, and what they feel while browsing e-commerce, opens new possibilities for customization and adaptation of recommendations, based on the real taste and interest of the user. Future studies should go in this direction.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The toolkit is part of the NextPerception project that has received funding from the European Union Horizon 2020, ECSEL-2019-2-RIA Joint Undertaking (Grant Agreement Number 876487). The authors would also thank the research fellow Lorenzo Marchesini for its support in the design and development of the eye gaze detection system.

References

[1] A. Talipu, A. Generosi, M. Mengoni, L Giraldi, "Evaluation of Deep Convolutional Neural Network architectures for Emotion Recognition in the Wild," in 2019 IEEE 23rd International Symposium on Consumer Technologies, 25-27, 2019. IEEE https://doi.org/10.1109/ISCE.2019.8900994

[2] C. Lallemand, G. Gronier, V. Koenig, "User experience: A concept without consensus? exploring practitioners' perspectives through an international survey," Computers in Human Behavior, **43**, 35–48, 2015. https://doi.org/10.1016/j.chb.2014.10.048

[3] J. Jabbar, I. Urooj, W. JunSheng, N. Azeem, "Real-time sentiment analysis on E-commerce application" in 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC) https://doi.org/10.1109/ICNSC.2019.8743331

[4] X. Fang, J. Zhan, "Sentiment analysis using product review data," Journal of Big Data, **2**(1), 5, 2015. https://doi.org/10.1186/s40537-015-0015-2

[5] S. Hedegaard, J. G. Simonsen, "Extracting usability and user experience information from online user reviews," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2013. https://doi.org/10.1145/2470654.2481286

[6] J. Nielsen, "The 90-9-1 rule for participation inequality in social media and online communities," 2016. Retrieved 20.07.2020, from https://www.nngroup.com/articles/participation-inequality/

[7] S. Jaiswal, S. Virmani, V. Sethi, K. De, P. P. Roy, "An intelligent recommendation system using gaze and emotion detection," Multimedia Tools and Applications, **78**(11), 14231-14250, 2019. https://doi.org/10.1007/s11042-018-6755-1

[8] I. Portugal, P. Alencar, D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review," Expert Systems with Applications, **97**, 205-227, 2018. https://doi.org/10.1016/j.eswa.2017.12.020

[9] A. Bielozorov, M. Bezbradica, M. Helfert, "The role of user emotions for content personalization in e-commerce: literature review," in International Conference on Human-Computer Interaction, 2019. https://doi.org/10.1007/978-3-030-22335-9_12

[10] A. Generosi, S. Ceccacci, M. Mengoni, "A deep learning-based system to track and analyze customer behavior in retail store," in 2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin), 2018. https://doi.org/10.1109/ICCE-Berlin.2018.8576169

[11] A. Generosi, A. Altieri, S. Ceccacci, G. Foresi, A. Talipu, G. Turri, ... & L. Giraldi, "MoBeTrack: a toolkit to analyze user experience of mobile apps in the wild," in 2019 IEEE International Conference on Consumer Electronics (ICCE), 2019. https://doi.org/10.1109/ICCE.2019.8662020

[12] E. Ganglbauer, J. Schrammel, S. Deutsch, M. Tscheligi, "Applying Psychophysiological Methods for Measuring User Experience: Possibilities, Challenges and Feasibility," in Workshop on User Experience Evaluation Methods in Product Development, August 25, 2009. Uppsala, Sweden.

[13] S. Hedegaard, J. G. Simonsen, "Extracting usability and user experience information from online user reviews," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2013. https://doi.org/10.1145/2470654.2481286

[14] V. Georges, F. Courtemanche, S. Senecal, T. Baccino, M. Fredette, P. M. Leger, "UX heatmaps: mapping user experience on visual interfaces," in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016. https://doi.org/10.1145/2858036.2858271

[15] R. Y. da Silva Franco, R. Santos do Amor Divino Lima, M. Paixão, C. G. Resque dos Santos, B. Serique Meiguins, "UXmood—A Sentiment Analysis and Information Visualization Tool to Support the Evaluation of Usability and User Experience," Information, **10**(12), 366, 2019. https://doi.org/10.3390/info10120366

[16] A. Altieri, S. Ceccacci, M. Mengoni, "Emotion-Aware Ambient Intelligence: Changing Smart Environment Interaction Paradigms Through Affective Computing," in International Conference on Human-Computer Interaction, 2019. Springer, Cham. https://doi.org/10.1007/978-3-030-21935-2_20

[17] A. Kartali, M. Roglić, M. Barjaktarović, M. Đurić-Jovičić, M. M. Janković, "Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches," in 2018 14th Symposium on Neural Networks and Applications(NEUREL), 2018. https://doi.org/10.1109/NEUREL.2018.8587011

[18] P. Ekman, V. F. Wallace. *Manual for the facial action coding system*. Consulting Psychologists Press, 1978.

[19] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," 1–25, 2018. https://doi.org/10.1109/TAFFC.2020.2981446

[20] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution" 2016. https://doi.org/10.1145/2993148.2993165

[21] C. F. Benitez-Quiroz, R. Srinivasan, Q. Feng, Y. Wang, A. M. Martinez, "EmotioNet Challenge: Recognition of facial expressions of emotion in the wild," 2017.

[22] D. W. Hansen, Q. Ji, "In the eye of the beholder: a survey of models for eyes and gaze," IEEE Trans Pattern Anal Mach Intell, **32**(3), 478–500, 2010. https://doi.org/10.1109/TPAMI.2009.30

[23] J. Sigut, S. A. Sidha, "Iris center corneal reflection method for gaze tracking using visible light," IEEE Transactions on Biomedical Engineering, **58**(2), 411-419, 2010. https://doi.org/10.1109/TBME.2010.2087330

[24] R. Valenti, T. Gevers, "Accurate eye center location and tracking using isophote curvature," in 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008. https://doi.org/10.1109/CVPR.2008.4587529

[25] H. Wu, Y. Kitagawa, T. Wada, T. Kato, Q. Chen, "Tracking iris contour with a 3D eye-model for gaze estimation," in Asian Conference on Computer Vision, 2007. https://doi.org/10.1007/978-3-540-76386-4_65

[26] X. Zhang, Y. Sugano, M. Fritz, A. Bulling, "Appearance-based gaze estimation in the wild," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2019. https://doi.org/10.1109/CVPR.2015.7299081

[27] S. Winkler, R. Subramanian, "Overview of eye tracking datasets," in 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX), 2019. https://doi.org/10.1109/QoMEX.2013.6603239

[28] J. Lemley, A. Kar, A. Drimbarean, P. Corcoran, "Convolutional neural network implementation for eye-gaze estimation on low-quality consumer imaging systems," IEEE Transactions on Consumer Electronics, **65**(2), 179-187, 2019. https://doi.org/10.1109/TCE.2019.2899869

[29] X. Zhang, Y Sugano, M. Fritz, A. Bulling, A., "It's written all over your face: Full-face appearance-based gaze estimation," in Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition Workshops, 2017. https://doi.org/10.1109/CVPRW.2017.284
[30] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, A. Torralba, "Eye tracking for everyone," in Proceedings of the IEEE conference on computer vision and pattern recognition 2016. https://doi.org/10.1109/CVPR.2016.239
[31] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012. https://doi.org/10.1145/3065386
[32] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in 2010 ieee computer society conference on computer vision and pattern recognition-workshops, 2010. https://doi.org/10.1109/CVPRW.2010.5543262
[33] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition 2016. https://doi.org/10.1109/CVPR.2016.308
[34] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
[35] D. E. King, "Dlib-ml: A machine learning toolkit." The Journal of Machine Learning Research, **10**, 1755-1758, 2009.
[36] M. D. Zeiler, "Adadelta: an adaptive learning rate method," arXiv preprint arXiv:1212.5701, 2012.
[37] L. Chamberlain, "Eye tracking methodology; Theory and practice," Qualitative Market Research: An International Journal, 2007. https://doi.org/10.1108/13522750710740862
[38] EmotioNet Challenge, Retrieved July 30, 2020 from https://cbcsl.ece.ohio-state.edu/EmotionNetChallenge/index.html

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 193-205 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Real-Time Identification and Classification of Driving Maneuvers using Smartphone

Munaf Salim Najim Al-Din*

College of Engineering and Architecture, Electrical and Computer Engineering, University of Nizwa, Nizwa, 616, Oman

ARTICLE INFO

Article history:
Received: 18 August, 2020
Accepted: 01 November, 2020
Online: 08 November, 2020

Keywords:
Driving behavior
Identification and classification
Smartphones

ABSTRACT

The fast-paced development of smart technologies and the prevalence of vehicles, created an urgent demand to study and improve safety issues related to driving. In order to reduce traffic accidents, driving behavior was found to be very important issues to study and investigate. Recently, the advent and widespread of smartphone platforms with advanced computing competence and embedment of a variety sensing elements have greatly contributed to the development of solutions that can detect, and evaluate driving behavior and skills. In this study the development of a real-time smartphone-based identification and classification system for highway driving maneuvers is presented. The proposed system has been designed to detect ten different maneuvers usually performed by drivers when driving on highways. The methodology is based on separating the identification and the classification processes. The identification process is performed by a hybrid pattern matching scheme that combines Dynamic Time Warping (DTW) and neural networks. While a second neural network has been used to classify maneuvers, severity based statistical and time features. The separation of the identification and classification processes simplifies and accelerates the learning processes of the neural networks and greatly improves both system's reliability and accuracy

1. Introduction

During the last century vehicles have prevailed all over the world and have greatly contributed to the development and prosperity of the world economy. According to [1], the number of registered vehicles in the world has increased from 1.015 billion in 2010 to nearly 1.5 billion in 2020. In spite of all their positive impacts, vehicles have negative consequences on environmental pollution, global warming and greenhouse gas emissions, congested roads and noises, fuel consumption, social impacts, road accidents and many others to list. These negative consequences have stringent the environmental and safety regulations and standards for automotive industry. Therefore, to be in compliance with the regulations and standards, innovative solutions and newly emerging technologies have been adopted and utilized to provide the utmost advanced, efficient, safest and comfortable transportation means. In these days, high-end vehicles are equipped with smart systems such as cruise control, advanced emergency braking, safety warning systems, GPS technology for navigation, fuel efficiency, and automated driving system. Furthermore, in the near future, the automotive industry is expected to evolve even further by the production of autonomous vehicles that operate in a fully connected and digitized environment. Such immense leap in transportation technologies is only made possible by intensive studies in the field of vehicular applications especially the studies of driving behaviors.

Although that advanced technologies such as smart safety systems have been integrated within modern, still fully human driven vehicles is and will continue to dominate traffic safety since almost all of vehicles in use at the present time are of conventional type. Conventional vehicles are those fully driven and controlled by human drivers, and have few or none of the new technologies and accessories. The impediment in utilizing the new technologies is due to prohibitive cost that cannot be afforded by most of vehicles' owners especially in developing countries in addition to insulation difficulties [2]. Based on this argument, safety and road accidents due to human errors are still urgent issues that need to be addressed and solved for the majority of conventional vehicles. Based on the World Health Organization reports, vehicle accidents are considered to be the ninth cause of death in the world in addition they are a leading cause of fatal injuries [3]. Previous road safety studies and statistics have showed that the majority of road accidents are attributed to both physiological and behavioral human factors [4]. Therefore, research in the field of driving behavior monitoring systems has gained great momentum and

*Corresponding Author: Munaf Salim Najim Al-Din, munaf@unizwa.edu.om

several research institutes have developed systems to monitor and analyze the driving patterns and activities. Beside their urgent importance these studies serve the research and developments of different applications such as the design of advanced driver assistance system, intelligent transportation systems, and autonomous vehicles [5-6]. Additionally, they helped logistics companies to reduce their operational costs through tracking and monitoring their vehicles [7], and insurance companies to evaluate the driver performance, usage and can be used to resolve insurance tariffs and disputes [8].

The study and analysis of the concepts related to driving processes and problems involve the utilization of Driving Monitoring and Assistance System (DMAS). DMASs are usually designed to at least provide two main tasks, data acquisition and driving maneuvers analysis. The first unit is responsible for automating the collection of driving data, such as driver's attentiveness, vehicle dynamic state, driver's actions and driving environment or surrounding. The second unit is designed to extract patterns and features from the driving data and then utilizes intelligent computational techniques to identify and classify the driving behavior. In general real-time acquisition of driving patterns and parameters can be achieved through various types of sensors, such as GPS, GNSSs, In-Vehicle sensors such as CAN-BUS signals, costumed designed systems, vision systems and smartphones. Among this range of sensing platforms smartphone-based DMAS have growing so rapidly in number and sophistication, due to the prevalence, continuous improvement of the computing power, advancement of integrated sensors and continuous reduction cost in smartphones. Therefore, smartphones are taking an important role in the field of Intelligent Transport Systems such as vehicle monitoring driver behavior analysis [9].

This paper is an extended version of the work published in [10] and it is intended to presents the development of a real-time smartphone-based highway driving maneuvers identification and classification system. The system consists of three layers namely, data acquisition layer, preprocessing layer and the identification and classification layer. It is well known that the estimation driving parameters with high degree of accurately greatly affects the detection and classification of driving maneuvers. Therefore, in this paper a new approach to estimate the speed of a moving vehicle derived from GPS and the IMUs is developed. The identification and classification layer consists of three main units, a pattern matching units, to identify the patterns of estimated driving parameters. The DTW technique is used to identify the pattern of each driving parameter during maneuver execution. The second unit consists of a single feed-forward neural network that will identify the type of the maneuver based on the combination of the identified patterns from the first unit. Finally the third unit is responsible for the manoeuver classification and it is performed by a second neural network.

2. Literature Review

In the past decade several techniques and methodologies have been proposed and used to recognize and classify driving maneuvers and then infer an evaluation for driving behavior. Up to date there is no agreement on the parameters and techniques that could uniquely be used for this purpose. This section presents a short review of the existing technologies and methodologies used for driving maneuvers identification and classification. The literature review illustrates only two major issues related to this field of research namely, the data collection and identification and classification techniques.

Depending on the purpose and functionalities that is required to be achieved, there are various types of data acquisition systems (DAS) developed in the past ten years. The first category of DAS relies on the in-vehicle sensors which broadcast their data through the CAN-BUS which can be captured through the OBD II port. These systems can only provide information that can be used to monitor vehicle' state such as speed, fuel consumption and engine RPM and some limited information related to the driver actions such as throttle pedals positions [11, 12]. Many research centers have developed customized monitoring systems that combine in-vehicle sensors with GPS unit, inertial measurement units, and video systems of the external road and traffic environment and inside the vehicle itself [13]. In general sensors are configured to perceive external and internal environments to provide contextual details about the driving and drivers states. The drawbacks of the simple plug and play monitoring systems are the limited data that can be extracted with. On the other hand, drawbacks of in-vehicle customized systems are their high cost, installation complexity, compatibility and customer acceptance [5]. The developments of smartphones in terms of computational power and integration of sensing devices such as Inertial Measurement Units (IMUs), GPS and many other sensors combined with their ubiquitous presence have led to wide prevalence of smartphones in vehicular applications. Nevertheless, there are shortcomings when employing smartphones as monitoring systems such noisy measurement, coordinates orientation due smartphone placement, battery life and they cannot access vehicles data by their own. Several techniques, as they will be explained later, have been developed to provide the required remedies to overcome these problems [14]. Smartphones have also been utilizes as an intermediate processing hub with the aid of special applications to collect vehicle's sensors data and perform the necessary computation. In general, these applications enable the smartphone to communicate with the OBD-II port, either through Bluetooth or WiFi, and to collect a number of vehicle's sensors data such as the speed, engine RPM and many other diagnosis parameters. The processing of these in-vehicle sensors' data streams can be combined with smartphone sensors to further enrich the analysis [15].

Based on the collected data or the estimated driving parameters and the maneuvers to be detected, different techniques have been suggested to identify the type of a maneuver and classify it. A straightforward approach used to process the driving data is based on specifying a number of thresholds for each maneuver, that are defined merely based on the range of collected driving data and the experience. Once the sensed signals exceed a predefined threshold for a specified period of time then a maneuver is detected [16-19]. The main disadvantage in using threshold-based approaches is related to the absence of both the sensitivity and specificity that are required to obtain accurate identification and classification since improper definition of thresholds can result in poor performance. In a different approach, researchers considered the vehicle driving as a reasoning process, hence rule-base and fuzzy logic systems have been used to model driving behavior in terms of driving maneuvers classification only, i.e. it is not possible to detect the

type of the maneuver by using fuzzy logic schemas [13,20-24]. The implementation of a classical fuzzy logic classifier depends on the definition and setting of fuzzy input variables, the rule-based inference system and the fuzzification and defuzzification processes. Therefore, fuzzy logic classifiers are not constantly precise, since the both the fuzzy rules and conclusions are based on postulation, so it may not be widely accepted. Furthermore, they don't have the capability of machine learning and the setting of the precise inference rules and, membership functions is a difficult task. To overcome the aforementioned limitations in both threshold-based and fuzzy logic classifiers several adaptive methods have been proposed to automatically generate thresholds and fuzzy rules from numerical data [20-24]. Adaptive systems have showed distinctive superiority over classical fuzzy systems due to their learning capabilities, their ability to be modified and extended and finally the ability to extract newly modified rules and fuzzy variables. As a consequence of this combination between the adaptive learning and human experience adaptive vigorous classification systems can be designed with high degree of reliability and accuracy.

Since the data and information used in the identification and classification of driving behaviors and maneuvers are time varying signals, pattern matching techniques, such as DTW, have been employed. In this approach the maneuvers are identified and classified by comparing the similarity of the collected sensors' signals or estimated driving parameters patterns with predefined templates. If the incoming measured sequence has adequate similarity to one of the standard templates, then the system will indicate that a specific driving maneuver has been executed or detected. The major advantage of pattern matching techniques is that the system can sense maneuvers in spite of any differences, in their amplitudes or durations, with predefined templates. Therefore, it would be possible to develop a single set of standard templates that be used for matching maneuvers for different drivers to the same template set [25-28]. The main drawback of pattern matching approach is excessive computational requirements since the algorithms compute the similarity differences for all elements in two signals. This problem will be more significant when the classification analysis is dependent on multidimensional time series. Furthermore, enormous effort is required to extract and select the reference templates and it is also very hard to gather all types of templates due to the exclusive driving styles and behaviors of drivers.

Very recently, machine learning algorithms have been adopted and implemented to identify and classify driving maneuvers. Machine learning algorithms, when are well trained can generate specific rules that will enable them to identify and classify abrupt irregularity on a large set of data. Different approaches have been suggested and implemented to extract features for training these approaches such as simple threshold-based discriminators, statistical values, time domain parameters, frequency domain parameters, and most importantly rules inferred from clustering algorithms. Machine learning algorithms are generally categorized into three different approaches namely the supervised learning, unsupervised learning and semi-supervised learning. In Supervised learning techniques a dataset is first collected and labeled to certain predefined classes. Then a learning model is chosen and trained with some of the labeled dataset to produce a predicted output in different testing samples. Several algorithms under the supervised learning category have been used in the identification and classification of driving maneuvers and behavior such as K-Nearest Neighbor, Naive Bayes, Decision Trees [29], linear regression [30], Support Vector Machines (SVM) [31-32] and Neural Networks [33]. Unsupervised learning methods are used when there are no pre-knowledge about ground truth labels. Therefore, they are used to classify collected unknown data into groups depending on common trends, attributes, patterns, or relationships. There are different unsupervised approaches such as clustering techniques and self-organizing maps. Each approach uses different algorithm for classifying data into groups, and these algorithms could be simple straightforward that can divide collected data based on the common attributes or similar trends in their features. Nevertheless, when performing a pre-processing step of the data, unsupervised learning techniques are usually used with other techniques to enable specific characteristics of the learning model,. The K-means clustering [34] and Principal Component Analysis algorithms [30] are samples of the approaches used in driving behavior analysis.

2.1. Data Collection

As it has been stated modern smartphones are integrated with sensors that can be utilized to collect vehicle driving data for driving maneuver analysis. Previous work [35, 36] showed that precisely preprocessed and calibrated smartphone sensors are becoming a competitive approach to the high price in-vehicle customized DAS. In this study ten frequent maneuvers usually executed by drivers during highway driving are going to be detected and classified. Four of these maneuvers namely, braking, acceleration and left and right lane changes will be analyzed for both straight and curved road segments, in addition to another two maneuvers which are the merging on or exiting a highway. In the preliminary stages of this study an online vehicle driving data recording system has be developed and installed on three different android-based smartphones. The sensors that have been used are the Accelerometer, Gyroscope and the GPS. The developed App has been test by comparing its performance with well-established available Apps like AndroSensor. The sampling rate for the IMUs is in the range between 50 and 100 Hz, from which the optimal sampling rate then will be obtained.

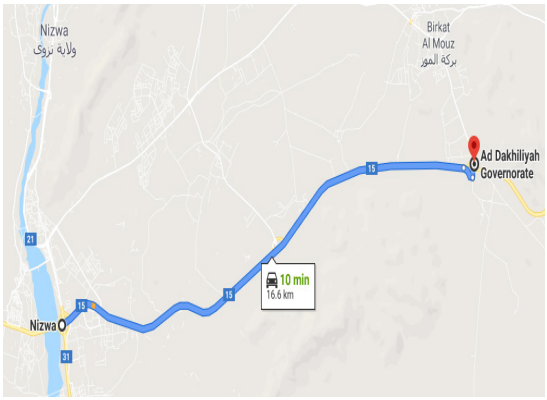


Figure 1: The route used to collected training data

The development of the system proposed in this study passed through two phases, the development and testing phase and the final real-time testing phase. In the first phase adequate data for training and testing the system were first collected. Ten drivers

with different vehicles were enlisted to drive along a highway route of 16km long that covers varying road conditions as shown in Figure 1. In order to have adequate data for training and testing each driver repeat each of the basic divining maneuvers for at least five times with different severity levels from the defensive to the aggressive level. Data collected from inertial sensors in these tests are first pre-processed, then segmented and finally labeled into types and classes.

In addition to the collected data by the smartphone the state of the surrounding vehicles and the state of the driver have been recorded by two video cameras. These two recorded videos are combined with the manually segmented vehicle's driving data to define each recorded maneuver's irregularity levels by three experts. Each expert examined all recorded maneuver and assigned a category or a score for them. Each extracted maneuver will be assigned with an inter-rater agreement score that is calculated from the scores provided by each expert. If the inter-rater agreement score is low then the maneuver is assigned with deviated scores by the experts and there is no agreement on the classification. Therefore, experts are requested to iterate their evaluation to reach a high inter-rater agreement score for most of the extracted maneuvers.

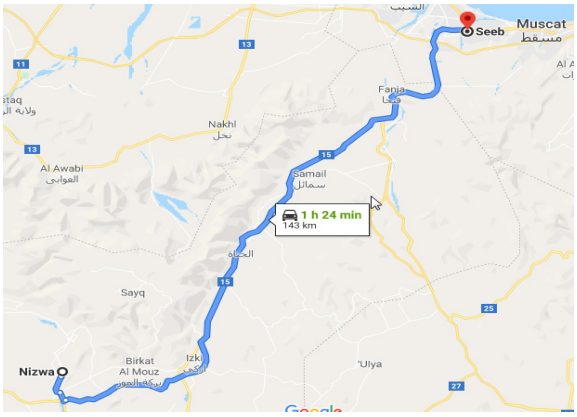


Figure 2: The route used to test data

One of the common problems that obstruct the development of reliable and accurate classification systems is the lack of adequate training and testing data. The lack of sufficient data for any classification problem could results in a noisy and instable class distribution. Therefore, for the sake of improving the system a second dataset was collected by installing a special app, developed by the author, in the smartphones of different 15 drivers, whom they drive on a daily base from the city of Nizwa to the capital Muscat over the route shown in Figure 2. The aim of collecting the second dataset is to enrich the training and testing data by performing a naturalistic driving through a route that is very dynamic and contains almost every different road types.

3. Calibration and Signal Processing

Smartphones' sensors are cheap low-grade sensors that have several deficiencies that may impose restrictions to their use in certain class of applications. In vehicular applications, it is hard to measure and guarantee smartphones accuracy and reliability due to five main prominent problems and limitations namely:

1- Sensor and algorithm complexity

2- Limited battery power.

3- The perturbation of the measurements due to the usage of the device that causes change in its orientation.

4- The low accuracy of smartphone sensors.

3.1. Sensors Calibration

Smartphones' IMUs have poor performance and are prone to error when especially used in measuring dynamic motion applications such as vehicle driving patterns. Therefore, sensors calibrations are used to correct sensors' deterministic errors. The most common types of these sources of errors are biases, scale factors and axes misalignment. In addition to the above mentioned deterministic errors, smartphones' sensors are also exposed to random noise and vibration, non-linearity due to thermal and magnetic effect and many others [37, 38]. To overcome these deficiencies a proper calibration process has been employed to eliminate Smartphones' sensors deterministic errors namely, fixed biases, scale factor errors, and misalignment errors.

The accelerometers are used to measure the instantaneous forces acting on the sensor. The measured acceleration along any axis consists of three components namely; linear acceleration, gravitational acceleration field and noise. On the other hand, the gyroscope measures the rate of change in the device's angular displacement, i.e. angular velocity, along the smartphone's three orthogonal axes. As mentioned errors in an IMU's measurements are contributed to three main categories: biases, scale-factor errors, and misalignment errors. The following equations (1) and (2) are used to model the accelerometer and gyroscope instantaneous measurement:

$$a_m(t) = S_a a_L(t) - g + b_a + n_a \quad (1)$$

$$\omega_m(t) = S_g \omega(t) + b_g + n_g \quad (2)$$

Here a_m, a_L, S_a, g, b_a and n_a represent the instantaneous measured acceleration component, the linear acceleration, scaling factor, gravitational acceleration, bias and noise respectively. While ω_m, ω, S_g, b_g and n_g represent the measured angular velocity along one of the three axes, the true angular velocity, scaling factor, bias and noise respectively. In this paper a new set of calibration procedures were proposed and used to estimate and calibrate smartphone's IMU sensors deterministic errors by considering the effects of vehicle vibration and noise [39-40].

3.2. Filtering

It is well known that data collected using IMU sensors contains substantial noise level, specially the accelerometer and the gyroscope readings. The noise in raw data can be contributed to several sources but mainly it is due to vibration noise and white Gaussian noise. There are different sources of vibration in vehicles such as road roughness, engine induced vibration, whirling of shafts, worn out parts, speed bumps, wheels, etc. In general, the data collected from the sensors contains random white Gaussian noise signals that oscillate with high frequency and typically have zero mean value. A sample for the measured noisy accelerometer and gyroscope signals are shown in Figure 3). In order to attain reliable and accurate recognition, the noise needs to be eliminated as much as possible. There are different approaches to filter

distorted signals; for example statistical filters [41], digital filters [27], Kalman Filter [42] and many others.

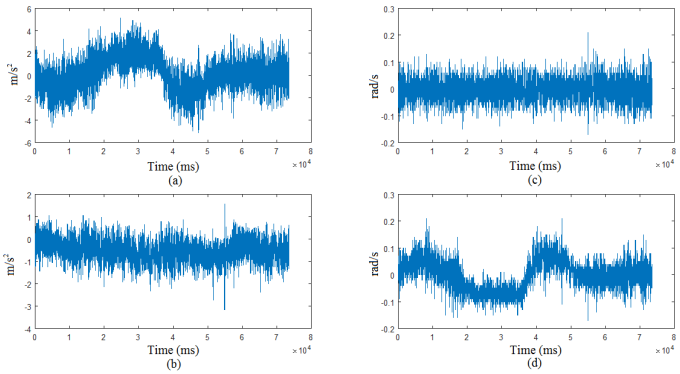


Figure 3: Noisy Smartphone Sensors Signals: (a) Accelerometer Signal x-Direction, (b) Accelerometer Signal y-Direction, (c) Gyroscope Signal x-Direction and (d) Gyroscope Signal z-Direction

In this paper three types of low-pass filters have been investigated namely; one-dimensional Kalman filter, simple moving average filter, and locally weighted running line smoother (LOSS). The performances of the aforementioned filters were compared to select the filter that effectively removes high frequency noise and maintain the real time variation of different driving patterns to the utmost extent.

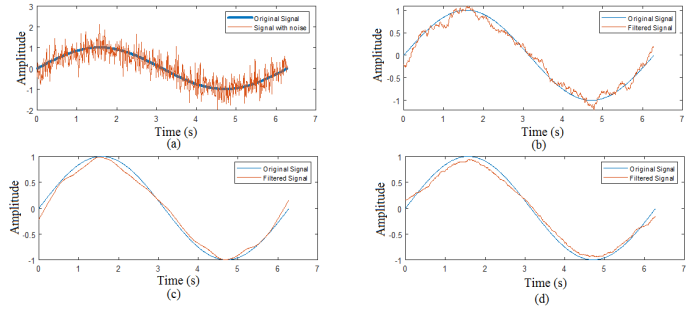


Figure 4: Filtering Results for Test Signal; (a) Original Test Signal; (b) Moving Average Filtering; (c) Kalman Filtering; (d) LOSS Filtering

The simplest filter is the moving average filter which filters the measured sampled signal by replacing each sampled value with the average of the neighboring sampled values defined within a given span or window. Kalman filter is a recursive predictive filter that is based on the use of state space techniques and recursive algorithms. The filter observes over time a series of measured data values, including noise, and then estimates new values that tend to be more accurate than those based on the actual measurement. The LOESS filter estimates the latent function in a point-wise fashion, where for each value of the measured sampled signal a new estimated value is found by using its weighted neighboring sampled (known) values. Detailed theory and implementation of this type of filters is presented in [43]. Due to the random irregular road pavement and vehicle vibration, it is found that the separation between of the real signals through total elimination of noise is not possible. Therefore, a test sinusoidal signal with imposed noise is used to assess the performance of the three filtering approaches. Figure 4 shows the results obtained for the three filters. It is clear that the LOESS technique can achieve very close results to Kalman filter where both effectively remove the nose while preserving the shape of the original signal. Figure 5 shows a sample of the measured accelerometer signals, from which the same conclusion regarding the two filters can be drawn. The LOSS filter has been used in the implementation of noise filtering unit of the complete system because of it is simpler to implement and faster to execute.

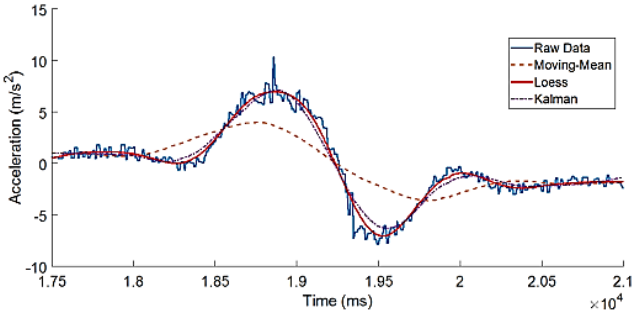


Figure 5: Raw and Filtered Accelerometer Signal

3.3. Smartphone Orientation

When smartphones are used to detect vehicle's driving patterns they could be placed at any location with any arbitrary orientation. Since sensors' measured signals are expressed in a specific frame that is static relative to the device, then one of the most interesting problems of smartphone-based vehicular applications is the separation of the dynamics of the smartphone from the dynamics of the vehicle. In this paper the smartphone is fixed to the windshield by a holder and it could have any orientation. Therefore, a reorientation correction process for the smartphone's coordinate systems is integrated as a pre-processing module to transform the measured sensors data to the vehicle coordinate system. This process is performed by a sequence of geometrical rotations using Euler Angles by assuming that the vehicle is horizontally aligned during the initial calibration period, so that the vehicle roll and pitch angles relative to a tangent frame both are zero. Assuming that the vehicle does not experience any acceleration, the smartphone's roll and pitch angles can be estimated from accelerometer measurements of the gravity vector as shown in Figure 6.

The determination of Euler angles has fully detailed in [44]. Upon the determination of rotational angles, the referenced coordinate system for the smartphone can be transformed to the vehicle-referenced coordinate system by multiplying with the reorientation matrices as follows:

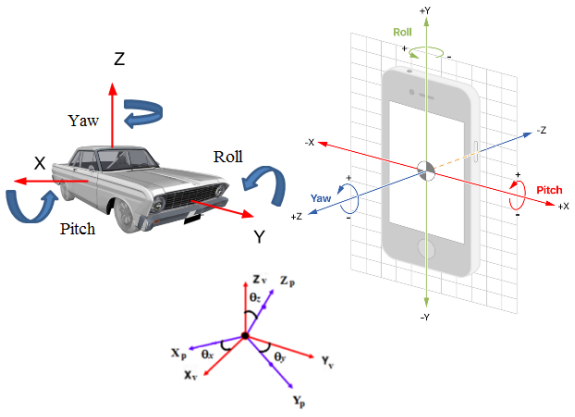


Figure 6: Smartphone and Vehicle Coordinate Systems

$$\begin{bmatrix} f_{xv} \\ f_{yv} \\ f_{zv} \end{bmatrix} = \mathbb{R} * \begin{bmatrix} f_{xp} \\ f_{yp} \\ f_{zp} \end{bmatrix} \quad (3)$$

where f_{iv} is the transformed measured signal to the vehicle reference, f_{ip} the measure signal in the phone reference and R is rotation matrix defined by:

$$\mathbb{R} = R(\theta_x) * R(\theta_y) * R(\theta_z) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & \sin\theta_x \\ 0 & -\sin\theta_x & \cos\theta_x \end{pmatrix} * \begin{pmatrix} \cos\theta_y & 0 & -\sin\theta_y \\ 0 & 1 & 0 \\ \sin\theta_y & 0 & \cos\theta_y \end{pmatrix} * \begin{pmatrix} \cos\theta_z & \sin\theta_z & 0 \\ -\sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4)$$

4. System Overview

The general architecture of the proposed system and its design methodology is shown in Figure 7). The system contains five main modules: raw data acquisition, pre-processing, driving parameters estimation, maneuvers detection unit and maneuvers identification and classification unit. The function of each module is briefly described as follows:

a. Android smartphones were used to capture vehicle's raw data through by using its IMUs, i.e. the accelerometer and gyroscope at a rate of 50 samples/second. Furthermore, smartphone's GPS data are also collected and used to provide vehicle's location and to correct speed estimation.
b. The pre-processing unit is designed to perform four functions namely; signals filtering, sensors error correction, transformation of sensors data to vehicle's coordinate system and finally estimates the driving parameters.
c. The manoeuvre detection unit is designed to detect and slice and executed manoeuvre in real-time by implementing the endpoint detection algorithm that identifies the starting and ending time.
d. The identification and classification unit is responsible to identify manoeuvre's type and its irregularity class. In this study the identification and the classification processes were performed separately by using a hybrid pattern matching and neural networks schemes.

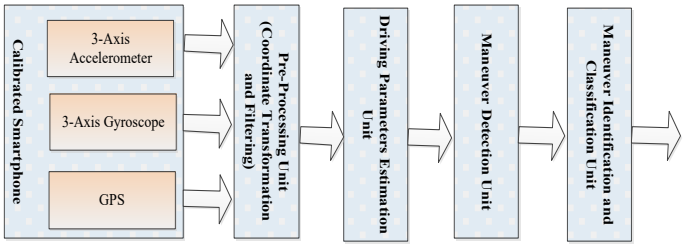


Figure 7: The Proposed System Architecture

5. Driving Parameters Estimation

A successful implementation of a driving maneuvers identification and classification system depends on the set of features and parameters that are used in the this process. In the literature there are two main approaches used to define the set of features and parameters namely, driver actions based approach and vehicle state based approach. In the first approach a number of essential driving signals can be collected from the vehicle's CAN-BUS such as vehicle speed, gas pedal pressure, brake pedal pressure, steering wheel angle, and acceleration. The second approach utilizes different approach by monitoring the vehicle's dynamic state in terms of longitudinal and lateral movements. These longitudinal and lateral movements can be detected and identified by using IMUs, GPS and the fusion of both the IMUs and the GPS. In this study it is believed that the description of vehicle' dynamical state in terms of actual driving parameters is more informative than other approaches and useful for future investigations. Therefore, the vehicle's acceleration, orientation and speed are used and represented as a time-series signals.

5.1. Longitudinal and Lateral Accelerations

The first and the simplest parameters that can be estimated directly are the longitudinal and the lateral acceleration components. The two components can be determined directly by subtracting the gravity the from the filtered accelerometer's Cartesian components, a_{xp}, a_{yp} and a_{zp}. The resulted singles are then transformed from the smartphone's to the vehicle's coordinate system by using Eqn. (4). The two components of the acceleration can be obtained by:

$$a_{lat} = a_{xv}\sin(\gamma) + a_{yv}\sin(\rho) \quad (5\text{-a})$$

$$a_{lon} = a_{xv}\cos(\gamma) + a_{yv}\cos(\rho) \quad (5\text{-b})$$

where, a_{lat} and a_{lon} are the lateral and longitudinal accelerations, a_{xv}, a_{yv} are the x and y components of the acceleration represented in the vehicle coordinate system and the angles γ and ρ are given by:

$$\gamma = \cos^{-1}\frac{a_{xv}}{\sqrt{a_{xv}^2+a_{yv}^2}} \quad \text{and } \rho = \cos^{-1}\frac{a_{yv}}{\sqrt{a_{xv}^2+a_{yv}^2}} \quad (6)$$

5.2. Vehicle's Orientation

The second estimated parameter is vehicle's heading or orientation. The attitude of a vehicle can be described by its angles of rotation along three axes as indicated in Figure 6). The attitude of any object can be modeled by different methods such as Direction Cosine Matrix, Quaternions and Euler angles. The first method has nine parameters to be estimated, hence it is difficult to implement. On the other hand, Quaternions has less parameters and it has advantages over Euler angles since it does not suffer from the gimbal lock problem, but still the Euler angles method is used in this study because it is more comprehensible and easier to decompose the rotations into separate degrees of freedom. Furthermore, since the smartphone is held in a fixed position inside the vehicle then the gimbal lock problem will not appear.

Euler angles can be found directly by using the measured gyroscope's data and integrate them with respect to time [45]. The accelerometer can also be used to estimate Euler angles, where simple trigonometric projections can be utilized to obtain both the Roll and the Pitch angles. Nevertheless, both methods could results in low prediction accuracy due to the stochastic bias variation in gyroscopes and the noisy accelerometer signals that results in an accumulated drift error [45]. To overcome the limitations of the direct approaches, sensor fusion techniques are developed to join all sensors' data to compensate the deficiency of each method. There are two filtering techniques namely; Kalman and complementary filters are widely used for sensor fusion. Because

of its simplicity and reliability, the digital complementary filter is utilized to estimate of the orientation. The filter uses a low-pass filter for the accelerometer's data to remove the high frequency noise and the dynamic bias, and utilizes a high-pass filter to suppress the drift from the gyroscope's data. Figure 8) shows a simple block diagram illustrating the implementation of a digital complementary filter. The filter can determine the angles using the following equation:

$$\theta_i = \alpha\left(\theta_{i-1} + \Delta\theta_{gi}\right) + (1-\alpha)\theta_{a\,i} \quad (7)$$

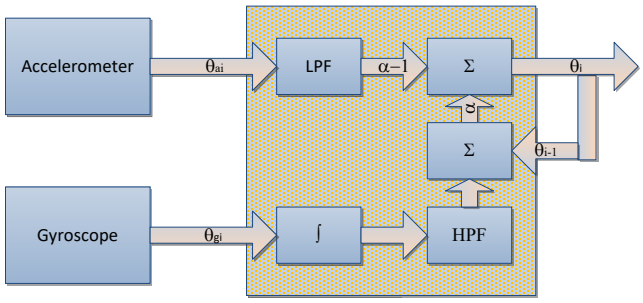


Figure 8: The Implementation of Digital Complementary Filter

In Eqn. (7), θ_i represent any of the estimated angles (Roll, Pitch or Yaw) and θ_{i-1} is the formerly determined angles, while the subscripts *g* and *a* designate the sensor type. The empirical constant α is selected to set the time constant of the error-signal calculations and stabilize the rate-sensor angle calculation.

5.3. Vehicle's Speed

Based on the kinematic equations, vehicle's speed can be found directly from the smartphone accelerometer data by integrating the corresponding acceleration component, longitudinal or lateral, as follows:

$$v_f(t) = v_i(t) + \int_0^t a(t)dt \quad (8)$$

where $v_f(t)$ is final speed, $v_i(t)$ is the initial speed and a(t) is vehicle's acceleration. The system is a discrete one, hence a sequence of the acceleration data is collected and equation (8) can be rewritten as:

$$v_f(n) = v_i(0) + \sum_{i=0}^{nT} \frac{1}{n} a(i) \quad (9)$$

where T is the sampling rate, n is the present nth sample and a(i) is the ith accelerometer value.

As it can be noted the basic idea of estimating the speed is very simple, and it is expected that with rigorous calibration and noise filtering of the raw data will improve the estimation process and a reasonable accuracy would be achieved. Nonetheless, speed estimation by the direct integration of accelerometer data suffer from significant deviations and unrealistic drifts because the accelerometer data is not pure and still contains some noise and other effects. This can be clearly shown in Figure 9), where the variation of the speed obtained from the GPS and those obtained by using the direct integration are shown. The doted green curve shows the estimated speed when the accelerometer data are not calibrated neither filtered while the dashed orange curve shows the trend in the estimated velocity for the preprocessed accelerometer data. As it can be seen, the error is accumulated over time and will obviously cause serious errors in the estimation of the speed. Therefore, a correction method is required to instantaneously correct the speed estimation by eliminating the accumulative error.

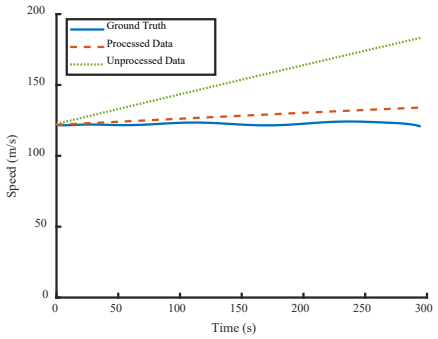


Figure 9: Comparison between Measured and Estimated Speed

In the literature there are mainly three methods that can be used to correct the accumulated error in numerical integration of the accelerometer data namely; high pass correction filter, Kalman filter and sensor fusion techniques. In this study, a simple but effective sensor fusion technique has been adopted and utilized to provide accurate velocity estimation [46].The basic idea of the technique is based on fusing the high frequency estimated speed with a low frequency ground truth speed obtained through the GPS. To illustrate the implementation of the vehicle speed estimator, consider Figure 10) that shows two equally spaced time intervals at which correction process take place at these instants. Suppose that the estimated speed and that measured by the GPS at time instant $t = T_A$ are v_{AE} and v_{AG} respectively, while those at the instant $t = T_B$ are v_{BE} and v_{BG}. The error of the speed at reference points then are given by:

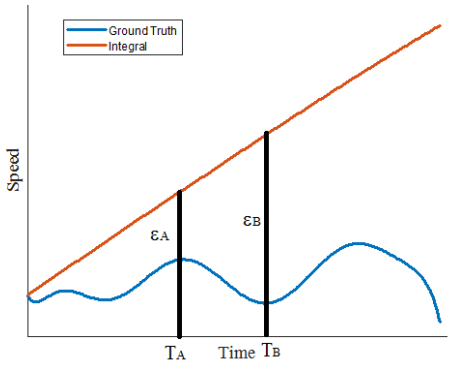


Figure 10: Vehicle Speed Correction Process

$$\varepsilon_A = v_{AG} - v_{AE} \quad \text{and} \quad \varepsilon_B = v_{BG} - v_{BE} \quad (10)$$

If we can consider that the rate of change in the accelerometer error is constant, i.e. that the accelerometer error varies linearly in this short interval, then this rate change can be expressed as follows:

$$\Delta\bar{a} = \frac{\varepsilon_B - \varepsilon_A}{T_B - T_A} \quad (11)$$

From Eqn. (11) the accumulative error at any instant (t) can be given by:

$$\emptyset = \int_{T_A}^{t} \Delta\bar{a}\ dt = \Delta\bar{a}kT_s \quad (12)$$

The constant k is the number of samples passed from the time instant T_A to t, and T_s is the sampling rate of the accelerometer. By knowing the accumulative error at any instant, then the estimated speed at instant (t) can be given by:

$$v_{tE} = v_{AG} + \sum_{i=T_A}^{t} \frac{1}{k} a(i) - \emptyset \quad (13)$$

Further enhancement been has been achieved by using exponential moving average correction to the error rate change as follows:

$$\Delta\bar{a}(i) = \alpha\Delta\bar{a}(i-1) + (1-\alpha)\frac{\varepsilon_B - \varepsilon_A}{T_B - T_A} \quad (14)$$

Figure 11 shows the estimated speed and the speed obtained through the GPS. It can be clearly seen that there is some small error between the estimation and the ground truth, which indicates that the estimator has excellent accuracy.

6. Maneuvers Detection

The main task that is required to be performed by any driving monitoring system is to reveal driving maneuver with reasonable level of reliability and accuracy. Table (1) presents a list of the maneuvers that are detectable by the proposed system. The system should be able to process the measured signals or the estimated parameters and attempts to detect the maneuver. The system detects the variation in the vehicle's longitudinal and lateral states by continuously monitoring the variation in the speed and orientation at a rate of 10Hz. Once a maneuvers is detected, the system can then proceed to recognize and classify it. There are different methods that can be used to detect the starting ending times of a maneuver [47-48]. In this paper the short-term energy endpoint detection algorithm with sliding window has been adopted, where the speed and orientation are continuously separated to event and non-event segments.

Table 1: Maneuvers Classes

1-	Acceleration straight road segment	2-	Acceleration curved road segment
3-	Braking straight road segment	4-	Braking curved road segment
5-	Left lane change straight road segment	6-	Left lane change curved road segment
7-	Right lane change straight road segment	8-	Right lane change curved road segment
9-	Merging into highway	10-	Exit from highway

The detection module consists of three key phases. In the first phase the signal is divided into non-overlapping windows of a 100ms period. The short-term energy is first computed for the data sequence in each window. For an infinite sequence of a discrete signal the energy is defined by:

$$\hat{x}[m] = x[m]\Lambda[n-m], n-N+1 \le m \le n \quad (15)$$

where Λ is a window function given by:

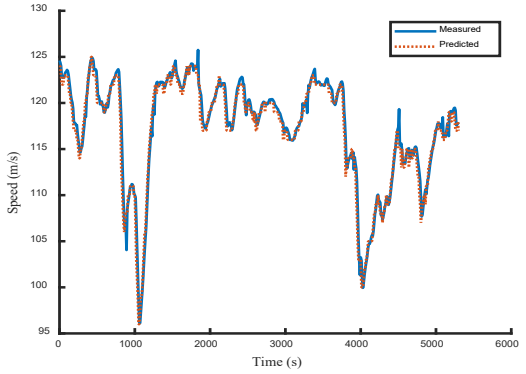
$$\Lambda[n-m] = \begin{cases} 1 & 0 \le n \le N-1 \\ 0 & \text{Otherwise} \end{cases} \quad (16)$$

The energy contained in this short interval then can be computed by:

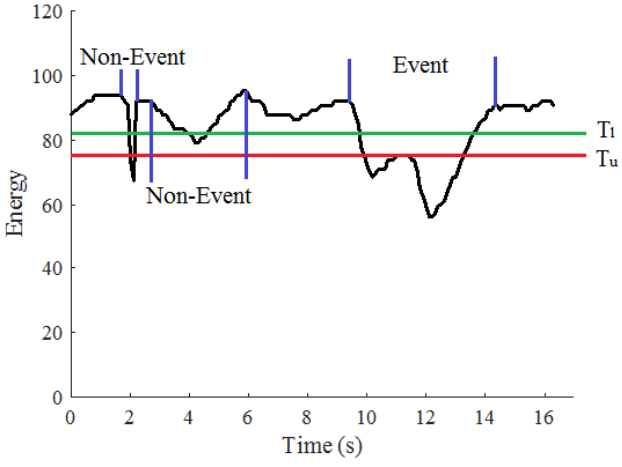
$$E_w = \sum_{m=n-N+1}^{n} (x[m]\Lambda[n-m])^2 \quad (17)$$

The value of the energy computed by Eqn. (17) is then compared to a threshold value, see Figure 12, and if the computed energy is lower than a specific threshold, then this frame is rejected and considered as a non-event segment. Otherwise, if the computed energy is higher than a lower threshold level T_l then the second stage will be executed.

Once the second stage is triggered, the starting time of the detected maneuvers is logged and the short-term energy is computed for a rectangular sliding window function as in Eqn. (17). Depending on the computed short-term energy in the second stage one of the following possibilities could be activated:



(a)



(b)

Figure 11: (a) Estimated and Measured Speed (b) Maneuver Detection using Short-Term Energy

- If the computed short-term energy is maintained to be less than the upper threshold T_u for less than 2 seconds or the energy fall back to a value lower than T_l, then this segment is treated as a non-event segment and the execution will resume from stage 1.

- If the computed short-term energy becomes higher than the upper threshold T_u for more than one second, then this segment will be treated as an event segment. In this case the system will continue computes the energy and compare it with the threshold until it drops back to the lower than T_l again. It is should be noted that if the energy falls down to be lower than T_u before the end of the first second then this segment will be regarded as a non-event segment, and the execution will start again from stage

7. Driving Maneuvers Identification and Classification

The Identification process is used to describe a driving maneuvers by a well-defined unique type. On the other hand the classification process evaluates the driving maneuvers to a certain class according to given criteria. In general there are two main approaches used for maneuver identification and classification, the first is based on the selection of a number of distinctive features and uses a machine learning technique, and the second approach is based on pattern matching algorithms. In this study 60% of the first dataset, described in section three, were used to study the time variation of each estimated driving parameters when a certain maneuvers has been executed. Figure 12) shows a sample of parameters' variations for a brake maneuvers. Intensive off-line observations and investigation were made to study and analyze the behavior of the parameters variation for each maneuvers type and to infer a general trend in their variations. It has been clearly noticed that the variation of the parameters follow common variation patterns and this fact agrees with what been presented in the literature. Figure 13) shows ideal approximated patterns that could be generated during any driving maneuver. As it can be notices that the driving parameters have limited number of patterns, for example the longitudinal acceleration, the lateral acceleration and the speed have three different patterns each, while the heading angle has five different patterns. Based on these facts, the identification and classification processes are performed separately where the identification process is performed by using a hybrid pattern matching neural network unit and the classification processed is performed by using another neural network.

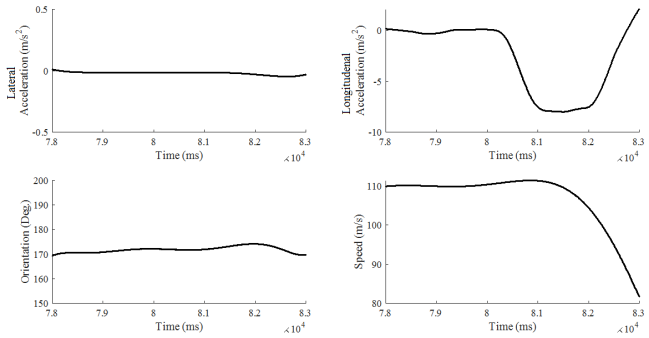


Figure 12: Driving Parameters Variation during a Break Manoeuvre

Figure 14) shows the basic units of the identification and classification module. The module consist of three units, a pattern matching units, to identify the patterns of estimated parameters, the second unit consist of a single feed-forward neural network that will identify the type of the maneuvers based on the combination of the identified patterns. Finally the third unit, which is responsible for the manoeuver classification, is performed by another neural network.

The pattern matching unit utilizes four DTW units, each one is designed to identify the pattern type for one of the parameters. The DTW is a pattern matching technique uses the discrete dynamic programming model to compute the alignment between two discrete time signals or sequences by assessing the similarity between them, regardless of the time/speed synchronization and distortion effects due to dynamic spatiotemporal differences.

Event Type	Longitudinal Acceleration	Lateral Acceleration	Heading Angle	Speed
Acceleration				
Braking				
Left- Lane Change				
Right- Lane Change				
Entry				
Exit				

Figure 13: Driving Parameters Variation for Different Maneuver

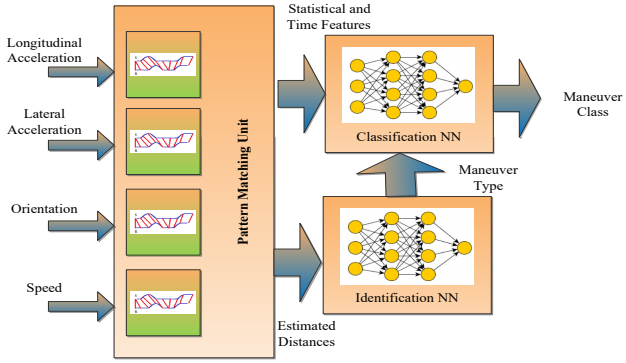


Figure 14: Structure of Identification and Classification Unit

To illustrate how the DTW algorithm works, consider two discrete signals, $R = \{r_1, r_2..., r_m\}$ which is the reference signal and an unknown signal $S = \{s_1, s2..., s_n\}$ that need to be identified, where m and n represent the number of samples in R and S, respectively. In Figure 15) the two time series that are highly similar, are shown. To align these two sequences an $(m \times n)$ local distance matrix is constructed by populating its elements with the Euclidean distances as follows:

$$d(i,j) = (R(i) - S(j))^2 \quad (18)$$

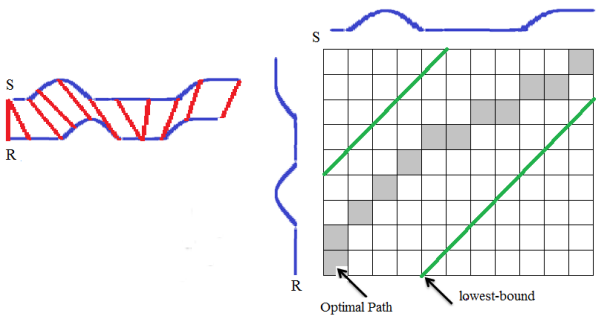


Figure 15: A DWT Matrix with a Warp Path

The correlation between the patterns of these two sequences is expressed by the cumulative distance measured over the optimal path within the distance matrix. For this purpose a warping matrix (D) is constructed and filled by searching for the minimum distance between the two sequences. This minimum distance can be found by computing the total distance D(i, j) for each entry (i, j) as:

$$D(i,j) = \begin{cases} d(1,1), & \text{if } i = j = 1 \\ d(i,j) + \min\{D(i-1,j), D(i,j-1), D(i-1,j-1)\}, & \text{otherwisw} \end{cases} \quad (19)$$

In the final stage the optimal warping path (P), as shown in Figure 15), and the DTW distance are to be computed. The Warping path is a set of adjacent matrix elements that detect the similarity between R and S and it represents the minimum distance between R and S. The matrix (P) contains K elements which range in value as:

$$\max(|R|, |S|) \leq K < (|R|, |S|) \quad (20)$$

Finally the optimal path Po is the one that minimizes the warping cost given by:

$$DTE(R,S) = \min\left\{\frac{1}{K}\sqrt{\sum_{k=1}^{K} P_k}\right. \quad (21)$$

To reduce the complexity of the DTW technique and thus speeding up the execution time, not all the warping paths are evaluated. To reduce the number of paths considered during the matching process four constraints are usually applied namely; monotonicity, continuity, boundary and windowing. To prevent the warping path from rolling back on itself, the monotonicity constrain is imposed to assure that any two adjacent elements, $p_k = (p_i, p_j)$ and $p_{k-1} = (\hat{p}_i, \hat{p}_j)$ will follow the inequalities $(p_i - \hat{p}_i) \geq 0$ and $(p_j - \hat{p}_j) \geq 0$. On the other hand, to guarantee a one-step advancement in the warping path, the continuity constrain is used to restrict the movement of the point (i, j) to the next point must be (i+1, j+1), (i+1, j), or (i, j+1). Finally, to assure that the warping path contains all points of both sequences, the boundary constrain is used to force the warping path to start from the top left corner and ends at the bottom right corner. Finally the windowing will prevent pathological mappings between two sequences by restricting one point in the first sequence to be mapped to a limited number of points in the other sequence; hence the path is likely to be around the diagonal without deviating much from it [49].

In order to identify the types of the maneuver by the DWT technique, a set of templates for each maneuvers are required as reference signals. To find a template signals for different set of sequences, one can consider that template sequence could be the longest common sequence, or the medoid sequence, or the average sequence. The selection of the reference pattern for each specific maneuvers class is not a simple task because the set of sequences collected for each maneuvers class have different time durations and amplitudes. In this study, the sequence with the minimum average distance from all other sequences for a specific class is selected as the reference template. The details of the averaging strategy are given in [50].

When dealing with single 1-D signal identification, the DTW technique will directly identify the test signal. Nevertheless, when it is required to identify a maneuvers or an action by using multiple measured signals, the DTW won't be able to identify its type directly. This is because each signal will need a special DTW unit to identify the pattern of the signal only. In order to predict the type of the maneuvers, the combination of the DTW distances were used as features to train a neural network that will be used for the final identification of the maneuvers.

Statistical and time metrics have been utilized as features to classify maneuvers' irregularities. In general maneuvers are categorized into one of three main types; Hard, Normal and Light. By studying all the possible patterns for all the classes, see Figure 13), it was found that there are seven distinctive patterns as shown in Figure 16). For each pattern time metrics and statistical values such as max, min, mean, standard of deviation and the variance are used as classification features. The segmented maneuvers collected in the first dataset were represented by a vector that contains the statistical and the time metrics for each driving parameter and the classification label. As it has been mentioned the classification labels are obtained from classification made by the experts. It is should be noted that the percentage of the rejected samples was less than 10%.

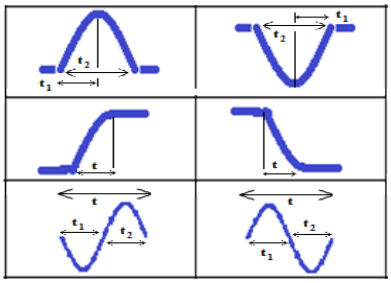


Figure 16: Time features for the common signals

8. System Validation

This section presents an experimental evaluation for the three main components of the proposed system namely, detection, the identification and classification units. To test the performance of the system, two different datasets have been used. In the first phase the first dataset described in section three has been used to train and conduct a preliminary testing of the system. In this phase a single smartphone (Samsung Galaxy S5) have been calibrated and used in all the tests. As been mentioned ten drivers with different car models were asked to conduct each of the basic maneuvers, listed in table (1), at least five times with different severity levels from the Light to the Hard level. From this dataset 1500 maneuvers were collected and segmented then each was combined with two videos, the first used to record the surrounding traffic state and the second used to record the driver's state, i.e. hands and feet. To obtain the ground truth classification of each maneuvers, experts were consulted to evaluate the abnormality level. Each assessor assigned abnormality level scores for the entire recorded maneuvers and the final score for every recorded maneuvers was calculated using the inter-rater reliability method.

For this initial dataset 60% of the collected samples were utilized to extract the information and the features that were used in the training phase. The information and features includes the computation of the DTW reference templates, the DTW's distances and the statistical and time metrics. The DTW reference

templates have been stored and used to measure a vector that includes the distances of the warping for each recorded driving parameter of a specific tested sample. These vectors are used to train the first neural network which is used to identify the type of the maneuvers, while another set of vectors that contain the statistical and time metrics features used to train the classification neural network.

The remaining 40% of the first dataset have been used to validate the performance of the system. Figure 17 shows results obtained when the 40% testing samples fed to the identification module. It can be seen clearly that the identification rate is very high and reached a 100% successful rate for some of the maneuver and the minimum identification rate scored is 85%, and this is obvious since the identification process combines all the similarities for the four parameters. Figure 18, shows the confusion matrix classification for the 40% testing samples. Again it can be noticed that the classification rate is very high nearly 95% for some cases. This high identification and classification rates are expected since the two processes have been separated.

In the second phase of this study, the author developed a complete Android-based APP that detects, identify and classify the highway driving maneuver in real-time. The architecture of the system is described in section five. The first test for the system was to check its ability to detect the starting and end of a driving maneuvers. The complete trips performed by the ten drivers, from which the first dataset was collected, were used just to detect the driving maneuver. Figure 19, shows the detection of maneuvers for a sample trip. It should be noted that the first 1500 maneuver collected in the first data set were used to determine the lower and upper thresholds for short-term energy for both the vehicle's speed and orientation. The results obtained by the system were compared with those already registered manually by the author. The system showed a high detection rate where more than 96% of the manually registered maneuver were detected successfully by the detection unit.

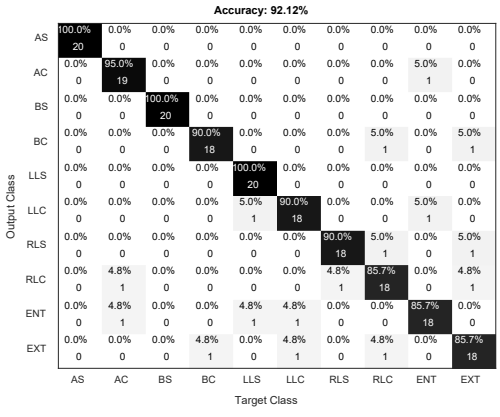


Figure 17: Confusion Matrix for Maneuvers Identification

The developed application allows two types of analysis namely, the real-time analysis and off-line analysis. The graphical interface of the developed application is designed to achieve user friendly interaction with the driver. All the driving maneuver are presented instantly to the driver using both audio and visual indicators. Each maneuvers type is represented by a special icon that it will be lighted up; along with a human voice that utters the name of the maneuvers when it is detected by the application. Furthermore, the icon's color will be changed in accordance to the class of the maneuvers. This attractive and user friendly interface helped in measuring the performance and the reliability of the system in the second stage of the study. As been mentioned in section three a second dataset was collected from a naturalistic driving performed by 15 drivers who are daily travel a distance of 140km back and forth. We have asked driver's companions to check the alerts produced by the application and check them. Data recoded by the companions and those stored in the SD card are then analyzed. Table (2), presents a comparison between the number of maneuver performed by each driver and those detected by the system. Note that the first number in any cell represents the performed maneuver, while the second number represents the detected maneuver. Figure 20 and 21 show the confusion matrix for the maneuvers' identification and classification respectively.

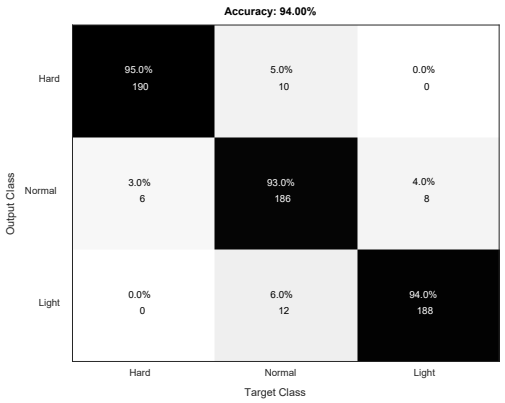


Figure 18: Confusion Matrix for Maneuvers Classification

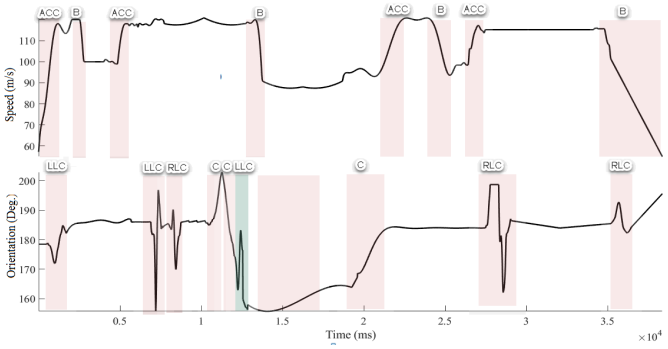


Figure 19: Driving Maneuvers Detection in a Complete Driving Trip

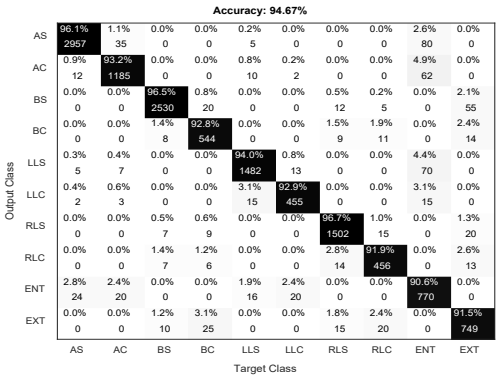


Figure 20: Confusion Matrix for Maneuvers Identification (Second Dataset)

Table 2: Comparison between Performed and Detected Maneuver for each Driver

Driver	ACC/S	ACC/C	B/S	B/C	LLC/S	LLC/C	RLC/S	37 : 34	Ent.	Exit
1	221: 210	95 : 87	184 : 176	48 : 42	112:109	37 : 33	114 : 111	40 : 36	60 : 57	60 : 56
2	211: 207	86 : 81	159 : 156	41 : 36	105:99	30 : 26	104 : 99	35 : 32	58 : 52	59 : 54
3	199: 189	85 : 81	154 : 142	43 : 38	108:100	41 : 36	107 : 102	39 : 36	59 : 57	59 : 57
4	230: 222	87 : 84	200 : 194	44 : 40	110:104	36 : 33	109 : 105	31 : 28	58 : 56	58 : 56
5	219: 213	93 : 86	194 : 186	46 : 42	114:109	40 : 36	111 : 106	34 : 29	60 : 56	60 : 57
6	216: 209	91 : 85	188 : 181	45 : 40	107:104	39 : 34	106 : 101	41 : 35	59 : 56	60 : 56
7	220: 212	89 : 83	166 : 162	41 : 36	109:108	38 : 32	110 : 104	37 : 34	59 : 55	59 : 55
8	223: 214	90 : 85	172 : 168	42 : 37	111:107	34 : 30	113 : 108	40 : 36	58 : 57	59 : 57
9	218: 210	94 : 86	190 : 187	47 : 40	112:110	37 : 34	105 : 100	35 : 32	55 : 51	55 : 52
10	210: 202	93 : 88	185 : 179	48 : 44	108:102	40 : 36	108 : 104	39 : 36	56 : 53	57 : 54
11	200: 193	92 : 84	184 : 180	43 : 39	115: 110	35 : 32	112 : 107	31 : 28	58 : 55	57 : 53
12	207: 201	90 : 84	180 : 173	44 : 39	104:99	39 : 36	110 : 106	34 : 29	60 : 56	60 : 56
13	215: 208	89 : 85	176 : 172	40 : 35	106:102	31 : 28	104 : 99	41 : 35	57 : 55	57 : 54
14	201: 196	88 : 83	185 :180	42 : 37	110:106	34 : 29	101 : 97	37 : 34	59 : 56	60 : 56
15	196: 191	95 : 89	190 : 186	45 : 41	114:108	41 : 35	110 : 104	40 : 36	57 : 52	56 : 51

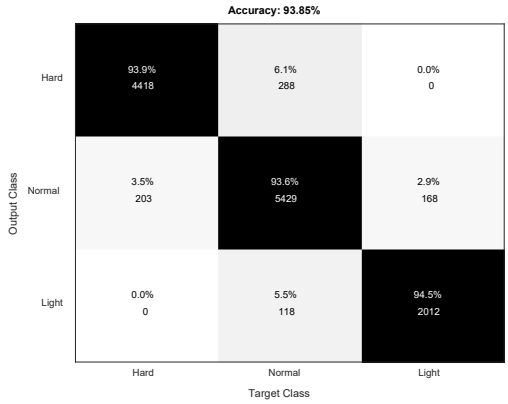


Figure 21: Confusion Matrix for Maneuvers Classification (Second Dataset)

9. Conclusion

This paper proposes a novel real-time smartphone-based highway driving maneuvers identification and classification system. The system collects vehicle's longitudinal and lateral movements using smartphone's IMU sensors and vehicle's location by using the GPS. All the smartphones used in this investigation were calibrated the collected smartphone's raw data were filtered and smoothed by using the statistical LOESS filter. Driving parameters namely vehicle's acceleration, orientation and speed were used in the identification and classification of the driving maneuver. Sensor fusion techniques have been used to estimate vehicle's orientation and speed.

It was noted that driving parameters for each event class have common patterns, thus the DTW technique combined with a feed-forward neural network are used for the identification. In the classification process a second neural network was employed, and it is trained to classify a maneuver classes. Two datasets were collected through a number of experiments conducted on highway routes. Results obtained in this study show an excellent detection, identification and classification rates were achieved.

References

[1] S.C. Davis and Robert G. Boundy, Transportation Energy Data Book, Ed. 38, Oak Ridge National Lab TN-United States, 2020.

[2] D. Searsona , G. Pontea , T Paul Hutchinsona , Robert Andersona and Mary Lydona, "Emerging vehicle safety technologies and their potential benefits: Discussion of expert opinions" in 2015 1st Australasian Road Safety Conference (ARSC2015), 1-11, 2015,

[3] World Health Organization: "Global status report on road safety 2018", WHO: Geneva, Switzerland, 2018.

[4] A. Haghi, D. Ketabi, M. Ghanbari and H. Rajabi, "Assessment of Human Errors in Driving Accidents; Analysis of the Causes Based on Aberrant Behaviors", Life Science Journal, **11**(9), 414-420, 2014.

[5] M.Q. Khan and Sukhan Lee, "A Comprehensive Survey of Driving Monitoring and Assistance Systems", Sensors **19**(11), 2574-2606 ,2019, doi: 10.3390/s19112574

[6] C.M. Martinez ; M. Heucke ; Fei:Yue Wang ; Bo Gao ; Dongpu Cao, "Driving Style Recognition for Intelligent Vehicle Control and Advanced Driver Assistance: A Survey", IEEE Transactions on Intelligent Transportation Systemsm **19** (3) , 666–676, 2018, doi: 10.1109/TITS.2017.2706978.

[7] M. VivaldiniI; Silvio R. I. PiresI; Fernando Bernardi de SouzaII, "Improving logistics services through the technology used in fleet management", Journal of Information Systems and Technology Management, **9**(3), 541-562, 2012, doi: 10.4301/S1807-17752012000300006

[8] M.J. Prasad, S Arundathi, Nayana Anil, Harshikha, B. S. Kariyappa, "Automobile black box system for accident analysis", in 2014 International Conference on Advances in Electronics Computers and Communications (ICAECC), 1-5, 2014, doi: 10.1109/ICAECC.2014.7002430.

[9] T.K. Chan ; Cheng Siong Chin ; Hao Chen ; Xionghu Zhong, "A Comprehensive Review of Driver Behavior Analysis Utilizing Smartphones" IEEE Transactions on Intelligent Transportation Systems, 21(10), 4444 - 4475, 2020, doi: 10.1109/TITS.2019.2940481.

[10] M. S. N. Al-Din, "Highway Driving Events Identification and Classification using Smartphone," in the 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), 1-6, 2018, doi: 10.1109/ICCSDET.2018.8821090.

[11] R. Kridalukmana, H. Y. Lu and M. Naderpour, "An object oriented Bayesian network approach for unsafe driving maneuvers prevention system", in the 2017 12th International Conference on Intelligent Systems and Knowledge

Engineering (ISKE), 1-6, 2017, doi: 10.1109/ISKE.2017.8258758.
[12] I. Silva and JoséEugenio Naranjo, "A Systematic Methodology to Evaluate Prediction Models for Driving Style Classification", Sensors, **20**(6), 1692-1713, 2020, doi: /10.3390/s20061692
[13] H. Malik, G. S. Larue, A. Rakotonirainy and F. Maire, "Fuzzy Logic to Evaluate Driving Maneuvers: An Integrated Approach to Improve Training," IEEE Transactions on Intelligent Transportation Systems, **16**(4), 1728-1735, 2015, doi: 10.1109/TITS.2014.2371061.
[14] T. K. Chan, C. S. Chin, H. Chen and X. Zhong, "A Comprehensive Review of Driver Behavior Analysis Utilizing Smartphones," IEEE Transactions on Intelligent Transportation Systems, **21**(10). 1-32, 2019, doi: 10.1109/TITS.2019.2940481.
[15] N. AbuAli, "Advanced vehicular sensing of road artifacts and driver behavior," in the 2015 IEEE Symposium on Computers and Communication (ISCC), 45-49, 2015, doi: 10.1109/ISCC.2015.7405452.
[16] L. M. Bergasa, D. Almería, J. Almazán, J. J. Yebes and R. Arroyo, "DriveSafe: An app for alerting inattentive drivers and scoring driving behaviors," in the 2014 IEEE Intelligent Vehicles Symposium Proceedings, 240-245, 2014, pp. doi: 10.1109/IVS.2014.6856461.
[17] M. Fazeen, B. Gozick, R. Dantu, M. Bhukhiya and M. C. González, "Safe Driving Using Mobile Phones," IEEE Transactions on Intelligent Transportation Systems, **13**(3), 1462-1468, 2012, 1462:1468, doi: 10.1109/TITS.2012.2187640.
[18] Yang Zheng, A. Sathyanarayana and J. H. L. Hansen, "Threshold based decision-tree for automatic driving maneuver recognition using CAN-Bus signal," in the 2014 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), 834-2839, 2014, doi: 10.1109/ITSC.2014.6958144.
[19] T. Pholprasit, W. Choochaiwattana and C. Saiprasert, "A comparison of driving behaviour prediction algorithm using multi-sensory data on a smartphone," in the 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 1-6, 2015, doi: 10.1109/SNPD.2015.7176249.
[20] H.R. Eftekhari and Mehdi Ghatee, "Hybrid of discrete wavelet transform and adaptive neuro fuzzy inference system for overall driving behavior recognition", Transportation Research Part F, **58**, 782-796, doi: 10.1016/j.trf.2018.06.044.
[21] H.R. Eftekhari, M. Ghatee,: "A similarity:based neuro:fuzzy modelling for driving behavior recognition applying fusion of smartphone sensors", Journal of Intelligent Transportation Systems, **23**(1), 72-83, 2019, doi: 10.1080/15472450.2018.1506338.
[22] C. Arroyo, L. M. Bergasa and E. Romera, "Adaptive fuzzy classifier to detect driving events from the inertial sensors of a smartphone," in the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), 1896-1901, 2016, doi: 10.1109/ITSC.2016.7795863.
[23] A. Aljaafreh, N. Alshabatat and M. S. Najim Al:Din, "Driving style recognition using fuzzy logic," in the 2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES 2012), 460-463l, 2012, doi: 10.1109/ITSC.2014.6957822.
[24] M. Najim, Ahmad Aljaafreh, Nashat Albdour and Maen Saleh:, "Driving Styles Recognition Using Decomposed Fuzzy Logic System", International Journal of Electrical, Electronics and Computer Systems, **16**(1) 820-824, 2013.
[25] D.A. Johnson and M. M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in the 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), 1609-1615, 2011, doi: 10.1109/ITSC.2011.6083078.
[26] C. Saiprasert, Pholprasit, T. and Thajchayapong, S., "Detection of Driving Events using Sensory Data on Smartphone", International Journal of Intelligent Transportation Systems Research, **15**(1), 17-28, 2015, doi:10.1007/s13177-015-0116-5.
[27] G. Singh, D. Bansal and Sanjeev Sofat, "A smartphone based technique to monitor driving behavior using DTW and crowdsensing", Pervasive and Mobile Computing, **40**, 56-70, 2017, doi: 10.1016/j.pmcj.2017.06.003.
[28] P. Tkachenko, J. Zhou, Davide Gagliardi and Luigidel Re, "On:line Maneuver Identification in Highway Traffic Using Elastic Template Matching", IFAC:PapersOnLine, **51**(15), 557-562, 2018, doi: 10.1016/j.ifacol.2018.09.214.
[29] M. Wu; Zhang, S.; Dong, Y., "A Novel Model-Based Driving Behavior Recognition System Using Motion Sensors", Sensors, **16**(10), 1746-1769, 2016, doi: 10.3390/s16101746.
[30] C. Chen, 1 Xiaohua Zhao, "Driver's Eco-Driving Behavior Evaluation Modeling Based on Driving Events", Journal of Advanced Transportation, **2018**, 1-12, 2018, doi: 10.1155/2018/9530470.
[31] J. Yu, Z. Chen, Y. Zhu, Y. Chen, L. Kong and M. Li, "Fine-Grained Abnormal Driving Behaviors Detection and Identification with Smartphones," IEEE Transactions on Mobile Computing, 1**6**(8), 2198-2212, 2017, doi: 10.1109/TMC.2016.2618873.
[32] J. F. J´unior, E. Carvalho, B. V. Ferreira, C. de Souza, Y. Suhara, A. Pentland, and G. Pessin, "Driver behavior profiling: An investigation with different smartphone sensors and machine learning," PLoS one, 12(4), e0174959, 2017. doi:10.1371/journal.pone.0174959
[33] P. Brombacher, J. Masino, M. Frey and F. Gauterin, "Driving event detection and driving style classification using artificial neural networks," in the 2017 IEEE International Conference on Industrial Technology (ICIT), 997-1002, 2017, doi:10.1109/ICIT.2017.7915497..
[34] U. Fugiglando et al., "Driving Behavior Analysis through CAN Bus Data in an Uncontrolled Environment," IEEE Transactions on Intelligent Transportation Systems, **20**(2), 737-748, 2019, doi: 10.1109/TITS.2018.2836308.
[35] Manon Kok, Jeroen D. Hol and Thomas B. Schon, "Using Inertial Sensors for Position and Orientation Estimation", Foundations and Trends in Signal Processing: **11**(1-2), 1-153, 2017, doi:10.1561/2000000094.
[36] I. Aicardi , P. Dabove , A. Lingua and M. Piras, "Sensors Integration for Smartphone Navigation: Performance and Future Challenges", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, **XL**(3), 9-16, 2014 .
[37] I. P. Prikhodko, B. Bearss, C. Merritt, J. Bergeron and C. Blackmer, "Towards self-navigating cars using MEMS IMU: Challenges and opportunities," in the 2018 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL), 1-4, 2018, doi:10.1109/ISISS.2018.8358141.
[38] J. Almazán, L. M. Bergasa, J. J. Yebes, R. Barea and R. Arroyo, "Full auto:calibration of a smartphone on board a vehicle using IMU and GPS embedded sensors," in the 2013 IEEE Intelligent Vehicles Symposium (IV), 1374-1380, 2013, doi: 10.1109/IVS.2013.6629658.
[39] M. S. Najim Al:Din, "Calibration and Pre-Processing Techniques for a Smartphone:Based Driving Events Identification and Classification System," in the 2018 IEEE Electron Devices Kolkata Conference (EDKCON), 396-402, 2018, doi: 10.1109/EDKCON.2018.8770487.
[40] Munaf Salim Najim Al:Din; Atef Saleh Al:Mashakbeh, "Development of a highway driving events identification and classification using smartphone", International Journal of Nanoparticles, **12**(1-2), 152-173, 2020, doi;10.1504/IJNP.2020.106006
[41] T. Pan, C. Kuo and M. Hu, "A noise reduction method for IMU and its application on handwriting trajectory reconstruction," in the 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 1-6, 2016, doi: 10.1109/ICMEW.2016.7574685.
[42] M. Wu, S. Zhang and Y. Dong, "A novel model:based driving behavior recognition system using motion sensors", Sensors, **16**(10), 1746-1769, 2016, doi: 10.3390/s16101746 .
[43] Wendy L. Martinez and Angel R. Martinez, "Exploratory Data Analysis with MATLAB", Chapman & Hall/CRC Press, 2005.
[44] F Bruwer and M Booysen, "Vehicle Acceleration Estimation using Smartphone:Based Sensors", in the 2015 34th Annual Southern African Transport Conference (SATC 2015), 118-129, 2015, doi: 10.13140/RG.2.1.3244.4640 .
[45] Y. Huang, Y. Rizal and M. Ho, "Development of attitude and heading reference systems," in the 2015 International Automatic Control Conference (CACS), 13-18, 2015, doi: 10.1109/CACS.2015.7378358.
[46] T. Lei, A. A. Mohamed, and C. Claudel, "An imu-based traffic and road condition monitoring system," HardwareX, **4**, e00045, 2018, doi: 10.1016/j.ohx.2018.e00045
[47] Xu Dawei,Wu Bian, Zhao Jianwei et al., "A real time algorithm for voice activity detection in noisy environment", Computer Engineering and Application, **24**(1), 115-117, 2003.
[48] Liang-Sheng Huang and Chung-Ho Yang, "A novel approach to robust speech endpoint detection in car environments," in the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings 1751-1754, 2000, doi: 10.1109/ICASSP.2000.862091.
[49] B. Johnen and B. Kuhlenkoetter, "A Dynamic Time Warping algorithm for industrial robot motion analysis," in the 2016 Annual Conference on Information Science and Systems (CISS), 18-23, 2016, doi: 10.1109/CISS.2016.7460470.
[50] F. Petitjean, J. Inglada, and P. Gancarski, "A global averaging method for dynamic time warping, with applications to clustering," Pattern Recogn., **44**(3), 678–693, 2011, doi: 10.1016/j.patcog.2010.09.013.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 206-210 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

The Effect of E-Service Quality on Customer Satisfaction and Loyalty (Case Study at E-Marketplace XYZ in Indonesia)

Hanny Juwitasary*, Christian Christian, Edi Purnomo Putra, Hilman Baskara, Mohammad Wildan Firdaus

School of Information Systems, BINUS University, Jakarta, 11480, Indonesia

ARTICLE INFO

Article history:
Received: 16 June, 2020
Accepted: 25 September, 2020
Online: 08 November, 2020

Keywords:
Information Systems
E-Service Quality
Customer Satisfaction
Customer Loyalty

ABSTRACT

Online transactions make it easy for people to get products or sell the products through online applications. The success and failure of online sales depends on how satisfied and loyal the customer is to the service of the product or business which can certainly influence and increase competition between the online sales industry. Based on that background we want to measure how much influence the quality of services provided by the Online Marketplace to customers has an impact on customer satisfaction and customer loyalty. The measurement method used is to use E-Service Quality, to determine customer satisfaction (Customer Satisfaction) and determine the level of customer loyalty (Customer Loyalty). This research uses a quantitative approach and uses random sampling techniques. Data collection using a questionnaire with data samples totaling 102 respondents. The results of this study indicate that there is a significant relationship between E-Service Quality, Customer Satisfaction and Customer Loyalty to customers that increase customer online shopping transactions so that companies are able to compete with other companies.

1. Introduction

By looking at the growing number of internet users in Indonesia[1], this has an impact on people's behavior in Indonesia towards buying and selling interests [2]. Based on data [3], the number of smartphone users which continues to increase from time to time seems to be in line with the increasing number of online store users from websites and mobile apps. The average amount of Indonesian consumer spending when shopping online from all segments of the shopping category can reach US $ 36 (around Rp. 481 thousand). This value also makes Indonesia occupy the position of second size basketball in Southeast Asia, far behind that of developed countries in Singapore which reached US $ 91 or around Rp1.3 Million [4]. Based on the data and according to experts the development of e-marketplaces in Indonesia is very fast [1], including e-marketplaces owned by XYZ. In 2018 it was stated that five e-marketplace applications that were frequently downloaded consisted of five such marketplaces. Dominating the top ranking is XYZ's e-marketplace. In this study we want increases customer online shopping transactions in the e-marketplaces owned by XYZ by knowing the indicators that influence satisfaction and loyalty can increase competition between the online selling industry so that companies can still compete with other companies.

*Corresponding Author: Hanny Juwitasary, hjuwitasary@binus.edu

2. Related Works

Service quality is the key to the success of a business [5], in addition to the quality of service that makes a key factor in the success of a business, loyalty also makes one a key factor in the success of a business [6].

In the previous study conduct by Tianxiang Sheng, the researcher testing e-service quality model from four dimensions, with the customers' satisfaction as the main variable, the study has analyzed the effects of four dimensions of service quality on customers' satisfaction and loyalty where all hypotheses are accepted. Other study by Akroush used SERVQUAL to analyze the effect of service quality on customer satisfaction and loyalty [7]. The result sho positive and significant influence of tangible, assurance, empathy, and reliability on consumer satisfaction, and customer satisfaction has a positive and significant effect on customer loyalty. customer satisfaction is the customer's perspective based on expectations and then the shopping experience thereafter [8], other said customer satisfaction is everyone's awareness that is differents. This feeling is obtained by comparing the expected service and the service obtained [9]. Customer loyalty can also be defined as the relationship that the customer maintains with the seller after making the first transaction [10].

The success and failure of online sales depends on how satisfied the customer[11], the relationship between satisfaction and loyalty is almost as intuitive[12]. In more recent research found satisfaction leads to loyalty[13], then knowing the indicators that influence satisfaction and loyalty can increase competition between the online selling industry[14]. By looking at these developments the author wants to conduct research with the e-service quality scale from Parasuraman's E-S-QUAL[15] scale in terms of the basic service on user satisfaction on XYZ e-marketplace applications that have an impact on customer loyalty [16][17].

3. Research Methods

Research will be conducted by considering the following matters: questionnaire, online searching, and document Study[1]. We used quantitative research for data processing, quantitative research is systematic scientific research on parts and phenomena and their relationships. The purpose of quantitative research is to develop and use mathematical models, theories and hypotheses related to natural phenomena. We have distributing questionnaires that have been filled out by respondents through Google Form which will later be processed and analyzed, we distributed the questionaire with simple random sampling method[1].

This research uses E-SQUAL MODEL[2], where the statement contained in the questionnaire is a description of each variable such as, Efficiency[15], Requirement Fulfillment, Accessibility, Privacy[3] that influence Satisfaction [4] and Loyalty[5]. The following are research models we can use:

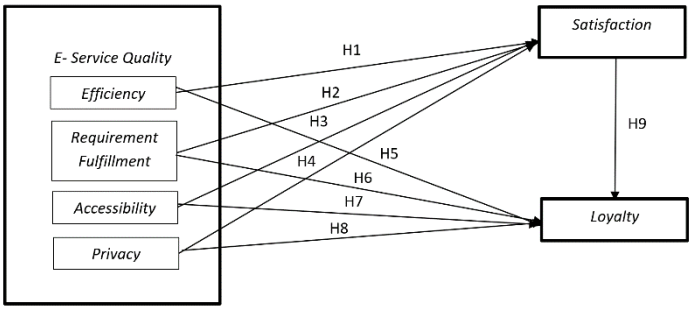


Figure 1: Research Model
Source: (T.Sheng, C.Liu [18])

Table 1: Demographic
Source: demographic from questionaire

Variable	N	Percentange
Gender		
Male	38	37%
Female	66	63%
Age		
11-15 years	2	2%
16-20 years	20	19%
21-25 years	55	53%
26-30 years	11	11%
31-35 years	3	3%
>= 36 years	13	13%
Domicile		
DKI Jakarta	25	24%
Banten	61	59%
Jawa Barat	11	11%
Jawa Tengah	2	2%
Jawa Timur	1	1%
Bali	3	3%
NTB	1	1%
Frequency Visit Online Marketplace		
Once in a week	78	75%
2-3 times in a week	26	25%

3.1. Validity and Reliability Test

By using a valid and reliable instrument, the research results are expected to be valid and reliable as well. If the instrument is not tested for validity and reliability, then the data collected with the instrument is doubtful [19]. A valid instrument means that the instrument can be used to measure what should be measured and can display what must be displayed [19].

Table 2: Validity and Reliability Test Results
Source: Questionnaire Testing Results with SPSS

Variable	Pernyataan	CI-TC	*Cronbach's Alpha*
E-Service Quality	EF1	0.768	0.949
	EF2	0.697	0.950
	EF3	0.558	0.951
	RF1	0.696	0.950
	RF2	0.549	0.951
	RF3	0.557	0.951
	AC1	0.649	0.950
	AC2	0.682	0.950
	P1	0.721	0.949
	P2	0.575	0.951
Customer Loyalty	CL1	0.538	0.951
	CL2	0.372	0.953
	CL3	0.679	0.950
Customer Loyalty	CL4	0.648	0.950
	CL5	0.638	0.950
	CS1	0.760	0.949
	CS2	0.623	0.950
	CS3	0.737	0.949
Customer Satisfaction	CS4	0.584	0.951
	CS5	0.540	0.951
	CS6	0.771	0.949
	CS7	0.645	0.950
	CS8	0.676	0.950
	CS9	0.710	0.950
	CS10	0.676	0.950
	CS11	0.719	0.949
	CS12	0.766	0.949
	CS13	0.790	0.949
	CS14	0.541	0.951
	CS15	0.604	0.951

Based on the results of the validity and reliability test in Table 2, researchers omitted 2 respondents in the validity test using SPSS with 2 respondents coming from the Banten region because the number of respondents in the Banten area was quite large so the sample of respondents the researchers used became 102. Those pieces the statement proposed for this study has an r table of 0.195 with a significant level of 5%, while the CI-TC of all statements are judged to be greater than the r table. The Cronbach's Alpha

value of all statements is above 0.6. It can be concluded that all these statements are valid and reliable as research instruments.

3.2. Convergent Validity Testing

According to [20] convergent validity is the extent to which measures are positively correlated with alternative steps of the same construct. The size loader is said to be high if the value of the collection must be 0.708 or higher.

Table 3: Convergent Validity Testing Result
Source: Questionnaire Testing Results with SPSS

	Accessibility	Efficiency	Loyalty	Privacy	Requirement Fulfillment	Satisfaction
AC1	0.867					
AC2	0.900					
CL1			0.635			
CL2			0.582			
CL3			0.828			
CL4			0.813			
CL5			0.817			
CS1						0.797
CS10						0.696
CS11						0.723
CS12						0.787
CS13						0.814
CS14						0.557
CS15						0.614
CS2						0.709
CS3						0.775
CS4						0.580
CS5						0.538
CS6						0.782
CS7						0.667
CS8						0.683
CS9						0.736
EF1		0.868				
EF2		0.905				
EF3		0.673				
P1				0.942		
P2				0.904		
RF1					0.893	
RF2					0.788	
RF3					0.751	

From Table 2 it can be seen that there are ten indicators that have values less than 0.708, namely indicators CL1 (0.635), CL2 (0.582), CS10 (0.696), CS14 (0.557), CS15 (0.614), CS4 (0.580), CS5 (0.538), CS7 (0.667), CS8 (0.683), and EF3 (0.673).

All indicators above refer to opinions [20], each indicator with a value between 0.40 to 0.70 does not have to be deleted, except when deleted, the indicator can increase the value of composite reliability.

3.3. Testing Discriminant Validity

Discriminant Validity shows the extent to which a construct is completely different from other constructs. Discriminant Validity implies that the construct is unique and can capture events that are not represented by other constructs in the model[21]. Cross Loadings is the first approach to assess the validity of discriminant indicators. Specifically, the external load of the indicators on the related construct must be greater than the other constructs [20].

Table 4: Testing Discriminant Validity Result
Source: Questionnaire Testing Results with SPSS

	Accessibility	Efficiency	Loyalty	Privacy	Requirement Fulfillment	Satisfaction
AC1	0.867	0.540	0.401	0.478	0.535	0.579
AC2	0.900	0.646	0.551	0.451	0.521	0.585
EF1	0.663	0.868	0.606	0.396	0.573	0.698
EF2	0.537	0.905	0.599	0.373	0.534	0.610
EF3	0.437	0.673	0.322	0.329	0.557	0.512
CL1	0.388	0.383	0.635	0.353	0.257	0.479
CL2	0.281	0.275	0.582	0.128	0.131	0.304
CL3	0.515	0.552	0.828	0.467	0.409	0.610
CL4	0.392	0.568	0.813	0.377	0.525	0.571
CL5	0.411	0.523	0.817	0.353	0.406	0.586
P1	0.517	0.442	0.482	0.942	0.451	0.724
P2	0.444	0.375	0.383	0.904	0.315	0.566
RF1	0.542	0.621	0.500	0.362	0.893	0.634
RF2	0.485	0.523	0.386	0.262	0.788	0.462
RF3	0.419	0.469	0.285	0.415	0.751	0.493
CS1	0.487	0.514	0.630	0.560	0.498	0.797
CS10	0.410	0.544	0.482	0.473	0.414	0.696
CS11	0.455	0.557	0.610	0.599	0.489	0.723
CS12	0.546	0.619	0.571	0.612	0.494	0.787
CS13	0.572	0.626	0.575	0.593	0.541	0.814
CS14	0.323	0.418	0.327	0.283	0.378	0.557
CS15	0.395	0.565	0.488	0.316	0.353	0.614
CS2	0.357	0.376	0.427	0.445	0.364	0.709
CS3	0.438	0.554	0.617	0.533	0.510	0.775
CS4	0.446	0.512	0.379	0.336	0.459	0.580
CS5	0.408	0.402	0.467	0.169	0.268	0.538
CS6	0.648	0.542	0.480	0.653	0.597	0.782
CS7	0.489	0.518	0.348	0.550	0.549	0.667
CS8	0.477	0.556	0.406	0.511	0.409	0.683
CS9	0.424	0.495	0.524	0.603	0.544	0.736

3.4. Testing the value of AVE (Average Variance Extracted)

The next convergent validity test is to test the average variance extracted (AVE) value of each variable.the value of AVE 0.50 or higher shows that, on average, constructs explain more than half of the indicator variants[20]. Conversely, if AVE is less than 0.50, it shows that, on average, there are variant items that are explained by the construct.

Table 5: Average Variance Extracted
Source: Questionnaire Testing Results with SPSS

Construct	Average Variance Extracted (AVE)
Accessibility	0.781
Efficiency	0.675
Loyalty	0.551
Privacy	0.852
Requirement Fulfillment	0.661
Satisfaction	0.494

It can be seen in Table 5 that there is an average variance extracted (AVE) value on the customer satisfaction variable which

has AVE value of 0.494. The author re-observes in Table 3 to find the outer loading value of customer satisfaction is low. In Table 3 there is a low outer loading value that belongs to customer satisfaction, that is, CS5 with a value of 0.538 and the writer deletes the CS5 indicator to increase the AVE value of customer satisfaction[21].

Table 6: Average Variance Extracted (after)
Source: Questionnaire Testing Results with SPSS

Constructs	Average Variance Extracted (AVE)
Accessibility	0.781
Efficiency	0.675
Loyalty	0.551
Privacy	0.852
Requirement Fulfillment	0.661
Satisfaction	0.510

Table 7: Cronbach's Alpha
Source: Questionnaire Testing Results with SPSS

Constructs	Cronbach's Alpha
Accessibility	0.721
Efficiency	0.756
Loyalty	0.793
Privacy	0.829
Requirement Fulfillment	0.744
Satisfaction	0.924

Table 8: Composite Reliability
Source: Questionnaire Testing Results with SPSS

Constructs	Composite Reliability
Accessibility	0.877
Efficiency	0.860
Loyalty	0.857
Privacy	0.920
Requirement Fulfillment	0.853
Satisfaction	0.935

After the researchers removed the CS5 indicator, it can be seen in Table 6 that AVE value owned by customer satisfaction increased from 0.494 to 0.510.

Table 9: Hypothesis Test Result
Source: Questionnaire Testing Results with SPSS

Hypothesis	Path	Original Sample	T-Statistic	Conclusion
H1	EF→CS	0.385	4.664	Significant
H2	RF→CS	0.193	2.323	Significant
H3	AC→CS	0.046	0.541	Less significant
H4	P→CS	0.444	5.954	Significant
H5	EF→CL	0.245	1.723	Less significant
H6	RF→CL	-0.056	0.446	Less significant
H7	AC→CL	0.091	0.754	Less significant
H8	P→CL	-0.034	0.263	Less significant
H9	CS→CL	0.519	3.029	Significant

After the researchers removed the CS5 indicator, it can be seen in Table 8 that AVE value owned by customer satisfaction increased from 0.494 to 0.510.

The following conclusions are obtained from the results of testing the hypothesis of the research model based on Table 9:

a. Hypothesis 1, Efficiency Factor (EF) has a significant positive effect on Customer Satisfaction (CS). That is because the value of the path coefficient and T-statistics are at a minimum, namely 0.384 and 4.664.
b. Hypothesis 2, Requirement Fulfillment (RF) factors have a significant positive effect on Customer Satisfaction (CS). That is because the value of the path coefficient and T-statistics are at a minimum, namely 0.193 and 2.323.
c. Hypothesis 3, The Accessibility (AC) factor is less influential on Customer Satisfaction (CS). That is because the path coefficient and T-statistics are below the minimum values, namely 0.046 and 0.541.
d. Hypothesis 4, Privacy factor (P) has a significant positive effect on Customer Satisfaction (CS), which is due to the value of the path coefficient and T-statistics are at a minimum value, namely 0.444 and 5.954.
e. Hypothesis 5, Customer Loyalty (CL) Efficiency Factor (EF). That is because the path coefficient and T-statistics are at a minimum value, 1,723.
f. Hypothesis 6, The Requirement Fulfillment (RF) factor is less significant to Customer Loyalty because the value of T-statistics is at a minimal value of 0.446.
g. Hypothesis 7, The Accessibility (AC) factor is less significant to Customer Loyalty (CL) because the value of T-statistics is at a minimum value of 0.754.
h. Hypothesis 8, Privacy Factor (P) Less significant to Customer Loyalty (CL) because the T-statistic value is at a minimum value of 0.263.
i. Hypothesis 9, The Customer Satisfaction (CS) factor is significantly positive towards Customer Loyalty (CL) because the T-statistics value is 3.029.

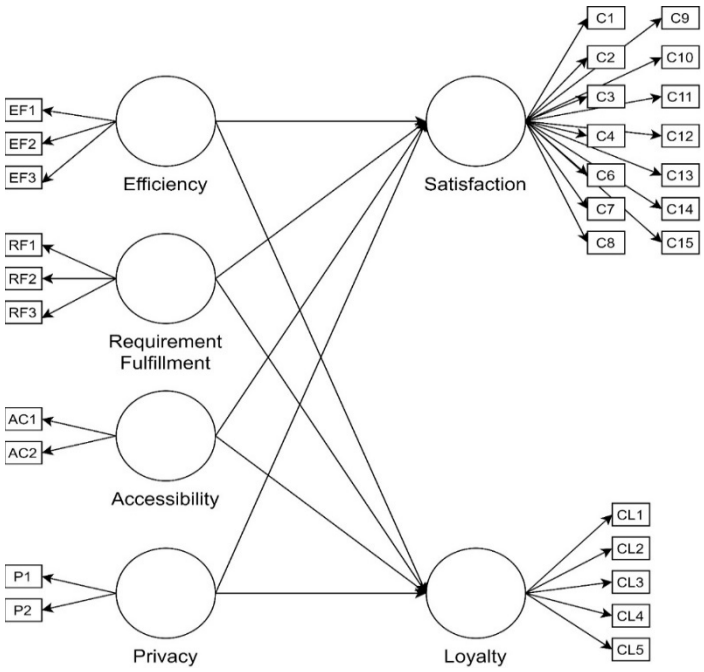


Figure 2: Results of Partial Research Model
Source: (author analysis)

In Figure 2, the relationship between variables based on the hypothesis received and has been proven to have a significant effect. The following discussion of the results of hypothesis testing.

4. Results

Based on the results of data processing related to the influence of e-service quality on customer satisfaction and customer loyalty in the XYZ e-marketplace, it can be concluded as follows:

1. There are 10 indicators namely customer satisfaction (CS4, CS5, CS7, CS8, CS14, CS15), efficiency (EF3), and customer loyalty (CL1, CL2) that do not meet the criteria because the value is less than 0.7 but of the 10 indicators only customer satisfaction (CS5) is omitted due to the low outer loading value that is 0.538 so that the AVE of customer satisfaction shows below 0.5. Therefore, researchers conducted outer loading testing again by removing the indicator on customer satisfaction (CS5) after the researchers re-tested the outer loading, the AVE value indicates that all indicators have reached values above 0.5 and met the required requirements.
2. The factors that influence this research on customer satisfaction are efficiency, privacy, and fulfillment requirements. But other variables such as accessibility have less effect on customer satisfaction. While the factors that influence customer, loyalty are only customer satisfaction. But other variables such as accessibility, efficiency, privacy, and fulfillment requirements have less effect on customer loyalty
3. There is a positive and significant effect on the efficiency, privacy, and customer satisfaction variables on customer loyalty variables with each value above 1.96
4. In measuring the significant level of relationship between variables it can be concluded that the accessibility variable has no significant effect on customer satisfaction. This has been proven by the results of the measurement of accessibility variable has a value of 0.541 <1.96, then so does the variable efficiency, requirements fulfillment, accessibility, and privacy has an insignificant effect on customer loyalty with each value below 1.96, which means not significant.

5. Discussion

E-Service Quality in the XYZ e-marketplace must be re-considered and must be improved again both in terms of the services provided to customers. The results of the study, e-service quality that has been done in the Accessibility section is something that must be considered again because of the low level of customer satisfaction in the given accessibility. On other hand e-service quality does not directly affect customer loyalty due to achieving a customer loyalty, the customer must feel satisfied first so that the customer can be loyal and shop again at the XYZ e-marketplace. The last service quality has a good impact on customer satisfaction, this is because when customers shop using the XYZ e-marketplace application the services provided by XYZ already meet customer needs but there is one indicator that is accessibility that is not achieve customer satisfaction.

6. Future Work

From the results of our research, researchers suggest:

1. In the future work the study should compare to another marketplace.
2. Due to the many insignificant hypotheses (5 of 9) it is recommended that future research use other variables to achieve much better research results.
3. Try to use other models such as UTAUT or TAM to do research about the impact of e-service quality on customer satisfaction and customer loyalty.
4. Adding the number of samples due to the small number of samples then it causes the results of testing the research hypothesis is not satisfactory, so it is advisable to increase the number of samples in future research.

References

[1] O. Yusuf, Pengguna Internet Indonesia Nomor Enam Dunia, 1–2, 2014.

[2] H.K. Soemartono, Survei APJII: Penetrasi Internet di Indonesia Capai 143 Juta Jiwa, Buletin Asosiasi Penyelenggara Jasa Internet Indonesia, Edisi-22 2, 3, 2018.

[3] R. Maulana, T. Perilaku Konsumen Belanja Online Indonesia Tahun 2018 Menurut iPrice, https://id.techinasia.com/tren-perilaku-konsumen-o, 2018.

[4] Peta E-Commerce Indonesia, 1, 2020. [Accessed: 10-Jun-2020].

[5] C.P. Anh, "Factors Influence Customer Satisfaction In Mobile Commerce", 6–52, http://urn.fi/URN:NBN:fi:amk-2015121621158, 2015.

[6] R.Q. Danish, A.A. Himayon, H.J. Iqbal, S. Raza, J. Shahid, "The Impact of Service Quality, Customer Satisfaction and Loyalty Programs on Customer's Loyalty: An Evidence from Telecommunication Sector," European Online Journal of Natural and Social Sciences, **7**(1), 40–47, 2018.

[7] Q. Chen, S. Rodgers, Y. He, "A Critical Review of the E-Satisfaction Literature," American Behavioral Scientist - AMER BEHAV SCI, **52**, 38–59, 2008, doi:10.1177/0002764208321340.

[8] J. Cronin. Jr, S. Taylor, "Measuring Service Quality - A Reexamination And Extension," The Journal of Marketing, **56**, 55–68, 1992, doi:10.2307/1252296.

[9] R.L. Oliver, "Whence consumer loyalty?," Journal of Marketing, **63**(SUPPL.), 33–44, 1999, doi:10.2307/1252099.

[10] M.N. Khuong, N.Q. Dai, "The Factors Affecting Customer Satisfaction and Customer Loyalty — A Study of Local Taxi Companies in Ho Chi Minh City, Vietnam," International Journal of Innovation, Management and Technology, **7**(5), 228–233, 2016, doi:10.18178/ijimt.2016.7.5.678.

[11] T.O. Jones, "Why Satisfied Customers Defect,", https://hbr.org/1995/11/why-satisfied-customers-defect, 1995. [Accessed: 10-Jun-2020].

[12] V. Zeithaml, A.P. Parasuraman, A. Malhotra, "A Conceptual Framework for Understanding E-Service Quality: Implications for Future Research and Managerial Practice," Marketing Science Institute Working Paper Report No, 2000.

[13] K. Laosethakul, W. Boulton, "Critical Success Factors for E-commerce in Thailand: Cultural and Infrastructural Influences," EJISDC The Electronic Journal on Information Systems in Developing Countries, **30**, 1–22, 2007, doi:10.1002/j.1681-4835.2007.tb00205.x.

[14] M. Akroush, S.D. Ali, I. Affara, "Service quality, customer satisfaction and loyalty in the Yemeni mobile service market," International Journal of Services, Economics and Management, **7**, 53, 2015, doi:10.1504/IJSEM.2015.076323.

[15] A. Parasuraman, V.A. Zeithaml, A. Malhotra, "E-S-QUAL a multiple-item scale for assessing electronic service quality," Journal of Service Research, **7**(3), 213–233, 2005, doi:10.1177/1094670504271156.

[16] M.N. Khuong, N.Q. Dai, "The Factors Affecting Customer Satisfaction and Customer Loyalty — A Study of Local Taxi Companies in Ho Chi Minh City, Vietnam," International Journal of Innovation, Management and Technology, **7**(5), 228–233, 2016, doi:10.18178/ijimt.2016.7.5.678.

[17] B. Ranjbarian, S. Fathi, Z. Rezaei, "Factors Influencing on Customers' E-Satisfaction: A case Study from Iran," Interdisciplinary Journal of Contemporary Research in Business, **3**(9), 1496–1511, 2012.

[18] T. Sheng, C. Liu, "An empirical study on the effect of e-service quality on online customer satisfaction and loyalty," Nankai Business Review International, **1**(3), 273–283, 2010, doi:10.1108/20408741011069205

[19] Sugiyono, "Metode Penelitian Kuantitatif Kualitatif dan R&D," 2011.

[20] A. Leguina, "A primer on partial least squares structural equation modeling (PLS-SEM)", International Journal of Research & Method in Education, **38**(2), 220-221, 2015, doi:10.1080/1743727X.2015.1005806

[21] A. Özer, M.T. Argan, M. Argan, "The Effect of Mobile Service Quality Dimensions on Customer Satisfaction," Procedia - Social and Behavioral Sciences, **99**, 428–438, 2013, doi:10.1016/j.sbspro.2013.10.511.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 211-220 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Trace-Driven Simulation of LoRaWAN Air Channel Propagation in an Urban Scenario

Eugen Harinda, Hadi Larijani *, Ryan M. Gibson

Glasgow Caledonian University, Networks and Cybersecurity, School of Computing and Built Environment, Glasgow, G4 0BA, United Kingdom

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 18 October, 2020
Online: 08 November, 2020

Keywords:
LoRaWAN
Trace-Simulation
Received Signal Strength
Deterministic models
Empirical Models
Urban Environment

ABSTRACT

Long-range, Low-Power Wide Area Network (LoRaWAN) is a very scalable solution for the Internet of Things (IoT). Due to the air channel environment's complexity, connectivity is a crucial parameter for successfully planning and deploying the IoT networks. Measurements and simulations have been used to evaluate LoRaWAN propagation models in the Urban environment, but it is a challenging task. While practical propagation evaluation has been prohibitively expensive, the theoretical modeling results have been less accurate. This paper uses real-world measurements and a trace-driven simulation technique to evaluate the RF propagation models' prediction performance for LoRaWAN 868 MHz propagation. First, a novel LoRaWAN trace-driven simulation of Glasgow city centre has been performed. Second, LoRaWAN 868 MHz measurements have been used to perform a critical analysis of LoRaWAN trace-driven Radio Frequency (RF) propagation models and validation. The processed trace dataset is composed of GPS coordinates, and the corresponding LoRaWAN received signal strength. The dataset has been extracted from 5017 datasets of LoRaWAN measurements taken from Glasgow city centre. A trace simulation program built-in ICS-Telecom was used to simulate LoRaWAN propagation in the real-world urban environment. Comparison of LoRaWAN simulation traces and the real-world data was performed to evaluate the prediction performance accuracy of Deygout 94, ITU-R 525/526, and COST-Walfish Ikegami (COST-WI) propagation models. All models over-estimated LoRaWAN trace-simulated RSS levels in comparison to collected measurement samples. While Deygout 94 prediction accuracy was higher with mean absolute error (MAE) at 0.83 dBm and standard deviation (SD) at 4.17 dBm, COST-WI performed poorly with MAE and SD at 2.87 dBm and 10.96 dBm respectively.

1 Introduction

This paper is an extension of work originally presented in The International Conference on UK-China Emerging Technologies, Glasgow, UK, August 2019 [1]. The purpose of this work was the application of a trace-based approach to evaluate the prediction capabilities of the conventional propagation models for LoRaWAN networks. As the Internet of Things (IoT) network is steadily growing, an understanding of conventional radio propagation models' prediction accuracy is key to the successful planning and deployment of Long-range, Low-Power Wide Area Network (LoRaWAN).

Unlicensed IoT technologies allow anyone with an idea to create and deploy IoT applications. Authors in [2] have predicted that a global number of IoT connected devices will rise to 75.44 from 15.4 billion in 2025. LoRaWAN [3] is one of the Low Power Wide Area Network (LPWAN) technologies for the interconnection of physical things. It uses Long-Range (LoRa) modulation technology [4] at the physical layer to achieve long-distance connectivity among the connected things. LoRaWAN end-devices connect to LoRa gateway using Radio Frequency (RF) propagation through the air channel and LoRaWAN network and application servers. Due to the complexity of the air channel environment, connectivity is a crucial parameter for the successful deployment of the IoT networks. To understand whether a LoRaWAN network meets the acceptable connectivity performance, a detailed investigation of the network using radio propagation prediction models is requisite to the network's deployment. However, Received Signal Strength (RSS) prediction in LoRaWAN networks has become a complex task due to prediction

*Corresponding Author: Hadi Larijani, Cowcaddens Road, Glasgow, G4 0BA, U.K, (0044)-141-331-3190, H.Larijani@gcu.ac.uk

performance differences in RF propagation models. These models' ability to accurately predict the RSS depends on how correctly they capture factors affecting RF signal propagation in the air channel.

Propagation models rely on natural and environmental factors to calculate signal power attenuation, path loss in the air channel. In the Line-of-Sight (LOS), the signal power attenuators include RF and the distance between end-devices and LoRa gateways. On the other hand, in cities, where Non-Line-of-Sight (NLOS) is dominant, environmental attenuators of signal power include human-made structures such as dense tall and short buildings, cars, and bridges. The path loss incurred by the propagating signal due to these obstructions is a consolidation of LOS and NLOS path loss due to multipath, scattering, diffraction, and reflection losses. To incorporate sufficient natural and environmental factors in the RF propagation models, numerous empirical propagation models [5] and deterministic propagation models [6] have been designed.

Different empirical and deterministic RF propagation modeling techniques highlight contrasts in prediction performance accuracy and application scenario. Empirical models were designed to predict the RSS in the event of insufficient knowledge of terrain profile data. These models were designed based on actual measurements taken from the propagation environment conditions. They are made of simple mathematical formulae with no reliance on the digital terrain data and are less accurate for a distance range of less than 1 *km* [7]. On the other hand, deterministic models largely depend on the digital terrain models to predict the RSS. These models take into consideration the free space and NLOS path loss. Accordingly, detailed knowledge of digital terrain models is a prerequisite for their prediction performance. Further contrasts exist within empirical and deterministic models in terms of parametric design consideration and prediction performance accuracy. Empirically, measurements and simulation methods are accurately used to investigate these models' ability to predict the RSS in radio networks.

Taking measurements from the real-world networks and simulation of the air channel propagation are vital methods used to evaluate the RF models. Studies [5], [6] and [8] have used measurements and simulations to evaluate the performance of various propagation models, but less has been done to validate their applicability on LoRaWAN networks in an urban environment. The studies in [9]–[13] have used only measurements to understand the behavior of air channel propagation on LoRaWAN network operations, but this is an expensive approach as best practices involve taking measurements at the later stages of network optimization. To validate the applicability of empirical propagation models for LoRaWAN in the city environment, our previous work in [14] showed less satisfactory empirical models' prediction performance against measurements. Trace-driven simulation can bridge the gap between measurement and simulation to evaluate RF propagation models' performance for LoRaWAN networks.

In this paper, we use real-world measurements and trace-driven simulation technique to evaluate the RF propagation models' prediction performance for LoRaWAN networks. Trace-driven simulation method can be used to evaluate computer-based models, but it works well if a trace contains the experience required by the trace simulation program [15]. The RF models are implemented in ICS-Telecom [16], and the simulation required a digital terrain model of Glasgow city centre. The simulation that uses a trace that contains the system's experience makes a more realistic simulation performance [17]. The study utilized a digital terrain model at 25 *m* and a dataset from Glasgow city centre. This dataset contained 5017 measurements with GPS coordinates and their corresponding LoRaWAN RSS. Trace-driven simulation results were compared against the real-world measurements to evaluate the prediction accuracy of Deygout 94, ITU-R 525/526, and COST-WI. All models over-estimated LoRaWAN trace-simulated RSS but showed superior prediction performance over empirical models in [14]. The result shows that the LoRaWAN network's trace-driven simulation with these models provides a more realistic RSS prediction than the conventional simulations.

The main contributions of this work are given in the summary below:

- A novel use of measurement and trace-driven simulation of LoRaWAN 868MHz propagation prediction for Glasgow city centre.
- The trace-driven simulation and extraction of LoRaWAN RSS traces spread over the digital terrain model of Glasgow city centre with the minimum, maximum range of trace signal levels, and the total number of trace measurements.
- Critical analysis of LoRaWAN trace-driven RF propagation models and validation with real-world measurements. Deygout 94 prediction accuracy was higher with *MAE* at 0.83 *dBm* and *SD* at 4.17 *dBm*.
- Visual comparative analysis of trace patterns between the LoRaWAN real-world measurement site in Figure 2. and trace-simulated site in Figure 5.
- Shown the functional relationship (Figure 4.) between trace-driven simulation and the data collection, data processing, and input to the prediction models.

This paper is arranged as follows: Section I briefly introduces LoRa technology, LoRaWAN network, conventional RF propagation models, motivation for performance and evaluation of RF models with trace-driven simulation, and the related work. Section II provides details about measurements and simulation. Section III contains the RF models' performance analysis and discussion. Finally, section IV presents the conclusion and prospective future work.

1.1 Overview of LoRaWAN and LoRa

LoRaWAN is one of the media access control protocol for low-power Wide Area Network (WAN). It is a long-range wireless communication technology that enables low-powered IoT devices and applications to communicate over the Internet [18]. LoRaWAN specification V1.0 [3] published in 2015, provides LoRaWAN details to describe the network communication protocol and architecture. The typical LoRaWAN network architecture is a star topology consisting of the end-devices, applications, gateway, and the server. These devices may be connected through the air or cables. LoRaWAN-enabled devices may be categorized as Class A, B, or C, depending

on the capacity for a trade-off between downlink latency and battery-life saving. Figure 1. describes a typical star-topology of LoRaWAN architecture. While LoRaWAN defines the architecture and system communication protocol, LoRa is a modulation technology defined at the physical layer.

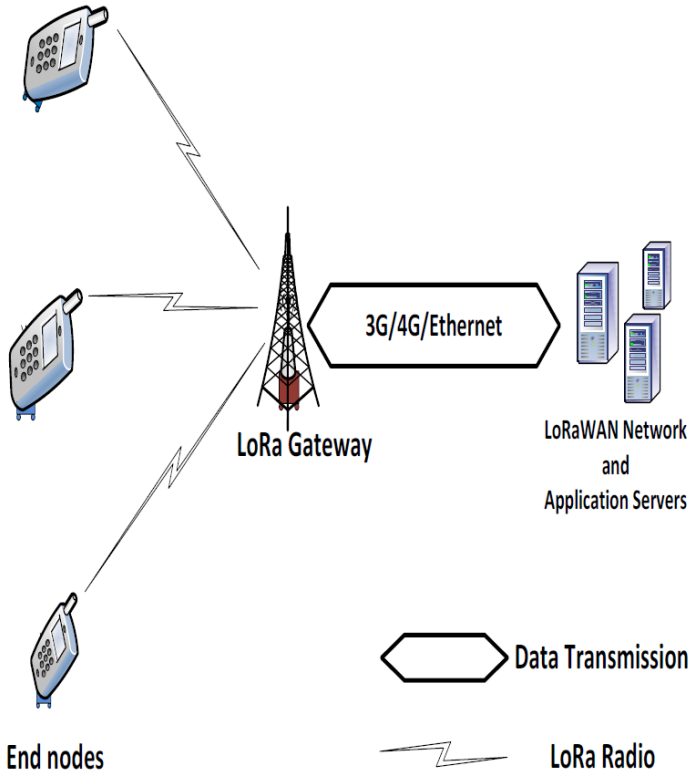


Figure 1: Basic LoRaWAN Network Architecture

LoRa is a spread spectrum technique adapted by Semtech [4] and implemented at the physical layer for digital wireless modulation. It was derived from chirp spread spectrum technology [19] to achieve a long-range radio communication in LoRaWAN networks while retaining the ability to operate at low-power levels. Furthermore, the LoRaWAN network employs different spreading factors (SF), another key technology to allow a trade-off between the coverage and data rates [20]. Higher SF configuration helps the LoRaWAN network to transmit further, but at the cost of low data rates, and vice-versa. Thus, the LoRaWAN network delivers the required connectivity to enable the LoRa end-devices to transmit data over the air channel to the gateway and then to the network and application servers.

1.2 Radio Propagation Models

RF propagation models[21] characterize the propagation of radio waves as function carrier frequency, distance, and other NLOS factors in the air channel environment. These RF models, deterministic, semi-empirical, and empirical, employ several attenuation factors, including LOS, diffraction, multipath, to estimate the RSS. These models perform connectivity analysis in radio networks to fast-track radio coverage information. Radio propagation effects are heavily site-specific and depend on configuration parameters such as terrain cover, operating frequency, distance, transmitter, and receiver antenna height. In this study, commonly used RF models in ICS-Telecom, Deygout 94, ITU-R 525/526, and COST-231 Walfish-Ikegami models are evaluated for RSS prediction accuracy in LoRaWAN network. These models are respectively, deterministic, semi-deterministic, and empirical.

1.3 Deygout 94

Deygout 94 model determines diffraction attenuation for N successive knife-edge or rounded obstacles. The model is based on the idea of the main signal blocking component [22], which exerts considerably high obstruction losses to the link between the transmitting and receiving ends. The overall diffraction loss, L_D, is given as:

$$L_D = \sum_{i=1}^{N} L_1(V_i) \tag{1}$$

where L_1 is the diffraction loss in case of one main obstruction between the transmitter and receiver, and V_i is the diffraction parameter for the ith knife-edge. The total path loss is the sum of Deygout 94 diffraction losses, free space, sub-path attenuation, and 3D losses that are implemented in ICS-Telecom.

1.4 ITU-R 525/526

The ITU-R 525/526 uses ITU-R P 525 [23] to calculate the free space loss and ITU-R P 526 to calculate attenuation due to diffraction [24]. The concept of ITU-R 526 calculation of diffraction losses is similar to Deygout 94. It calculates sub-path and diffraction losses using the Delta Bullington diffraction model [25]. The ITU-R 525 calculates the basic free space transmission loss, P_{A_0} between isotropic antennas, as follows:

$$P_{A_0} = 32.4 + 20log_{10}(d) + 20log_{10}(f) \tag{2}$$

where f is frequency (MHz) and d is the distance (km).

1.5 COST-231 Walfish-Ikegami

COST 231 Walfish-Ikegami (COST-WI) the model [26] is a combination of Walfish and Ikegami models. Unlike the deterministic model that requires digital terrain models, including the buildings layer, COST-WI is an empirical model that requires characteristic values. It improves the path loss prediction by considering more data to characterize medium-sized flat urban environments [27]. Compared to other empirical models, the COST-WI parameters include the heights h_b of the building, the separation between buildings d, the widths of roads w, and the angle θ to the direct radio path. The range of fundamental parameters considered is 800-2000 MHz for frequency, 1-3 m, 4-50 m, 0.02-5 km for distance, and end-device and gateway antenna height, respectively. The model calculates path loss due to the LOS and NLOS. The mathematical formulae for LOS and NLOS are respectively defined in (3) and (4) below.

$$P_{LOS} = 42.64 + 26log_{10}(d) + 20log_{10}(f) \tag{3}$$

If NLOS is the case, then path loss calculation is a combination of attenuation due to free space P_{LOS}, the rooftop to street diffraction, and the scattering L_{rts}, and the multiple screen diffraction loss L_{msd}. The totality of this (NLOS) path loss is described mathematically [27**?**] as follows:

$$P_{NLOS} = P_{LOS} + L_{rts} + L_{msd} \tag{4}$$

The L_{rts}, the diffraction loss from the rooftop to the street is determined as in the following formula:

$$L_{rts} = -16.9 - 10log_{10}(w) + 10log_{10}(f) + 20log_{10}(h_b - h_r) + L_{ori} \tag{5}$$

where, w is width of the roads, h_b and h_m are the height of building and end-device mobile station respectively. The street orientation correction factor, L_{ori}[28] is given as:

$$L_{ori} = \begin{cases} -10 + 0.35\alpha & \text{for } 0° < \alpha < 35° \\ 2.5 + 0.0755(\alpha - 35) & \text{for } 35° < \alpha < 55° \\ 4 - 0.0114(\alpha - 55) & \text{for } 55° < \alpha < 90° \end{cases} \tag{6}$$

where α is the street orientation angle, L_{msd}, the multiscreen loss, represents diffraction loss from multiple obstacles. The following mathematical representation determines it:

$$L_{msd} = L_{bsh} + K_a + K_d log_{10}(d) + K_f log(f) - 9log_{10}(s_b) \tag{7}$$

where the correction factors, L_{bsh} and K_a represent path loss when the gateway is above and below the rooftops, respectively. s_b represents buildings separation. The terms K_d and K_f quantify the diffraction loss as a factor of the distance, d and frequency, f, and are defined in[26] as follows:

$$L_{bsh} = \begin{cases} -181log(1 + h_t - h_b) & h_t > h_b \\ 0 & h_t \leq h_b \end{cases} \tag{8}$$

$$K_a = \begin{cases} 54 & h_t > h_b \\ 54 - 0.8(h_t - h_b) & h_t < h_b \text{ and } d_{km} \geq 0.5km \\ 54 - 1.6(h_t - h_b)d & d_{km} < 0.5km \end{cases} \tag{9}$$

$$K_d = \begin{cases} 18 & h_t > h_b \\ 18 - 15(h_t - h_b)/h_b & h_t \leq h_b \end{cases} \tag{10}$$

$$K_f = -4 + \begin{cases} 0.7(f_{MHz}/925 - 1) & \text{for medium-size city and suburban} \\ 1.5(f_{MHz/925-1}) & \text{for metropolitan centres} \end{cases} \tag{11}$$

1.6 Motivation for Performance and Evaluation of RF Models with Trace-driven Simulation

The evaluation of how the air channel affects the propagation of wireless communication signals is a challenging task. In cities, the physical environment, including buildings, bridges, people's motion, and vehicles, adversely affect propagating RF signals. To accurately evaluate the effects of the physical environment in cities, simulation methods and models have become an integral part of wireless communication systems' planning and deployment. However, it has been challenging to identify which RF propagation models, from their significant number, produce realistic predictions of the RSS. The assumption that simulation models can be applied to modeling propagation effects in different environments, protocols, and applications have often been proved less accurate.

The literature is yet to contain either the RF propagation prediction models or an established understanding of the actual RF propagation prediction models' performance for the LoRaWAN propagation system in cities. Studies that attempted to evaluated LoRAWAN propagation system performance in different urban scenarios used the analysis of the real-world measurements [9]–[13]. However, the use of field measurement to characterize the impact of the air channel on RF propagation is usually costly as it involves real systems. In wireless communication networks, taking real-world measurements is usually the last phase of network planning and deployment, aiming to optimize a wireless network system performance. Instead, the network planning tools use propagation prediction models to estimate path loss due to air channel conditions. In our previous work [14], we evaluated conventional empirical propagation models' performance to determine applicability for predicting LoRaWAN RSS in an urban scenario. Although the study showed that COST-WI had better prediction performance for LoRaWAN RSS than other models considered for the study, this result was over-predicted by 6.48 dBm. Therefore, this indicates that empirical RF models lack accuracy in attempts to capture paths loss over the rooftops, diffraction, and reflections arising from the buildings and streets set up, and these factors are better handled with deterministic models.

In this paper, we intend to evaluate Electromagnetic Wave (EM) based deterministic models to understand the applicability of these models to predict LoRaWAN RSS in urban areas. The EM deterministic models' ability to accurately consider the phenomena of path loss over the rooftops, diffraction, reflections from buildings, dispersion, and wave running are essential features for dense urban outdoor RF propagation [29]. Deterministic models require the use of accurate and high-quality terrain models of the environment. However, in the presence of low-quality terrain models, the use of trace-driven simulation for evaluation of deterministic and semi-deterministic models produced more accurate LoRaWAN RSS prediction than empirical RF models. This study can give insight into the effectiveness of trace-simulation with standard propagation models for evaluation of IoT connectivity with LoRaWAN networks at 868 MHz in the NLOS urban environment.

1.7 Related Work

Trace-driven simulation is a technique that utilizes both measurements and simulation for evaluation and prediction of the performance of the computer-based systems [30]. All the trace-driven studies reviewed in this paper used a software probe to collect measurement trace data except [31], which used device-specific data collection from the testbed. Trace-driven simulation technique has been introduced in [32] as a technique that uses analysis of measurement trace data to redo similar experimental circumstances in the simulations. Trace-driven simulation has been used to perform a realistic simulation.

Trace-driven simulation use analysis of the experimental trace data in an attempt to recreate simulation of the experimental conditions. Authors in [33] stated that sufficient and accurate measurement datasets could directly serve as a basis of propagation performance prediction or fitting the RF models. Trace-driven simulation was used in [34] to model the error behavior of the WaveLAN wireless channel. Chia et al [35] proposed an Automata-Synchronized Replay (EAR) system software to perform trace-based simulation of the measured packets received over the Wireless Local Area Network. The EAR was used to control trace packet, synchronize the environmental effects in trace packet with the packets, and other signals received from the real-world environment. The system implementation results show that the EAR achieved an event reproduction rate of 92.45% under a fading environment. Trace-driven simulation can be used to evaluate wireless link models.

In [36], the authors presented a Wireless Link Simulator (WiLinkSim). This simulator implemented models based on mathematical distribution to correctly represent wireless errors (bit error rate) due to the air channel's effects. WiLinkSim utilizes trace-driven models for realistic simulation of Rayleigh and multipath fading and during the deployment of a real-world network. G. Judd and P. Steenkist [37] indicated that trace-driven simulation is one of the accurate methods to model RF signal propagation. Authors in [38] used measurement error traces to configure stochastic error models to reproduce the signal-to-noise-plus-interference ratio errors of the received wireless signal. The accuracy for Bernoulli, Gilbert-Elliot, and chaotic map error models' ability to reproduce or estimate the measured dataset was performed by comparing the measured traces with models' simulated traces. The results showed that Gilbert-Elliot and chaotic map error models attained enough accurate reproduction of the measured traces.

2 Measurement Setup and Simulation

In the first step, the measurement trace dataset was collected, and then, the RF propagation prediction models were trace simulated using the measurement processed trace data. The measurement trace dataset was collected from an experimental LoRaWAN network installed in Glasgow City using LoRaWAN end-devices and Servers. The processed trace data needed for trace-simulation was selected from the measured data. This trace data served as input data to the trace simulation program. Below is the detailed description of data collection, data processing, and trace-simulation setup.

Table 1: Measurement and Trace-simulation Parameters.

S/N	Parameters	Values
1	Operating frequency Band	868 *MHz*
2	Bandwidth	125 kHz
3	End-device transmit power	14 *dBm*
4	Number of Gateway	3 m
5	Gateway antenna heights	30, 27 and 27 *m*
6	End-device antenna height	1.5 *m*
7	Measurement spreading factor (SF)	7-12
8	Trace-simulation spreading factor (SF)	7, 8 and 12
9	Maximum transmit and receive distance	2275 *m*

2.1 Data Collection Setup

Capturing the field measurements is essential in characterizing air channel impact on LoRaWAN system applications under real-world conditions. We performed an experimental field study to acquire real-world LoRaWAN measurements from Glasgow city, Scotland, the UK. On the map, the city's latitude and longitude are 55.8642° N and 4.2518° W, respectively. The city is built at the bank of River Clyde and constitutes tall and huge buildings and open spaces. Data collection was performed in this city's environment to characterize the effects the NLOS has on the air channel for RF propagation of the LoRaWAN network. The leading equipment used to collect data is the LoRaWAN end-device and three gateways. The former was a Multitech mDot module [39] with LoRa and LoRaWAN protocol, regulated by a single board Raspberry Pi computer, whereas the later constituted Kerlink gateways, enabled with Semtech LoRa SX1301 chip [40]. The Raspberry Pi computer was connected to a 3G/GPS module to keep location records for the received from transmitting end-device operated at 868 *MHz* and 14 *dBm*.

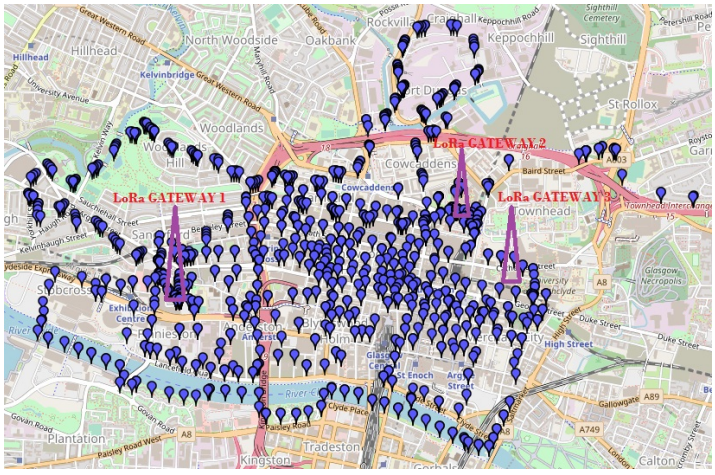


Figure 2: A Google map showing the location of the three LoRa Gateways installed in Glasgow city centre and measurements waypoints. The blue markers on the map represent GPS and their corresponding LoRaWAN RSS waypoints obtained during the measurements

The gateways operated at three locations. The first gateway at Glasgow Caledonian University, 30 *m* on top of George More building. The second gateway at Strathclyde University, 27 *m* on top of James Weir building, and the third gateway placed at 27 *m* on top of

Skypark. The distance between gateway at Skypark and James Weir building is 2005 *m* apart, and both buildings are respectively, 590 *m* and 1900 *m* from George More building. Figure 2. is the display of a 1:300 Google map showing the topography of Glasgow city centre used for the LoRaWAN RSS measurements. It indicates the location of the three LoRa Gateways and measurement waypoints. The three gateways utilize the cable and mobile networks to connect to the network server to record packets received from the transmitting end-device. The mobile end-device transmissions operated the 868 *MHz* and 14 *dBm*. Table 1. shows details of LoRaWAN transmission parameters. The packets were sequentially transmitted, and success or failure to receive an ACK message was recorded. These packets altered between 7-12 the spreading factor. We collected the data in a backpack logger unit at the walking speed in Glasgow city. The data collected in the network server and backpack logger unit were combined and processed in a database. This data shall be known as measurement trace data.

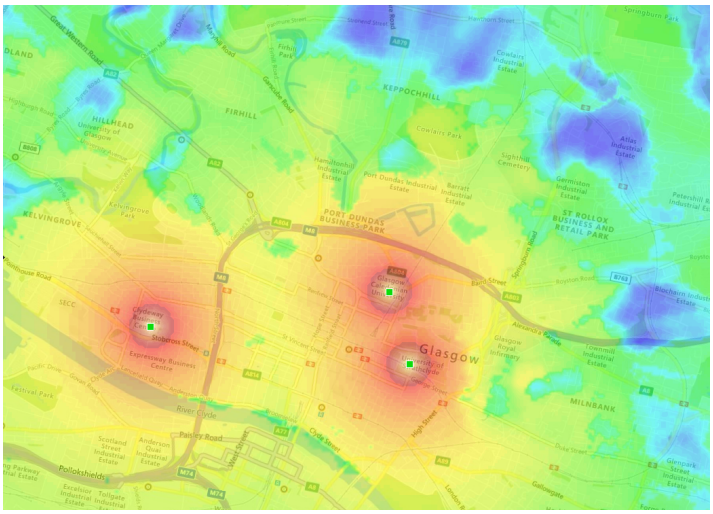


Figure 3: LoRaWAN 868 *MHz* simulation without traces over a Bing digital terrain model of Glasgow city centre. The three green squares in a reddish/brown color represent the simulated LoRa gateways. Colors represent the variation of signal levels, with blue indicating the weakest and reddish the strongest signals.

2.2 Data Processing and Input to Models

To prepare LoRaWAN system measurements for a trace simulation program, we processed measurement trace data. Measurement trace data contained feature data and metadata not required by the trace simulation program. A total of 5008 measurement trace data collected in the backpack logger unit and network server was loaded into the database. We used the command line manipulation commands to pare measurement trace data down to locations coordinates and the RSS suitable trace simulation. Consequently, manipulation of measurement raw data resulted in 5008 datasets for latitude, longitude, and corresponding RSS required for the ICS-Telecom trace simulation function.

The trace simulation program is built from LoRaWAN and RF propagation system models. LoRaWAN system configuration parameters in Table I. and RF propagation models are introduced in section I. The LoRaWAN parameters in this table are based on the LoRaWAN band regulation in the European Union and the physical position of gateway and end-device antenna heights used during the measurements. Similarly, the maximum transmission is the distance between the transmission and reception attained during the measurements. However, due to lack of mobility during the simulation, the SF was heuristically fixed at 7, 8, and 12. The trace simulation program imported the processed trace data through the measurement import function to perform a trace-driven simulation. It executes the LoRaWAN and RF propagation environment using the processed trace data. The trace-driven simulation program's output, which is of interest to this work, was the trace-driven RSS. This output is compared with the RSS in the processed trace data for performance analysis. Figure 4. shows a block diagram of the trace-driven simulation model. This model's main parts are; data collection, data processing, and LoRaWAN and RF propagation conceptual models.

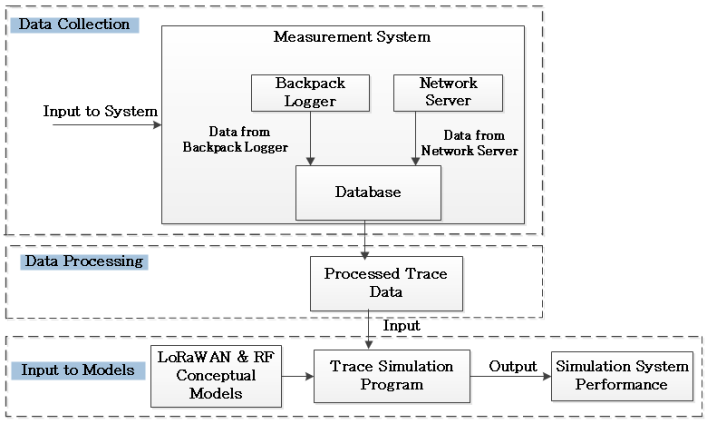


Figure 4: Block Diagram of Trace-driven simulation

2.3 Trace-Driven Simulation

The trace-driven simulation requires a combination of system measurements and simulation [41]. In this paper, trace-simulation set up involved importing processed trace data and configuration of the Glasgow city centre digital terrain model. When the trace simulation program starts, it loads the processed measurement trace data and the LoRaWAN RF propagation system model to reproduce the same experimental conditions in the ICS-Telecom. Authors in [**?**] showed that trace-simulation produces more realistic simulation results than conventional simulation. In [42], the authors provided a detailed account of the advantages of trace-simulation over the conventional simulation. To simulate Glasgow city centre terrain, we imported and configured a 1:300 Bing terrain model of Glasgow city centre at 25 *m* resolution into ICS-Telecom. However, low-resolution terrain maps such as 25 *m* provided a digital elevation model and the city's clutter to the simulator, leaving out the building layer. This input facilitated the simulation of the LoRaWAN system in Glasgow city centre. Over the terrain map, we placed three system gateway models. Each gateway model location configuration on the terrain map matched the latitude and longitude of system gateways used during the LoRaWAN measurements. Like the measurement system, the distance between LoRaWAN gateway models at Strathclyde University and Skypark on the terrain map represents 2005 *m* apart. The respective distance between these

gateway models from Glasgow Caledonian University represents 590 m and 1900 m on the ground. Figure 3. shows the LoRaWAN 868 MHz simulation system set up without traces over a Bing digital terrain model of Glasgow city centre. The next step was to load the processed trace data for trace-driven simulation of the LoRaWAN 868 MHz system.

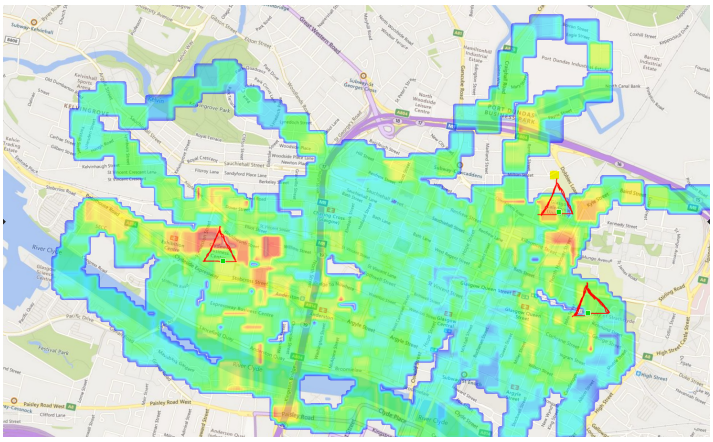


Figure 5: A trace-driven simulation of LoRaWAN 868 MHz over a Bing digital terrain model of Glasgow city centre. The three green squares at the base of red triangles represent the simulated LoRa gateways. Colors represent trace-driven RSS data. The blue color refers to the weakest signals, and reddish/brown refers to the strongest signals. The trace-driven data follows waypoints of the real-world measurements.

Through the measurement import function, the trace simulation program imported the processed trace data. Presented to the digital terrain map are LoRaWAN system measurement locations, a set of latitude and longitude coordinates. Figure 5. shows a trace-driven simulation of LoRaWAN 868 MHz over a 1:300 Bing digital terrain model of Glasgow city centre. It presents the trace-simulated RSS in the previously identified location with the processed trace data, which contains latitude, longitude, and corresponding RSS. At the base of the trace simulation program, there are LoRaWAN and RF propagation system models. These models are the designed base of the trace-driven simulation program. LoRaWAN and RF conceptual system models comprise LoRa modulation, LoRaWAN network, RF propagation models, and the terrain model. The configuration of the LoRaWAN simulation system parameters is similar to LoRaWAN system parameters shown in Table I. On the other hand, the configured RF propagation models are described in Section I. After the trace-simulation, extraction of the predicted trace-driven LoRaWAN RSS followed for each of the RF models.

The extraction of trace-simulated LoRaWAN RSS from the LoRaWAN simulation system provided a trace-simulated dataset required for prediction performance analysis. The minimum trace signal level was set to -140 dBm and the maximum trace signal level set to -61 dBm. The total number of trace data was set to 5008. The trace-simulated LoRaWAN RSS are spread over the digital terrain map of Glasgow city centre in Figure 5. and the color code shows the varying signal strength, which depends on different factors, including clutter obstructions in the environment. The predicted trace-driven dataset for each propagation model was extracted from the LoRaWAN simulation system digital terrain map. Alternatively, the predicted trace-driven dataset can be exported from the simulation system. To validate the obtained trace-simulated dataset, we compared the RF propagation models' trace-simulated LoRaWAN RSS with the LoRaWAN system RSS, measured from Glasgow city centre.

3 Performance Analysis and Discussion

To evaluate models' prediction performance accuracy, we validated LoRaWAN trace-simulation with measurement trace data. Common mathematical error methods measured the prediction accuracy of the models. The error performance metrics used to represent the error between measurement and simulation are the *Mean Square Error (MSE)*, Δ'_{yi}, *Mean Absolute Error (MAE)*, $|\Delta y|$, and *Standard Deviation (SD)*, σ_e. In this paper, N indicates the total number of measured or simulated data samples, whereas Δ_{yi} denotes the difference between predicted and measured datasets. While *MAE* is used as a measure of RF models prediction error, *SD* measures the average distance between measurement and the predicted dataset. C. J. Willmott and K. Matsuura [43] examined the ability of *MAE* and *Root-Mean-Square Error (RMSE)* to represent the average error performance of a model. The authors concluded that *MAE* is a preferred measure of one dimension average error performance compared to *RMSE*, which is suitable for more than one error performance of a model. However, in [44], authors argued that both *MAE* and *RMSE* may be used to evaluate the average error prediction accuracy of the models. In this study, the *MAE* is considered to be an intuitive measure of average error prediction accuracy in RF models.

$$\Delta_{yi} = (RSS_{predicted} - RSS_{measured}) \tag{12}$$

$$\Delta'_{yi} = \frac{1}{N}\sum_{i=1}^{N}(RSS_{predicted} - RSS_{measured})^2 \tag{13}$$

$$|\Delta y| = \frac{1}{N}\sum_{i=1}^{N}|\Delta_{yi}| \tag{14}$$

$$\sigma_e = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\Delta_{yi} - |\Delta_y|)^2} \tag{15}$$

Table 2: Error Performance Metrics

Parameters	Deygout94	ITU-R525/526	COST-WI
MSE	18.7812	20.978	24.2155
MAE	0.83	1.01	2.87
SD	4.17	5.84	10.96

The comparative analysis of RF models prediction accuracy showed that all the three investigated models over-estimated the RSS measurements. However, Deygout 94 registered higher prediction accuracy with *MAE* at 0.83 dBm. The *MAE* for ITU-R 525/526 and COST-WI was 1.01 dBm and 2.87 dBm, respectively. In terms of prediction error, Deygout 94 prediction had the lowest *Mean Square Error, MSE* at 18.7812 dBm, followed by the ITU-R 525/526 and COST-WI with *MSE* at 20.978 dBm and 24.2155

dBm respectively. Furthermore, Deygout 94 exhibited the lowest *standard deviation,* σ_e at 4.17 *dBm*, followed by the ITU-R 525/526 and COST-WI at 5.84 *dBm* and 10.96 *dBm* respectively. This result implies that Deygout 94, a deterministic model, performs better RSS predictions than the ITU-R 525/526 and COST-WI models. COST-WI poor prediction performance is attributed to its reliance on the buildings' information. Table II. shows models' prediction error performance metrics calculated in equations 12-15 for MSE, MAE and σ_e using measured and trace-driven dataset in the NLOS city environment. Overall, it can be observed that Deygout 94 obtained low prediction error performance.

Density and vertical obstructions presented to the application through a high-resolution DTM is crucial for modeling RF propagation [45]. The DTM at 25*m* resolution that presented the NLOS obstructions over Glasgow city centre terrain landscape into the ICS-Telecom lacked detailed buildings' information. The graphical representation in Figure 6. shows the relationship between the LoRaWAN 868 MHz measurement dataset and models trace-simulation performance in the NLOS conditions in Glasgow city centre. The measurement and trace-driven simulated data are plotted as a RSS function, in *dBm* and distance, d in meters. The visible gap between measurements and trace-driven simulation may be attributed to the low resolution of the DTM, which fails to represent actual NLOS conditions in Glasgow city centre to the Trace Simulation Program. However, trace-driven RSS trend against the distance follows RSS measurements with Walfish-Ikegami keeping more distance and Deygout 94, making a closer drift with measurements. The clustered data visible on the graph indicates many RSS collected at specific locations due to reduced NLOS obstructions. On the other hand, the flattened regions are an indicator of either low variations of LoRaWAN RSS as signal reception reduce or completely blocked due to an increase in density and the height of buildings in the City's air channel.

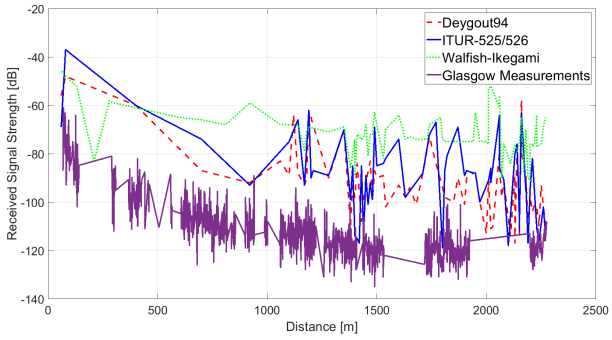


Figure 6: Models trace-simulated performance comparison with LoRaWAN 868 MHz measurements. The vertical axis plots real-world LoRaWAN RSS measurements, *dBm* and models' trace-simulated LoRaWAN RSS data, *dBm*. The horizontal axis represents the distance, d in m.

The trace-driven simulated models improved prediction accuracy performance compared to empirical models in the previous study [14]. Both studies were carried out for LoRaWAN 868 MHz in Glasgow city centre, and the same measurement dataset was used for validation. Table III. shows the comparison of prediction accuracy of empirical models, in cyan color heading and trace-simulated models, in the yellow color heading. The improvement in trace-simulated predictions compared to empirical RF models can be attributed to measurement traces and differences in model design and application. The spreading of LoRaWAN RSS over the Glasgow city centre map shows how trace-driven simulation leads closer to measurements.

Table 3: Comparison of Emprical and Trace-simulated Models Prediction Accuracy

Parameters	*Deygout94*	*ITU-R525/526*	*COST-WI-Traced*	**Okumura-Hata**	**COST231**	**COST-WI**
MSE	18.7812	20.978	24.215	48.84021	57.2056	96.746
MAE	0.83	1.01	2.87	5.564	6.131	7.413
SD	*4.17*	*5.84*	10.96	9.158	11.425	7.454

The analysis of measurements and simulation mapping over Glasgow city centre's map shows significant variation in the RSS spreading. The simulation mapping of RSS in Figure 3. spreads all over the map compared to the mapping of RSS measurement, which follows the data collection path. On the other hand, LoRaWAN trace-driven simulation mapping in Figure 5. attains a close match to the mapping of RSS measurements in Figure 2. when the minimum and maximum trace signal level and the total number of trace dataset is set. The closer pattern between the measured and trace-driven RSS mapping matches the claims that trace-driven simulation performs realistic simulation [32]. This pattern shows that the trace-driven simulation technique can perform a realistic evaluation of RF models for the planning and deployment of LoRaWAN in cities.

4 Conclusion and Future work

We have created a trace-driven simulation of LoRaWAN 868 MHz for Glasgow city centre. Trace-driven simulation techniques using the real-world measurement trace dataset make it possible for RF propagation models to produce realistic RSS predictions of LoRaWAN radio links. This paper has used measurements to perform trace-driven simulation over the DTM of Glasgow city centre. We have used LoRaWAN trace-driven RF propagation models to analyse and evaluate several RF propagation models' prediction performance accuracy for LoRaWAN 868 MHz. The RF models considered in this study are Deygout 94, ITU-R 525/526, and COST-WI. The trace simulation program, built in the ICS-Telecom, imported 5017 processed trace datasets measured from an experimental LoRaWAN 868 MHz network deployed in Glasgow city centre. The processed trace dataset contained the latitude, longitude, and the corresponding measurement RSS. The comparative performance analysis of trace-driven LoRaWAN simulation traces and the real-world measured LoRaWAN processed trace dataset showed that RF propagation models produce realistic LoRaWAN RSS predictions.

All models over-estimated LoRaWAN trace-simulated RSS levels in comparison to real-world collected data samples. However, a comparison of trace-driven simulated models and empirical RF models shows that trace-simulated RF models produce a more accurate prediction. Deygout 94 prediction accuracy was the higher with *MAE* at 0.83 *dBm* and *SD* at 4.17 *dBm*. COST-WI cannot

be used for simulation of LoRaWAN coverage estimation in an urban environment unless high-resolution DTM models are used in ICS-Telecom. Currently, most studies use measurements to evaluate LoRaWAN performance. This approach is expensive, and trace-driven simulation can be a better option. Although it requires DTM, the use of even low-resolution DTM produces results better than conventional models. The random neural network could be used to design a more accurate and inexpensive RF propagation model in the future.

References

[1] E. Harinda, H. Larijani, R. M. Gibson, "Trace-Driven Simulation for LoRaWan868 MHz Propagation in an Urban Scenario," in 2019 UK/ China Emerging Technologies (UCET), 1–5, 2019, doi:10.1109/UCET.2019.8881854.

[2] R. Taylor, D. Baron, D. Schmidt, "The world in 2025 - predictions for the next ten years," in 2015 10th International Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT), 192–195, 2015, doi:10.1109/IMPACT.2015.7365193.

[3] T. E. T. K. I. O. N. Sornin (Semtech), M. Luis (Semtech), "LoRaWAN™ Specification," LoRaWAN™ Specification 1.1 — LoRa Alliance™, **2015**, 2015.

[4] LoRa®Alliance, "A technical overview of LoRa® and LoRaWAN™," https://lora-alliance.org/sites/default/files/2018-04/what-is-lorawan.pdf/, 2015, [Online; accessed 25-September-2018].

[5] V. S. Abhayawardhana, I. J. Wassell, D. Crosby, M. P. Sellars, M. G. Brown, "Comparison of empirical propagation path loss models for fixed wireless access systems," in 2005 IEEE 61st Vehicular Technology Conference, **1**, 73–77 2005, doi:10.1109/VETECS.2005.1543252.

[6] E. Greenberg, E. Klodzh, "Comparison of deterministic, empirical and physical propagation models in urban environments," in 2015 IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems (COMCAS), 1–5, 2015, doi:10.1109/COMCAS.2015.7360394.

[7] C. Braithwaite, M. Scott, UMTS network planning and development: design and implementation of the 3G CDMA infrastructure, Elsevier, 2003.

[8] O. Fratu, A. Martian, R. Craciunescu, A. Vulpe, S. Halunga, Z. Zaharis, P. Lazaridis, S. Kasampalis, "Comparative study of Radio Mobile and ICS Telecom propagation prediction models for DVB-T," in 2015 IEEE Interna-tional Symposium on Broadband Multimedia Systems and Broadcasting, 1–6, 2015, doi:10.1109/BMSB.2015.7177260.

[9] R. Sanchez-Iborra, J. Sanchez-Gomez, J. Ballesta-Viñas, M.-D. Cano, A. Skarmeta, "Performance Evaluation of LoRa Considering Scenario Conditions," Sensors, **18**(3), 772, 2018, doi:10.3390/s18030772.

[10] A. Augustin, J. Yi, T. Clausen, W. Townsley, "A Study of LoRa: Long Range & Low Power Networks for the Internet of Things," Sensors, **16**(9), 2016, doi:10.3390/s16091466.

[11] L. Feltrin, C. Buratti, E. Vinciarelli, R. De Bonis, R. Verdone, "LoRaWAN: Evaluation of Link- and System-Level Performance," IEEE Internet of Things Journal, **5**(3), 2249–2258, 2018, doi:10.1109/JIOT.2018.2828867.

[12] A. J. Wixted, P. Kinnaird, H. Larijani, A. Tait, A. Ahmadinia, N. Strachan, "Evaluation of LoRa and LoRaWAN for wireless sensor networks," in 2016 IEEE SENSORS, 1–3, 2016, doi:10.1109/ICSENS.2016.7808712.

[13] J. Petajajarvi, K. Mikhaylov, A. Roivainen, T. Hanninen, M. Pettissalo, "On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology," in 2015 14th International Conference on ITS Telecommunications (ITST), 55–59, 2015, doi:10.1109/ITST.2015.7377400.

[14] E. Harinda, S. Hosseinzadeh, H. Larijani, R. M. Gibson, "Comparative Performance Analysis of Empirical Propagation Models for LoRaWAN 868MHz in an Urban Scenario," in 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), 154–159, 2019, doi:10.1109/WF-IoT.2019.8767245.

[15] J. W. C. Fu, J. H. Patel, "Trace driven simulation using sampled traces," in 1994 Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences, **1**, 211–220, 1994, doi:10.1109/HICSS.1994.323170.

[16] I. Telecom, "ATDI," 2014.

[17] S. W. Sherman, J. C. Browne, "Trace Driven Modeling: Review and Overview," in Proceedings of the 1st Symposium on Simulation of Computer Systems, ANSS '73, 200–207, IEEE Press, 1973, doi:10.1109/HICSS.1994.323170.

[18] D. Singh, O. G. Aliu, M. Kretschmer, "LoRa WanEvaluation for IoT Communications," in 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 163–171, 2018, doi:10.1109/ICACCI.2018.8554713.

[19] B. Reynders, S. Pollin, "Chirp spread spectrum as a modulation technique for long range communication," in 2016 Symposium on Communications and Vehicular Technologies (SCVT), 1–5, 2016, doi:10.1109/SCVT.2016.7797659.

[20] F. Van den Abeele, J. Haxhibeqiri, I. Moerman, J. Hoebeke, "Scalability Analysis of Large-Scale LoRaWAN Networks in ns-3," IEEE Internet of Things Journal, **4**(6), 2186–2198, 2017, doi:10.1109/JIOT.2017.2768498.

[21] S. Faruque, Radio frequency propagation made easy, Springer, 2014.

[22] J. Deygout, "Multiple knife-edge diffraction of microwaves," IEEE Transactions on Antennas and Propagation, **14**(4), 480–489, 1966, doi:10.1109/TAP.1966.1138719.

[23] R. ITU, "ITU-R P. 525-3," Calculation of Free-Space Attenuation, ITU-R Recommendation, **2016**.

[24] P.-. ITU-R, "A path-specific propagation prediction method for point-to-area terrestrial services in the VHF and UHF bands," 2014.

[25] C. Lee, S. Park, "Diffraction Loss Prediction of Multiple Edges Using Bullington Method with Neural Network in Mountainous Areas," International Journal of Antennas and Propagation, **2018**, 2018, doi:10.1155/2018/8737594.

[26] D. Har, A. M. Watson, A. G. Chadney, "Comment on diffraction loss of rooftop-to-street in COST 231-Walfisch-Ikegami model," IEEE Transactions on Vehicular Technology, **48**(5), 1451–1452, 1999, doi:10.1109/25.790519.

[27] J. Milanović, G. Šimac, K. Mazor, "Performance analysis of empirical propagation model for long term evaluation (LTE) network," in 2016 International Conference on Smart Systems and Technologies (SST), 69–74, 2016, doi:10.1109/SST.2016.7765635.

[28] P. Frangos, D. Jaggard, S. Sautbekov, G. Georgiev, S. Savov, "Propagation of electromagnetic waves in terrestrial environment for applications in wireless telecommunications," 2014, doi:10.1155/2014/297082.

[29] W. J. Krzysztofik, "Radio Network Planning and Propagation Models for Urban and Indoor Wireless Communication Networks," 2018.

[30] H. Lucas Jr, "Performance evaluation and monitoring," ACM Computing Surveys (CSUR), **3**(3), 79–91, 1971, doi:https://doi.org/10.1145/356589.356590.

[31] G. T. Nguyen, R. H. Katz, B. Noble, M. Satyanarayanan, "A trace-based approach for modeling wireless channel behavior," in Proceedings of the 28th conference on Winter simulation, 597–604, 1996, doi:10.1145/256562.256764.

[32] P. Owezarski, N. Larrieu, "A trace based method for realistic simulation," in 2004 IEEE International Conference on Communications (IEEE Cat. No.04CH37577), 4, 2236–2239 Vol.4, 2004, doi:10.1109/ICC.2004.1312915.

[33] G. Hampel, D. Abush-Magder, A. Diaz, L. M. Drabeck, M. J. Flanagan, J. M. Graybeal, J. D. Hobby, M. MacDonald, P. A. Polakos, J. Srinivasan, H. Trickey, L. Zhang, G. Rittenhouse, "The new paradigm for wireless network optimization: a synergy of automated processes and human intervention," IEEE Communications Magazine, **43**(3), S14–S21, 2005, doi:10.1109/MCOM.2005.1404593.

[34] G. T. Nguyen, B. Noble, R. H. Katz, M. Satyanarayanan, "A trace-based approach for modeling wireless channel behavior," in Proceedings Winter Simulation Conference, 597–604, 1996, doi:10.1145/256562.256764.

[35] C. Ku, Y. Lin, Y. Lai, P. Li, K. C. Lin, "Real traffic replay over WLAN with envi-ronment emulation," in 2012 IEEE Wireless Communications and Networking Conference (WCNC), 2406–2411, 2012, doi:10.1109/WCNC.2012.6214199.

[36] H. Bai, M. Atiquzzaman, "Error modeling schemes for fading channels in wireless communications: A survey," IEEE Communications Surveys Tutorials, **5**(2), 2–9, 2003, doi:10.1109/COMST.2003.5341334.

[37] G. Judd, P. Steenkiste, "Using Emulation to Understand and Improve Wireless Networks and Applications," NSDI'05, 203–216, USENIX Association, USA, 2005, doi:10.5555/1251203.1251218.

[38] M. Wellens, M. Petrova, J. Riihijarvi, P. Mahonen, "Building a better wireless mousetrap: need for more realism in simulations," in Second Annual Conference on Wireless On-demand Network Systems and Services, 150–157, 2005, doi:10.1109/WONS.2005.6.

[39] MultiTech, "Long Range LoRa® Modules (MTDOT Series)," https://www.multitech.com/brands/multiconnect-mdot, 2019, [Online; accessed 3-November 2020].

[40] LoRa®Alliance, "Wireless and sensing products datasheet," https://www.semtech.com/uploads/documents/sx1301.pdf, 2017, [Online; accessed 25-September-2019].

[41] S. W. Sherman, "Trace driven modeling: An update," ACM SIGSIM Simulation Digest, **7**(4), 87–91, 1976, doi:10.1145/1013610.807302.

[42] S. W. Sherman, J. Browne, "Trace driven modeling: Review and overview," in Proceedings of the 1st symposium on Simulation of computer systems, 200–207, 1973, doi:10.5555/800241.807240.

[43] C. J. Willmott, K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," Climate research, **30**(1), 79–82, 2005, doi:0.3354/cr030079.

[44] T. Chai, R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature," Geoscientific model development, **7**(3), 1247–1250, 2014, doi:https://doi.org/10.5194/gmd-7-1247-2014.

[45] R. J. Peckham, J. Gyozo, Digital terrain modelling, Springer, 2007.

Acronyms

IoT	Internet of Things
LPWAN	Lower Power Wide Area Network
RF	Radio Frequency
ITU-R	International Telecommunication Union
COST-WI	COST (European Cooperation in Science and Technology - Walfish Ikegami
LoRaWAN	Long-range, Low-Power Wide Area Network
ITU-R	International Telecommunication Union
LOS	Line-of-Sight
NLOS	Non Line-of-Sight
RSS	Received Signal Strength
ICS-Telecom	Image Cartography System - Telecommunication
SF	Spreading Factor
MAE	Mean Absolute Error
SD	Standard Deviation
RMSE	Root Mean Square Error
EM	Electromagnetic Waves
EAR	Automata- Synchnronised Display
WiLinkSim	Wireless Link Simulator
DTM	Digital Terrain Model

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 221-227 (2020)
www.astesj.com *Special Issue on Multidisciplinary Innovation in Engineering Science & Technology*

ASTES Journal
ISSN: 2415-6698

Multi-Directional Light Sensing Using A Rotating Sensor

Hoang Anh Dung[*,1], Nguyen Manh Cuong[2], Nguyen Phan Kien[3]

[1]*Faculty of Electronics Technology and Communication, Hanoi Open University, Hanoi, 100000, Vietnam*

[2]*Military Technical Academy, Hanoi, 100000, Vietnam*

[3]*School of Electronics and Telecommunications, Hanoi University of Science and Technology, Hanoi, 100000, Vietnam*

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 22 October, 2020
Online: 08 November, 2020:

Keywords:
Light Intensity Sensors
Smart Sensor
The Internet of Things
Intelligent Lighting Control
Control and Automation

ABSTRACT

The performance of indoor illumination control in many applications, such as in an intelligent building, relies on the quality of the light sensors. In many cases, the light level is not uniform and depends on the direction of the illumination source. It usually requires multiple sensors set up in different directions to gather the overall light level. We propose a system that can provide multi-directional light sensing data by rotating a single sensor. This approach overcomes the problem of static sensors network by dynamically changing the measuring angular of the light sensor. We present a sensor system prototype using the ESP8266 controller board, BH1750 light sensor, stepper motor, and 3D printed rotation base mechanism. The system can calculate the sensing angle and transmit sensing data to the monitoring unit or Internet of Things platforms for visualization and analysis. The testing results in normal workrooms show that the rotating sensor can measure the light level in different directions and detect the direction of the main illumination source. Even blocking some directions, the sensor still is able to accurately measure and provide sensing information on the remaining directions. Our sensor system is useful in both whole lighting and local lighting control applications.

1 Introduction

Light sensing information has a vital role in illumination control applications as in industrial lighting or a sensor-actuator (S-A) building. The characteristics and capability of light sensors highly affect lighting control performance. The indoor illuminance depends on both artificial and natural light. The light measuring is not just based on the photodetector in the sensor, but also other factors like sensor direction and occupant activities. Typically, the S-A system uses multiple light sensors or a network of sensors that are located distributively. As mentioned in [1], there are challenges in the use of a network that contained a large number of sensors such as unreliable sensor nodes (fault tolerance), difficult to add a new sensor or to change sensors direction (scalability), and late respond of light changing events (low latency). For a single sensing node, the sensor could be blocked by unaware objects or illuminated by an offensive light beam (not the lighting source), causing an inaccurate measurement. We proposed in [1] a rotating light sensor system that helps to reduce these issues. This paper presents a new design of the system with more details of the evaluation results.

The system proposed in [1] used a single light sensor rather than multiple sensors, and the sensor can rotate freely to measure the light illuminance from numerous directions. The sensor also recognizes the measuring angle as a part of the sensing data. In [1], we introduced a demonstration version of the sensor system that had a relatively large size. We present here a new version with a reduction in the overall size. For validation purposes, we add a static sensor to the system. The significant advantage of our system is that it is able to provide light sensing data at different angles using just a single sensor compared to a system that uses multiple sensors. The rotating sensor can detect the temporary light changing at a specific angle or determine which measuring direction may have been blocked. These features allow the sensor to accurately tracking the light information in real indoor applications. The system can also work on the Internet of Things (IoT) [2] applications since it can connect and transfer data to the popular IoT platforms.

The paper is organized as follows. The following section provides the related works. The methodology, design, and features of the rotating sensor system are presented in Section 3. The performance analysis and experimental results are provided in Section 4. The last section concludes the paper.

[*]Corresponding Author: Hoang Anh Dung, Email:dungha@hou.edu.vn

2 Related Work

Light sensing data is used in the illumination control system is often come from either a single sensor or a network of sensors. The use of light sensors networks has become more popular, especially in medium and larger size applications. Using a single sensor may not provide enough sensing information, while a network of sensors has the challenges of fault tolerance [3], scalability [4], and low latency [5]. We have observations about these challenges based on our work in light sensing and lighting control research. Firstly, the fault tolerance feature in the light sensors network requires the system to work correctly when some sensors may have been failed. It is difficult to satisfy this requirement when the light sensors are distributed at a different location. For the scalability, there is a challenge in the application that requires to increase illumination areas such as in the commercial center or industrial working floors. In such a case, the component cost would increase, and the sensor system should be portable enough to adapt to the extensions. Lastly, the static sensor system may have a low latency issue system when it places the sensor in a fixed direction and orientation. The static system may not quickly sense the changing of illumination level in the zones uncovered by the sensor. In [6], the lighting control system used the wireless sensor networks data to control the illumination to reduce the energy costs. A significant deployment and maintenance costs of the sensors was reported. In addition to the components cost, there was another complexity in the controlling system related to the variation of the light level caused by the sunlight [6]. The study in [7] indicated the effect of other natural factors like sun movement, clouds, and shadows on changing the light levels in the testing applications. The effect of distance and direction of the sensors and light sources on the sensing data was studied in [8]. They studied the controlling of both overall and local illumination. The controlling plan uses the local light sensing data to provide illumination upon the user's needs. A limitation is that the system requires a light sensor to be carried with the user to measure the local light level. These local measurement data were used in the illumination controlling plan to meet the user requirement.

For the location-based illumination control, the whole lighting provides background light, and the local lighting provides concentration light [9]. In this system, the control system uses light sensing data to control both background and concentration illumination. The work in [10] had evaluated the effect of the local light sensing on controlling the whole lighting devices. It also requires to determine the location zone of the user to provide the personalized light control [11, 12].

In [13], the author studied the distributed light control system to accommodate both the user occupancy and the natural daylight level. A large number of light sensors were used in their system to examine the relationship between the occupation bases and the relevant light sensors. The testing result also indicates the dependence of the light control performance on the accuracy of the local and whole light sensing information.

We have not seen such a system using a rotating sensor in light sensing applications. In the field of radio tomography, a study presented in [14] introduced a rotating RF sensor system to sense the radio system from multiple directions and improve the accuracy of tomographic imaging. As we stated in [1], since the light ambient is heterogeneous, it is necessary to create a light sensor system that can measure light from different angles and reduce the problem of multiple sensor networks. Such a system is useful to increase the quality of intelligent lighting control applications.

3 Design of the new Sensor System

The rotating sensor motioned in [1] was a demo prototype. We have designed a new version of the sensor system that could be used as a real measurement tool. An image of the new sensor system is presented in Figure 1. The new design took care of the following factors:

- Improving the rotation system.
- Reduce the distance between the sensor and the rotation module.
- Optimizing the overall diameter of the sensing system.
- Adding a static sensor for referencing.

3.1 Multi-Directional Light Measurement Vector

Consider a visible light source s, the Lumens from this source can be calculated based on the I_s intensity [15] as:

$$\Phi_s = 4\pi I_s. \tag{1}$$

For the isotopic light source, where the intensity is the same in all directions, the measurement of the luminous flux on a unit area, also known as or light level in a specific area or illuminance, is the ratio of the Lumens Φ over the area of the measuring surface. However, in the indoor environment where the source is not isotopic, for a flat radiating surface of a sensor, the calculation of the illuminance need to consider cosine the observation angle (θ) with respect to the surface normal as shown in the Equation (2)

$$E = \frac{\Phi_s \cos\theta}{A}. \tag{2}$$

In the lighting control system, the observed ΔE, changing of illuminance on the light sensor surface, are used to control light bulb brightness:

$$\Delta E = \frac{\Delta\Phi_s \cos\theta}{A}. \tag{3}$$

In our proposed system, the light level is measured in multiple directions. Hence, we use a vector of illuminance as:

$$\mathbf{E} = \begin{bmatrix} E_0 \\ E_1 \\ \vdots \\ E_{N-1} \end{bmatrix}, \tag{4}$$

where N is the number of the measuring directions and E_n is the measured illuminance at the direction $n \in [0, N-1]$.

There are two major advantages of the sensing data provided by a rotating light sensor comparing to a static sensor. First, the light intensity measured by a static sensor depends on a single sensing

direction, the angle θ as in Equations (2) and (3), while the rotating sensor provides a combination of illuminances from multiple directions. In the real application, such as lighting control, the measuring direction could be blocked by an object that makes the sensing value of the static sensor inaccurate. The rotating sensor can be able to void this issue as it measures the light from different angles. Second, the rotating sensor uses a vector of illuminance (4) that can indicate the main light source direction. Furthermore, we can apply the gradient calculation (∇) on the vector of illuminance to measure the rate and direction of change in the light intensity.

3.2 *Features of the Rotating Sensor System*

The new sensor system is shown in Figure 1 has an overall diameter of 100 mm. This design is small enough to use in most indoor applications. Table 1 summarizes the features of the sensor system. The system uses one light sensor to dynamically measure multi-directional light intensity. In this system, there is no need to use many sensors as in a sensor network. By rotating the light sensor, our system may avoid the issues when the static measuring angle has been blocked by an object.

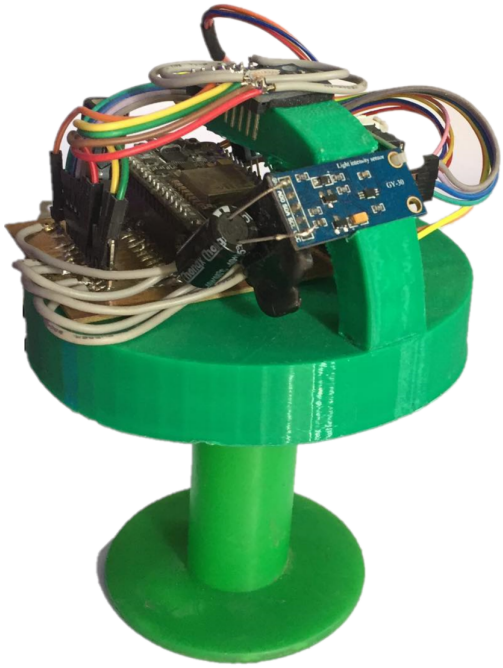


Figure 1: A sample image of the new rotating light sensor system. The light sensor is attached to the side of the curved handle of the rotation base structure.

Table 1: Features of the sensor system

System Features	Value
Light Sensor	BH1750
Stepper Motor	28BYJ-48
Stepper Motor driver board	ULN2003
Microcontroller Module	ESP8266
Overall Dimension (mm)	100
No. of Measuring Directions	8 and 16
Sensor Measuring Resolution (Lux)	1 (high) or 4 (low)

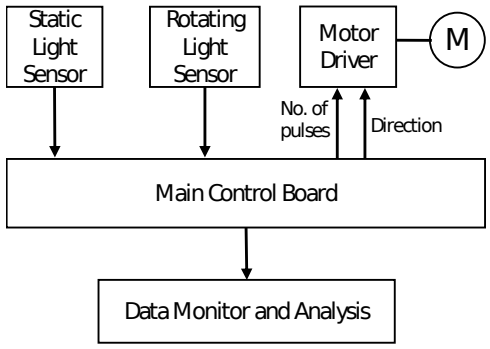


Figure 2: Diagram of the sensor system.

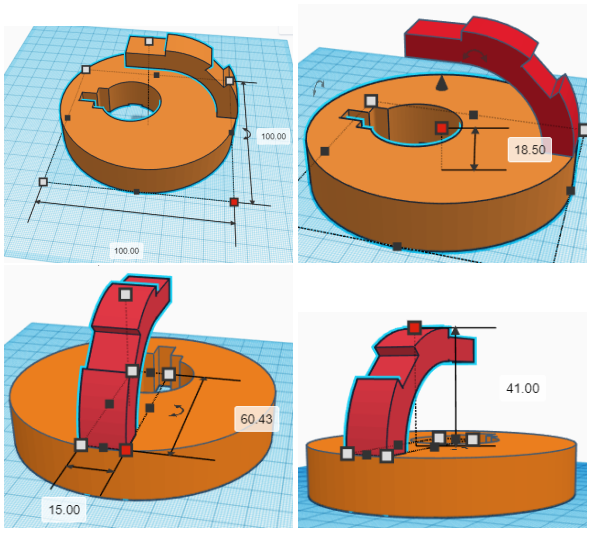


Figure 3: Sample views of the rotation base structure design.

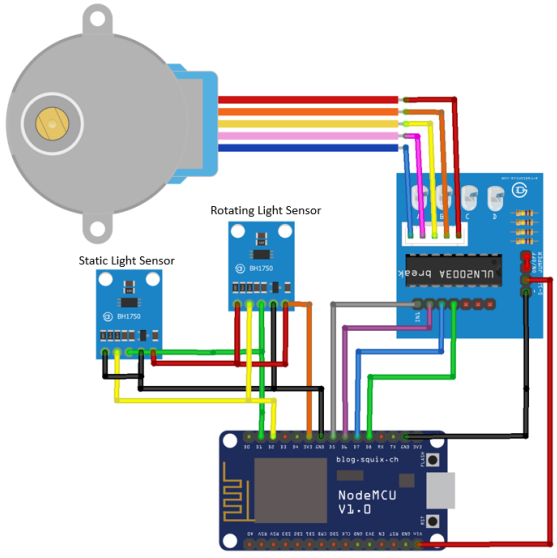


Figure 4: Conceptual logic circuit of the sensor system.

As shown in the system diagram in Figure 2, the major components of the system are the light sensor, stepper motor, motor diver board, and the main control board. The system uses a micro-

controller to control the rotation, sensing the light illuminance, and transfer data to the monitoring platform. In this design, we have improved the rotation system and created a new base structure of the sensor. Figure 3 shows several views our the new base structure. The structure was 3D printed and used in the sensor system.

As a modified version in [1], the new system uses an additional static sensor to provide a reference as requested. In this study, we consider the static sensor is the sensor that has a fixed sensing surface normal vector, thus the observation angle θ for any lighting source stays unchanged (see Equation 2). The rotating sensor is the sensor that its observation angle can be changed. By rotating the sensor in an axis different than the normal vector of its sensing surface, we can change its observation angle.

The logic circuit of the new system is shown in Figure 4. The sensors communicate with the microcontroller using the I^2C protocol. Both static and rotating sensors are the BH1750 [16] digital light intensity measuring module. The sensor module has an internal converter to convert the analog signal to the digital value. The module also contains a preprocessor to process the digital value and return the measuring value in the range of 1 to 65535 Lux.

The circuit uses a ULN2003 stepper motor driver to provide a drive interface between the microcontroller and the stepper motor. The ULN2003 driver pins, IN1, IN2, IN3, and IN4, are connected to the microcontroller digital pins. The power supply pins (VCC and GND) of the ULN2003 driver board are connected to the microcontroller 5V pin and GND pin, respectively. The stepper motor, model 28BYJ-48, runs in 32 steps mode and has a built-in gear with a 64 to 1 reduction ratio to produce 2048 steps per revolution. The step-angle relation is calculated as:

$$\alpha = \frac{i \times 360}{2048}, \tag{5}$$

where α represents the angle, and i is the number of controlling step of the motor. Sample images of the motor and the driver board were presented in [1]. In this new design, the motor is attracted to the center hole of the base structure shown in Figure 3. The motor can rotate the whole structure hence rotate the light sensor. For the microcontroller, we used the ESP8266 [17] board that runs in the Arduino environment. This module has a Wi-Fi microchip with a full TCP/IP stack to transfer the data. We program the system in the Arduino code using the C++ programming language.

Table 2: Relation of the sensing directions, angle interval, and motor steps interval

No. of Sensing Directions	$\Delta\alpha$	Steps Interval
8	45°	256
16	22.5°	128
32	11.25°	64
64	5.625°	32

3.3 Measuring Multi-Directional Light Illuminance

The main procedures of the illuminance measurement are kept similar to in [1] except for the additional signal of gathering the static sensor data. As mention in [1], for the applications that require the data of light intensity in different directions, the users may have to manually set up several sensors at different orientations. We developed an algorithm to automatically rotate the sensors, calculate the angles, and gather the illuminance in multiple directions. Figure 5 presents the major steps of the algorithms. In the beginning, the program set up the initial direction of the sensor. The sensor can rotate in a full 360° revolution, hence any angle can be set as the initial sensing direction. Then the sensor system starts collecting the angle and illuminate data. In each iteration of the major loop shown in Figure 5, the control program rotates the sensor by increasing the motor step, calculates the sensing angle, gets the light level information, and transmits the data to the monitoring software. The counting of the motor steps is used to calculate the angle based on the Equation (5). When the motor reaches a full circle (at 2048 counted steps), the counting is reset to Zero, and the rotating direction is reversed. The sensor system can measure light in any direction or at a group of interval directions. Table 2 summarizes some combinations of the sensing directions, angle interval ($\Delta\alpha$), and the motor steps interval. The visualizations of 8 sensing directions with 45° apart and 16 directions with 22.5° apart are shown in Figure 6.

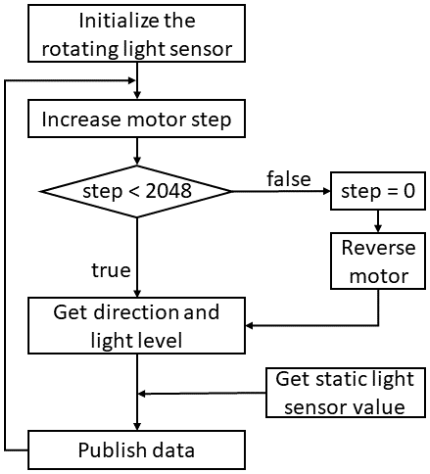


Figure 5: Flowchart of the control algorithm of the light sensor system. The major procedures are to control the rotating light sensor. The command to get static light sensor data is optional.

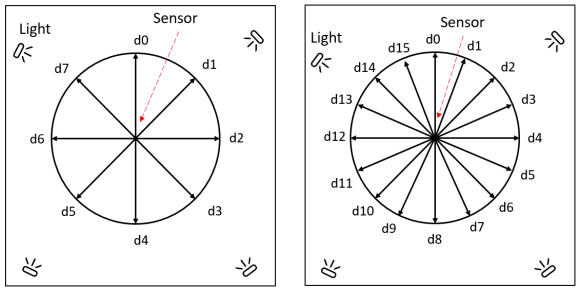


Figure 6: Sample views of the rotating light sensor with different number sensing angles. **Left**: 8 directions from $d0$ to $d7$ with $\Delta\alpha = 45°$. **Right**: 16 directions from $d0$ to $d15$ with $\Delta\alpha = 22.5°$. The lights are for presentation purpose only.

4 Performance Analysis of the Sensor System

Indoor lighting control is the major application of our proposed sensor system. We have tested the system in a workroom with the effect of the popular parameters on the sensor, like the changing of light power and blocking object. We present here the experiments

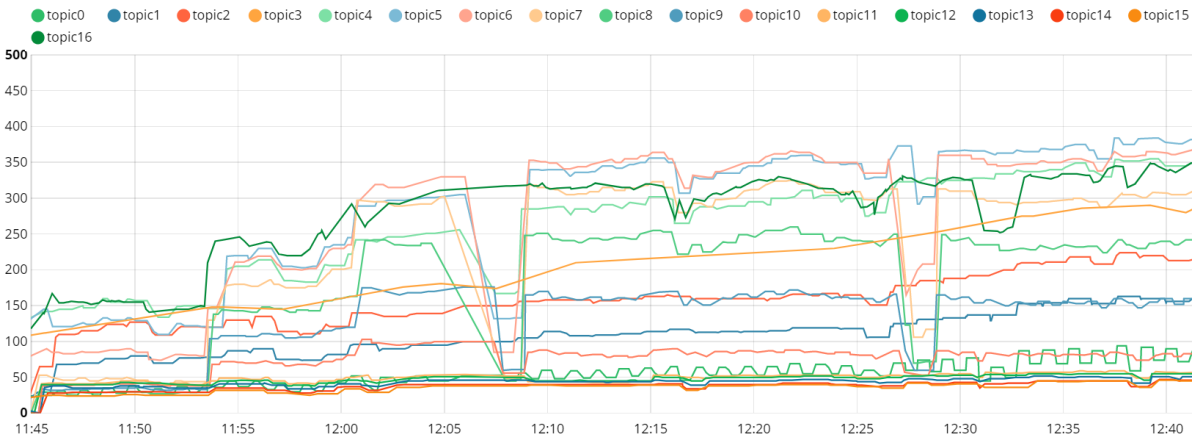


Figure 7: Snapshot of monitoring the light sensing data on the Ubidots website.

and results for testing the ability of the system to detect the main light direction, detect the blocked direction, and accurately measure light intensity.

4.1 Measuring and Transmitting Light Data

In [1], we presented the experiment that showed how the sensor measures the light level from multiple directions. For this new design, in addition to the multi-directional light sensing data, the system also collects the sensing data of a static reference sensor. We tested our device on transferring the data to the Ubidots [18] IoT platform. The data was transferred remotely through the Wifi signal. Figure 7 shows a part of the testing data monitored on the Ubidots website with multiple light sensing signals in one plot. We also have tested our system with the ThingSpeak [19] platform that has integrated MATLAB toolbox for analyzing the data.

For the data presented in Figure 7, the rotating sensor measured the light in 16 directions with $\Delta\alpha = 22.5°$ (Table 2 contains more detail of the measuring angles). It can be seen that our system is able to measure and provide light sensing data from multi-directional. In Figure 7, the rotating sensing data is labeled as a topic channel for each direction from *topic*0 to *topic*15. The static sensor channel is labeled as *topic*16. When monitoring the data, the channel topics (directions), the *Lux* values, and the measuring times can interactively be seen by selecting a data point on the graph. As shown in the figure, different sensing direction has a different value. In the beginning, when all lights were off, all sensing data was low (roughly under 150 *Lux*). Then, after we turned on all the lights (around 11:55 am timestamp, see Figure 7), the system recorded the increase of light levels in all directions. We blocked some directions of the rotating sensor at two later movements. As of the result, there were two failings of the light level at the blocked directions (in the 12:05 - 12:10 pm and the 12:25-12:30 pm periods, see Figure 7). In this experiment, the static sensor (*topic*16) was not blocked at any time, and it was used to provide the reference sensing data.

4.2 Detecting Main Light Direction

The distribution of light sources may generate an unbalance light level in the room. Our sensor can detect the major light direction that information helps to rearrange the illumination sources or indicate the location having the best light level. In this experiment, we place the rotating sensor at the center of a workroom that has several light sources. The light sources were placed around the room at different locations and distances to the sensor. We regularly turn on the lights to test the sensor performance. The sensor was set to measure in 16 directions ($d0 \ldots d15$, with $\Delta\alpha = 22.5°$). Figure 8 displays some radar charts of the measured results. For each chart, the data value as a length on angular spokes represents the magnitude of sensing value in a corresponding direction. The line connects the data values represses one full rotating revolution (360°).

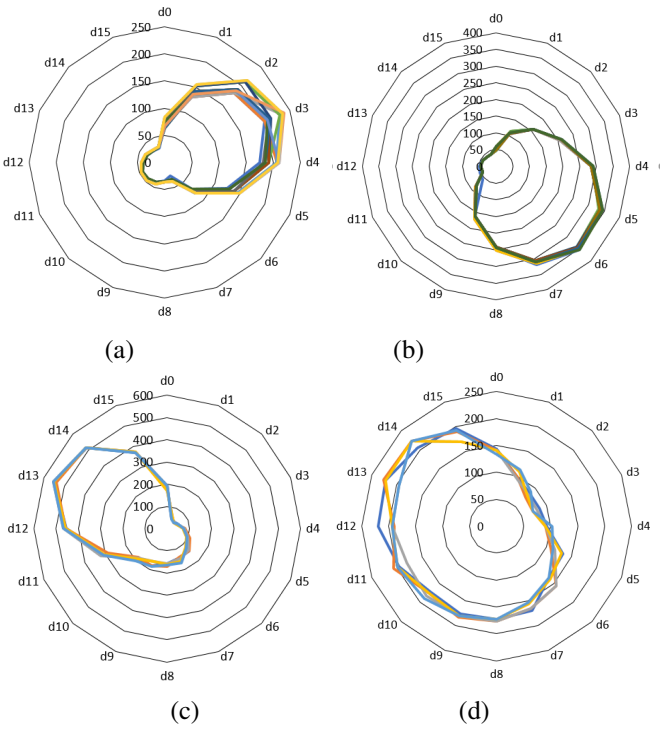


Figure 8: Radar charts of light sensing data on different main illumination directions.

The charts in Figure 8 associate with four scenarios of the lighting in the testing room. For an instant, Figure 8 (a) shows the *Lux* values measured in directions $d2$, $d3$, and $d4$ are much higher than the rest. In that case, only the light source at the direction of $d3$ was on, hence there is a peak sensing data in $d3$ (about 250 *Lux*) and two relevant highs on the adjacent angles $d2$ and $d4$. Figure 8 (b) shows the major level of light in $d5$, $d6$, and $d7$ with the peak value

of 350 *Lux*, and Figure 8 (c) shows the main level in $d12$, $d13$ and $d14$ with above 500 *Lux* peak value. These results plotted in the radar charts admit that the rotating sensor can detect the direction of the main light source. This result also can be observed from our vector of the illuminance shown in Equation 4 since the direction of the vector is influenced by the major scalar member values. Figure 8 (d) presented a test case where the light level is more consistent on the left side of the sensor ($d9$ to $d15$). In the case of a uniform light level, we may obtain consistent light sensing values in all directions.

4.3 Detecting Blocked Directions

We have performed an experiment to show how the rotating sensor can detect the temporary blocking directions with sudden changes in the sensing data. During this test, the rotating sensor was set up to run first, and then some sensing directions were temporarily blocked. Figure 9 plots the testing result. Before the time $T1$, the sensor indicates the major light level in directions $d4$, $d5$, $d6$, $d7$ and $d8$ with the high value above 300 *Lux* ($d6$). In the period of $T1$ and $T2$, when the sensor was blocked from the major light source direction, there was a sudden fall in the sensing value of the relevant angles. As plotted in Figure 9, all sensing values of $d4$ to $d8$ directions have steeply decreased to around 50 *Lux* while other values mostly unchanged. After the $T2$ moment, when the blocking object moved out, the sensing values from the main light directions got back to normal. It can be seen that we can use the directions having a temporary steeply reduction to indicate which sensing angle of the sensor was blocked. In the experiment presented in [1], we reported the result of detecting a sudden change of the natural background light, such as when opening the doors to get more sunlight.

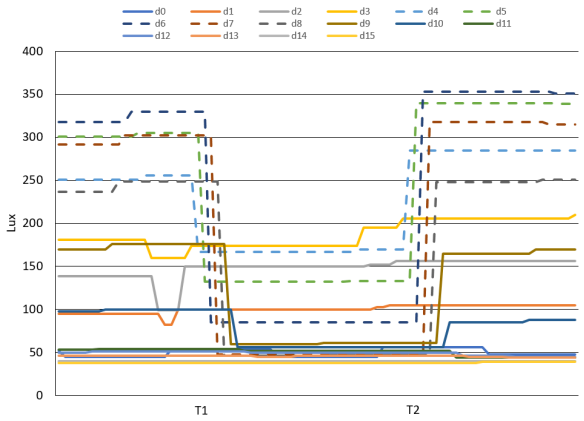


Figure 9: Sensing data of the rotating sensor when blocking some sensing angles.

4.4 Accuracy Analysis

The demand level of illumination to be controlled depends on the change in the sensing value. When using a single sensor or a network of sensors that was set at a static direction, it is possible that an unaware object temporarily blocks the sensor. Such a case makes a temporary change in the measuring data, leading to an inaccurate controlling of the illumination. The rotating sensor provides light sensing values from different directions that can help to avoid this problem.

We performed an experiment that used both static and rotating sensor. The artificial lighting sources were kept unchanged, but the static sensor would be blocked at a certain time during the experiment. Figure 10 shows the *LUX* values of the static sensor at a fixed angle and the rotating sensor rotated in eight directions (from $d0$ to $d7$). In this experiment, the rotating sensor measures light at every 45°. In the beginning, the static sensor data was around 450 *Lux* (±20), while the rotating sensor had different values at different directions with the high value in $d4$ of above 500 *Lux* and $d3$ around 430 *lux*. The directions $d0$, $d5$, and $d7$ had sensing values in between 320 and 350 *Lux*, and the direction $d1$, $d2$, and $d6$ had data in the range of 270 to 300 *Lux*. Then, we blocked the measuring direction of the static sensor. As shown in Figure 10, there is a sharp fall in the value of the static light sensing data down to 100 *Lux* and lower. If an application used just the static sensor to control the lights, it would incorrectly record a reduction in the light level and would control an unnecessarily increase of the illumination power. There was no need to change the illumination since the lights were stable throughout this experiment (except for a small variance in the natural light). For the rotating sensor, the sensing signals in the directions $d0$, $d1$, $d6$, and $d7$ were almost stable (around 300 *Lux* and 350 *Lux*). There was a small reduction in the direction $d2$ and $d5$ and a quire decreasing in the direction $d3$ and $d4$, but these sensing values were still around 250 *Lux*. When we blocked the static sensor, we approached the sensor location through the direction $d3$ and $d4$ that caused the changing of the data in these (and surrounding) directions. If we use the rotating sensing data of multiple directions for the lighting application, the control unit could use the data of the stable directions and decide to keep the illumination level the same. The prevention of an unnecessary increase in illumination power is useful to save energy costs.

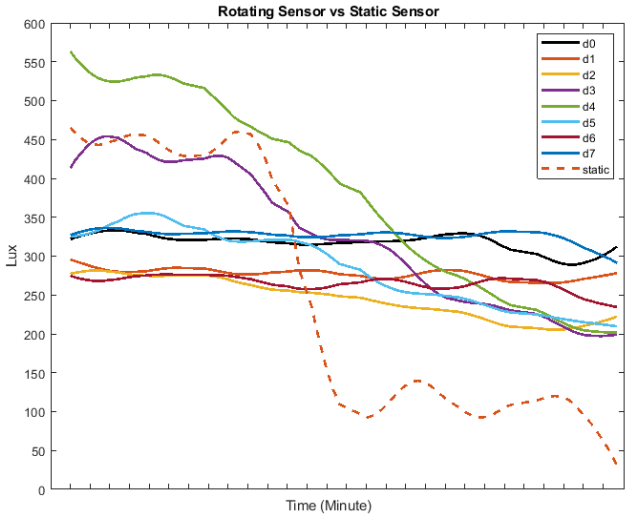


Figure 10: Light sensing values of the static sensor and the rotating sensor. The static sensor was blocked at a certain time, causing a reduction in the sensing value. The time ticks in the horizontal exist are presented at 60 seconds interval.

5 Conclusion

We have presented a rotating light sensor system that can measure light level in multi-direction rather than using server sensors positioning at different orientations. The system uses a stepper motor to rotate the sensor and track the motor steps to calculate the measuring angle. The light level values can be directly transferred to the moni-

toring system for further process. Besides, the sensor can connect with the popular IoT platforms, such as Ubidots and Thingspeak, to remotely visualize and analyze the sensing data. We also presented a new design of the sensor system with an optimal compact structure. The performance of the sensor system has been evaluated through testing experiments. The results show that the proposed sensor system is able to measure light level in different directions, detect the main light source direction, and determine the directions having temporary light changing. Moreover, even some sensing directions may be blocked, the rotating sensor still accurately measures the light level in other directions. In the future, we intend to develop a three-dimensional (3D) rotation light sensor. A motion tracking sensor shall be used to keep track of the 3D orientation. We plan to optimize the rotation system and perform more experiments of using the rotating light sensor in the light control applications.

References

[1] D. A. Hoang, T. T. Tung, C. M. Nguyen, K. P. Nguyen, "Rotating Sensor for Multi-Direction Light Intensity Measurement," in 2019 International Conference on System Science and Engineering (ICSSE), 462–467, 2019, doi: 10.1109/ICSSE.2019.8823447.

[2] A. P. Plageras, K. E. Psannis, C. Stergiou, H. Wang, B. B. Gupta, "Efficient IoT-based sensor BIG Data collection–processing and analysis in smart buildings," Future Generation Computer Systems, **82**, 349–357, 2018, doi: https://doi.org/10.1016/j.future.2017.09.082.

[3] S. Chouikhi, I. El Korbi, Y. Ghamri-Doudane, L. A. Saidane, "A survey on fault tolerance in small and large scale wireless sensor networks," Computer Communications, **69**, 22–37, 2015, doi:https://doi.org/10.1016/j.comcom.2015.05.007.

[4] C. Dandelski, B. Wenning, D. V. Perez, D. Pesch, J. M. g. Linnartz, "Scalability of dense wireless lighting control networks," IEEE Communications Magazine, **53**(1), 157–165, 2015, doi:10.1109/MCOM.2015.7010529.

[5] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, H. Dai, "A survey on low latency towards 5G: RAN, core network and caching solutions," IEEE Communications Surveys & Tutorials, **20**(4), 3098–3130, 2018, doi:10.1109/COMST.2018.2841349.

[6] V. Singhvi, A. Krause, C. Guestrin, J. H. Garrett Jr, H. S. Matthews, "Intelligent light control using sensor networks," in Proceedings of the 3rd international conference on Embedded networked sensor systems, 218–229, ACM, 2005, doi:https://doi.org/10.1145/1098918.1098942.

[7] J. Lu, D. Birru, K. Whitehouse, "Using Simple Light Sensors to Achieve Smart Daylight Harvesting," in Proceedings of the 2Nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, BuildSys '10, 73–78, ACM, New York, NY, USA, 2010, doi:10.1145/1878431.1878448.

[8] L. Yeh, C. Lu, C. Kou, Y. Tseng, C. Yi, "Autonomous Light Control by Wireless Sensor and Actuator Networks," IEEE Sensors Journal, **10**(6), 1029–1041, 2010, doi:10.1109/JSEN.2010.2042442.

[9] M. Pan, L. Yeh, Y. Chen, Y. Lin, Y. Tseng, "A WSN-Based Intelligent Light Control System Considering User Activities and Profiles," IEEE Sensors Journal, **8**(10), 1710–1721, 2008, doi:10.1109/JSEN.2008.2004294.

[10] D. Caicedo, A. Pandharipande, "Distributed Illumination Control With Local Sensing and Actuation in Networked Lighting Systems," IEEE Sensors Journal, **13**(3), 1092–1104, 2013, doi:10.1109/JSEN.2012.2228850.

[11] K. Warmerdam, A. Pandharipande, "Location data analytics in wireless lighting systems," IEEE Sensors Journal, **16**(8), 2683–2690, 2015, doi: 10.1109/JSEN.2015.2509982.

[12] X. He, A. Pandharipande, "Location-Based Illumination Control Access in Wireless Lighting Systems," IEEE Sensors Journal, **15**(10), 5954–5961, 2015, doi:10.1109/JSEN.2015.2449276.

[13] N. van de Meugheuvel, A. Pandharipande, D. Caicedo, P. van den Hof, "Distributed lighting control with daylight and occupancy adaptation," Energy and Buildings, **75**, 321 – 329, 2014, doi:https://doi.org/10.1016/j.enbuild.2014.02.016.

[14] M. Bocca, A. Luong, N. Patwari, T. Schmid, "Dial it in: Rotating RF sensors to enhance radio tomography," in 2014 Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), 600–608, 2014, doi:10.1109/SAHCN.2014.6990400.

[15] C. Nave, "Hyperphysics," 2020.

[16] M. electronics, "BH1750 ROHM Semiconductor Datasheet," 2020.

[17] E. Systems, "ESP8266 Overview," 2020.

[18] Ubidots, "Ubidots Internet of Things and Cloud tools," 2020.

[19] thingspeak, "ThingSpeak IoT analytics platform," 2020.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 228-238 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Vehicle Rollover Detection in Tripped and Untripped Rollovers using Recurrent Neural Networks

Kailerk Treetipsounthorn[1], Thanisorn Sriudomporn[2], Gridsada Phanomchoeng[*,2,3], Christian Dengler[4], Setha Panngum[1], Sunhapos Chantranuwathana[2,5], Ali Zemouche[6,7]

[1]*Department of Computer Engineering, Chulalongkorn University, Bangkok, 10330, Thailand*

[2]*Department of Mechanical Engineering, Chulalongkorn University, Bangkok, 10330, Thailand*

[3]*Applied Medical Virology Research Unit, Chulalongkorn University Bangkok 10300, Thailand*

[4]*Chair of Automatic Control, Technical University of Munich, Munich, Germany*

[5]*Smart Mobility Research Unit, Chulalongkorn University, Bangkok 10300, Thailand*

[6]*University of Lorraine, CRAN UMR CNRS 7039, Cosnes et Romain, 54400, France*

[7]*EPI Inria DISCO, Laboratoire des Signaux et Systèmes, CNRS-CentraleSupélec, Gif-sur-Yvette, 91192, France*

ARTICLE INFO

Article history:
Received: 01 September, 2020
Accepted: 01 November, 2020
Online: 10 November, 2020

Keywords:
Rollover Index
Tripped Rollover
Untripped Rollover
Recurrent Neural Network
Feedforward Neural Network
Tanh
LSTM
GRU

ABSTRACT

Comparing to other types of vehicle accidents, fatality rate of tipped rollover accidents shows significant number. Thus, tripped rollover prevention systems are important in order to keep driver safe. In other hands, different rollover indices are defined to handle the risk. The variable unknown parameters of each index, for instance, current load of the vehicle or center of gravity, are considered as a difficulty. In this work, the recurrent neural networks, which are designed to work on sequential data in order to provide data estimation without additional estimation algorithm, are investigated in purpose to estimate the tripped and untripped rollover index. The vehicle simulation software with industrial standard CarSim is applied to validate the result. The Tanh recurrent neural network is stated in the result to be the most accurate tripped rollover index estimator for the uncertain parameters, for example, sprung mass and the height of the center of gravity. The suitable input features for tripped and untripped rollover index and neural network structure are verified. To prevent and provide warning of rollover, an advance future prediction can also be designed for the future tripped and untripped rollover prediction.

1. Introduction

This work is an extension based originally on the groundwork presented in the 9th IEEE International Conference on Robotics, Automation and Mechatronics (CIS-RAM 2019) [1]. This paper will focus on development of recurrent neural networks for tripped rollover while Ref. [1] focus on only untripped rollover. The number of rollover accidents has grown continuously according to the change in vehicles' size, which are both weight and dimension, over the last 10 years. The majority type of fatality on the road is pointed to rollover accidents. In 2010, 9.1 million cars, which includes passenger cars, SUVs, pickups and vans, crashed and 2.1% of those crashes were caused by rollovers collected by National Highway Traffic Safety Administration's records (http://www.safercar.gov). From those 2.1% of crashes, there were over 7,600 deaths which was up to 35% of total deaths from vehicle accidents [2]. In addition, 20% of vehicle-crash deaths in Europe are from the rollover [3]. In Germany, 2-5% and 90% of vehicle accidents and deaths from the accidents are caused by vehicle rollover accidents in 2005 respectively [4].

The vehicle rollover accidents are categorized as tripped and untripped rollovers. A rollover accident that have contacted with an external obstacle, such as a pothole or guardrail, is called a tripped rollover. Untripped rollovers are defined the other way

*Corresponding Author: Gridsada Phanomchoeng, Department of Mechanical Engineering, Chulalongkorn University, Bangkok, 10330 Thailand, gridsada.phanomchoeng@gmail.com

round. It happens because of the effect of high lateral acceleration on a dangerous turn, i.e. sharp turn, and it is unrelated to external input [5-6]. Tripped rollovers tend to happen more frequently than untripped rollovers with 95% and 5% of total rollovers according to the statistic provided by NHTSA [2]. From the significant number of deaths from rollovers compared to the total number of high-way traffic deaths, vehicle safety must be improved to prevent the vehicle from rollover accidents.

Hence, several automotive manufacturers, e.g., Ford and Volvo have advanced with active rollover prevention systems which aim to prevent untripped rollover accidents [7]. Analysts and researchers likewise take interest largely in differential braking systems e.g. Anti-lock Braking Systems (ABS), Electronic Braking Systems (EBS) and Electronic Stability Programs (ESP) [3, 7-13]. The baking systems perform vehicle's yaw rate and speed reduction by adjusting each wheel's sequence of brake patterns. Also steer-by-wire technology [13-17] e.g. active front steering (AFS), active rear steering (ARS), or four-wheel steering (4WS) are to be used in case of the balancing of performances among yaw stability, violence, and rollover prevention and an active suspension system [18-20] can be used to avoid rollover when the wheels lifted off the road. Then, an active roll bar stabilization system [20-22] uses an active stabilizer for active torque distribution between front and rear axles of the vehicle. This can reduce the roll motion. In addition, active torque management [23] has the capability to prevent rollover accidents. Many control techniques are developed to enhance these systems.

Precise detection of vehicle rollover is mandatory in order for these active rollover prevention systems to fully function. To the present, there are multiple impending vehicle rollover detections developed with computational rollover index. The rollover indices are developed by various approaches. For example, Static Stability Factor (SSF) was developed to measure the rollover resistance of a given vehicle design configuration [24]; ref. [25] introduced rollover indices that consider kinetic energy, then Rollover Prevention Energy Reserve (RPER) is started off to look into rollover risks by taking energy function as the main account [7]; ref. [26] introduced a dimensionless rollover index that takes roll dynamics into consideration by utilizing phase plane analysis; zero-moment point also is developed to predict the possibility of rollover [9]; lateral Load Transfer Ratio (LTR) is popularly discussed in numerous studies [5, 7, 14-18]; to name a few. It is seen that a lot of rollover indicators are developed based on LTR such as Multiple Rollover Indices (MRIs) for trailer vehicles and Predictive Lateral Transfer Ration (PLTR) [9]. Lateral Load Transfer Ration is actually one of the approaches that take several factors into account, for example, roll angle, roll center and height of center of gravity of a vehicle. Many techniques have been established to estimate these factors. For example, extended Kalman filter technique or nonlinear observer/estimator has been built up to improve the accuracy of the lateral Load Transfer Ratio. Other than that, ref. [27] introduced an algorithm that utilizes the time-to-rollover (TTR) metrics to assess the time until rollover as a warning option. With this approach, an advance real-time prediction of rollover risk can be provided. Moreover, this technique can be used with several other rollover indices simultaneously. Despite a number of rollover detection techniques developed until now, only untripped rollovers are detected.

Ref. [5, 6, 28] has introduced a new rollover index which incorporates external road inputs as a factor to detect not only the untripped but also tripped rollover conditions. The external obstacle road inputs are estimated based on a vehicle dynamic model and nonlinear observer. The simulation and experimental results are presented using a 1/8th scaled vehicle.

Since the indicator to detect rollover risk is very important for improving the rollover prevention systems and rollover warning system [29-31], it is necessary to develop the rollover indicator for detecting both tripped and untripped rollovers. It is shown that the traditional rollover index is not effective enough to identify tripped rollovers and there are only few focused on tripped rollovers that are published. In addition, most accurate rollover indicators relay on dynamic models which require known parameters such as vehicle mass or height of center of gravity of a vehicle. Therefore, here proposed is a neural network indicator developed to identify tripped and untripped rollovers. Neural networks known for adaptable and nonlinear information processing capability perform fittingly in the areas of predication, expert system, and mode identification [32-37]. By taking the approach to be introduced afterward into account, the estimation algorithms to estimate unknown parameters such as roll angle, height of center of gravity of a vehicle, or vehicle mass are not required. These parameters may be changed all the time due to the number of passengers and load. Also, the technique can be applied in many situations. This research is the extended technique in [1, 38] which is focused on untripped rollover to detect tripped rollover.

This research has the following outline: Section 2 will introduce the vehicle rollover and the rollover index for detecting tripped and untripped rollovers. Next, the recurrent neural network technique for detecting tripped and untripped rollover is presented in Section 3, and the methodology to apply the recurrent neural network to the rollover problem is presented in Section 4. The process of evaluation of the suitable neural network, the suitable input features, the suitable neural network structure and the future tripped rollover prediction are included in this section. Then, Section 5 will present the tripped rollover results. Finally, Section 6 will discuss the conclusion to the findings.

2. Vehicle Rollover Index

Traditional rollover index posts as most popular technique to detect vehicle rollover among many techniques [5-7, 14-18, 28-33] used nowadays.

2.1. Traditional Rollover Index

In order to detect lift-off condition of the wheels [6, 7], the rollover index is used as a real-time indicator. Lateral Load Transfer Ratio (LTR) is known as another name of the index. The standard rollover index is designed for predicting incidents that are caused with the lift-off conditions of the wheels for untripped accidents. A number of researches use the basic model to derive their rollover indices. A scaled lateral acceleration, a function of roll angle [12], or a function of lateral acceleration and roll angle [9, 18] is considered as a base argument for these rollover indices.

The rollover index by the conventional definition is portrayed by

$$R = \frac{F_{zr} - F_{zl}}{F_{zr} + F_{zl}}, \quad -1 \le R \le 1 \tag{1}$$

where F_{zr} is right normal tire force and F_{zl} is left normal tire force. $R = 1$ ($F_{zl} = 0$) or $R = -1$ ($F_{zr} = 0$) is the value when a rollover incident is defined. F_{zr} equals to F_{zl} and $R = 0$ is when the vehicle travels straight forward. If $F_{zr} = 0$, then $R = 1$ and this will cause the lift-off occurrence of the left wheel. If $F_{zl} = 0$, then $R = -1$ and this will cause the lift-off occurrence of the right wheel.

Based on the one-degree-of-freedom model, the summation and difference of tire forces $F_{zr} + F_{zl}$ and $F_{zr} - F_{zl}$ can be calculated. An implementable rollover index processed with the one-degree-of-freedom model is demonstrated by [1, 6]

$$R = \frac{F_{zr} - F_{zl}}{F_{zr} + F_{zl}} = \frac{2m_s a_y h_R}{mgl_w} + \frac{2m_s h_R tan\phi}{ml_w} \tag{2}$$

where m_s is Sprung mass, $m = m_s + m_u$, m_u is unsprung mass, a_y is lateral acceleration, h_R is Center of gravity (C.G.) height from roll center, g is acceleration due to gravity, l_w is track width, and ϕ is roll angle.

Lacking influence of the vertical road input and other external inputs limit the rollover index in (2) with only the untripped rollover detection. Lateral acceleration and roll angle influence this rollover index. However, the roll angle, which is complicated to measure, is removed in some published paper's rollover index calculation.

2.2. Rollover Index for Tripped and Untripped Rollover

The influence of road inputs and other external forces are considered in rollover index both for tripped and untripped rollovers in order to indicate the tripped rollovers caused by external input tripping. Figure 1. shows a vehicle rollover model where the influence of right road inputs, z_{rr}, left road input, z_{rl}, and an unknown lateral force input, F_{lat}, at an arbitrary height, h_{lat}, from the roll center are presented. It also demonstrates the normal tires forces, F_{zr} and F_{zl}.

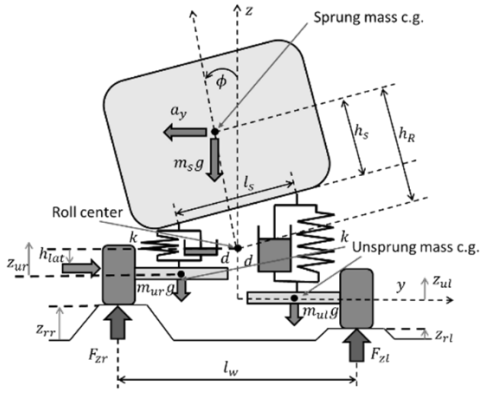


Figure 1: Tripped and Untripped Rollover Model

The four-degree-of-freedom vehicle model provides the derivation of the rollover index in the case of tripped rollover and the derivation is presented in [6]. Equation (3) shows the rollover index for tripped and untripped rollover which is shown in [6].

$$R = \frac{m_u(\ddot{z}_{ur} - \ddot{z}_{ul}) - \frac{2}{l_s^2}(I_{xx} + m_s h_R^2)(a_{zl} - a_{zr})}{m_u(\ddot{z}_{ur} + \ddot{z}_{ul}) + m_s\ddot{z}_s + mg} + \frac{\frac{2}{l_s}m_s a_y h_R cos\phi + \frac{2}{l_s}m_s g h_R sin\phi}{m_u(\ddot{z}_{ur} + \ddot{z}_{ul}) + m_s\ddot{z}_s + mg} \tag{3}$$

where $\ddot{z}_{ur}$ is right unsprung mass acceleration, $\ddot{z}_{ul}$ is left unsprung mass acceleration, $\ddot{z}_s$ is sprung mass acceleration, l_s is length between the left and right suspensions, I_{xx} is roll moment of inertia, a_{zl} is right vertical accelerometer measurement, a_{zr} is left vertical accelerometer measurement, and $m = m_s + 2m_u$.

The second term of (3), the function of lateral acceleration and roll angle, is similar to (2) when comparing the traditional rollover index in (2) and rollover index for tripped and untripped rollovers in (3). To be more clarify, the second term of (3) is the traditional rollover index for detecting untripped rollover, and the first term of (3) is the additional term used for detecting tripped rollover.

There are some difficulties in implementing the (3) to determine tripped rollover due to the parameters and variables that are unable to identify, e.g. the location of roll center, center of gravity (C.G.) height or roll angle. The estimations of these unknowns are required to complete the estimation algorithms. Furthermore, the difference of dynamics between an actual vehicle and the four-degree-of-freedom model might lead to errors. The recurrent neural network is implemented to detect tripped and untripped rollovers in order to optimize the estimation. After training the data, the traditional estimation algorithm to estimate parameters and detect rollover are not required in the model anymore.

3. Recurrent Neural Networks (RNNs)

Recurrent neural network is a popular special type of neural networks. In this paper, sequential data will be sensor data over a period while in other cases it can be, for example frames in video, or words in a text. When a new input is introduced to the sequence, a hidden state vector, h, which is an added memory of the recurrent neural network, is updated. To avoid features losses while training neural networks, the crucial data of the previous sequence is maintained in this hidden state. The recurrent neural network has the ability to work with various lengths of sequence data in addition to the benefit in compression of past information.

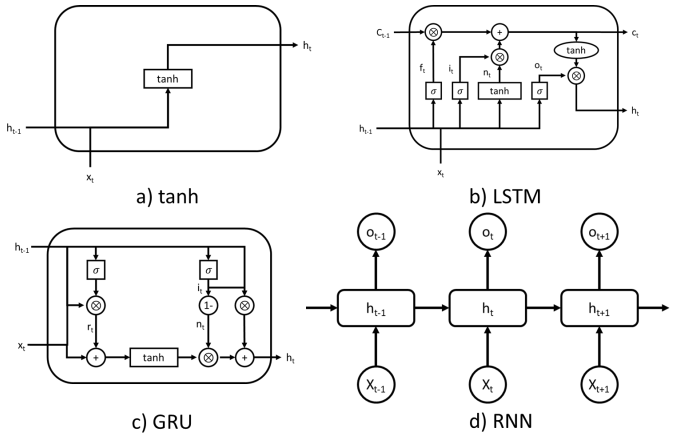


Figure 2: The Structural of Recurrent neural network a) tanh architecture b) LSTM architecture c) GRU architecture d) RNN Connection

There are four types of neural networks used to evaluate and create the neural network models for detecting tripped and untripped rollover. The four types of neural networks are 1) feedforward neural network (FNN), 2) Tanh neural network, 3) long short-term memory (LSTM) neural network [39], and 4) gates recurrent unit (GRU) neural network [40]. The detail of implementing these neural networks can be found in [39-46] and the short introduction of them is shown in [1]. The structural of recurrent neural network is shown on Figure 2.

4. Methodology

The data which is not normally determined in a vehicle such as the wheel's forces exerted on the ground by the vehicle wheels is mandatory to compute the precise rollover index in (1). Moreover, Ref. [6, 28] determine the tripped and untripped rollover index based on half vehicle model (4 Degree of Freedom). The variables, which are used, are right unsprung mass vertical acceleration, $\ddot{z}_{ur}$, left unsprung mass vertical acceleration, $\ddot{z}_{ul}$, right vertical accelerometer measurement, a_{zr}, left vertical accelerometer measurement, a_{zl}, lateral acceleration, a_y, sprung mass acceleration, $\ddot{z}_s$, roll angle, ϕ, roll rate, $\dot{\phi}$, and right suspension compression, $(z_s - z_{ur} + \frac{l_s}{2} sin\phi)$. They also assume that the parameters such as sprung mass, m_s, unsprung mass, m_u, C.G. height from roll center, h_R are constant and not changed.

To detect tripped rollover with various conditions, the good measurement inputs are needed to verify for recurrent neural networks. Also, a feedforward neural network is compared with recurrent neural networks in this case to validate the advantages of employing recurrent neural networks rather than a static neural network. The feedforward neural network uses the same inputs as the recurrent neural networks for processing the data and a comparable number of parameters. Moreover, the structure of neural networks is verified for determining the suitable one.

There is a reconstruction of the relationship between the inputs and the rollover index from a dataset processed in a simulation constraint. The next part is presented with the details on the dataset used for training, also the structure of the neural networks and the training setup will be described

4.1. Dataset generation for tripped and untripped rollover

The software CarSim, a standard industry vehicle simulation software, is used to process all the data concerned. In a first step, 22,000 simulations with non-selectively sampled height of the center of gravity and sprung mass of the car are simulated with a variety of tripped rollover scenarios. The height of the center of gravity is uniformly studied between 0.6 and 1.2 meters, based on [47] as a background which also take an estimate of additional loads into account. The sprung mass is uniformly generated between 2100 to 3000 kg, a mass comparable to loaded SUVs. Next, a trajectory of the scenarios is generated based on Fig 2. The trajectory consists of the 50 meters of straight path and the circular arc path of the random radius in the range from 50 to 200 meters based on highway data of Bangkok, Thailand. Samples 80 to 180 km/h uniform distribution will determine the target velocity of the vehicle. Then, the CarSim is going to control the real velocity over the trajectory in a feedback loop to remain adjacent to the speed that is set goal. Also, the steering angle of the vehicle is close-loop controlled to follow the trajectory.

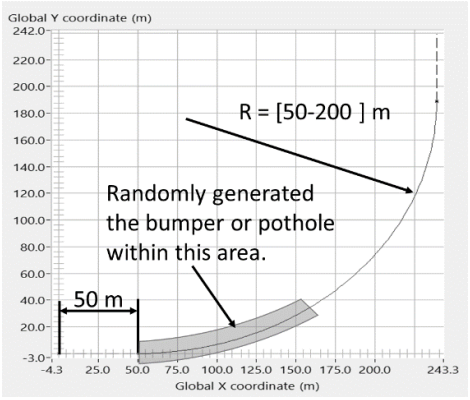


Figure 3: CarSim Trajectory Setup

To create tripped rollover scenarios, the bumper and pothole are randomly generated along the circular arc path of the trajectory as shown in Figure 3. The shape of the bumper and pothole is randomly created from the sine function with the amplitude between -0.15 to 0.15 m and the frequency between 0.625 to 1.25 Hz. Only one bumper or pothole is created for each trajectory. Also, it is created only half lane of the trajectory to make only half side of the vehicle strike the bumper or pothole. For the left turn trajectory, if the bumper is generated, the bumper is created on the left side of the vehicle and if the pothole is generated, the pothole is created on the right side of the vehicle. Then, for the right turn trajectory, if the bumper is generated, the bumper is created on the right side of the vehicle and if the pothole is generated, the pothole is created on the left side of the vehicle. These increase the chance for tripped rollover.

The normal forces on the left and right wheel, F_{zl}, F_{zr}, are provided by a simulation, so each estimation is able to process to determine its rollover index. The rear wheel vertical forces are used to identify the rollover index. However, the front wheels, employing the said methodology together with a small variation from elastic deformations in the wheels and wheel suspensions, can also specify the rollover index as well. The rollover index computed with (1) is the neural networks' training target, calculated according to (1). From there, when two wheels are lifted off road, the sequences stop. That is because the rollover index is defined only for the case of at least one wheel lifting the ground. And it will contribute to a sequence range of length between 0.325 to 19.925 seconds. In every simulation case, there will be a collection of data of the vehicle variables at the sample rate of 40 Hz. However, each simulation dataset also possesses insignificant and irrelevant information of the subject vehicle or those that are unable to be scaled in an actual subject. So, the unnecessary data are needed to be discarded. Then, for the 22,000 files of h5 files (1.64 GB), they are splitted into 3 groups. There are a training set of 16,000 files, a validation set of 4000 files, and

a test set of 2,000 files. The training set is used for training. Other than that, it is mandatory for the validation set to detect possible overfitting of the model while the test set is used to evaluate the error of the neural networks.

4.2. Training of the recurrent neural networks for tripped rollover

There are three major factors of neural networks for detecting tripped and untripped rollover needed to be verified. The first factor is which neural network should be suitable, the second factor is which inputs should be good for neural networks, and the last one is which structure of the neural networks should be implemented. To verify these three factors, three training sets are simulated. Then, future tripped and untripped rollover prediction also is verified with simulation.

4.2.1 Training for determining a suitable neural network type

For this task, four types of neural networks will be compared and given information. Table 1 is presented with the detail of the neural networks which are constructed to take into account approximately the same quantity of parameters. About 5,000 parameters are to be taken to comparison between each neural network. In Table 1, "$N°$ neurons" shows the number of neurons in each layer in the horizontal axis, and also takes into account the input layer (n inputs) and output layer (1 output). The recurrent neural networks, the second entry in "$N°$ neurons" represents the number of hidden states. "$N°$ params" specifies the number of parameters in the neural network. In this column, the added number will be those of all elements in its weight matrices and bias vectors.

Firstly, a feedforward neural network (FNN) will be introduced. FNN is made up of three hidden layers using the $tanh$ activation function and a linear output layer as shown within [1]. Secondly introduced is a neural network (Tanh) which is a recurrent neural network, that employs a $tanh$ recurrent layer. And followed the recurrent layer aforementioned are a $tanh$ hidden layer and an output layer. Thirdly, LSTM neural network which utilizes a long short-term memory layer, again followed by a $tanh$ layer and linear output layer. And for the fourth one, GRU neural network that adopts a static Tanh layer and a linear output layer, works with the gated recurrent unit. Then, the neural networks training diagram is shown on Figure 4.

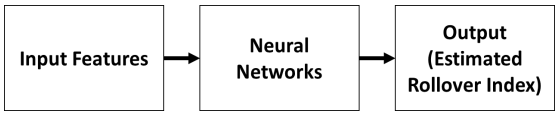


Figure 4: The Neural Networks Training Diagram

Table 1: Summary of The Neural Networks' Structure

NN Type	$N°$ neurons	$N°$ param
FNN	n 48 48 42 1	48n+4501
Tanh	n 46 46 1	46n+4417
LSTM	n 30 16 1	120n+4353
GRU	n 32 32 1	96n+4353

Table 2: Input Variables for Neural Networks

Symbolic	Descriptions
$v_{z_{wc}}$: $[v_{z_{wc_{L1}}}; v_{z_{wc_{L2}}}; v_{z_{wc_{R1}}}: v_{z_{wc_{R2}}}]$	Vertical speed of each wheel: Left front, $v_{z_{wc_{L1}}}$, Left rear, $v_{z_{wc_{L2}}}$, Right front, $v_{z_{wc_{R1}}}$, Right rear, $v_{z_{wc_{R2}}}$, (km/h)
A_{xyz}: $[A_x; A_y; A_z]$	Longitude acceleration, A_x, Lateral acceleration, A_y, Vertical acceleration, A_z, (g)
AA_x	Roll Acceleration of sprung mass body (rad/sec^2)
AV_{xyz}: $[AV_x; AV_y; AV_z]$	Roll rate, AV_x, Pitch rate, AV_y, Yaw rate, AV_z, (deg/sec)
$A_{z_{sm}}$	Vertical acceleration of sprung mass C.G. (g)
v_x	Longitude speed (km/h)
$Steer_{DM}$	Steer angle from driver (deg)
$CmpS$: $[CmpS_{L1}; CmpS_{L2}; CmpS_{R1}; CmpS_{R2}]$	Compression of spring each wheel: Left front, $CmpS_{L1}$, Left rear,$CmpS_{L2}$, Right front, $CmpS_{R1}$, Right rear,$CmpS_{R2}$, (mm)

Table 3: Input Combination Training for Determining a Suitable Neural Network Type

No.	Input Features	n	FNN	Tanh	LSTM	GRU
1	$[v_{z_{wc}}; A_{z_{sm}}; A_{xyz}]$	8	4885	4785	5313	5121
2	$[v_{z_{wc}}; A_{z_{sm}}; AA_x]$	6	4789	4693	5073	4929
3	$[v_{z_{wc}}; A_{z_{sm}}; AV_{xyz}]$	8	4885	4785	5313	5121
4	$[v_{z_{wc}}; A_{z_{sm}}; A_{xyz}; AA_x]$	9	4933	4831	5433	5217
5	$[v_{z_{wc}}; A_{z_{sm}}; A_{xyz}; AV_{xyz}]$	11	5029	4923	5673	5409
6	$[v_{z_{wc}}; A_{z_{sm}}; AA_x; AV_{xyz}]$	9	4933	4831	5433	5217
7	$[v_{z_{wc}}; A_{z_{sm}}; A_{xyz}; AV_{xyz}; AA_x]$	12	5077	4969	5793	5505

For the inputs to the neural networks, it is still unclear to select them to predict the tripped rollover. Thus, to initially investigate this task, the inputs are selected based on (3) [6] and the full vehicle suspension model. In addition, the selected inputs can be measured by sensors [6, 28]. The list of the inputs is shown on Table 2. The input combinations are also evaluated in this case. The input combinations for training in this section is shown on Table 3.

Next, the neural networks are trained over 3000 epochs using backpropagation through time. When the process is completed, since the result beyond that came out slower on the validation and training loss. The mean squared error is used as the loss function:

$$L = \frac{1}{N_{traj}} \sum_{i=1}^{N_{traj}} \frac{1}{T_i} \sum_{t=1}^{T_i} \left(R_{ro,i,t} - y_{i,t}\right)^2 \quad (4)$$

with $R_{ro,i,t}$ the rollover index as calculated using the cases data in trajectory i at time step t and $y_{i,t}$ the scalar output of one of the neural networks for the same trajectory and time step.

The leading software that is used for computing the neural networks is the Knet module [49] by the Julia programming language. The program runs on the notebook, Acer Nitro, with CPU i7-7700HQ (2.8 GHz), 8 GB Ram, GeForce GTX1050, and

Window 10 (64 bits). In the meantime, Training is done on a consumer level GPU is used to perform the network training for about 1 hour per model by using average quantity of parameters. This altogether resulted in more than 240 hours of training time. The Ram usage during the training is approximately 6 to 7 GB. And there also implemented the Adam optimizer along with the suggested settings in [50]. Then, the gradient was computed on mini-batches of 2,048 sequences.

4.2.2 Training for determining suitable inputs for the neural network

Since the performance of the neural network is required and the sensor cost is needed to be minimized, the inputs to the neural network need to be optimized. The extra inputs, suspension compressions as using in [28] for each wheel, are included to the list for evaluation. Then, the input combinations are evaluated to determine the suitable inputs for predicting rollover. The input reduction is also included in this test for reducing number of using sensors. The lists of the inputs are shown in Section 5.2 Table 5.

Next, only the best neural network type from the section 4.2.1 are trained. The training process and the loss function are implemented similar to the section 4.2.1.

4.2.3 Training for determining suitable neural network structure

In this task, the neural network improvement is investigated. Eight different structures of the best neural network from the section 4.2.1 are compared. The neural network structure is modified by increasing number of the layers and the neurons. Also, the optimized inputs from the section 4.2.2 are used in this task. Then the neural network structures are also summarized in Table 6. Next, the training process and the loss function are implemented similar to the section 4.2.1.

4.2.4 Training for future tripped and untripped rollover prediction

The future tripped and untripped rollover perdition uses the vehicle parameter measurement to prediction rollover index of a vehicle in the future time based on the training model. It can provide an extra time to prevent or warn the danger of the rollover.

In this case, the output of each data set is shifted by time while the input is reduced to identify with to the length of the output. The shifted time varies from 0.1 second up to 1 second. Then, the best model from Section 4.2.3 are used and, the training process and the loss function are implemented similar to the section 4.2.1

5. Results

The training as descript in Section 4 are applied to the data set. The final losses on the train, validation and test datasets are presented in this section.

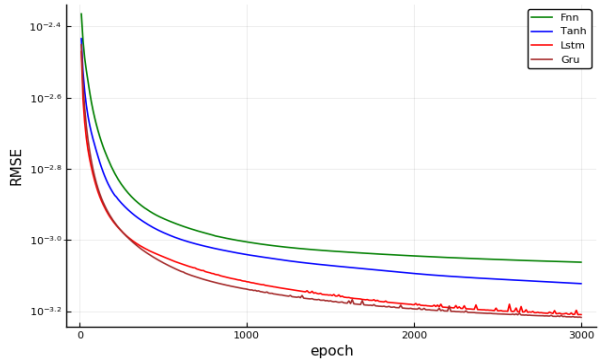
5.1. The suitable neural network type

The test losses in different neural networks on the different input features are shown on Table 4. Also, the example of loss on the train dataset and validation dataset over the epochs are presented in Figures 5a and 5b, subsequently. The training and validation losses of all different neural networks and all input features shows that the overfitting does not occur for all cases because the validation loss decreases as the training loss decreases and the losses from validation dataset are small as the losses from training dataset. Also, the underfitting did not occur for all cases because the neural network models can detect the rollover index as shown in Figures 6 and 7.

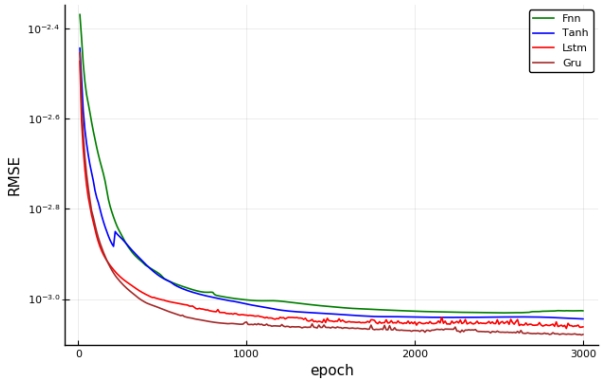
Table 4: Remaining Loss of Different Neural Networks on Test Data Set

No.	Input Features	n	Average RMS error of 2000 Data Set $\times 10^{-4}$			
			FNN	Tanh	LSTM	GRU
1	$[v_{z_{wc}}; A_{z_{sm}}; A_{xyz}]$	8	15.9	10.5	11.3	27.2
2	$[v_{z_{wc}}; A_{z_{sm}}; AA_x]$	6	228	33.3	96.3	156
3	$[v_{z_{wc}}; A_{z_{sm}}; AV_{xyz}]$	8	17.7	13.2	14.4	28.9
4	$[v_{z_{wc}}; A_{z_{sm}}; A_{xyz}; AA_x]$	9	12.5	10.3	11.1	17.4
5	$[v_{z_{wc}}; A_{z_{sm}}; A_{xyz}; AV_{xyz}]$	11	9.82	7.91	9.98	22.8
6	$[v_{z_{wc}}; A_{z_{sm}}; AA_x; AV_{xyz}]$	9	17.1	22.7	13.3	43.0
7	$[v_{z_{wc}}; A_{z_{sm}}; A_{xyz}; AV_{xyz}; AA_x]$	12	9.09	7.54	9.47	21.5

Note: the highlighted cell of the table represents the smallest value of the table.



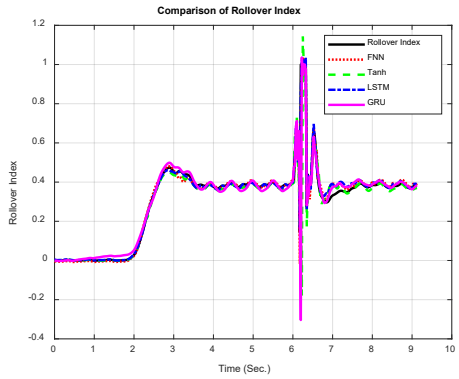
a) Training Loss of No. 7 Table 4



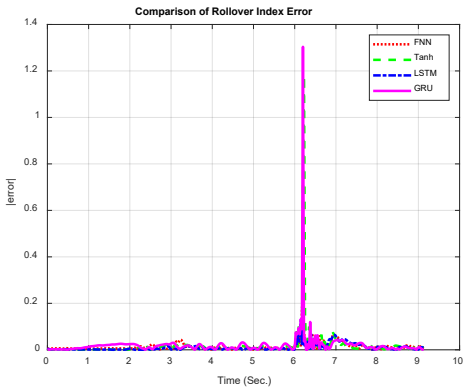
b) Validation of No. 7 Table 4

Figure 5: The Example of Training and Validation Loss

By comparing four types of the neural networks on Table 4, GRU recurrent neural network and FNN seems have the large test losses for all different input features. Since the tripped rollover happen in the very short time So, there is few tripped data occurred. This makes GRU which has a too long-term memory cannot learn the model well. Also, FNN contributes to a large outstanding error because the formula only processes the valuation with the current input x_t.

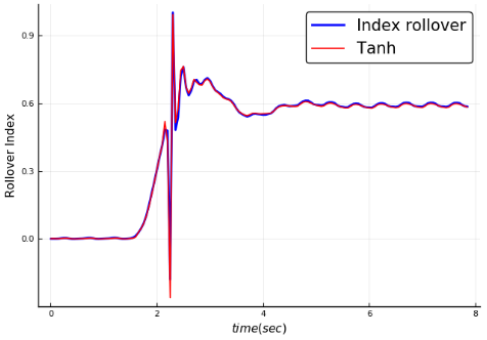


a) Comparison of Rollover Index and Rollover Index from Neural Networks

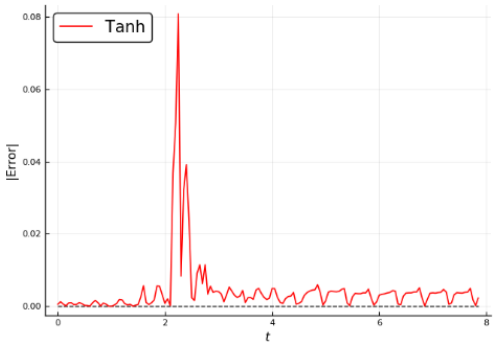


b) Comparison of Rollover Index Error

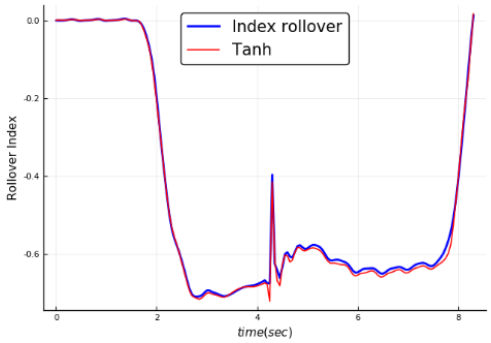
Figure 6: The Example of Rollover Index in the Case of ($m_s = 2722\ kg$, $h_R = 0.847\ m$, $v_x = 82\ km/h$)



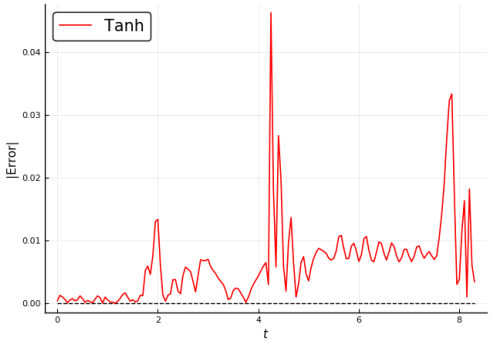
a) Comparison of Rollover Index in the Bumper Case



b) Comparison of Rollover Index Error in the Bumper Case



c) Comparison of Rollover Index in the Pothole Case



d) Comparison of Rollover Index Error in the Pothole Case

Figure 7: The Example of Rollover Index in the Bumper Case ($m_s = 2331\ kg$, $h_R = 0.777\ m$, $v_x = 118\ km/h$) and Pothole Case ($m_s = 2912\ kg$, $h_R = 0.730\ m$, $v_x = 141\ km/h$)

However, Tanh recurrent neural network has the lowest final loss on test loss excepted for No. 6 Table 4 with the input features, $[v_{z_{wc}}; A_{z_{sm}}; AA_x; AV_{xyz}]$ because it has not too long term memory. While the test loss from LSTM recurrent neural network is larger than it from Tanh recurrent neural network except for No. 6 Table 4.

For an empirical comparison, the rollover index (black solid line in Figure 6a) according to (1) and the neural network estimation are shown in Figure 6a for the selected section of the test trajectory. The vehicle takes a cornering at about 2 seconds and during cornering it strikes a bumper at about 6 seconds. This caused a rear wheel to leave the road while its counterpart stayed intact on the ground. The estimated rollover index for the LSTM, Tanh, and FNN are similar and barely distinguishable, whereas the estimation of GRU shows a clear offset and a large error at around 6 seconds. The absolute value of the estimation error is presented subsequently in Figure 6b. As shown, the highest values in the estimation error for the rollover index is seen following the subject vehicle returns from rollover condition back to its normal condition that all of the wheels are on the ground at around 6.1 seconds in the trajectory. However, situational peaks are more prominent, while those of GRU are even more grand.

Figure 7 shows the rollover index and rollover index error of the Tanh recurrent neural network of different vehicle parameters on the bumper and pothole trajectories. The Tanh recurrent neural network can predict the rollover index even though the vehicle parameters are changed because the neural network can handle the vehicle parameter changing. It can learn and adapt the parameters inside the neural network and then predict the rollover index.

Figure 5a shows the case that the vehicle with the parameters, $m_s = 2722\ kg$, $h_R = 0.847\ m$, $v_x = 82\ km/h$, takes a cornering at about 1.8 seconds and during cornering it strikes a bumper at about 2.4 seconds. The maximum error of the estimation happens at about 2.4 seconds. Next, Figure 5b shows the case that the vehicle with the parameters, $m_s = 2912\ kg$, $h_R = 0.730\ m$, $v_x = 141\ km/h$, takes a cornering at about 1.8 seconds and during cornering it strikes a pothole at about 4.4 seconds. Then, the maximum error of the estimation happens at about 4.4 seconds.

In summary, the suitable neural network for tripped rollover is Tanh recurrent neural network since it has ability to retain information of the past and handle the vehicle parameter changing. Also, the lowest test loss happens when using Tanh recurrent neural network with No. 7 Table 4. with the input features, $[v_{z_{wc}}; A_{z_{sm}}; A_{xyz}; AV_{xyz}; AA_x]$.

5.2. The suitable input features for the neural network

Since the error of the Tanh neural network is needed to be improved, the suspension compressions for each wheel, longitude speed, and steering angle are added and evaluated. The test losses of these cases are shown in Table 5. The results show that the suspension compressions for each wheel, longitude speed, and steering angle are significant input features for the neural network. They can reduce the test losses up to 2.93×10^{-4}. Also, once the the suspension compressions for each wheel, longitude speed, and steering angle have been included in the input features, the input, AA_x, seems not effect to the test losses.

In case that the number of suspension compressions is reduced to be only front or rear side, the test losses is increased as shown in No. 3 and 4 Table 5. No. 3 Table 5. has higher the test losses because the actual rollover index is determined from the rear wheel vertical forces.

In summary, the suitable input features for tripped rollover for Tanh neural network is $[v_{z_{wc}}; A_{z_{sm}}; A_{xyz}; AV_{xyz}; v_x; steer_{DM}; CmpS]$ which are agree with [28]. The example of rollover index in this case is shown in Figure 8.

Table 5: Remaining Loss of the Tanh Neural Network on Different Input Feature on Test Data Set

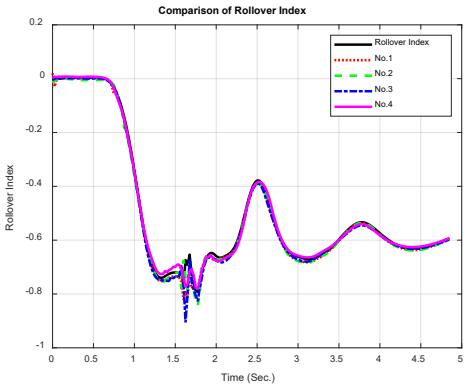
No.	Input Features	n	$N°$ param	Average RMS error of 2000 Data Set $\times 10^{-4}$
1	$[v_{z_{wc}}; [A]; CmpS]$	17	5,199	4.61
2	$[v_{z_{wc}}; [A]; AA_x; CmpS]$	18	5,245	4.82
3	$[v_{z_{wc_{L1}}}; v_{z_{wc_{R1}}}; [A]; CmpS_{L1}; CmpS_{R1}]$	13	5,061	7.41
4	$[v_{z_{wc_{L2}}}; v_{z_{wc_{R2}}}; [A]; CmpS_{L2}; CmpS_{R2}]$	13	5,061	6.58

*$[A] := [A_{z_{sm}}; A_{xyz}; AV_{xyz}; v_x; steer_{DM}]$

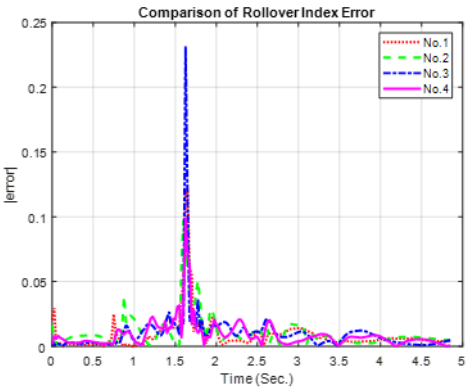
Note: the highlighted cell of the table represents the smallest value of the table.

5.3. The suitable neural network structure

The error of the Tanh neural network is still needed to be improved. The different types of neural network as shown in Table 6. are evaluated with 3 set of input features, case No. 1, 3, and 4 from Section 5.2.



a) Comparison of Rollover Index



b) Comparison of Rollover Index Error

Figure 8: Example of Rollover Index ($m_s = 2583\ kg$, $h_R = 0.829\ m$, $v_x = 101\ km/h$)

The results show that the neural networks in the case no. 7 Table 6 with the input #1 have the lowest test losses of 3.66×10^{-4}. In addition, the neural networks in the case no. 5 Table 6 with the input #3 has the test losses of 4.33×10^{-4} and the neural networks in the case No. 8 Table 6 with the input #2 has the test losses of 7.30×10^{-4}.

In summary, for the lowest test losses, the suitable Tanh neural network structure is the case No. 7 Table 6 which is (n 46 46 50 30 1). Moreover, for the low test losses and low number of input features, the suitable Tanh neural network structure is the case No. 5 Table 6. The test losses in this case is different from the test loss from the best one only 0.67×10^{-4}.

Table 6: Remaining Loss of Different Tanh Neural Network Structure on Test Data Set

No.	Neural Networks' Structure	Average RMS error of 2000 Data Set $\times 10^{-4}$			
		$N°$ param	Inputs #1	Inputs #2	Inputs #3
			n=17	n=13	n=13
1	n 12 12 1	12n+337	7.87	12.30	9.18
2	n 24 20 1	24n+1145	5.76	9.91	7.11
3	n 46 46 1	46n+4417	4.61	7.41	6.58
4	n 46 46 30 1	46n+5811	5.54	7.90	7.35
5	n 46 46 30 30 1	46n+6741	4.34	10.30	4.33
6	n 46 46 30 50 1	46n+7381	4.43	7.32	6.14
7	n 46 46 50 30 1	46n+8281	3.66	7.83	5.71
8	n 46 46 50 50 1	46n+9321	4.14	7.30	5.43

* Inputs #1: $[v_{z_{wc}}; [A]; CmpS]$

Inputs #2: $[v_{z_{wc_{L1}}}; v_{z_{wc_{R1}}}; [A]; CmpS_{L1}; CmpS_{R1}]$

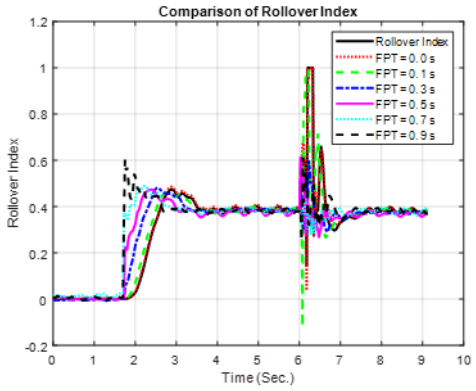
Inputs #3: $[v_{z_{wc_{L2}}}; v_{z_{wc_{R2}}}; [A]; CmpS_{L2}; CmpS_{R2}]$

Note: the highlighted cell of the table represents the smallest value of the table.

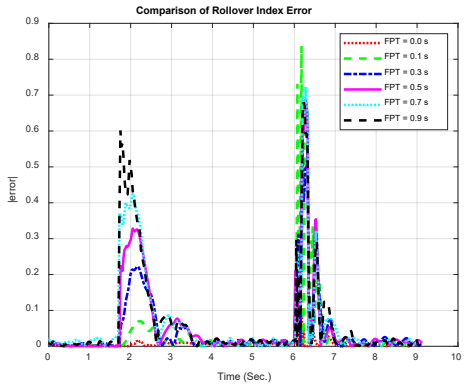
5.4. Future tripped and untrippped rollover prediction

By using the Tanh neural network in No. 7 Table 6, the error of the models shown in Table 7. Also, the example of the rollover index perdition is shown on Figure 9. The results show that once the shift time increases, the error increases. The difference between shift time 0.0 and 0.1 second is approximately 25×10^{-4}. The error seems large. However, if consider the future tripped rollover prediction plot on Figure 9a, the plot shows that they seem correctly predict the trend of rollover index up to 0.5 seconds in future. The prediction lines look like the actual rollover index but it is shelfed by the time.

In summary, this predict is very useful and can provide an extra time prediction for rollover prevention and warning system.



(a) Comparison of Rollover Index



(b) Comparison of Rollover Index

Figure 9. The Example of the Rollover Index in the Case of ($m_s = 2722\ kg$, $h_R = 0.847\ m$, $v_x = 82\ km/h$)

Altogether, the results exhibit that the Tanh recurrent neural network is able to predict the tripped rollover index with a larger preciseness than others once it is considered with the undetermined parameters, for instance, the height of the center of gravity and sprung mass. So, once processing the training data, it is necessary to take the said undetermined values into consideration. The suitable input features for tripped rollover index should be $[v_{z_{wc}}; A_{z_{sm}}; A_{xyz}; AV_{xyz}; v_x; steer_{DM}; CmpS]$ with Tanh neural network structure of (n 17 46 46 50 30 1). Moreover, the future tripped rollover prediction can provide an extra time prediction for rollover prevention and warning system.

Table 7: Remaining Loss of the Tanh Neural Network on Test Data Set

Future Perdition Time (Sec)	Average RMS error of 2000 Data Set $\times 10^{-4}$
0.0	3.66
0.1	29.2
0.2	35.6
0.3	43.4
0.4	55.5
0.5	83.4
0.6	133.4
0.7	214.1
0.8	328.2
0.9	464.5
1.0	616.6

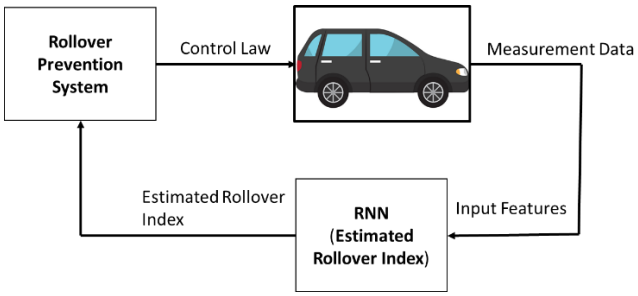


Figure 10: Application Diagram of The Neural Network

The application of the developed neural network is shown on Figure 10. The measurement data from a vehicle is sent to the recurrent neural network which can compute the estimated rollover index within 0.03 ms. Then, the estimated rollover index is sent to rollover prevention/warning system to calculation the control law to make the vehicle action and reduce the roll risk.

6. Conclusions

An essential factor of the carrying out the rollover preventive systems and warning ones is the assessment of the rollover risk. The rollover risk pointer or the rollover index is estimated by employing the feedforward neural network together with three other recurrent neural networks. It is seen that when in occurrence of any uncertain parameters, for instance, undetermined height of the center of gravity of the subject, the Tanh recurrent neural network shows higher performance juxtaposed with others. Calculations of the loss on a test set and an empirical evaluation on test trajectories support in verifying the estimation quality. The outcomes of the test are encouraging for further research, and the performance assessment of the recurrent neural networks in an actual vehicle would be an interesting subject of further research.

The sensory inputs that are chosen play a crucial part in the estimation of the rollover index. The results show that the suitable input features for tripped rollover index should be $[v_{z_{wc}}; A_{z_{sm}}; A_{xyz}; AV_{xyz}; v_x; steer_{DM}; CmpS]$ which are agree

with the previous research. Also, the suitable Tanh neural network structure of is (n 46 46 50 30 1). Moreover, the future tripped rollover prediction can correctly predict the trend of rollover index up to 0.5 seconds in future. This will be useful for rollover prevention and warning system.

Acknowledgment

This research was funded by Ratchadaphiseksomphot Endowment Fund Chulalongkorn University, grant number: CU_GI_62_05_21_01.

References

[1] C. Dengler, K. Treetipsounthorn, S. Chantranuwathana, G. Phanomchoeng, B. Lohmann, and S. Panngum, "Vehicle rollover detection using recurrent neural networks," in The 9th IEEE International Conference on Robotics, Automation and Mechatronics (CIS-RAM 2019), Bangkok, Thailand, Nov. 2019.

[2] US Department of Transportation, Type of Rollovers, National Highway Traffic Safety Administration (NHTSA), Washington, DC, August 2017. [Online] Available: http://www.safercar.gov

[3] W. Xinye, GE. Xiaohong, L. Shuyou, and H. Hongwu, "Research on Vehicle Rollover and Control," in 2010 2nd International Conference on Advanced Computer Control, 2010.

[4] R. L. Martin, C. Ilse, "Robust indirect-defined envelope control for rollover and lateral skid prevention," Control Engineering Practice, **61**, 149-162, 2017.

[5] J.J. Rath, M. Defoort, and K.C. Veluvolu, "Rollover index estimation in the presence of sensor faults, unknown inputs, and uncertainties," IEEE transactions on intelligent transportation systems, **17**(10), 2949-2959, 2016.

[6] G. Phanomchoeng, and R. Rajamani, "New rollover index for the detection of tripped and untripped rollovers," IEEE Transactions on Industrial Electronics, **60**(10), 4726-4736, 2013. DOI: 10.1109/TIE.2012.2211312

[7] R. Rajamani, and P. Damrongrit Neng, "New paradigms for the integration of yaw stability and rollover prevention functions in vehicle stability control," IEEE Transactions on Intelligent Transportation Systems, **14**(1), 249-261, 2013. DOI: 10.1109/TITS.2012.2215856

[8] K. Warwas, and A. Krzysztof, "An Application of Neural Networks to Control Stability of an Articulated Vehicle in Real Time," in Information Systems Architecture and Technology: Proceedings of 36th International Conference on Information Systems Architecture and Technology–ISAT 2015–Part I, 135-146. Springer International Publishing, 2016.

[9] B. Zhu, P. Qi, Z. Jian, and G. Litong, "Integrated chassis control for vehicle rollover prevention with neural network time-to-rollover warning metrics," Advances in Mechanical Engineering, **8**(2), 2016. https://doi.org/10.1177%2F1687814016632679

[10] K. Warwas, and K. Augustynek, "An Application of Neural Networks to Control Stability of an Articulated Vehicle in Real Time," in Information Systems Architecture and Technology: Proceedings of 36th International Conference on Information Systems Architecture and Technology – ISAT 2015 – Part I, 135-146, 2016.

[11] S. Solmaz, M. Akar, R. Shorten, "Adaptive Rollover Prevention for Automotive Vehicles with Differential Braking," in Proceedings of the 17th World Congress The International Federation of Automatic Control Seoul, Korea, July 6-11, 2008.

[12] S. Lee, Y. Fitri, K. Misawa Kasahara, and M. Yasuchika, "Rollover prevention with predictive control of differential braking and rear wheel steering," in Robotics, Automation and Mechatronics (RAM), 2013 6th IEEE Conference on, 144-149. IEEE, 2013.

[13] Z. Yu, J. Li, F. Dai, and S. Li, "Anti-Rollover Control Based on Fuzzy Differential Braking for Heavy Duty Commercial Vehicle," in Conference: 2016 4th International Conference on Machinery, Materials and Computing Technology, 2016.

[14] S. Yim, "Design of a preview controller for vehicle rollover prevention," IEEE Transactions on Vehicular Technology, **60**(9), 4217-4226, 2011. DOI: 10.1109/TVT.2011.2169687

[15] S. Solmaz, C. Martin, and S. Robert, "A methodology for the design of robust rollover prevention controllers for automotive vehicles with active steering," International Journal of Control, **80**(11), 1763-1779, 2007. https://doi.org/10.1080/00207170701473987

[16] R.A. Azim, and F.M. Malik, "Rollover mitigation controller development for three-wheeled vehicle using active front steering," Mathematical Problems in Engineering, 2015.

[17] N. Zhang, G.M. Dong, and H.P. Du, "Investigation into untripped rollover of light vehicles in the modified fishhook and the sine maneuvers. Part I: Vehicle modelling, roll and yaw instability," Vehicle System Dynamics, **46**(4), 271-293, 2008. https://doi.org/10.1080/00423110701344752

[18] S. Yim, P. Yongjin, and K. Yi. "Design of active suspension and electronic stability program for rollover prevention," International journal of automotive technology, **11**(2), 147-153, 2010.

[19] D. Konik, "Dynamic drive-the new active roll stabilization system from BMW group-system description and functional improvements," in Proc. 5th Int. Symp. Avec 2000, Ann Arbor, MI 2000.

[20] S. Yim, K. Jeon, and K. Yi, "An Investigation into Vehicle Rollover Prevention by Coordinated Control of Active Anti-roll Bar and Electronic Stability Program," International Journal of Control Automation and Systems, **10**(2), 2012. https://doi.org/10.1080/002071707014739887

[21] D.J. Sampson, and C. David, "Active roll control of single unit heavy road vehicles," Vehicle System Dynamics, **40**(4), 229-270, 2003.

[22] J. Yoon, W. Cho, and J. Kang, "Design and evaluation of a unified chassis control system for rollover prevention and vehicle stability improvement on a virtual test track," Control Engineering Practice, **18**(6), 585-597, 2010.

[23] D. Piyabongkarn, J.Y. Lew, R Rajamani, and J.A. Grogg, "Active driveline torque-management systems," IEEE Control Systems, **30**(4), 86-102, 2010.

[24] M.C. Walz, Trends in the static stability factor of passenger cars, light trucks, and vans. No. HS-809 868. 2005.

[25] S.B. Choi, "Practical vehicle rollover avoidance control using energy method," Vehicle System Dynamics, **46**(4), 323-337, 2008.

[26] J. Yoon, K. Dongshin, and Y. Kyongsu, "Design of a rollover index-based vehicle stability control scheme," Vehicle system dynamics, **45**(5), 459-475, 2007.

[27] B.C. Chen, "Warning and control for vehicle rollover prevention," No. 30-00934 UMI. Ann Arbor, MI: University of Michigan, 2001.

[28] G. Phanomchoeng, and R. Rajamani, "Real-Time Estimation of Rollover Index for Tripped Rollovers with a Novel Unknown Inputs Nonlinear Observer," IEEE/ASME Transactions on Mechatronics, **19**(2), 743-754, 2014.

[29] R. Rajamani, and N. Piyabongkarn, "New Paradigms for the Integration of Yaw Stability and Rollover Prevention Functions in Vehicle Stability Control," IEEE Transactions on Intelligent Transportation Systems, **14**(1), 249-261, 2013.

[30] X. Cai, and J. Wu, "Hazardous Chemicals Vehicles Rollover Pre-warning System Based on CPS," in Proceedings of the 2012 IEEE Sixth International Conference on Software Security and Reliability Companion, 2012.

[31] J.R. Kerrigan, N.J. Dennis, D.P. Parent, and S. Purtsezov, "Test system, vehicle and occupant response repeatability evaluation in rollover crash tests: The deceleration rollover sled test," International Journal of Crashworthiness, **16**(6), 583-605, 2011.

[32] Bo.C. Chen, and H. Peng, "A Real-time Rollover Threat Index for Sports Utility Vehicles," in Proceedings of the Amencan Control Conference, 1999.

[33] Bo.C. Chen, and H. Peng, "Rollover Warning Of Articulated Vehicles Based On A Time-To-Rollover Metric," in Proceedings of the 1999 ASME International Congress and Exposition, Knoxville, TN, November 1999.

[34] B.L. Boada, M.J. L. Boada, L. Vargas-Melendez, and V. Diaz, "A robust observer based on H1 filtering with parameter uncertainties combined with Neural Networks for estimation of vehicle roll angle," Mechanical Systems and Signal Processing, **99**, 611–623, 2018.

[35] X. Glorot, and A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," in Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS) 2010.

[36] L. Wang, N. Zhan, and H. Du, "Real-time Identification of Vehicle Motion-Modes using Neural Networks," Mechanical Systems and Signal Processing **50-51**, 632–645, 2015.

[37] E.N. Sanchez, L.J. Ricalde, R. Langari, and D. Shahmirzadi, "Rollover Prediction and Control in Heavy Vehicles via Recurrent Neural Networks," 43rd IEEE Conference on Decision and Control, December 14-17, 2004.

[38] K. Treetipsounthorn, and G. Phanomchoeng, "Real-Time Rollover Warning in Tripped and Untripped Rollovers with A Neural Network," in The 4th International Conference on Control Science and Systems Engineering (ICCSSE 2018), Wuhan, China, Aug. 2018.

[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, **9**(8), 1735–1780, 1997.

[41] R. Dey and F. M. Salemt, "Gate-variants of gated recurrent unit (gru) neural networks," in 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS). IEEE, 1597–1600, 2017.

[42] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks,"

in Proceedings of the fourteenth international conference on artificial intelligence and statistics, 315–323, 2011.
[43] P. J. Werbos et al., "Backpropagation through time: what it does and how to do it," in Proceedings of the IEEE, **78**(10), 1550–1560, 1990.
[44] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in International Conference on Machine Learning, 1310–1318, 2013.
[45] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," arXiv preprint arXiv:1412.3555, 2014.
[46] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in International Conference on Machine Learning, 2342–2350, 2015.
[47] C. Winkler, K. Campbell, and C. Mink, "Center of gravity height: a round-robin measurement program," A Technical Report to The Motor Vehicle Manufacturers Association MVMA Project No. 0167A, The University of Michigan, Transportation Reserch Institute, 1991.
[48] C. E. Rasmussen, Gaussian Processes for Machine Learning. MIT University Press Group Ltd, 2005.
[49] D. Yuret, "Knet: beginning deep learning with 100 lines of julia," in Machine Learning Systems Workshop at NIPS 2016, 2016.
[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," CoRR, abs/1412.6980, 2014. [Online] Available: http://arxiv.org/abs/1412.6980

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 239-246 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Hybridization of Improved Binary Bat Algorithm for Optimizing Targeted Offers Problem in Direct Marketing Campaigns

Moulay Youssef Smaili[*,1], Hanaa Hachimi[1,2]

[1]*Systems Engineering Laboratory, ENSA, KENITRA, 14020, Morocco*

[2]*Systems Engineering Laboratory LGS, BOSS Team, Sultan Moulay Slimane University, Av Med 5, Po Box 591, Beni Mellal, Morocco*

ARTICLE INFO

Article history:
Received: 19 August, 2020
Accepted: 21 October, 2020
Online: 10 November, 2020

Keywords:
Binary Bat Algorithm
Genetic Algorithm
Hybridization
Low level Hybridization
Customer Targeting

ABSTRACT

One of the biggest business problems of marketers, is to optimize the return on investments of direct Marketing campaign. This main purpose which can only be ensured by targeting the appropriate customer. The main challenge faced by companies when advertising, is to configure properly a campaign, by choosing the right target, so A high user acceptance rate is ensured to advertisements. However, when dealing with an important size of data, the important specification to consider is the combinatorial aspect of the problem and the limitation of the approach based on mathematical programming methods. In this work, and considering the optimization of targeting offers as an of NP-hard problems, we concluded that the use of a meta heuristic algorithm is more suitable to use a classical (exact) method. We choice to use an improved bat Algorithm hybridized with Genetic Algorithm. The results of computational experiments confirmed that the proposed algorithm gives competitive results.

1. Introduction

There are two axes which characterize research in metaheuristic: exploration and exploitation. Exploration guarantees a diversified search in the space regions; however, the exploitation seeks to ensure the intensification of the search to find a better solution. The goal is to have a balance between these two contradictory objectives [1]. In the literature, many works make a difference between local search methods, instead of single solution-based and population-based methods. In this work, and unlike previous works, we deal with the targeted offers problem in direct marketing campaigns using population-based methods, and we focus on two algorithms; the Genetic Algorithm which can easily escape from the local minimum, and the Bat Algorithm (BA) which has the characteristic to quickly converge towards a local solution. Based on the imitation the concept of natural choice and heredity, the Genetic Algorithm is considered one of the most famous evolutionary search techniques. It was effective in optimizing several types of problem where a deterministic polynomial solution is infeasible. The Bat Algorithm, inspired by the foraging behavior of micro-bats, has been developed by Yang in 2010 [2], and it can be considered as an effort to gather positive side of previous metaheuristic algorithms. Thanks to its simplicity, flexibility, and easy implementation, the Bat Algorithm works well with complicated problems. The main feature of this algorithm is the quick convergence at a very initial stage by keeping a balance when switching from exploration to exploitation. In order to maximize their benefits, firms use "marketing campaigns" as a fundamental tool, along two main axes; by attracting new customers, so-called acquisition or by prospecting additional revenues with their existing customers by proposing new products to them; and this is called retention. Our approach focuses on the second axis (retention). To maximize the marketing profit to determine the combination: which product to offer and for which customer, under a set of constraints by using a new improved Bat Algorithm hybridized with Genetic Algorithm. The final goal is to choose the right product, to offer to the right customer, at the right time to maximize the marketing Return on Investment (ROI). This goal is not easy to implement because of the fact that firms have multiple products to offer under a complex set of business constraints [3]. In this work, and due to the discrete aspect of this kind of problems, the use of a binary version of the Bat Algorithm is essential. The standard BA was originally designed for the optimization of non-discrete problems [4]. The objective is to find a solution with a discrete value. The sought solution is represented by a matrix which contains binary values; 1 if product j is offered to customer i, and 0 otherwise. This paper is organized as follows,

[*]Corresponding Author: Moulay Youssef Smaili, 00212664349580, Email: smailli.moulayyoussef@uit.ac.ma

section 2 describes some related works in the same area; works about the targeted marketing problem are mentioned first, followed by works which process the Bat Algorithm's binary aspect in a discrete search space. In section 3, we describe the problem of Market Targeting and the modeling adopted to search a solution for the problem. In section 4, we describe two metaheuristic algorithms; namely the Binary Bat Algorithm (BBA), and the Genetic Algorithm. In this section we present our proposed algorithm: Improved Genetic Binary Bat Algorithm (IGBBA) developed to encounter the shortcomings of the two first algorithms while searching for adapted solutions. Experimental results are presented in section 5 for the three algorithms with a comparison of their performance.

In [3], author presented an optimization problem and replaced the classic database Marketing problem intended to generate the optimal profit under a set of business constraints. Approximately 2.5 million customers, eleven unique offers, and five investments are considered. The solution provided an approximately optimal solution to the ideal capacity assignment problem. The result is a decision describing which customer to target and which product to offer to him. In [5], author presented eight algorithms to solve the problem of maximizing the Return on Investment of targeted marketing. Seven of these algorithms are considered as exact methods and the eighth one is a heuristic method: the Tabu Search (TS). In the results, the TS showed the best performances. In [6], author took up the problem as described by author in [5], focusing on solving the problem by developing an efficient technique. The approach followed by author in [6] consists of hybridizing two metaheuristics: Greedy Randomized Adaptive Search Procedure (GRASP), and General Variable Neighborhood Search (GVNS). Three steps characterize the method followed by in [6]; the first one verified the efficiency of the GRASP algorithm proposed to initialize a good solution. In the second step, author kept the same parameters, but this time, they called the VND procedure for a local search. In the third step the complete proposed algorithm GGVNS was tested and it provided better results. In [7], author proposed a multi-objective DMLS metaheuristic for searching for sets of non-dominated solutions, in order to maximize both the promotion campaign's total profit and the risk-adjusted return. In order to remedy to the concept of volatility, author, in [7] proposed a greedy randomized initial solution builder, able to perform local searches considering different neighborhood exploration techniques, and a genetic method is designed. All the works cited above used metaheuristic based on local search methods to ensure an optimization of the return on investment, while respecting the business constraints. On the other side, the use of a Binary Bat Algorithm will allow us to deal with the target marketing optimization problem by looking for solutions made up of discrete (binary) values instead of looking for continuous solutions. In designing the Bat Algorithm in its binary version, how to update velocity and position must be adapted in order to guarantee the binary aspect of the research. In [8], the goal is to use the concept of velocity and position of bats in real space to move each bat in a discrete space with 0 and 1 values only. The idea was to change the position of micro-bats with the probability of their velocity. To achieve this objective, the technique used is based on a transfer function which allows a transformation of the value of the velocity to a probability, to update the position. this amounts to saying that this transfer function calculates the probability of changing the position from 0 to 1 and vice versa.

2. Problem Description

Direct Marketing is a tool that allows firms to promote their products directly to customers, and measure results quickly. One of the most important benefits of Direct Marketing is "Upgrading firm's loyalty strategies" in order to maximize the companies' Return on Investment. In [5], and In [3] before, authors presented the formulation of the product targeting problem as a mixed-integer programming (MIP) problem including more business constraints, and it can be rewritten as follows:

Given a set of m customers $C = [c_1, c_2 ... c_m]$, and a set of n products' offers $P = [p_1, p_2 \ldots p_n]$, the objective is to maximize the Return on Investment under these business constraints:

- The corporate hurdle rate: each company defines its hurdle rate (HR) to make sure that the Return on Investment is equal, at least, to a value of HR.
- During the campaign, the budget of each product is limited.
- A limitation is imposed on the total number of products offered to each customer.
- And there is also a Minimum Quantity Commitment (MQC), which is the number of units of a product to be offered in order for that product to be part of the campaign. It means that no customer will receive an offer of a product which is not part of the campaign. If the product belongs to the proposed ones then at least $Pj > 0$ customer will receive an offer.

A solution is represented by a binary array $R_{|C|\times|P|}$, where C indicates the set of available costumers, and P represents the possible products to be used in the campaign. If a given cell $s_{i,j}$ |i $\in$ C, j $\in$ P is equal to "1" (true), the product j will be offered to the customer i; otherwise, the value would be "0" (false).
There are two basic parameters of the customer lifetime value's computation; p_{ij} is the probability that customer i responds positively to an offer of product j, and DFV_{ij} is the return to the firm when customer i responds positively to the offer of product j [11, 12]. the problem can be modeled as follows:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(r_{ij}-c_{ij})x_{ij} - \sum_{j=1}^{m} fc_j y_j \qquad (1)$$

where:

- r_{ij} is the expected revenue of the company from the offer of product "j" to customer "i", and $r_{ij} = p_{ij} DFV_{ij}$
- c_{ij} is the cost generated while offering product "j" to customer "i"
- M_i is the upper limit value of number of different products to offer to a customer "i"
- P_j is the MQC limit associated with product "j"
- B_j is the budget, associated to the product "j" during campaign.
- fc_j is the cost of using product j in the campaign
- O_j is the MQC bound of the product j,
- HR is the hurdle rate specific to each company

The goal is to maximize the fitness function (1), by finding the optimal combination of the two matrices X and Y:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix} \text{ and } Y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

where:

$$x_{ij} = \begin{cases} 1 \text{ If product j is offered to customer i,} \\ 0 \text{ Otherwise} \end{cases}$$

and,

$$y_j = \begin{cases} 1 \text{ If product j is used in the campaign,} \\ 0 \text{ Otherwise} \end{cases}$$

The constraints, to which the problem is subject to, can be modeled as follows:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij}x_{ij} \geq (1+HR)\left[\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} + \sum_{j=1}^{n} fc_j y_j\right] \quad (2)$$

$$\sum_{i=1}^{m} c_{ij}x_{ij} \leq B_j \quad \text{for each j = 1…, m} \quad (3)$$

$$\sum_{j=1}^{n} x_{ij} \leq M_i \quad \text{for each i = 1…, m} \quad (4)$$

$$\sum_{i=1}^{m} x_{ij} \leq m y_j \quad \text{for each j = 1…, n} \quad (5)$$

$$\sum_{i=1}^{m} x_{ij} \geq P_j y_j \quad \text{for each j = 1…, n} \quad (6)$$

$$x_{ij}, y_{ij} \in \{0,1\} \quad \text{for each i =1…m and j=1 …, n} \quad (7)$$

Figure 1 below, illustrates an example of a problem with an advertising campaign optimization. The matrix represents a solution $S_{|C|x|P|}$ concerning 5 products to offer to 10 customers, so as to optimize the Return on Investment by taking into account business constraints (maximum products to offer for each customer, fixed cost and budget for each product). In summary, this figure represents the offers to be adopted by each customer. The first column on the left, represents the 10 customers who are the subject of the campaign with the maximum product values to offer to each Customer. The first line at the top represents the different product offers with the fixed cost of each product and the budget allocated during the campaign. the values 1 means that product j is offered to customer i during the campaign and 0 if not.

3. Methodology

There are several types of algorithm hybridization; The hybridization used in this paper contains both sequential and parallel hybridization. The sequential one concern using GA in the initialization of a feasible solution. The parallel hybridization is used for perfecting the solution found in each BA iteration by an application of the GA limited in terms of number of iterations.

3.1. Building initial solution using Genetic Algorithm (GA):

One of the very important phases when dealing with the customer targeting problem, is the creation of an initial good quality solution which must be feasible by respecting the business constraints.

$$(P1 - P2 - P3 - P4 - P5)$$

$$\begin{pmatrix} M_1 = 3 \\ M_2 = 3 \\ M_3 = 2 \\ M_4 = 2 \\ M_5 = 2 \\ M_6 = 2 \\ M_7 = 3 \\ M_8 = 2 \\ M_9 = 2 \\ M_{10} = 3 \end{pmatrix} \begin{pmatrix} 1\ 0\ 0\ 1\ 1 \\ 0\ 0\ 1\ 0\ 0 \\ 1\ 1\ 0\ 0\ 1 \\ 1\ 0\ 1\ 0\ 1 \\ 0\ 1\ 0\ 0\ 0 \\ 1\ 0\ 0\ 1\ 0 \\ 0\ 1\ 1\ 0\ 1 \\ 1\ 0\ 1\ 1\ 1 \\ 1\ 0\ 0\ 0\ 1 \\ 1\ 0\ 1\ 0\ 1 \end{pmatrix}$$

Figure 1: A solution example for the optimization problem of customer targeting

The classical initialization of the Bat Algorithm, and which is based on randomization, creates a non-feasible or poor-quality initial solution. To solve this problem, we opted to build the initial solution using GA with a reduced population, and a number of reduced chromosomes. Genetic Algorithms are considered as the most known heuristic technique of search and optimization, and are inspired by natural evolution. They have proven to be effective in several complex problems [3]. GA modeled the decision variable of a search problem by a string composed of alphabets with a certain cardinality. Each chain is considered a candidate solution, and it is called a chromosome, while the alphabets that make up the chromosomes represent the genes. In the problem of marketing campaigns' optimization, a chromosome is considered as a string representation of solutions to the problem. In this kind of problems, the classical GA is imposed and represents the solutions by a bit string. Chromosomes composed of bit strings is a representation formed of genes with binary values allele. These values take 1 if a product is offered to the customer and 0 otherwise. The gene may represent a decision to offer a product j to a specific customer i. The final goal is to optimize the fitness function. It is a calculation which determines to what extent the chromosome can represent a solution to a given problem. GA is also characterized by the notion of population; It treats an artificial population; these are strings in a binary alphabet usually [9]. GA is composed of 6 steps to find an optimal solution for the problem; the objective in targeted offers problem is to maximize the return on investment by finding an optimal solution of equation (1):

1. Initialization

   A randomized procedure is usually used when generating the initial population of the candidate solution. However, if we randomly choose the composition of the chromosomes with a weighting of 50% for each value 0 and 1, we will obtain a non-feasible solution which do not respect the business constraints. For this, we have opted in this paper for a weighting distribution of 80% when a product j is offered to customer i, and 20% when it is not. With this method we have the certainty that our initial solution respects the business constraints but it will not necessarily represent an initial good quality solution.

2. Evaluation

   When the population is initialized first time (population initialized), the fitness values of each candidate solutions are evaluated to determine how fit an individual is.

3. Selection

   The selection is a step that keeps the best solutions by memorizing copies of those solutions with the highest fitness values. This concept imposes the mechanism of maintaining the most suitable among the candidate solutions

4. Crossover

   Crossover is the most impactful phase in GA. It is a phase which consists in combining two or more parent solutions in order to generate a better solution.

5. Mutation

   The GA ensures via this stage a diversity of the population in order to escape a local convergence, by modifying the solution in an arbitrary way while acting locally.

6. Termination

   The algorithm terminates if there is no improvement in the solution quality, when time is reached, or when an acceptable solution is obtained.

The following is a pseudo code of Genetic Algorithm (GA):

Algorithm 1: Genetic Algorithm

```
Randomly create initial population
Calculate the fitness function for all individuals.
while Iteration < MaxGeneration do
        Select the best-fit individuals for reproductions;
        Proceeding to offspring by execution crossover and mutation phases;
        Evaluating the fitness function of new chromosomes;
end while
Return the best solution in the population
```

Figure 2: Pseudo Code of Genetic Algorithm

3.2. Binary version of Bat Algorithm

The Bat Algorithm (BA) is one of the most recent metaheuristic algorithms which uses a technique of imitation of micro-bats behavior to find an optimal solution. Previous work has proven the effectiveness of this algorithm compared to old metaheuristics like Genetic Algorithm (GA) or the Particle Swarm Optimization [8,10]. However, the continuous solutions generated by the BA are not valid for certain types of problem. The objective of our paper is to find a discrete solution (1 when the product "j" is offered to the customer "i" and 0 otherwise). For that reason, in this paper, the use of a binary version of Bat Algorithm is imposed [4]. In order to adapt this algorithm to the context of the Targeted Offers problem in Direct Marketing Campaigns, a binary version of Bat Algorithm will be used and which is proposed by authors [8,10]. This version is called Binary Bat Algorithm (BBA). By modifying the micro-bats positions values from 0 to 1 and vice versa, the BBA moves micro-bats in a binary search space [8, 10]. In [11,2], When designing the BA, author respected the two main techniques used by microbats to find their prey. Micro bats decrease loudness and increase the pulse rate of sound emitted during chase. These characteristics have been modeled as follows [11,2]: Each micro-bat i is characterized by its position pos_i^t, and its velocity vel_i^t at a specific iteration (time). Exploring new solutions is performed by updating positions of micro-bats as following:

$$vel_i^{t+1} = vel_i^t + (x_i^t - Glob^{best})Freq_i \quad (8)$$

$$pos_i^{t+1} = pos_i^t + vel_i^{t+1} \quad (9)$$

$Glob^{bes}$ represents the best solution achieved until this iteration, and $Freq_i$, which represents the frequency of the i-th microbat, is obtained following this equation:

$$Freq_i = Freq_{\min} + (Freq_{\max} - Freq_{\min})\beta \quad (10)$$

where β is randomly generated from a uniform distribution in [0,1]. Thanks to these equations the exploitability aspect is ensured in the Bat Algorithm. However, the exploitation side is guaranteed by a random walk procedure as follows:

$$pos_{new} = pos_{old} + \varepsilon L^t \quad (11)$$

L^t is the loudness of the bats' emitted sound, in order to perform an exploration instead of exploitation. L^t is the average loudness of all the micro-bats at this iteration step: $L^t = \langle L_i^t \rangle$, and ε is a random number in [-1,1] [2,8,9].

To keep balance between exploration aspect and exploitation aspect during the search process, loudness L^t, and pulse emission rate r_i should be updated only if the candidate solution is improved as the iterations proceed. They are updated as following [2,10]:

$$L_i^t = \alpha L_i^t \quad (12)$$

$$r_i^t = r_i^0 \left[1 - \exp^{(-\gamma t)}\right] \quad (13)$$

where α and γ are predefined constants. The search principal in BA, is once a micro-bat has found the prey's location, the loudness will be decreased while the pulse rate will be increased. However, The Bat Algorithm is designed, at the origin, to search in a continuous space of solutions. The algorithm, in its original version is not applicable when the set of solutions is discrete, because during an optimization in a binary search space, by treating only values 0 and 1, the Eq. (9) is no longer valid to update the position of the microbats. To encounter this obstacle, authors in [8, 14] proposed a new solution: the idea is to use a transfer function with the aim of updating the position of micro-bats with the probability of their velocities. In practice the transfer function has the role of mapping velocity values to probability values. In other words, the transfer functions are used to move the microbats in a binary space instead of a discrete space. The most important factor for designing this function is to force micro-bats with high velocities to switch their positions. To do this, a transfer function, called v-shaped function, and position updating rule are proposed in [8,10], and described in Equations (14) and (15) as follows:

$$VEL(vel_i^k(t)) = \left| \frac{2}{\pi} \arctan\left(\frac{\pi}{2} vel_i^k(t) \right) \right| \quad (14)$$

$$pos_i^k(t+1) = \begin{cases} (pos_i^k(t))^{-1} & \text{if rand} < VEL(vel_i^k(t+1)) \\ pos_i^k(t) & \text{if rand} \geq VEL(vel_i^k(t+1)) \end{cases} \quad (15)$$

$pos_i^k(t)$ and $vel_i^k(t)$ are the position and velocity of microbat "i" at the iteration "t", where k is the problem dimension. $(pos_i^k(t))^{-1}$ is the complement of $pos_i^k(t)$. Figure 3 presents the general steps of the Bat Algorithm in its binary version.

3.3. Proposed Hybridized Bat Algorithm

While dealing with the problem of targeted marketing maximization, we note that the Genetic Algorithm takes a considerable time to find a solution but with a minimum chance of blocking in a local minimum, while the Bat Algorithm has the advantage of quickly converging on a solution to the problem. However, BA may easily be trapped in a local minimum. Taking into account the random aspect of the initialization procedures, the two algorithms suffer from the poor quality of the initial solution. To overcome these obstacles, we propose to hybridize the Binary Bat Algorithm, to benefit from convergence in advanced iterations, with the Genetic Algorithm that generates better population from good parents. These results are close to a global optimum.

Algorithm 2: Standard Binary Bat Algorithm

Initializing the population of micro-bats: POS_i(i=1, 2, ..., n)=rand(0 or 1) and VEL_i=0
Define initial pulse frequency value $Freq_i$
while (num_iteration < Max generation iterations) **do**
  Update both frequency and velocities values
  Determining transfer function value by Eq. (14)
  Update micro-bats' positions by executing Eq. (15)
  if (rand > r_i) **then**
    Randomly choice a solution ($Glob^{best}$) among the best solutions.;
    Modifying some of the dimensions of position vector with some of $Glob^{best}$ dimensions;
  end if
  Randomly flying and Generating a new solution
  if (rand < Li and f(posi) > f(Globbest) and Constraint(Globbest)=True) then
  Accept the new solutions;
  end if
  Evaluating the fitness function for all microbats in the population
  Increase r_i and reduce L_i
  Getting $Glob^{best}$ after ranking the micro-bats
end while
Return the best solution.

Figure 3: Pseudo Code of Standard Bat Algorithm

The initial solution in the proposed algorithm is generated by performing the Genetic Algorithm for the first time, with a reduced number of chromosomal parents; this will quickly generate an initial solution of better quality than standard algorithms. The exploration and exploitation are controlled by pulse emission rate, in the case of standard Binary Bat Algorithm [12,13]. By proceeding iterations, this factor takes on larger values, which causes a progressive loss of the exploitation capacity of the algorithm. by introducing linear decreasing inertia weight factor, the exploitation capability of the algorithm is improved while the exploration capability will be improved by the hybridization of both algorithms. Figure 4 presents the steps of the proposed algorithm:

Algorithm 3: Hybridized Binary Bat Algorithm

```
Randomly create GA initial population
Calculate the fitness function for all individuals
while Iteration < MaxGeneration do
      Select the best-fit individuals for reproductions;
      Proceeding to offspring by execution crossover and mutation phases;
      Evaluating the fitness function of new chromosomes;
end while
Return Initial solution to execute Binary Bat Algorithm
Initializing the population of micro-bats: POS_i(i=1, 2, ..., n)=rand(0 or 1) and VEL_i=0
Define initial pulse frequency value Freq_i
while (num_iteration < Max generation iterations) do
   Update both frequency and velocities values
   Determining transfer function value by Eq. (14)
   Update micro-bats' positions by executing Eq. (15)
   if (rand > r_i) then
      Randomly choice a solution (Glob^best) among the best solutions.;
      Modifying some of the dimensions of position vector with some of Glob^best dimensions;
   end if
  Randomly flying and Generating a new solution
  if (rand < Li and  f(posi) > f(Globbest) && Constraint(Globbest)=True) then
  Accept the new solutions;
  end if
  Evaluating the fitness function for all microbats in the population
  Increase r_i and reduce L_i
  Getting Glob^best after ranking the micro-bats
end while
Return the best solution.
```

Figure 4: Pseudo Code of Hybridized Genetic Binary BA (IGBBA)

4. Computational Experiments and Discussion

The proposed algorithm was developed in Matlab 2015on a Lenovo T470s computer Core i7, with 2.6 GHz clock speed and 20 GB RAM. The linear solution was obtained using CPLEX (IBM solution), and the dataset, proposed and generated by author in [5], is used for comparing the three algorithms:

- The cost value c_{ij} may take 1,2 or 3 (Randomly)
- The return of the firm r_{ij} is an integer generated between 0 and 16 randomly.
- The corporate hurdle rate HR can take 5%, 10%, and 15% values.
- In this paper, there are three different values of the number of customers: 100 customers for small category S, 1000 customers for medium category M, and 10,000 customers for large category L.
- For each category of customers, we have two different numbers of product n: 5 and 10 products.
- For each combination (category of customer and number of products), a minimum-quantity commitment bound P_j is randomly generated between $\left[\frac{\sum_i M_i}{n}\right]$ and $\left[2\frac{\sum_i M_i}{n}\right]$.
- The fixed cost fc_j is generated randomly between $\frac{O_j}{2m(1+HR)}\sum_i\left[r_{ij}-(1+HR)c_{ij}\right]$ and $\frac{O_j}{m(1+HR)}\sum_i\left[r_{ij}-(1+HR)c_{ij}\right]$
- For budget B_j, three values are adopted: the two values of $\left[O_j\frac{\sum_i c_{ij}}{m}\right]$ $\left[2\frac{\sum_i M_i}{n}\frac{\sum_i c_{ij}}{m}\right]$ and a random integer between $\left[O_j\frac{\sum_i c_{ij}}{m}\right]$ and $\left[2\frac{\sum_i c_{ij}}{n}\right]$.
- The upper bound M_i is selected between 1 and n/5.

To test the performance of the proposed algorithm, we report the same dataset and conditions from computational experiments in [5] and [6]. The GAP is calculated when a feasible solution is found by the adopted algorithm. By analyzing Table 1, we clearly deduce that the proposed IGBBA algorithm gave better solutions compared to GA and BBA.

The GAPs of IGBBA are significantly lower, and the BBA appears as the second-best algorithm by comparing the GAPs. Figure 5, which is a graphic representation of the fitness function using the three discussed algorithms, shows the fast convergence of the BBA and the IGBBA compared to the Genetic Algorithm. As reported in Table 1, and for the three categories of the S3 format, IGBBA is clearly better than the methods used in the previous works, with a clear difference in performance based on the GAP calculation. However, for the medium and large categories, the IGBBA is considered very competitive and sometimes better than the other algorithms.

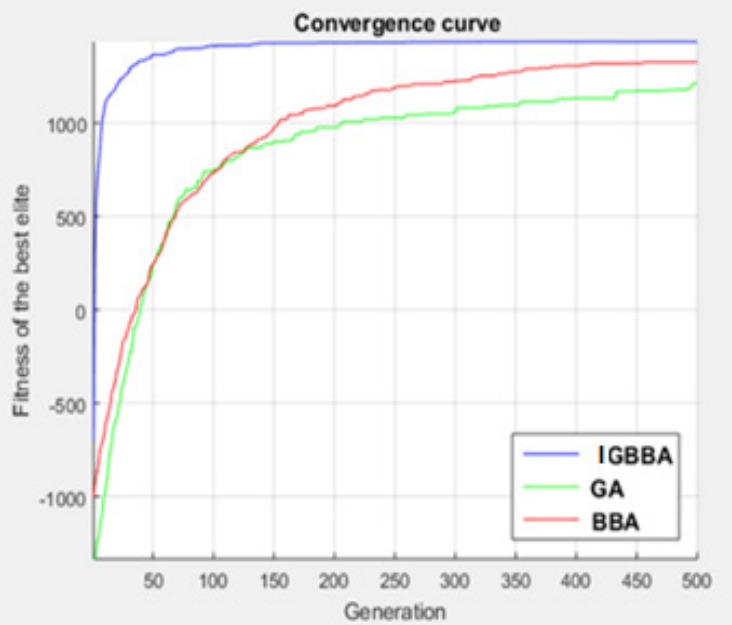


Figure 5: Convergence of fitness function

Table 1: Experiment results

	Small			Medium1			Medium2			Large		
	5	**10**	**15**	**5**	**10**	**15**	**5**	**10**	**15**	**5**	**10**	**15**
			Results experiments of [5]									
Initial Solution	19.2	17.85	18.34	17.9	16.2	18.36	18.92	19.23	17.84	26.262	24.866	25.655
Tabu Search	6.86	6.52	7.76	7.22	8.54	7.6	9.75	**9.58**	9.11	10.866	11.04	10.23
γ			**Results experiments of [6]**									
0	10.68	6.98	7.77	9.72	9.12	12.15	11.29	9.9	11.15	12.999	12.29	12.15
0.2	6.9	7.14	7.92	7.22	8.67	7.69	10.26	9.59	9.34	11.224	10.94	10.74
0.4	6.79	6.45	7.58	8.47	9.1	7.89	9.68	9.76	9.91	10.866	11.22	10.11
0.6	6.77	**6.5**	7.49	7.24	**8.42**	7.63	9.85	9.78	9.44	11.511	11.93	10.22
			IGBBA results									
Initial Solution (GA)	23.8	19.05	19.75	20.1	18.8	17.21	19.45	22.88	18.66	27.332	25.667	26.31
IGBBA	**6.69**	6.51	**7.01**	**7.11**	8.55	**7.12**	**9.45**	9.77	**9.07**	**10.866**	**11.01**	**10.11**

5. Conclusion

The BA which is one of the most efficient algorithms was appointed to deal with that type of problem; it was used in binary version given the discrete nature of the solution space. In order to be able to search in a discrete space, a transformation of the positions of the micro-bats, to discrete values, is essential. The hybridization can provide a more efficient behavior and a higher performance when dealing with large-scale problems. This combination carried out in this paper has shown its effectiveness by remedying the disadvantage of the risk of jamming in a minimum local. As extension of our work, new constraints can be proposed, like reducing products costs where customer respond positively to a big number of products or where the number of customers respond positively to a specific product j. On the other hand, an optimization of IGBBA algorithm can be proposed to reduce the execution time. We addressed a very important issue in the marketing strategy of companies in order to optimize their return on investment during advertising campaigns. We have taken up the modeling done by author in [5] with the business constraints that must be respected to maximize the profit of the company. To test its proposal, author, in [5], created a dataset randomly, respecting a certain logic of the upper and lower limits of each parameter; a job that will produce several dataset categories according to different sizes depending on the number of customers, number of products and the value of Hurdle rate. A dataset on which this paper was based to test our proposal and compare it with the results of [5] and the result of other works. Since the problem is considered an NP-hard one, the use of metaheuristics is required; we have treated the problem based on methods based-population differently to the works cited above and which are all based on neighborhood methods.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] R. Faia, T. Pinto, Z. Vale, J.M. Corchado, "Hybrid particle swarm optimization of electricity market participation portfolio," 2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017 - Proceedings, 2018-January (703689), 1–8, 2018, doi:10.1109/SSCI.2017.8285218.

[2] X.S. Yang, "A new metaheuristic Bat-inspired Algorithm," Studies in Computational Intelligence, **284**, 65–74, 2010, doi:10.1007/978-3-642-12538-6_6.

[3] M.D. Cohen, "Exploiting response models - Optimizing cross-sell and up-sell opportunities in banking," Information Systems, **29**(4), 327–341, 2004, doi:10.1016/j.is.2003.08.001.

[4] M.Y. Smaili, H. Hachimi, "Metaheuristic for optimizing targeted offers in Direct Marketing," 6th International Conference on Optimization and Applications, ICOA 2020 - Proceedings, 4–8, 2020, doi:10.1109/ICOA49421.2020.9094454.

[5] F. Talla Nobibon, R. Leus, F.C.R. Spieksma, "Optimization models for targeted offers in direct marketing: Exact and heuristic algorithms," European Journal of Operational Research, **210**(3), 670–683, 2011, doi:10.1016/j.ejor.2010.10.019.

[6] T.A. Oliveira, V.N. Coelho, M.J.F. Souza, D.L.T. Boava, F. Boava, I.M. Coelho, B.N. Coelho, "A hybrid variable neighborhood search algorithm for targeted offers in direct marketing," Electronic Notes in Discrete Mathematics, **47**, 205–212, 2015, doi:10.1016/j.endm.2014.11.027.

[7] V.N. Coelho, T.A. Oliveira, I.M. Coelho, B.N. Coelho, P.J. Fleming, F.G. Guimarães, H. Ramalhinho, M.J.F. Souza, E.G. Talbi, T. Lust, "Generic Pareto local search metaheuristic for optimization of targeted offers in a bi-objective direct marketing campaign," Computers and Operations Research, **78**, 578–587, 2017, doi:10.1016/j.cor.2016.09.008.

[8] S. Mirjalili, S.M. Mirjalili, X.S. Yang, "Binary bat algorithm," Neural Computing and Applications, **25**(3–4), 663–681, 2014, doi:10.1007/s00521-013-1525-5.

[9] J. McCall, "Genetic algorithms for modelling and optimisation," Journal of Computational and Applied Mathematics, **184**(1), 205–222, 2005, doi:10.1016/j.cam.2004.07.034.

[10] R.M. Rizk-Allah, A.E. Hassanien, "New binary bat algorithm for solving 0–1 knapsack problem," Complex & Intelligent Systems, **4**(1), 31–53, 2018, doi:10.1007/s40747-017-0050-z.

[11] S.S. Guo, J.S. Wang, X.X. Ma, "Improved Bat Algorithm Based on Multipopulation Strategy of Island Model for Solving Global Function Optimization Problem," Computational Intelligence and Neuroscience, **2019**, 2019, doi:10.1155/2019/6068743.

[12] X. Huang, X. Zeng, R. Han, "Dynamic Inertia Weight Binary Bat Algorithm with Neighborhood Search," Computational Intelligence and Neuroscience, **2017**, 2017, doi:10.1155/2017/3235720.

[13] S. Yilmaz, E.U. Kucuksille, "Improved Bat Algorithm (IBA) on Continuous Optimization Problems," Lecture Notes on Software Engineering, **1**(3), 279–283, 2013, doi:10.7763/lnse.2013.v1.61.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 247-255 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

IT GRC Smart Adviser: Process Driven Architecture Applying an Integrated Framework

Meriyem Chergui[1,*], Aziza Chakir[2]

[1]*Computer sciences Department, Higher National School of Electricity and Mechanics, Hassan II University, Casablanca, 20000, Morocco*

[2]*Economic sciences and management Department, Faculty of Economic and Social Legal Sciences, Hassan II University, Casablanca, 20000, Morocco*

ARTICLE INFO

Article history:
Received: 04 August, 2020
Accepted: 21 October, 2020
Online: 10 November, 2020

Keywords:
IT Governance
IT GRC frameworks
Artificial Intelligence
Knowledge management
Information System
Business Processes
IT processes

ABSTRACT

This article is a continuation of the work presented in the Third World Conference on Smart Trends in Systems Security and Sustainability (WorldS4). In this version we focus on matching IT and business processes from different management levels by using IT-Governance (ITG) Risk and Compliance frameworks in a smart way. In fact, every information system (IS) has two main interfaces with two key systems: Operating System (OS) and Decision System (DS). Every system has his own frameworks and best practices for efficient management. IT Governance is the ability to control IT strategy implementation by ensuring business and IT alignment for profitable digital services. Also, the variety of IT Governance frameworks for each hierarchical level in the enterprise, should be efficiently deployed together in order to have coherent recommendations for IS, OS and DS users. It is why we use COBIT 5 as a strategic IT Governance framework able to match and integrate other frameworks and best practices dealing with IT services, IT risks and compliance as well. We based the proposed IT GRC smart adviser on artificial intelligence and knowledge management as technical axes. This article presents, IT Governance frameworks, their matching problems and related works. Then it proposes a smart architecture with new functionalities to match COBIT processes with other frameworks in order to cover different management levels. The technical contribution of this article is to propose and implement an IT GRC smart adviser based on many frameworks and best practices to optimize and improve IT Strategies. The simulation was presented, the obtained results were compared to human experts' decisions and big similarities were observed. The aim of this article is to integrate knowledge from many IT GRC frameworks for better IT governance in a smart and easy way.

1. Introduction

Information system Governance is nowadays different from classical management of software and hardware. Many users with different profiles are interacting directly or indirectly with it. They use IS to enrich their data sources through operation systems deployed in different gadgets for simple use. They use it to answer daily questions as well such as production rate, new clients number and providers' expectations. The created data is the most valuable resources ever. It enables CEO and other Business offices to take the right decision. But good IT governance allows to take this decision in the right time through the right IT solution.

The real challenge of top managers is to use digital solution to create value. They should also manage risk, and comply with regulations. The decision is no longer IT responsibility. It achieves business and technical offices as well. The use of one IT governance framework in its document version is no longer enough: It is true that they contain very advanced know-how and feedback, but it requires experts, time and a large budget to take advantage of them. These three elements: time, budget and knowledge are very expensive in the current market due to its severe competition and the law of the most powerful.

As a result, digital tools are mandatory to facilitates the governance of the information system. It is also a need for

*Corresponding Author: Meriyem Chergui, Email: chergui.meriyem@gmail.com

performance management system to ensure the right and efficient use of these frameworks.

Consequently, in this new multi-user, multi-partner environment, the CIO's missions remain information system strategy definition, IT resources management and technical problems resolution as well. Meanwhile, with Big data and cloud computing matters, we notice the emergence of new technical and business risks that should be managed effectively by IT and business together. To perform these tasks, many IT governance frameworks should be deployed in easy to use digital solutions as we did in [1]. But the main problem is how to integrate them in the same personalized repository for every company. How can we design an IT Governance, risk and compliance framework that matches the company's needs with optimal time, minimum budget and good software quality insurance? How can we integrate Data Governance into IT Governance in a fluid and value-creating way by taking in consideration Governance critical success factors [2]?

To answer these question, we propose an intelligent IT Governance model based on COBIT 5 as a business driven repository for corporate and IT governance. In fact, COBIT 5 offers with a common, non-technical and technology-independent language a management framework able to align governance and management standards. By providing a simple architecture to structure guidance documents, Cobit 5 contains knowledge previously distributed in different standards like COBIT 4, Risk IT, Val IT, ITAF and BMIS. It also allows practical matchmaking with different management levels' frameworks like ITIL, CMMI, ISO 27000 family and COSO.

In this article, we propose an update of the smart IT Governance adviser architecture [1] to allow IT services, IT risk management and compliance efficient management, in addition to matching stakeholders' motivations with real business goals [2].

We will add new layers communicating with the semantic model, to link stakeholders and digital solutions via IT Governance ontology designed and created in [3]. We also propose a smart matching between every COBIT process and its correspondence with other frameworks. In IT Governance jargon we call that "Appling a unique and integrated standard principal" for better alignment as presented in [4].

After the introduction in the first section, the second section presents the most pertinent IT Governance Risk and Compliance frameworks and then exposes an overview of their matchings with COBIT 5. The third section gives a brief state of art of Knowledge management as a technical axe. The fourth section analyses briefly related works. The fifth section presents the smart IT Governance adviser updates and the knowledge management system architecture. The fifth section, discusses the new model, its uses and its extensions, the sixth section details the model implementation and results analysis. In the last section, the article is concluded and work perspectives are listed.

2. IT GRC frameworks

In the ITG literature, numerous frameworks are presented, with advantages and limitations of each of them. Thus, each framework offers specific recommendation according to its domain. In practice, ITG should makes these repositories coexist for better results, since they present a necessary complementarily for pertinent decisions. We present some of the main standards in terms of enterprise management, services, projects management and IT security to show how IT governance frameworks drive accountability and process maturity [5]. In what follows, we present the most used frameworks namely ITIL, CMMI, ISO27001 and IT BSC.

2.1. ITIL

IT Infrastructure Library (ITIL), is an approach documentation for IT service management, to support business organizations users. It is created by the British government in 1990 [6]. In 2004, ITIL v2 was created with 9 books, focusing on the link between technology and business, and based on processes to provide the right services to customers. Many companies are using ITIL to improve IT services control and good results are as for processes improvements. ITIL service is a way to deliver value to a customer in an easy way and with low cost and risk. In 2007, ITIL V3 was implemented on five best practice books, offering supplements by sector or by market as well as generic models (process maps). This version is not subject to certification, and it offers continuous improvement of the services offered to IT Department clients. The latest version is ITIL V4. its main advantage is the integration of new agile practices and Devops method. It has developed 34 practices, in 3 themes instead of 26 processes in 5 categories.

- 14 in "General Management practices",
- 17 in "Service management practices"
- 3 in "Technical management practices".

ITIL V4 edition book introduces new like Service Value System (SVS) and the four-dimensional model.

2.2. ISO 27001

ISO 27001 is a framework to implement information security management system (ISMS). Released in November 2005, it also provides a model for operation, monitoring, and review the ISMS. It also enables the creation and implementation of good IT service management, a s presented in [7]. The risk management approach ISO 27001 uses is a top-down and neutral method. It is a planning process with six parts:

- Security policy definition
- WSIS scope definition.
- Risk assessment performance.
- Risk identification.
- Goals and controls implementation
- Applicability declaration.

ISO 27001 deals with responsibility management, internal audit and continuous improvement without specific information security controls. It should be noted that in the accompanying code for each practice, a checklist must be taken into consideration.

The last version is ISO27001 with 2014 and 2015 corrections. It highlights information security management systems requirements with special focus on information technology and security techniques.

2.3. IT Balanced Scorecard

Balanced Scorecard (BSC) [8] is the framework explaining company vision and strategy. It suggests action plans to give internal processes feedback with continues improvement. The BSC, according to its authors Robert Kaplan and David Norton, allows for traditional financial results, but these results highlight the past, which was normal in the industrial era, with long-term investments. There are insufficient customer relationships. These financial elements are insufficient, however, to control companies in the digital age, which should be based on their future value by investing in customers, suppliers, employees, processes, technology and innovation. ". The BSC offers a management tool, organized on four aspects:

- Financial aspect: define the financial benefit provided to shareholders
- Customer aspect: market and customer expectations;
- Internal process aspect: potential improvements to be made to the company's internal processes, which bring value;
- Future construction aspect: mobilization of human resources and infrastructure for improvement and learning.

It is assumed that this approach leads to a good vision of corporate Governance.

2.4. CMMI

The Capability Maturity Model (CMM) [9] is a project management framework developed by the Software Engineering Institute (SEI) of Carnegie Mellon University. It describes software development maturity practices and principals. With an ad hoc path, CMMI proposes to software development organizations a process improvement capability. IT contains a list of models like MM Software, CMM systems engineering and integrated development of CMM products. These models were merged and extended into CMM integration known as "CMMI" with either a stage view or continuous view. As for the staged view, it offers five organizational maturity level (initial, managed, defined, quantitatively managed and optimization). As for the continuous view, it includes six levels of processing capacity (incomplete, executed, managed, defined, quantitatively managed and optimization).

2.5. IT GRC Frameworks matching with COBIT

For better IT Governance, IT department must deal with each task by deploying it in its real context and link it with a project or a process. IT processes are the starting points while governing an information system. For example: deploying an ERP; computer backup organization; servers' security; or new IT supplier integration;

An IT process is defined as a set of activities around a specific event to measure results for customers [10].

While analyzing the IT GRC frameworks before, we remark that process-driven aspect is almost in all them. IT changes approaching components into active components inside the organization. The IT process is measurable in effectiveness and efficiency by defining activities to implement, responsibilities and controls. At this level, each repository focuses on one or more concepts to find more priority and to zoom on either governance risk or compliance.

Meanwhile, process management is highly recommended in the agile environment we are living in. This type of management allows [11]:

- Requirements understanding
- Processes added value;
- Process performance and efficiency measurement;
- Positive communication between stakeholders.
- Top Management easy and measurable decisions.

Technically speaking, IT process is easily designed as a simple or complex function or procedure with inputs and outputs. In this article encapsulate best practices and frameworks processes into computerized executable functions.

a) ITIL –COBIT

COBIT identifies what the CIO should do, ITIL prescribes how it should be done to optimize the use of IT resources. ITIL and COBIT have several common points (for example the BAI10 process in COBIT is equivalent to configuration management process in ITIL).

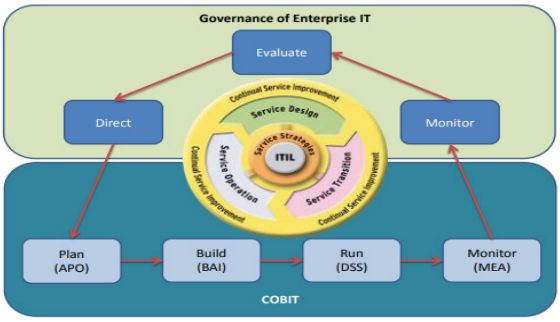


Figure 1: Matching COBIT with ITIL

COBIT 5 proposes the convenient IT investment to link priorities with objectives in both IT and business context (cascade principle). We have already seen it in [1]. The cascade of COBIT 5 objectives makes it possible to align IT objectives and company objectives, which makes it possible to refine and prioritize IT processes. the cascade of COBIT 5 objectives allows stakeholders to answer the following two questions:

- How to prioritize areas of interest when applying ITIL v3 processes?

- How to define the right maturity level and the efficient processes for better business decisions?

One last thing, COBIT 2019 is also correctly matched with ITIL v4 which allow us to upgrade our system regarding future IT GRC Frameworks versions.

b) CMMI –COBIT

Traditional mapping is designed to be used in one direction. Indeed, a user searching for an element in COBIT model finds the element or elements linked to the appropriate CMMI model. Others by the reverse probably exist in several places on the map (several elements of CMMI for a COBIT element) which is not easy to isolate to determine what is linked exactly. These traditional cards are unidirectional and they are valid for the correspondence of all IT GRC repositories. In this perspective, COBIT 5 CMMI Practices Pathway Tool, has emerged as a common Excel tool made by ISACA and CMMI, intended for users who want to navigate freely from COBIT 5 to CMMI and vice versa[12]. The dynamic and permanent correspondence between CMMI and COBIT makes it possible to limit maturity in the different levels of governance, which allows self-assessment and continuous development.

c) ISO27001 -COBIT

There is a mapping between COBIT 5 and ISO 27001 domains: “COBIT 5 for information security” this repository presents a mapping guide aiming at implementing the domain area of COBIT process with ISO27001 controls[13].

3. Knowledge Management

3.1. Definition

Knowledge Management (KM,) is a multidisciplinary research area offering methods, techniques and assistant tools: creation, acquisition, sharing and use of knowledge in an organization, with a view to 'improve individual, collective or organizational learning [14]. Knowledge is defined in a complementary way by raising one of his systemic aspects namely:

- Ontological (what is knowledge),
- Functional (what it does),
- Genetics (genetics here has the meaning of genesis not inheritance)
- Transformational (where it comes from and what it becomes)
- Teleological (why).

In artificial intelligence context, different types of knowledge are identified:

- Procedural knowledge to explain how can we solve a given problem is solved. It indicates how to complete a task;
- Declarative knowledge: it is statements describing objects and concepts and focusing on problem resolution indications.
- Meta Knowledge is specific domain knowledge explanation. For example, what is suitable at a given situation e to solve a given task;
- Heuristic knowledge: also called surface knowledge. It is empirical knowledge acquired by an expert through his past experience, it is used to guide reasoning;
- Structured knowledge: describes a mental model of the expert in the form of a structure (concepts, sub-concepts, objects, etc.).

3.2. Design methods

There are 3 approaches of Knowledge management modeling:

a) Bottom-up modeling approach:

The bottom-up or data-driven approach is based on a step of elicitation of expert knowledge (interviews, document analysis, task analysis, etc.) followed by a conceptualization step. Example: KOD and MIKE [14]. One of the advantages of the bottom-up approach is that it leaves experts free to express their perceptions and their tasks without constraining them. In addition, developed models correspond to existing points of view. However, the high cost in terms of time and expertise in the elicitation process, the difficulty of re-use as well as the complexity of the validation process are major drawbacks of the approach.

b) Top-down modeling approach:

This approach is called Model-Driven Approach. The development of knowledge models in this approach consists of finding pre-existing generic models and adapting them to the field and the application concerned. Two sub-approaches can be distinguished in this family: the select-and-modify approach and the modeling-by-composition approach from library elements. Example: CommonKADS [15], MASK [16]. The main advantages of this approach in addition to providing a support for knowledge acquisition are models generality and reusability. But adapting an existing model to a specific application is a very difficult task. In the literature, we also find mixed approaches which tried to take advantage of the two approaches like [17] and [18].

3.3. COBIT knowledge Base

Applying a single integrated framework is the third principle of COBIT 5 that proposes a knowledge management for COBIT alignment with other standards and frameworks for enterprise complete coverage. As shown in figure 2, Cobit 5 has a knowledge base of current guidance and contents collected from old versions and other standards and frameworks. This knowledge base enriched by enablers such as data, IT processes, resources and ethics allow different stakeholders to find answers to their questions about best practices to deal with different IT Governance situations in different levels (strategy, services providing, risk management…etc.). In fact, simplicity, guidance materials structuring and consistent production are the main advantages of this architecture that integrates different frameworks. knowledge in one shot [19]. IT Governance repositories are a set of knowledge sources capitalized by experts for better Information system management in different types of organizations. In this perspective, we are working on a smart IT Governance Knowledge

Management system that will be interfaced with the adviser in future version.

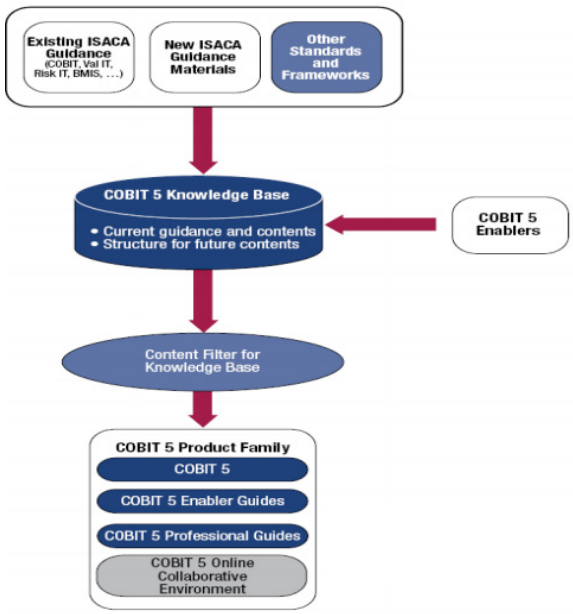


Figure 2: COBIT 5 Knowledge base

4. Related works

To assess the state of art of IT GRC platforms, a rigorous evaluation of the fourteen main suppliers [20] is presented in table 1. A comparison of the strengths of each of the solutions was established from the Forrester description, functionalities presented on the official sites and the demos available on the net, as long as the solutions are all proprietary.

Table 1: GRC existing platforms

Vendor	Product	Version	Strong points
ACL	ACL	ACL Analytics 12.5 ACL Analytics 12.5 Exchange 6.5	- Financial control analysis and testing -Audit solution - Community of practitioners -Simple user interface, mobile support and analytical integration.
Enablon	Enablon 8	V8.4	-Operational risks - Legal and regulatory aspect - Environmental, health and safety management - Corporate social responsibility initiatives, litigation, a responsible supply chain, air quality management or public service data.
IBM	IBM Ope,Pages	V7.3	-Financial control management -Operational risk management
LogicManager	LogicManager	V17	-Humain advisors availability
			-Fastest average deployment times
MetricStream	MetricStream M7 GRC PLatform	V7.0	-Energy, health care, insurance, food and beverage and automotive -Good user interface
Nasdaq	BWise	V5	-Middle market and information security organizations
NAVEX Global	PolicyTech NAVE Engage RiskRate EthicsPoint	Sep 1, 2017 July 31, 2017 August 30, 2017 August 15, 2017	-Compliance -Data Management
Riskonnect	Riskonnect Integrated Risk Management	GRC 2017.2	-Support 72 languages -GRC workflow and dashboards -Adapted to Salesforce.com clients
RSA Archer	RSA Archer	V6.2	- IT GRC - Financial risks and regulatory change management
Rsam	Rsam GRC	V9.2	-Data collection from different platforms -Risk analysis and management
SAI Global	Compliance 360	Sept 7, 2017	-Health care and IT GRC
SAP	SAP solutions for GRC	V10.1	- Big data and predictive analytics -Audit tool
ServiceNow	ServiceNow GRC	Jakarta	-IT Services management -IT GRC -Risks to information security and IT operations
Thomson Reuters	Thomson Reuters Connected Risk	Sept 7, 2017	-Risk and Compliance -Financial control management -Managing regulatory changes

The comparative analysis and the measurement of the effectiveness of these solutions in relation to the ITG dependency factors allow us to deduce that:

- In most cases, it is GRC solutions without IT GRC functionalities,
- The majority of these solutions even if they have IT GRC functionalities, it is ERP solution module to be deployed with the rest of the modules.
- The cost of setting up is high (by consulting the list of their references on the market and their targets).

- Need for external advice for better operation (possibly advice from the seller of the solution).
- There is a discrepancy with regard to the ITG implementation in the IS of various companies (size, turnover, type, maturity, location, etc.).
- They do not concern all of Information System stakeholders and all areas of the company at the same time.

5. Proposed model

This project started by EAS team [21] to propose an IT GRC digital solution with five intelligent modules for business and IT alignment by mapping processes. It also manages IT risks and measures business impacts. This was done by choosing for every process the right framework(s). There were many limitations for this architecture namely as far as execution time is concerned and also many communication problems among distributed systems which cause the decrease of IT Governance service's quality. IT is the reason why in[1], we implement the strategic level via a smart advisor to optimize the IT Governance semantic model. This solution main advantage was the "All in one" service and the easy use for final users. The advisor is a simple chatbot to which IT manager, Business Manager, CEO or even a project manager can ask IT Governance questions and he is answered simply and quickly.

In this version, we kept the same semantic engine model that interprets user real time questions using test mining. The same IT Governance ontology [3] to match concepts and relations but we replace our knowledge base by COBIT knowledge base. Cobit knowledge base, as shown before, is fed by Cobit but also the other repositories (ITIL, CMMI and ISO27001 at first) with processes matching.

This change allows the semantic engine not only to answer via Cobit guidelines but also ISO27001 for risk management, ITIL for services Governance and CMMI for IT project management in implicit way. The unique Knowledge base improves the performance of the system and ensures the efficiency of responses by arriving at a more delicate granularity for each process.

Indeed, the smart adviser is mainly made of an expert system applying cascading principles (1st principal of Cobit 5) to unify objective and value creation definition among different stakeholders. By updating the semantic engine, we also unified guidelines in hierarchical way.

Every stakeholder will find deeper guidelines depending on what he is asking for: a top manager will be satisfied by Cobit suggestions but CIO or a business manager will need more details that he will find properly in ISO27001, ITIL or even CMMI. So, by deploying the cascading principle and the single unified principle, the IT GRC smart adviser deep dive enterprise needs to suggest more operational recommendations.

The business driven architecture has a chatbot to ask stakeholder questions about eventual risks, resources management, value creation…etc) according to users' answers the smart advisor cascade his business request into enterprise goals. Then, depending on the questioned profile, the smart advisor asks for more details to match enterprise goals with IT goals.

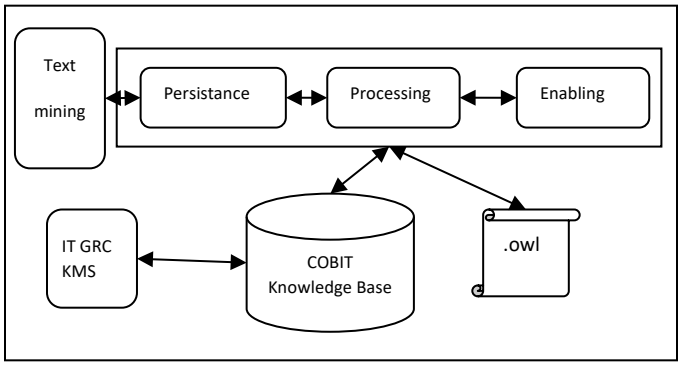


Figure 3: Semantic engine update

Every company description and characteristics should be previously configured in the knowledge base facts to define IT Governance road map. The proposed solution also allows cascading IT decision according to COBIT 5, through the semantic engine as shown in figure 2. After that, IT related goals are matched to enablers in a new abstraction level compared to [13]. The proposed semantic engine is now a five layers' system:

- Text mining layer annotate stakeholder answers.
- Persistence component storages stakeholder needs in the knowledge base
- Processing component compares stakeholder's request to Cobit 5 suggestions from knowledge base.
- Enabling component chooses the best facilitator (s) to implement.
- IT GRC KMS is a knowledge management system to feed Cobit Knowledge base with matched processes and guidelines.

As for Cobit Knowledge base, it is the set of rules and facts related to each COBIT process and domain one by one with: CMMI, ISO27001 and ITIL. In this level we are not updating the ontology since the system will interpret user questions through COBIT cascading.

6. Discussion

In previous work [1], The smart adviser contains a matchmaker agent between the semantic engine described above and IT GRC business driven configuration.

The matchmaker verifies first the company static parameters to initiate the IT GRC road map. It also enables the stakeholder to ask questions and have answers from the semantic engine in Chatbot way. Meanwhile, matchmaking research and priority algorithm detailed in [22] match every request with best guidelines according to a corresponding rate. The indexing algorithm does the following:

- Text indexing according to the words that compose it.
- Similarity test with inverted index.
- Returning similar texts with rank for every user query.
- Knowledge extraction from Cobit Knowledge Base.

Table 3: Proposed solution and state of the art comparison

Solution	Simple user interface	Risks management	- Legal and regulatory aspect	Human adviser	Deployment times	information security	IT GRC	Workflow and dashboards	Data management
ACL	X								
Enablon		X	X						
IBM		X							
LogicManager				X	X				
MetricStream	X								
Nasdaq						X			
Riskonnect	X							X	
RSA Archer		X	X				X		
Rsam		X							X
SAI Global							X		
SAP								X	X
ServiceNow		X	X				X		
Thomson Reuters		X	X						
NAVEX Global			X						X
Proposed solution	X	X	X	X	X	X	X	X	X

The smart advisor update is done in this step: Once similarity ranks are fixed, an inference engine questioned the rules base according to a maximization algorithm. The rules base communicates simultaneously with the fact base and the inference engine to give the convenient answer according to COBIT first and later to ITIL, CMMI and ISO27001 (see figure 3). The answer is returned to the matchmaker agent who sends it to the active adviser. One last thing is that two or three advisors could be handled for the same request successively for IT Governance, risk management or project management as well.

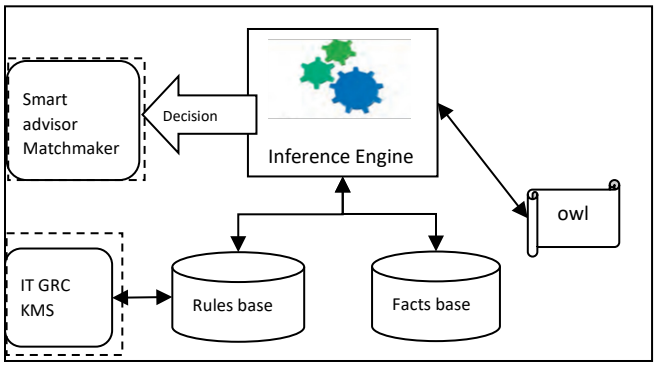


Figure 4: Cobit Knowledge Base architecture

In table 2 we compare the proposed IT GRC smart adviser to solutions presented in table 1:

To summarize, the update of the smart IT Governance adviser understand stakeholders' needs by annotating their questions and finding answers in Cobit 5 cascading model, via a text mining layer. This component is mainly managed by the processing entity connected with the Cobit knowledge base to confront them at this last base of rules to elaborate a roadmap with many granularity levels (IT Governance, risk management, services managements and IT project management. Cobit knowledge base is mainly made of facts bases and rules base linked to IT GRC Knowledge management system (that we will detail in a future work). Rules base is regularly actualized through IT GRC new repositories and expectations. As for Facts base, it is actualized by the organization information system parameters. The inference engine feeds the smart advisor with the appropriate decisions after semantic analysis and primitive matching. In addition, for every level facilitator implementation is also given, with a balanced squared card dimension, the result road map enables a prioritization of processes implementation.

As for the processing and analysis steps compared to [1], we:

- update the semantic engine,
- implement the COBIT knowledge base,
- link the KB with the matchmaker agent,
- implement the chatbot
- annotate answers
- test the adviser
- compare its results with a human expert's decisions
- evaluate the adviser performance

7. Implementation

To validate experimentally the proposed architecture, we implement and improve different versions of the platform. Let's give an overview of the IT Governance platform evolutions:

The first and second versions were java simulators, the third, fourth and fifth version were web application where, in addition to the proposed functionalities, communicate with the global EAS-IT GRC platform[23]. In the first version, we simply validated the communication of different intelligent agents with predefined business goals. In the second version[24], we added to the first version kernel, the knowledge base where we stored the basic information and the different mediation rules and we implemented an agent launch interface to simulate exchanges. We have carried out tests to validate results. In the third version, we have implemented almost all of the features offered with interfaces of potential users;

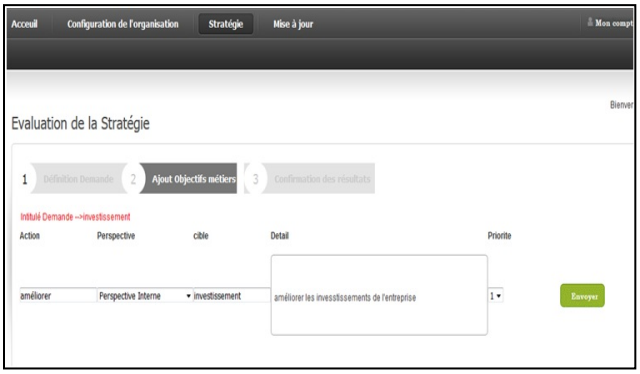
- IT Governance Ontology of information systems governance that we deployed in the mediation layer for the interpretation of offers and requests;
- Global comparison algorithm to improve the expert mediation system;

- Update repository with version 5 of Cobit;
- Request logging and learning to develop the proposed offers.

As for the fourth version presented in [1], we simplify user web interface and encapsulate IS Layer in a configuration agent for better device coordination in multiple access case. We also optimize the semantic engine model, to extract IT Governance knowledge.

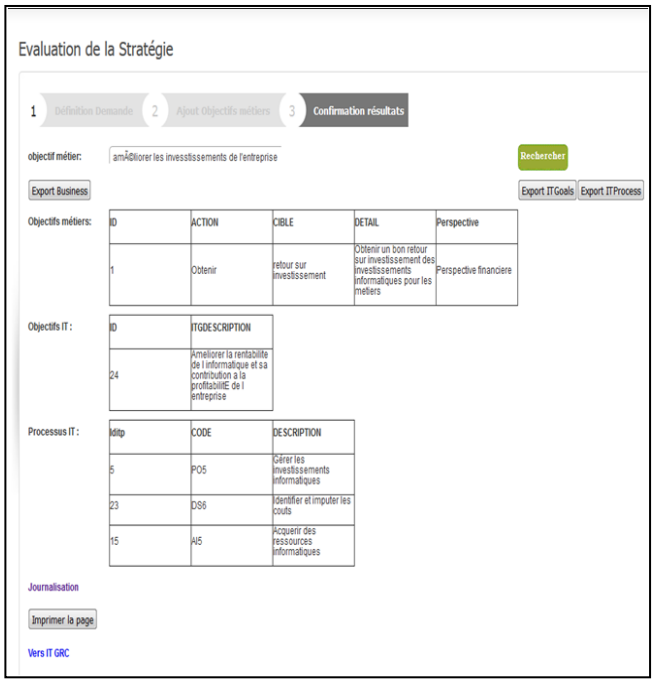
In the fifth version, presented in this paper, we add two main contributions namely the semantic engine knowledge management updater and the knowledge base detailed architecture.

Here is the fifth version IT Governance user query interface:



After a secure authentication and configuration questions interface, figure 5 shows user requests about the IT strategy or business strategy related to digital issues. An order of priority for every request is defined by the user (value from 1 to 5).

Figure 6 shows results as recommended business objectives, IT objectives and IT processes to focus on. A report is generated with other details such as facilitators, controls and metrics, maturity models, key activities and the responsibility matrix.

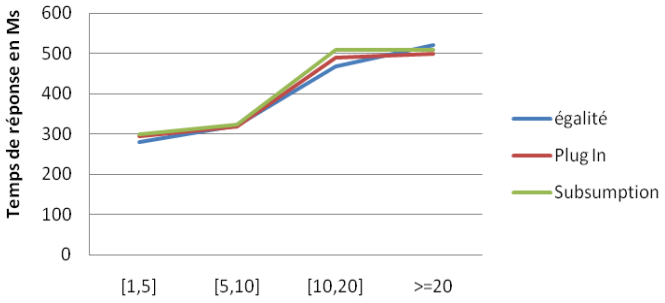


The application presents the detail of the request and the corresponding offers before allowing the user to edit the report, this primary summary allows him to decide whether it is necessary to start detailed ITG processing or obtained results are sufficient.

Many functional tests have been done. The corresponding framework of every objective is calculated via the priority function of the semantic engine. It also depends on the attribute and search criteria in the knowledge base. Let's compare as example the results of the platform with experts' decision about the request: "obtain relevant information and make strategic decisions in relation to risks" (experts proposed framework is ISO27001).

The results of the platform are presented in table 3. The request corresponds to several business objectives (OB). Each OB is detailed on one or more IT objectives and each IT objective is explained with several IT processes. According to results in table 3, we can deduce that the adviser chooses ISO27001 like the expert. IT objectives are recommended with priority gap due to semantic engine priority calculation; an IT Objective could be proposed by the expert but appears in the solution with a lower priority value.

As for evaluating the adviser response time to enhance its performance to give business managers rapid recommendations. It was made on a PC with Windows 10 operating system Intel Core i7 processor, 2.66 GHz and 16 GB of memory. Three comparison matchmaking modes were tested, namely equality, plug-in and Subsumption. We varied the number of requests according to four intervals [1, 5] [5, 10] [10.20] and> 20. The number of results per request were also taken into consideration.



Obtained results are presented in figure 7. The adviser is efficient since obtained results are in milliseconds. Governance advisor [1] is then presented trough a smart semantic model updater with IT Governance knowledge system architecture to deal with frameworks matching problem. These two contributions were added to implement other IT Governance frameworks for a 360° IT Strategy view: strategic alignment, risk and resources management, performance management and compliance. We chose COBIT 5 as IT Governance generic framework and semantic Web, and knowledge management as technical background to implement this solution.

As for limitations, the actual version of IT GRC adviser does not deep dive compliance recommendations. Compliance processes design should be done according to regional regulations. The same remark is available for specific risks namely financial and health risks analysis. As for the state of progress and perspectives of this work, we are working on:

- IT Governance Knowledge management system
- Functional tests and proof of concepts.

These perspectives enable the IT GRC advisor to be more performing so as to achieve industrialization level. The difference is also minimal between the 3 matchmaking mods which means that it is indifferent to requests complexity. The correlation of response time with requests number is linear per part, which means the algorithms complexity is reasonable with the increase of iterations.

8. Conclusion and perspectives

In this article, we raise IT Governance frameworks matching problem. After an introduction, we present IT Governance frameworks and their correspondences with COBIT5 as strategic level repository. We also listed existing IT Governance platforms with their strong and week points analysis. An update of the IT Governance advisor [1] is then presented trough a smart semantic model updater with IT Governance knowledge system architecture to deal with frameworks matching problem. These two contributions were added to implement other IT Governance frameworks for a 360° IT Strategy view: strategic alignment, risk and resources management, performance management and compliance. We chose COBIT 5 as IT Governance generic framework and semantic Web, and knowledge management as technical background to implement this solution.

As for limitations, the actual version of IT GRC adviser does not deep dive compliance recommendations. Compliance processes design should be done according to regional regulations. The same remark is available for specific risks namely financial and health risks analysis. As for the state of progress and perspectives of this work, we are working on:

- IT Governance Knowledge management system
- Functional tests and proof of concepts.

These perspectives enable the IT GRC advisor to be more performing so as to achieve industrialization level.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. Chergui, A. Chakir, H. Medromi, "Smart IT governance, risk and compliance semantic model: Business Driven architecture," in Proceedings of the 3rd World Conference on Smart Trends in Systems, Security and Sustainability, WorldS4 2019, 2019, doi:10.1109/WorldS4.2019.8903997.

[2] I. Alhassan, D. Sammon, M. Daly, "Critical success factors for data governance: a theory building approach," Information Systems Management, **36**(2), 98–110, 2019.

[3] C. Meriyem, S. Adil, M. Hicham, "IT Governance ontology building process: example of developing audit ontology," International Journal of Computer Techniques, **2**(1), 134–141, 2015.

[4] T. Huygh, S. De Haes, A. Joshi, W. Van Grembergen, "Answering key global IT management concerns through IT governance and management processes: A COBIT 5 View," in Proceedings of the 51st Hawaii International Conference on System Sciences, 2018.

[5] A. Joshi, L. Bollen, H. Hassink, S. De Haes, W. Van Grembergen, "Explaining IT governance disclosure through the constructs of IT governance maturity and IT strategic role," Information & Management, **55**(3), 368–380, 2018.

[6] H. Gunawan, "Strategic Management for IT Services Using the Information Technology Infrastructure Library (ITIL) Framework," in 2019 International Conference on Information Management and Technology (ICIMTech), IEEE: 362–366, 2019.

[7] A.S.O. Parra, L.E.S. Crespo, E. Alvarez, M. Huerta, E.F.M. Paton, "Methodology for Dynamic Analysis and Risk Management on ISO27001," IEEE Latin America Transactions, **14**(6), 2897–2911, 2016, doi:10.1109/TLA.2016.7555273.

[8] J.-H. Shin, J.-S. Choi, I.-H. Kim, "Development of IT BSC-based Assessment System to Measure BIM Performance for Architectural Design Firms," Journal of the Architectural Institute of Korea Planning & Design, **32**(1), 3–12, 2016.

[9] U. Durak, A. D'Ambrogio, T. Gerlach, "Applying IEEE Recommended Practice for Distributed Simulation Engineering and Execution Process for Modeling and Simulation Based Airborne Systems Engineering," in AIAA Scitech 2020 Forum, 896, 2020.

[10] T. Clauss, "Measuring business model innovation: conceptualization, scale development, and proof of performance," R&D Management, **47**(3), 385–403, 2017.

[11] N. Slack, A. Brandon-Jones, Operations and process management: principles and practice for strategic impact, Pearson UK, 2018.

[12] M. Ramakrishnan, A. Shrestha, A. Cater-Steel, J. Soar, "IT service management knowledge ecosystem–literature review and a conceptual model," 2018.

[13] R. Almeida, R. Lourinho, M.M. da Silva, R. Pereira, "A model for assessing COBIT 5 and ISO 27001 simultaneously," in 2018 IEEE 20th Conference on Business Informatics (CBI), IEEE: 60–69, 2018.

[14] H. de J.G. Antunes, P.G. Pinheiro, "Linking knowledge management, organizational learning and memory," Journal of Innovation & Knowledge, **5**(2), 140–149, 2020.

[15] S.-A.K. Kaya, N. Messaoudi, M. Bouhadida, A. Bennour, "Le Knowledge Management: Socle de Construction de Memoire de Projet," Réformes Economiques et Intégration En Economie Mondiale, **424**(3911), 1–21, 2016.

[16] M.S. Saleh, O. Ismail, A. Kamel, H. Hassan, "From CommonKADS to SOA Environment: An Adaptation Model," Arabian Journal for Science and Engineering, **43**(12), 7605–7619, 2018.

[17] P.G.C. Hector, J.-L. Ermine, V. Ribiere, A. Bennet, "A knowledge-based development model for primate cities of the developing world," International Journal of Knowledge-Based Development, **9**(4), 386–419, 2018.

[18] V. Bombardier, P. Charpentier, B. Almecija, "Design of a log sorting X-ray system based on domain ontologies," IFAC-PapersOnLine, **49**(12), 479–484, 2016.

[19] J.J. Kirton, B. Warren, "G20 Governance of Digitalization," International Organizations Research Journal, **13**(2), 16–41, 2018.

[20] R. Murphy, "The Forrester Wave™: Governance, Risk, And Compliance Platforms, Q1 2016," Cambridge, USA, 2016.

[21] M. Chergui, H. Medromi, A. Sayouti, "Intelligent audit of information systems by inter- organizational workflow," 2014 International Conference on Next Generation Networks and Services (NGNS), 1–6, 2014, doi:10.1109/NGNS.2014.7090888.

[22] M. Chergui, A. Chakir, H. Medromi, M. Radoui, A new approach for modeling strategic IT governance workflow, 2017, doi:10.1007/978-981-10-1627-1_22.

[23] A. Chakir, M. Chergui, H. Medromi, A. Sayouti, "An approach to select effectively the best framework IT according to the axes of the governance IT, to handle and to set up an objective IT," 2015 Third World Conference on Complex Systems (WCCS), 1–8, 2015, doi:10.1109/ICoCS.2015.7483311.

[24] A. Chakir, M. Chergui, J.F. Andry, "A smart updater it governance platform based on artificial intelligence," Risk, **8**, 9, 2020.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 256-260 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Multi Operated Virtual Power Plant in Smart Grid

Yevhen Fediv, Olha Sivakova, Mykhailo Korchak*

Lviv Polytechnic National University, Department of Electric Power Engineering and Control Systems, Institute of Power Engineering and Control Systems, Lviv, 79012, Ukraine

ARTICLE INFO

Article history:
Received: 23 August, 2020
Accepted: 14 October, 2020
Online: 10 November, 2020

Keywords:
Virtual power plant
Smart grid
AC voltage controller
GTO thyristor
Reactive power
IoT

ABSTRACT

In order to balance the power in intelligent distribution networks (Smart Grid or Microgrid), it is proposed to organize a «multi operated virtual power plant». The resources of active and reactive power for which can be obtained using an AC voltage controller with a phase-angle control for regulation of operating modes of the ohmic load of consumers, for example, distributed systems of electric space heating or electric water heating, etc. A method is proposed and the results of the analysis of the phase-angle control modes by gate turn off (GTO) thyristors of the AC voltage regulator, which provides the generation of virtual reactive power by consumers of active power, are presented. The process equipment of such a virtual power plant is fully suitable both for the dynamic production of a virtual resource of active power to balance, for example, the power of dynamic distributed renewable energy sources (virtual power plant (VPP) mode), and to regulate reactive power to ensure adequate voltage levels and increase stock stability of operation of electric load units (virtual reactive power plant (VRPP) mode). References 24, figures 8.

1. Introduction

This article is an extension and continuation of the work presented in the materials of the IEEE conference [1], which shows the possibility of power electronics to obtain a resource of reactive power from consumers-regulators with an active nature of consumption. Improving the efficiency of power supply, development in the direction of intellectualization according to the concept of Smart Grid is based on modern advances in power electronics [2-4].

In addition to balancing active power with frequency stabilization, an important operational task in Smart Grid systems is to ensure proper voltage levels in power grid units, increase the stability margin of power generation and consumption systems, and optimize operating costs in the transportation and distribution of electricity.

The basis for solving these problems is to ensure the balance of reactive power in the system, the availability of appropriate resources and the ability to quickly involve them at different levels of transmission, conversion and distribution of electricity [5, 6]. In the Smart Grid concept, the problem of reactive power compensation is highlighted as key. Different methods of reactive power compensation are considered in [7, 8].

*Corresponding author: Mykhailo Korchak., misha.korkirov@gmail.com +380981712257

The universal and most efficient centralized source of reactive power at present is the Static Compensator (Statcom) device [9], which is primarily intended for flexible power transmission systems or powerful industrial power supply systems.

Problems of using Statcom in low-voltage networks and their interaction with distributed energy sources are considered in articles [10,11], where special attention is paid to the use of photovoltaic inverters as reactive power sources.

In [12-14], the use of static compensator D-STATCOM for unbalanced radial power distribution systems is investigated in order to ensure voltage stability within acceptable limits. This review of known works addresses issues related to the optimal choice of power, control methods, installation location, etc. of reactive power sources as physical objects installed by the power supplier. However, currently in the Smart Grid an important role is given to virtual energy sources

In the classical sense, a VPP is a distributed power plant that combines the resources of active power generation from distributed energy resources (DER) as a reserve at the disposal of transmission and distribution operators [15]. VPPs in Microgrid are able to increase resource efficiency and maintain the necessary balance of generation and consumption in the power system. Distributed renewables and drives in VPPs are not virtual but real

resources. In fact, virtual (feigned) generation of electricity will be equivalent to refusing or limiting its consumption. For example, by integrating water heating or electric heating systems into a grid and remotely controlling them with electricity consumption, you can align the active power consumption schedule at a particular distribution network node without significant consequences for household or industrial consumers themselves.

Another possibility is the participation of customers in virtual power plants to provide reactive energy. [16-20] discusses the provision of reactive power services from DER using a market mechanism and a plan to convert Microgrid into VPP, which is beneficial for consumers, prosumers and utility operators.

Involving distributed sources to solve the problem of reactive power balancing, by analogy with a virtual power plant, can be combined into virtual reactive power sources. In [21, 22] the problem of reactive power is not allocated separately but included in the list of tasks for the organization of VPP. [23] presents the concept of a dynamic VRPP as an ancillary service to reasonably combine decentralized power distribution sources into a coordinated pool to provide reactive power for the needs of power transmission system operators. Common in the known concepts is the involvement in the VRPP real distributed sources of reactive power.

2. Purpose and task of the research

The aim of the work is to study the problem of balancing active and reactive power in distribution networks with the involvement of consumers' resources in the form of integrated virtual power plant (VPP-VRPP), where reactive power resource is obtained by regulating the receivers used for VPP**.** This is especially true for facility power systems that are not encouraged to install their own sources to compensate for reactive power.

The ranges of complex regulation by means of power electronics of consumers with active character of loading for reception of a virtual resource of active and reactive power by regulation according to needs of operators of power supply systems are substantiated in the work.

3. Finding a solution to the problem

Figure 1 shows the equivalent circuit diagram of one phase VPP-VRPP, which shows the AC voltage regulator, which includes a source of sinusoidal voltage with internal inductance, a block of power dual-operation (GTO) thyristors VS1 and VS2, a limiting diode suppression diode) VD, as well as resistive resistance R of water heater or electric heating.

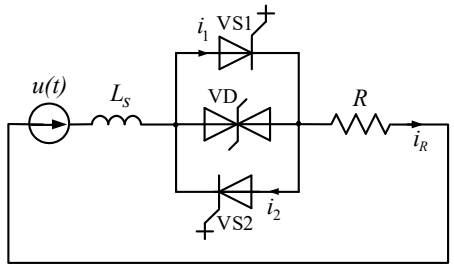


Figure 1: Equivalent power supply circuit of the heater

To protect against the effects of rigid switching of thyristors inductive overvoltage emissions in the interval between the closing of the thyristors, a two-way TVS diode VD is designed.

To obtain the energy characteristics in the steady state VPP-VRPP, we assume that the thyristors VS1, VS2 are ideal switches, not taking into account the diode VD, inductance $L_S = 0$, and voltage of the power supply system is sinusoidal as shown in (1)

$$u(t) = U_m \sin(\omega t + \varphi_u) \quad (1)$$

where U_m is power supply voltage amplitude; φ_u is the initial phase of the supply voltage; ω is angular velocity.

To obtain the power in the circuit of the resistor, controlled by key semiconductor elements, we have to analyze two characteristic modes.

3.1. Mode of consumption of virtual reactive power from the power supply network.

This mode is characteristic for traditional single-operation (semi-controlled) thyristors or SCR (Semiconductor Controlled Rectifier). For its realization thyristors VS1 and VS2 is being opened by the controlling impulses with a delay of an angle α relative to the natural angle of their entry into operation $\alpha_0 = 0 \pm \pi k$, where $k = 0, 1, 2\ldots$, and closing them at the natural time of closing $\beta_0 = \pi + \pi k$ (Fig. 2).

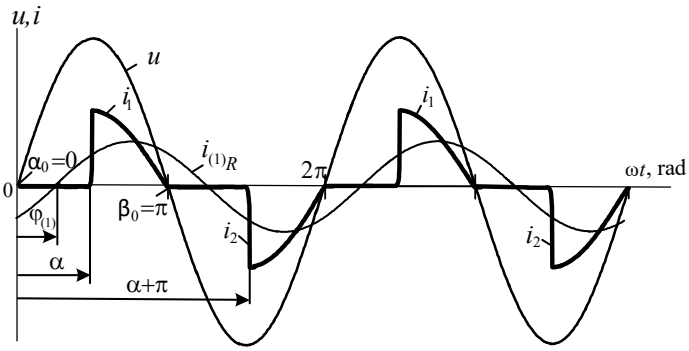


Figure 2: VRPP current and voltage time diagrams in the mode of reactive power consumption

As a result, we obtained the shift of the main harmonic of the source current $i_{(1)R}$ towards the delay relative to the voltage of the grid. It is equivalent to the consumption of virtual reactive power by electrical circuit, in which reactive elements are not presented.

It should be noted that the value of the reference times of the thyristor control angles (natural value of the angles of opening and closing of the thyristors) are characteristic of the purely active nature of the circuits with key elements.

To evaluate the energy effect of the above method of regulating the operation of the electric receiver, we decompose the current i_R in the Fourier series. *n*-harmonic component of the relative current is descripted in expression (2)

$$i_n(\omega t) = I_{nm*} \cdot \sin(n\omega t + \varphi_n) \quad (2)$$

can be obtained as (3)

$$I_{nm*} = \frac{I_{nm}}{I_b} = \sqrt{A_{n*}^2 + B_{n*}^2} \tag{3}$$

where I_b is the current amplitude in the mode of full conductivity of thyristors according to (4)

$$I_b = I_{m,R} = U_m / R \tag{4}$$

A_{n*} and B_{n*} are the coefficients for the decomposing to the Fourier series is written using (5) and (6):

$$\begin{aligned} A_{n*} &= \frac{1}{\pi}\int_{\alpha}^{\pi} i_R(\omega t)\cos n\omega t\, d(\omega t) + \frac{1}{\pi}\int_{\pi+\alpha}^{2\pi} i_R(\omega t)\cos n\omega t\, d(\omega t) \\ &= \frac{1}{2\pi}\left\{\frac{\cos(n\pi)+\cos\left[(1-n)\alpha\right]-\cos(2n\pi)+\cos\left[(1-n)(\pi+\alpha)\right]}{1-n}\right. \\ &\left.+\frac{\cos(n\pi)+\cos\left[(1+n)\alpha\right]-\cos(2n\pi)+\cos\left[(1+n)(\pi+\alpha)\right]}{1+n}\right\} \end{aligned} \tag{5}$$

$$\begin{aligned} B_{n*} &= \frac{1}{\pi}\int_{\alpha}^{\pi} i_R(\omega t)\sin n\omega t\, d(\omega t) + \\ &\frac{1}{\pi}\int_{\pi+\alpha}^{2\pi} i_R(\omega t)\sin n\omega t\, d(\omega t) \\ &= \frac{1}{2\pi}\left\{\frac{-\sin\left[(1-n)\alpha\right]-\sin\left[(1-n)(\pi+\alpha)\right]}{1-n}\right. \\ &\left.+\frac{\sin\left[(1+n)\alpha\right]+\sin\left[(1+n)(\pi+\alpha)\right]}{1+n}\right\} \end{aligned} \tag{6}$$

For the 1st harmonic ($n=1$) obtained equations (7) and (8):

$$A_{1*} = \frac{1}{2\pi}\left[\cos(2\alpha)-1\right] \tag{7}$$

$$B_{1*} = \frac{1}{2\pi}\left[\sin(2\alpha)+2(\pi-\alpha)\right] \tag{8}$$

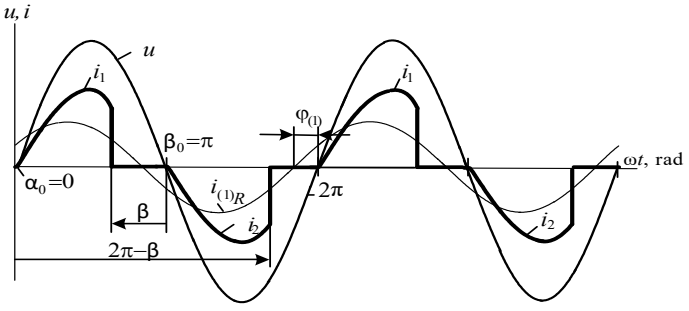


Figure 3: VRPP current and voltage time diagrams in the mode of reactive power generating

3.2. Mode of virtual reactive power generation in the power supply network.

This mode is dual to the previous one. For its realization thyristors VS1 and VS2 is being opened in the natural moments of their entry into operation $\alpha_0 = 0 + \pi k$, where $k = 0,1,2\ldots$, and closing with an advance on the angle β relative to the natural angle β_0 of closure (Fig. 3). As a result, we obtained the shift of the main harmonic of the source current $i_{(1)R}$ in the direction of advance relative to the grid voltage, which is equivalent to generation of reactive power by resistance into the power supply system.

The formulas obtained by the authors for the decomposition coefficients in the Fourier series of the resistor current take the form (9) and (10).

$$\begin{aligned} A_{n*} &= \frac{1}{\pi}\int_{0}^{\pi-\beta} i_R(\omega t)\cos n\omega t\, d(\omega t) \\ &+\frac{1}{\pi}\int_{\pi}^{2\pi-\beta} i_R(\omega t)\cos n\omega t\, d(\omega t) \\ &= \frac{1}{2\pi}\left\{\frac{1-\cos\left[(1-n)(\pi-\beta)\right]-\cos\left[(1-n)(2-\beta)\right]-\cos(n\pi)}{1-n}\right. \\ &\left.+\frac{1-\cos\left[(1+n)(\pi-\beta)\right]-\cos\left[(1+n)(2\pi-\beta)\right]-\cos(n\pi)}{1+n}\right\} \end{aligned} \tag{9}$$

$$\begin{aligned} B_{n*} &= \frac{1}{\pi}\int_{0}^{\pi-\beta} i_R(\omega t)\sin n\omega t\, d(\omega t) \\ &+\frac{1}{\pi}\int_{\pi}^{2\pi-\beta} i_R(\omega t)\sin n\omega t\, d(\omega t) \\ &= \frac{1}{2\pi}\left\{\frac{\sin\left[(1-n)(\pi-\beta)\right]+\sin\left[(1-n)(2\pi-\beta)\right]}{1-n}\right. \\ &\left.-\frac{\sin\left[(1+n)(\pi-\beta)\right]+\sin\left[(1+n)(2\pi+\beta)\right]}{1+n}\right\} \end{aligned} \tag{10}$$

For the 1st harmonic ($n=1$) we obtained equations (11) and (12):

$$A_{1*} = \frac{1}{2\pi}\left[1-\cos(2\beta)\right] \tag{11}$$

$$B_{1*} = \frac{1}{2\pi}\left[2(\pi-\beta)+\sin(2\beta)\right] \tag{12}$$

4. Generalization of the results of the analysis of control regimes

In both of the above modes of thyristor control with a change in the value of the control angles and there is a simultaneous change in both active and reactive power, the value of which can be calculated by expressions (13) and (14).

$$P_{1*} = \sqrt{A_{1*}^2 + B_{1*}^2}\cos(\varphi_{1u} - \varphi_{1i}) = I_{1m*}\cos\varphi_1 \tag{13}$$

$$Q_{1*} = I_{m1}\sin\varphi_1 \tag{14}$$

where, P_{1*}, Q_{1*}

$\sqrt{A_{1*}^2 + B_{1*}^2} = I_{1m*}$ – the amplitude of the fundamental harmonic of the current, $\varphi_1 = \arctan(B_1/A_1)$ – the initial phase of the fundamental harmonic of the current, and $\varphi_{1u} = 0$.

Figure 4 shows the results of the calculation of VPP-VRPP control characteristics by reactive power. The maximum generation of reactive power is $|Q_{1*}| \approx 0,32$ p.u. different signs are reached at the control angles of the boards $\alpha = \beta = \pi/2$. In case of regulating the VPP-VRPP by changing the opening angles $\alpha = \text{var}$ of thyristors and constant closing angles $\beta = \text{const} = 0$ from the mains consumes excess reactive bias power ($Q_1 > 0$), and by constant angles of thyristor opening ($\alpha = \text{const} = 0$) and adjusting the angle $\beta = \text{var}$ change we get the mode of generating reactive bias power to the mains ($Q_1 < 0$).

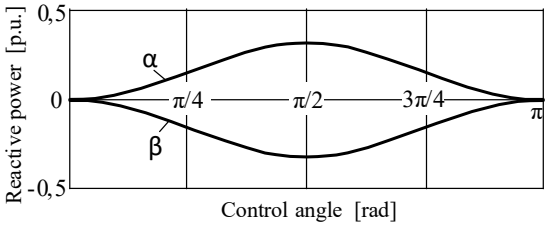


Figure 4: VPP-VRPP control characteristics by reactive power

The obtained characteristics are symmetric with respect to the angle $\alpha = \beta = \pi/2$, but in the mode of reactive power control the preference should be given to the range of angles $\alpha = \beta = 0 \ldots \pi/2$ of regulation of the boards, because in the range of angles $\alpha = \beta = \pi/2 \ldots \pi$ we get a significant harmonic current distortion (Fig. 5)

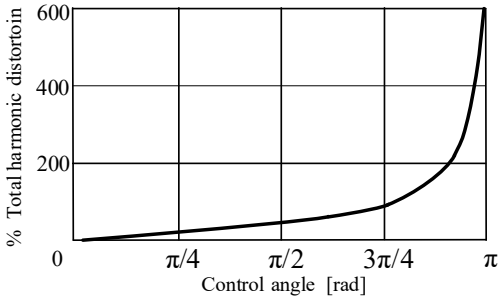


Figure 5: Coefficient of harmonic current distortion

Calculated by expression (15)

$$THD_I(\%) = \frac{I_{dis}}{I_1}100 = \frac{\sqrt{\sum_{n\neq 1} I_n^2}}{I_1}100, \qquad (15)$$

The value of the harmonic distortion coefficients of the current consumed from the mains for the above-recommended range of adjustment of the control angles is $THD_I = (0...36,6)\,\%$. For the mode of full conductivity of thyristors, there is no distortion, and for the modes $\alpha = \beta = \pi/2$ the harmonic distortion of VPP-VRPP current is $THD_I \approx 36,6\,\%$.

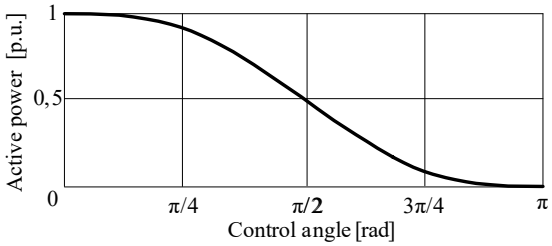


Figure 6: VPP-VRPP control characteristic by active power

The change in the reactive power of the VPP-VRPP in the process of regulation is accompanied by a simultaneous change in the active power consumed by the electric receiver (Fig. 6).

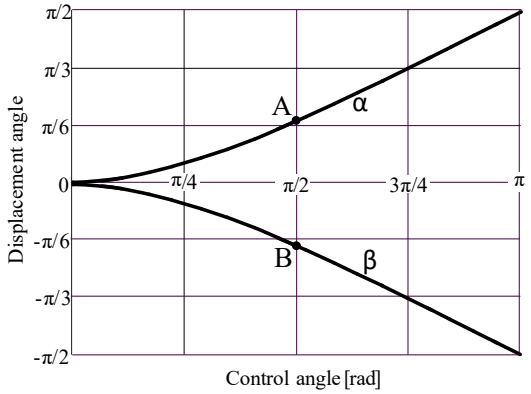


Figure 7: Displacement angle for two ways of regulation (modes) VRPP

As you can see from Fig. 4 and fig. 6, in the process of regulation, with increasing value of reactive power, the value of active power decreases, which generally contributes to the tasks to be solved by virtual power plants, the control effect on the active $dP/d\alpha < 0$, $dP/d\beta < 0$ and reactive $dQ/d\alpha > 0$, $dQ/d\beta > 0$ power of the resistive electrical receiver is different in signs.

We obtained modes in which the generation of reactive power to balance it in the load center is accompanied by a simultaneous decrease in active power consumption, which contributes to both the normalization of the voltage level (VRPP task) and active power balancing (VPP task).

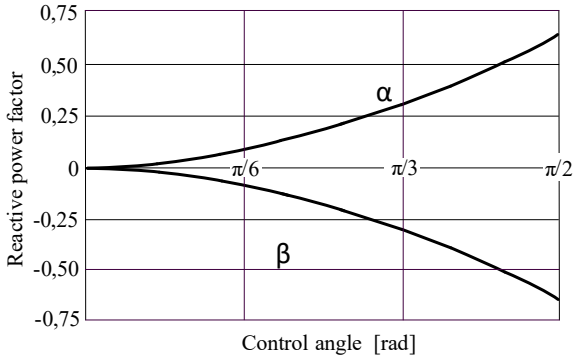


Figure 8: Reactive displacement power factor

In the mode of maximum reactive power $\alpha = \beta = \pi/2$, the value of the displacement angle of the main harmonic of the current is $\varphi_1 \approx \pm 32,5^\circ$ (points A and B on the corresponding characteristics of Fig. 7), and the coefficient of reactive displacement power (Fig. 8) $tg\varphi_1 = \pm 0,64$.

5. Conclusions

The article considers a comprehensive approach to solving the problems of active and reactive power balances in Microgrid networks, using hidden (virtual) resources, which are organized in the form of a virtual active-reactive power plant (VPP-VRPP). Using distributed resources of consumers and their active role in the operation of Microgrid.

Resources of active and reactive power are received from one source - ohmic receivers of consumption with involvement of means of power electronics as regulating elements.

By feeding the ohmic load from voltage regulators with GTO thyristors, the distribution network operator, if necessary, reduces the consumption of active power and adjusts the value and sign (consumption from grid or grid generation) of reactive power.

The range of change of control angles of thyristors of a voltage regulator is chosen for reasons of reduction of negative influence of nonlinear electric loading on quality of the electric power in Microgrid.

The use of integrated multi-controlled VPP-VRPP can be recommended for damping transients and ensuring the stability of Microgrid operation in modes with frequency and voltage value close to critical, to prevent avalanche-like changes.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors thank both the authorities of the Lviv Polytechnic National University especially leadership of the Institute of Power Engineering and Control Systems for their support during the preparation of this paper. The authors also thank the anonymous reviewers, whose comments significantly improved the content of the paper.

References

[1] Y. Fediv, O. Sivakova, M. Korchak, "Model of Virtual Source of Reactive Power for Smart Electrical Supply Systems," in 2019 IEEE 20th International Conference on Computational Problems of Electrical Engineering (CPEE), 1-4, 2019, doi:10.1109/CPEE47179.2019.8949159.

[2] G. Benysek, M.P. Kazmierkowski, J.Popczyk, R. Strzelecki, "Power electronic systems as a crucial part of Smart Grid infrastructure – a survey," Bulletin of the polish academy of sciences technical sciences, **59(**4), 2011, doi:10.2478/v10175-011-0058-2.

[3] B.K. Bose, "Power Electronics in Renewable Energy Systems and Smart Grid: Technology and Applications," Wiley-IEEE Press, 2019.

[4] B. K. Bose, "Power Electronics in Smart Grid and Renewable Energy Systems," in Proceedings of the IEEE, **105**(11), 2007-2010, 2017, doi:10.1109/JPROC.2017.2752538.

[5] J. Ali, S. Massucco, G. Petretto, "Reactive power provision to TSO/DSO by aggregators and conventional generators," 2017 IEEE International Conference on Smart Grid Communications (SmartGridComm), 486-491, 2017, doi:10.1109/SmartGridComm.2017.8340740.

[6] K. Wang, M.L. Crow, "Electricity Transmission, Distribution and Storage Systems," Woodhead Publishing Series in Energy, 2013.

[7] J. Ali, S. Massucco, F. Silvestro, "STATCOM application for voltage profiling of a distribution grid with high penetration of distributed energy resources," Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MEDPOWER 2018), 1-7, 2018, doi:10.1049/cp.2018.1883.

[8] J. Ali, S. Massucco, F. Silvestro, "Aggregation Strategy for Reactive Power Compensation Techniques-Validation," Energies **2019**, 1-13, 2019, doi:10.3390/en12112047.

[9] Q. Wang, B. Wang, W. Xu, J. Xu, "Research on STATCOM for reactive power flow control and voltage stability in microgrid," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2474-2479, 2018, doi:10.1109/ICIEA.2018.8398126.

[10] S. Hasan, R. Luthander, J. de Santiago, "Reactive Power Control for LV Distribution Networks Voltage Management," in 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe *(*ISGT-Europe*)*, 1-6, 2018, doi:10.1109/ISGTEurope.2018.8571817.

[11] D. Sailaja, V.B. Kumar, "A Voltage Regulator for Power Quality Improvement in Low-Voltage Distribution Grids," Journal of Applied Science and Computations, **5**(7), 461-468, 2018, DOI:10.1109/TPEL.2017.2693239.

[12] A.R. Guptaa, A. Kumar, "Impact of D-STATCOM placement on improving the reactive loading capability of unbalanced radial distribution system," Procedia Technology 25, 676-683, 2016, doi: 10.1016/j.protcy.2016.08.170.

[13] W. Guoyou, Z. Xilin, "Study on hybrid reactive power compensation in low-voltage distribution system," International Conference on Energy and Environmental Protection (ICEEP 2016), 618-622, 2016, doi:10.2991/iceep-16.2016.106.

[14] L. E. Christian, L. M. Putranto, S.P. Hadi, "Design of Microgrid with Distribution Static Synchronous Compensator (D-STATCOM) for Regulating the Voltage Fluctuation," 2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE), 48-52, 2019, doi:0.1109/SEGE.2019.8859860.

[15] M.M. Othman, Y.G. Hegazy, A.Y. Abdelaziz, "A Review of Virtual power plant Definitions, Components, Framework and Optimization," International Electrical Engineering Journal (IEEJ), 2010-2024, **6**(9), 2015.

[16] M. Korchak, O. Sivakova, Y. Fediv, «Virtual source of reactive power in electricity supply systems of household consumers,» Energy Engineering and Control Systems, 9–14, **5**(1), 2019 doi:10.23939/jeecs2019.01.009.

[17] D. Pudjianto, P. Djapic, G. Strbac, B. Stojkovska, A. R. Ahmadi, I. Martinez, "Integration of distributed reactive power sources through Virtual Power Plant to provide voltage control to transmission network," 25thInternational Conference on Electricity Distribution (CIRED-2019), 1-5, (2055), 2019. http://dx.doi.org/10.34890/933.

[18] J. Ali, S. Massucco, F. Silvestro, A. Vinci, «Participation of Customers to Virtual Power Plants for Reactive Power Provision,» Conference: 2018 53rd International Universities Power Engineering Conference (UPEC), 1-6, 2018, doi:10.1109/UPEC.2018.8541978.

[19] K. L. Anaya, M. G. Pollitt, "Reactive Power Procurement: Lessons from Three Leading Countries," Cambridge Working Papers in Economics, 2018.

[20] J. Naughton, H. Wang, S. Riaz, M. Cantoni, P. Mancarella, "Optimization of Multi-Energy Virtual Power Plants for Providing Multiple Market and Local Network Services," Electric Power Systems Research, 189, 2020, doi:10.1016/j.epsr.2020.106775.

[21] Levent Yavuz, Ahmet Önen, S.M. Muyeen and Innocent Kamwa, "Transformation of Microgrid to Virtual Power Plant - a Comprehensive Review," IET Generation Transmission & Distribution, **13**(11), 1994-2005, 2018, doi:10.1049/iet-gtd.2018.5649.

[22] R. Bertram, A. Schnettler, "A control model of virtual power plant with reactive power supply for small signal system stability studies," 2017 IEEE ManchesterPowerTech, Manchester, 1-6, 2017, doi:10.1109/PTC.2017.7980838.

[23] F. Muuß, G.A.N. Hemdan, M. Kurrat, D. Unger, B. Engel, "Dynamic virtual reactive power plant in active distribution networks," IEEE Eindhoven PowerTech (PowerTech, 697-702, 2015, doi**:**10.1109/PTC.2015.7232356.

[24] IEEE Standard Definitions for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced, or Unbalanced Conditions, IEEE Std., 1459-2010.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 261-268 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

A Fuzzy Controller Based SAPF for Power Quality Enhancement of Distribution System Integrated with Wind Energy Source

Vikas Kumar Sharma*, Lata Gidwani

Electrical Engineering Department, Rajasthan Technical University, Kota, 323021, India

ARTICLE INFO

Article history:
Received: 08 September, 2020
Accepted: 01 November, 2020
Online: 10 November, 2020

Keywords:
Wind energy
Distribution System
Power quality
Fuzzy controller
APCT
SAPF

ABSTRACT

Wind energy is emerged rapidly as the most important and viable sustainable power source due to the mature technology and wide availability. The integration of wind energy sources and utilization of nonlinear loads having different characteristics results in several challenges in a distribution system. One of which is power quality issue .Hence, in this work, the impact of power quality issues in a distribution system integrated with the wind generation system is studied in the uses of a shunt active power filter (SAPF) and presence of non-linear load conditions. The realization of SAPF is carried out using active power component theory (APCT) for reference current extraction, proportional-integral (PI) controller, and fuzzy logic controller (FLC) for dc-link voltage regulation and a basic hysteresis band current controller method for the extraction of switching pulse for operational the inverter. To enhance power quality in the proposed system, a modified fuzzy logic-controlled (MFLC) is developed to improve the transient performance of shunt APF. The proposed model is developed in MATLAB and results are given so as to show the presentation of the performance of this compensation technique in terms of mitigation of system harmonics, reactive power compensation, and enhancing the power factor.

1. Introduction

This Electrical energy has become an integral part of human life and its use is increasing globally by leaps and bounds to improve the living standard. Global concern of utilization of green and clean energy sources for sustainable power systems and shortage of fossil fuel reserves create opportunities for renewable energy sources (RES) of generation. In this manner, the expanding dangers to the earth from the vitality created by the broad utilization of fossil fuels have left us with the thought of renewable energy sources (RES) as the main conceivable arrangement. The demand for sustainable power sources has highly increased due to a lack of fossil fuel, environmental impact and global warming problems of the world. The classification and trends of wind energy chnology has been discussed [1-4] in detail. Wind power is the most developing, promising, and viable sustainable power source that has been dissipated over the word, and utilization of this energy source has been expanding vary large scale. The integration of distributed generation (DG) is directed to increments in vitality productivity and a decrease in emission. The ever-increasing load demand and scarcity of electric power have attracted the attention of power engineers to evolve the new techniques for the effective utilization of electrical energy. Such evolutions have increased the use of power electronics devices and controllers in manifold over the last two decades for more efficient operation of the electrical equipment and to transfer more power through the power networks, but they degrade the power quality in the distribution network connected with wind generation source.

The PQ problem of grid-connected wind energy conversion has been discussed [5-8] by many authors in past years. In [9], the author discussed the power quality issues related to flicker, reactive power, current harmonic distortion, stability of wind generator during the faulty conditions. Further, there are a lot of disturbances that create PQ problems such as harmonics distortion, voltage interruption, transient disturbances, frequency variations, etc. Out of these disturbances the main polluting impact will be created by harmonic distortion but at the same time, it does not mean that power conditioning is only confined to harmonic filtering. In [10], the author introduced the power electronics devices to operate on the sinusoidal voltage but they inject harmonics in the power system networks and increase reactive power demand. The injected harmonics are responsible for the distortion of voltage and current waveshapes. These nonlinear

*Corresponding author: Vikas Kumar Sharma., vikasvs1985@gmail.com

loads are a big reason of harmonics, and extra neutral currents in the power system network. In [11], the author reviewed the different sources of customer- generated harmonics include AC/DC converters adjustable speed drives, switching power supplies, AC phase control circuit, cyclo converters, magnetizing current of transformer, arc furnaces, welding machines, traction systems, switch-mode power supplies (SMPS), computers and programmable logic controllers (PLC's), etc. Moreover, the single-phase loads like personal computers (PCs), fluorescent lights, printers, Xerox machine, television, etc. are also responsible for excessive neutral current distribution system.

The capability and use of SAPF has been investigated in [12-14]. A SAPF could be used as the potential solution to mitigate the utility current disturbances and enhancing power quality in power distribution system integrated with wind generation system. The shunt connected custom power devices are especially useful to maintain power factor, balancing of load, limitation of harmonics and voltage regulation at distribution voltage level, etc. The SAPF operated in the current control mode is capable of injecting negative harmonic current into the grid. Hence, in today's scenario, active power is popular and preferred over the passive filter. Shunt APF has the ability to mitigate the current harmonics from the system but series has sued for voltage harmonics. In [15], the author discussed for fast and precise estimation of reference current, predictive and adaptive controllers based on the neural network have been proposed. The quick estimation of compensating currents a lot of artificial and soft computing techniques was used. Because of the added advantage of ANN to compute the Fourier series coefficient the authors have preferred this over others improved hysteresis current controller technique is overcome wide variation of switching frequency, hence, reduce the switching frequency. In [16], the author implemented a lot of on the hysteresis based current controller in literature like as adaptive frequency domain control of PWM strategy, three-dimensional PWM technique, etc.

In [17], the author developed a standalone wind-diesel hybrid system, in which an active shunt filter is introduced to mitigate the harmonics of load current. In [18-19], they are presented the recent work on the grid integration of renewal energy with SAPF for current harmonic elimination. The dc-link voltage control is achieved by either PI or fuzzy logic technique in SRF based SAPF for proposed system. In [20], the author discussed the purpose of analysis the harmonic reduction and SAPF control technique using harmonic extractor-based method to renewable energy source integrated with grid. In [21**]**, the author proposed the SRF reference current method based shunt active power filter with PI controller to mitigate the current harmonics from the grid connected wind at PCC simultaneously non-linear load conditions. In [22], the author implemented the ICC controller is proposed for shunt active filter for enhance power quality in renewable source interfaced grid system. In [23], the author proposed the system performance parameter like output voltage, frequency and THD level at variable wind speed and nonlinear load conditions. The validity of proposed SAPF is analyzed with fuzzy logic based PWM technique. In [24], the author implemented a technique for reduce of harmonics in hybrid SAPF connected distribution network to enhance the power quality. The optimal controller was made by the improved optimization algorithm.

The recent literatures on the grid integration of wind energy with SAPF mainly focus on the study of current harmonics reduction by the grid current THD analysis. However, a smaller number of publications had been reported for proposing distribution grid level implementation of SAPF for power quality enhancement with the integration of wind energy source. Hence, there is a need to investigate the methods which may be implemented at the grid level to mitigate the power quality problem due to wind energy into the distribution grid. The performance of APC- PI theory with direct and indirect current control methods are compared and found that in both conditions, the neutral current compensation is almost same but the remarkable improvement is observed in power factor, reactive power compensation and load balancing with indirect control method. It is proved through proposed simulation that the indirect current control method performance is better than direct current control for the distribution grid integrated with wind generation system. So, proposed Indirect APC-MFLC methods are suitable to compensate harmonics and reactive power, neutral current during unbalanced loading conditions. The system is capable of maintaining unity power factor as well as sinusoidal source currents. Hence, the power quality of distribution system is improved.

The major contributions of this paper are integration and design of APCT based SAPF in proposed system for enhances the power quality problem. The proposed MFLC controller strategy is validated through APC theory by extensive simulation results for enhancement in overall power quality problem of the proposed system. This paper is classified in five sections, first is the introduction in Section 1. Section 2 mathematical modeling and detail of the proposed model and proposed SAPF control method in Section 3, simulation result and discussion presented in Section 4. The conclusion is shown in Section 5.

2. Proposed Model

A PMSG based wind energy connected through AC-DC-AC converter to distribution grid with a non-linear load connected to load bus shown in Figure 1. A shunt APF connected with the system for enhancement of power quality in non-linear loading conditions and PMSG integrated with the grid. The proposed system is not only capable of transfer the wind energy to the grid, but will also act as a shunt active power filter (SAPF) to mitigate the current harmonics and regulate reactive power feed by the nonlinear loads. Table 1 provides parameter of wind turbine and PMSG.

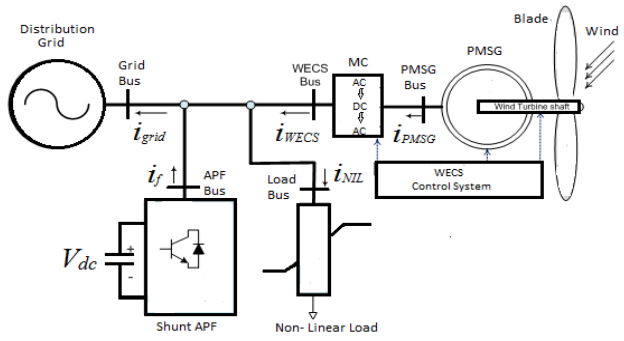


Figure 1: Block Diagram of Distribution Grid- Integrated Wind Generation System with Shunt APF and Non-linear Load.

A. Modeling of Wind Turbine System

The wind speed is sum of these components.

$$V_w(t) = V_B(t) + V_N(t) + V_G(t) + V_R(t) \quad (1)$$

where, V_w is the wind speed, all of them in "meter/ second" and time t is in "second Here this wind speed signal have four component, i.e. the basic speed of wind (V_B), a fluctuating wind (V_N) that describes a steady increase in wind speed, gust component (V_G) and a gradient wind component (V_R). The mechanical energy given wind turbine is determined as follow [3]

$$P_{\text{wind}} = \frac{1}{2} C_p(\lambda, \beta) \pi r^2 \rho_{\text{air}} V_w^2 \quad (2)$$

Power coefficient (C_p) is the ratio of mechanical power generated by the turbine to the power available in the wind. Which λ is the tip speed ratio, C_p is given as [3]

$$C_p = c_1\left(c_2 \frac{1}{\alpha} - c_3\beta - c_4\beta^x - c_5\right) \quad (3)$$

β is a function of the pitch angle of rotor blades, when β is equal to zero.

$$\frac{1}{\alpha} = \frac{1}{\lambda + 0.08\beta} - \frac{0.035}{1+\beta^2} \quad (4)$$

Therefore, the tip speed ratio can be given as [3]

$$\lambda = \omega_a r / V_w \quad (5)$$

The mechanical torque turbine is defined as follow [3]

$$T_{mech} = \frac{0.5 C_p \pi r^2 \rho_{air} V_w^2}{\omega_a} \quad (6)$$

The driven train is treated as one concentrated mass model to time efficiency and precision. It is defined as:

$$\frac{d}{dt}\omega_a = \frac{1}{j}(T_{mech} - T_{elect} - B_x\omega_a) \quad (7)$$

where, B_x is damping coefficient, ω_g is angular speed (rad/sec), T_{elect} is the electromechanical torque (Nm),T_{mech} is the aerodynamic torque and J equivalent rotational inertia of the generator (kg-m^2)

B. Modeling of PMSG

The direct axis and quadrature axis voltage equation of generator is define as

$$V_d \equiv -R_s i_d - L_d \frac{di_d}{dt} + \omega_e L_q i_q \quad (8)$$

$$V_q \equiv -R_s i_q - L_q \frac{di_q}{dt} - \omega_e L_d i_d + \omega_e \mu_g \quad (9)$$

ω_e is the rotor speed in rad/s of the generator, describe by

$$\omega_e = \omega_a P \quad (10)$$

The electromagnetic torque of PMSG is given by

$$T_{elect} = \frac{3}{4} P\left[\mu_g + \left(L_d - L_q\right) i_d\right] i_q \quad (11)$$

Then, the angular speed of generator is defined as [4]:

$$\frac{d}{dt}\omega_a = \frac{1}{j}\left(T_{mech} - T_{elect} - B_x\omega_a\right) \quad (12)$$

Table 1: parameter of wind turbine & generator

Parameter	Value
Generated Power, (P)	2 MW
Stator Resistance (R_s)	0.7305 mΩ
Stator d-axis Inductance (L_d)	1.21 mH
Stator q-axis Inductance (L_q)	2.31 mH
Permanent Magnet Flux (μ_g)	6.61 Wb
Inertia (J_{eq})	10000 kg.m^2
Radius of Blades (r)	38 m
Air Density (ρ)	1.205m^3

3. Proposed Control Method for SAPF

The Voltage Source Inverter has seen widely utilized in shunt APF setup to minimize power quality issues. It is also utilized for energy storage during ideal system conditions, but the charging and discharging of this capacitor controlled utilized a control theory. Hence, only the active filter fit for producing counter harmonic current and feeding the reactive power to point of common coupling (PCC). Figure 2 presents the configuration of the proposed shunt APF with a distribution system integrated into the wind energy conversion system (WECS). The main aim of proposed method is to evaluate the control strategy for current harmonics reduction, compensation of reactive power and balancing of load current at PCC when the supply current is distorted.

3.1. Active Power Component (APC) or Unit Vector Template Theory

Shunt APF can effectively handle the harmonics and reactive power problems simultaneously. Presently shunt APFs are designed to encounter all the harmonics which is generated by nonlinear loads but the distribution system is highly unbalanced, so it is required to mitigate the neutral current burden and make the source current sinusoidal for the calculation of the desired compensation current. The unit vector templates are determined as

$$\begin{aligned} U_a &= \frac{v_{Sa}}{V_{SP}} = sin\theta \\ U_b &= \frac{v_{Sb}}{V_{SP}} = \sin(\theta - 120^0) \\ U_c &= \frac{v_{Sc}}{V_{SP}} = \sin(\theta + 120^0) \end{aligned} \quad (13)$$

In a three-phase system, it is represented as i_a^*, i_b^*, and i_c^*. Neutral current ought to be zero, along these references neutral current taken as zero in a balanced system. Capacitor voltage error goes

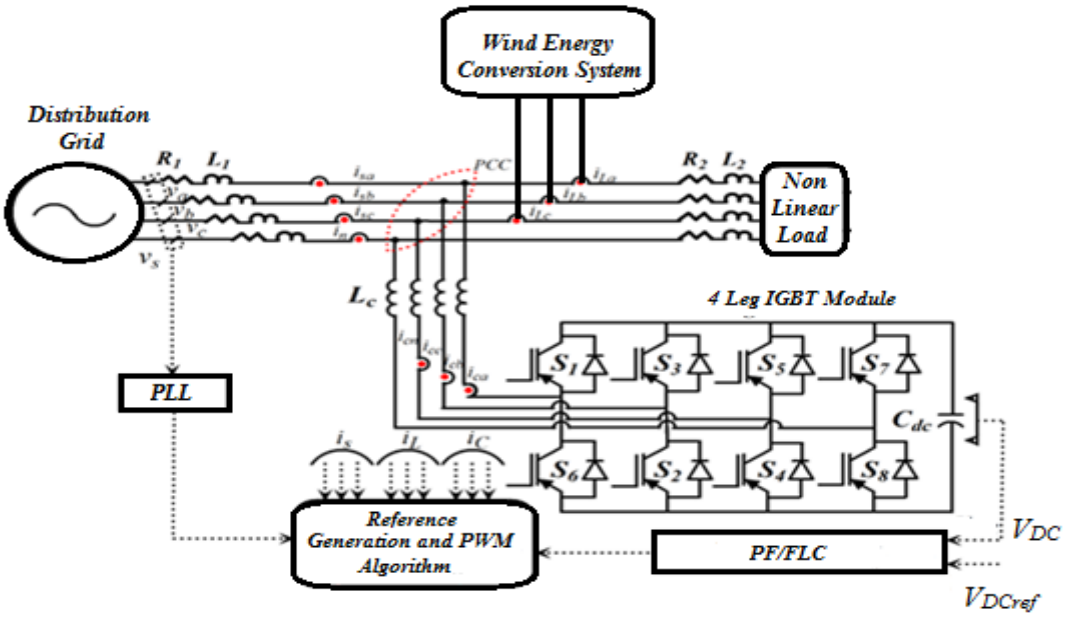


Figure 2: Proposed Shunt APF Configuration with Distribution System

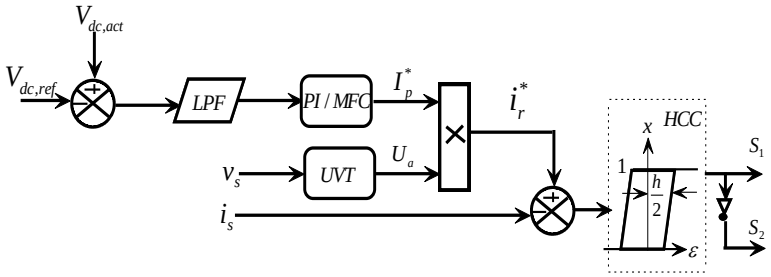


Figure 3: Block Diagram of UVT/APCT Algorithm with Indirect Current Control Method

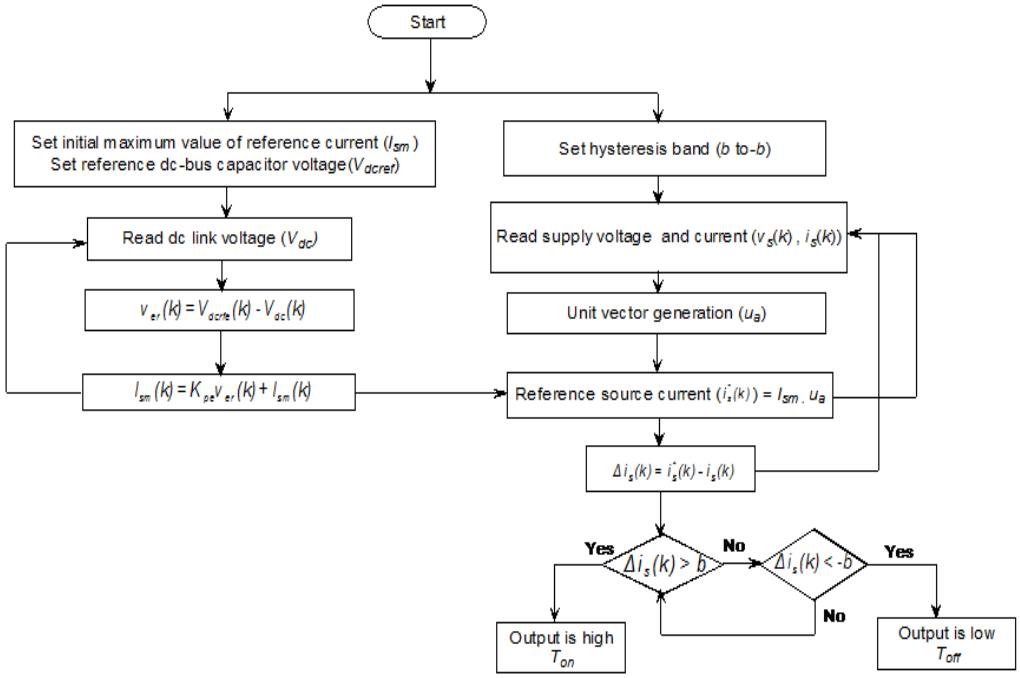


Figure 4: Flowchart of Active Power Component Theory (APCT) Algorithm

through the low power filter. The filtered dc voltage is contrasted with the reference value. Reference current can be composed as

$$i_a^* = I_m U_a,\ \ i_b^* = I_m U_b,\ \ i_c^* = I_m U_c,\ \ i_n^* = I_m U_n \qquad (14)$$

As present in the Figure 3 that capacitor voltage error goes through the LPF, which remove the high-frequency ripple from the voltage. This filtered dc voltage is compared with the reference value. Now error between the reference and grid current is processed by utilizing the following equations

$$\left.\begin{array}{c} i_a^* = I_m U_a \\ i_b^* = I_m U_b \\ i_c^* = I_m U_c \\ and \\ i_n^* = I_m U_n \end{array}\right\} \qquad (15)$$

Figure 4 presents the flowchart of the APCT algorithm with indirect current control method.

3.2. DC voltage Controller

DC voltage controller calculates response of the model and increasing proportional above a value can generate swaying. Integral control expels the error from the output. To actualize the PI controller, the dc link voltage is estimated the reference value. The transfer function of the PI-controller is given from [18] as follows

$$H(s) = (k_i / S) + k_p \qquad (16)$$

For the usage of the fuzzy rule over the PI controller in close loop filter, the dc voltage signal is compared with reference signal and reference error will be produced as follow [18]:

$$e = V_{\text{dc, reference}} - V_{\text{dc, actul}} \qquad (17)$$

The reference for nth sampling time is taken as a change of error in fuzzification process is determining as follows [18]-[19]

$$C_{e(n)} = e(n) - e(n-1) \qquad (18)$$

3.3. Proposed Fuzzy Ligic controller

The proposed modified fuzzy logic controller (MFLC) of the system is used to control the response of the dc voltage. They can be defined, as

$$x = \left(x_i, U_i, \left\{T_{x_i}^1, T_{x_i}^2, \cdots, T_{x_i}^{k_i}\right\}, \left\{\mu_{x_i}^1, \mu_{x_i}^2, \cdots, \mu_{x_i}^{k_i}\right\}\right) \quad \text{for } i = 1, 2, \cdots, n \qquad (19)$$

$$y = \left(y_i, V_i, \left\{T_{y_i}^1, T_{y_i}^2, \cdots, T_{y_i}^{l_i}\right\}, \left\{\mu_{y_i}^1, \mu_{y_i}^2, \cdots, \mu_{y_i}^{l_i}\right\}\right) \quad \text{for } i = 1, 2, \cdots, m \qquad (20)$$

where x & y is the input & output linguistic variable form a fuzzy output space. The size of a term set, $T_{(xi)} = K_i$ is called as the fuzzy partition number of X_i. In the case of a two – input-one-output proposed MFC system, If $T_{(x1)} = 9$ and $T_{(x2)} = 9$, than the number of fuzzy control rules is 9×9. A bell shaped relationship function has been used.

$$\mu_{xi}(x) = \exp\left(-\frac{(x - m_i)^2}{\sigma_l^{\ 2}}\right) \qquad (21)$$

This modified fuzzy controller rules are developed and approved dependent on AND Boolean logic operator. The new rules are characterized by fundamental stability position and moment in time response. To use of fuzzy sets, we restructure operators of the standard set method to fit the particular relationship functions of fuzzy logic for a value between zero & one. Here two variables are utilized as a contribution of MFLC, first is Error E and second change of error C_e. The moment in time response has been divided into four regions A_1, A_2, and A_3, A_4 with two arrangements of point, first is the peak value index (b_1, b_2), and second cross over-index (c_1, c_2). The fuzzy operator AND is formed from a grouping of fuzzy prepositions. Based on the above theory, rules are defined as ZE, PN NB, NS NM, PS, PB, and PB, PM for processing the dc-link response shown in Table 2. Figure 5 shows the moment in time response of a stable closed loop-system.

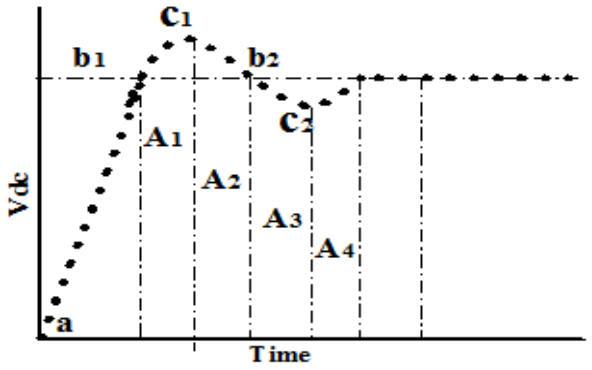


Figure 5: Moment in Time Response of a Stable Closed Loop System.

Table 2: Mechanism of Proposed modified fuzzy rule

Error/ Change error (E)/(Ce)	**NB**	**NM**	**NS**	**NN**	**ZE**	**PN**	**PS**	**PM**	**PB**
NB	NB	NB	NB	NB	NB	NB	NB	NM	ZE
NM	NB	NB	NM	NM	NM	NM	NS	ZE	PM
NS	NB	NM	NM	NS	NS	NS	ZE	PS	PM
NN	NB	NM	NS	NN	ZE	ZE	PS	PM	PB
ZE	NB	NS	NS	ZE	PN	PN	PS	PM	PB
PN	NS	NB	NS	PN	PS	PS	PS	PM	PB
PS	NM	NS	ZE	PS	PS	PS	PM	PM	PB
PM	NB	ZE	PS	PM	PM	PM	PM	PB	PB
PB	ZE	PM	PB	PB	PB	PB	PB	PB	PB

4. Result and discussion

The performance analysis of the distribution grid integrated with wind generation system is study with proposed SAPF under the use of different controllers for unbalanced loading condition. The impact of PQ issues in a distribution grid integrated with wind generation system is investigated in the presence of SAPF. The effectuation of SAPF is carried out using active power component theory (APCT) for reference current generation with direct and indirect current control method. Proportional integral (PI) controller is proposed with modified fuzzy logic controller (MFLC) for dc link voltage regulation and a basic hysteresis band current controller method. For study state analysis purposes, the load change with an increase at t=1.2 sec, and again it is changed with decreasing at t= 2.3 sec. The RMS load current at t=0.2 sec is in phase "a" 139.2 Amp, phase "b" 103 Amp, and in phase "c" 68.8 Amp. This result presents the unbalanced condition in the distribution system. Without compensation the neutral current flow in 73.3 Amp, the reactive power drawn is 20 kVAR, power factor degradation is below 0.8, and load current THD is 22.50%. A SAPF with a hysteresis-based APC technique is used to enhance the performance of the system according to certain power quality standards. In this technique, the reference signal is compared with actual utility load current.

4.1. Case 1: Direct current control with APCT-PI

In the direct current control method with APCT-PI, the switching pulses are generated through hysteresis controller by comparing the sensed three phase shunt APF current with their computerized generated reference current signal. The performance of the direct current control method with the APCT technique as shown in figure 6. The following observation has been made out after compensation reactive power flow between souse and load is 7 KVAR, natural current is around 7 Amp and system power factor is nearly 0.94.

4.2. Case 2: Indirect current control with APCT-PI

In the indirect current control method, the PWM switching pulses are obtained by comparing the sensed three phase current with their reference current which are generated by the reference current extraction methods. The performance of the in-direct current control method with the APCT technique is shown in figure 7. The following observation is made out after compensation reactive power flow between souse and load is 6 KVAR, natural current is around 6.8 Amp and system power factor is nearly 0.996. Hence the performance of the in-direct current control loop with unit vector template theory is found better than direct control so modified FLC implanted on indirect current method.

4.3. Case 3: Indirect current control with Modified Fuzzy Logic controller with APCT

It provides the batter result in a comparison of harmonics mitigation, reactive power compensation, and power factor improvement. From the figure 8, following observation, has been made out after compensation reactive power flow between souse and load is 3.7 KVAR, natural current is around 4.7 Amp and system power factor is nearly 0.999. Hence the performance of the in-direct current control loop with MFLC based APC theory is found suitable for SAPF. The comparisons of harmonics mitigation are shown in figure 9.

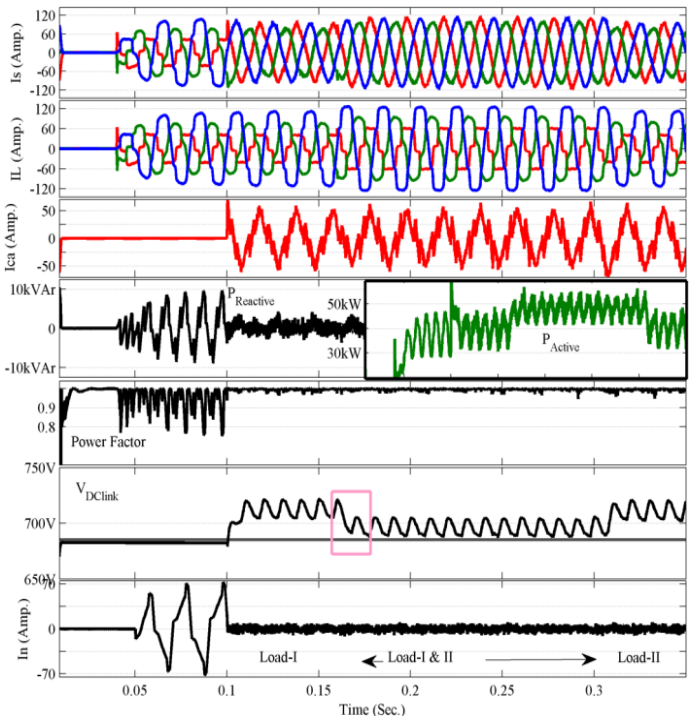


Figure 6: Indirect current control method shown as I_{sourse}, I_{load}, $I_{netural}$, $P_{react.}$, PF, V_{DC} and I_n with APC-MFLC theory.

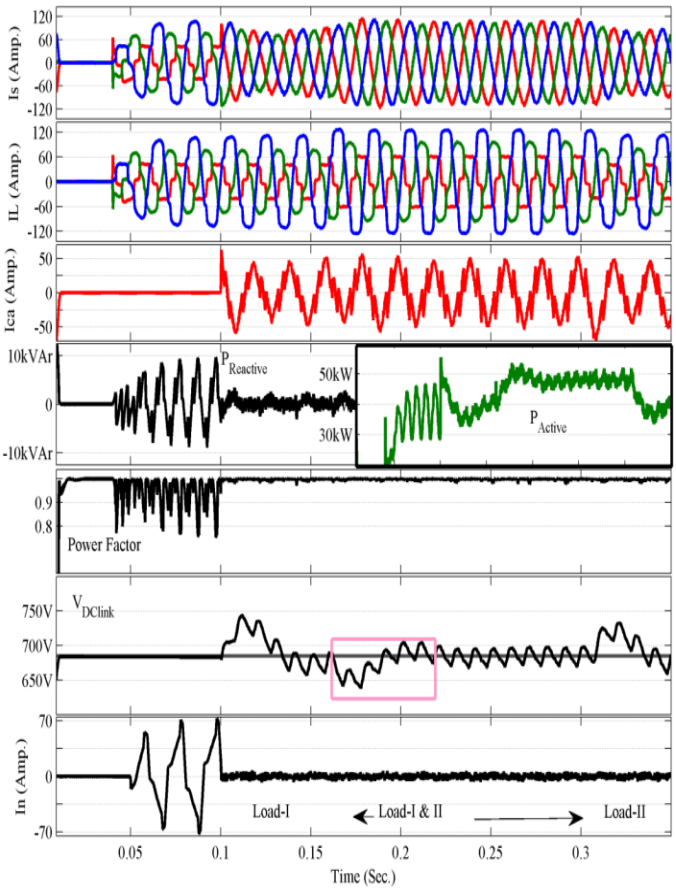


Figure 7: Direct current control method shown as I_{sourse}, I_{load}, $I_{netural}$, $P_{react.}$, PF, V_{DC} and I_n with APC-PI theory.

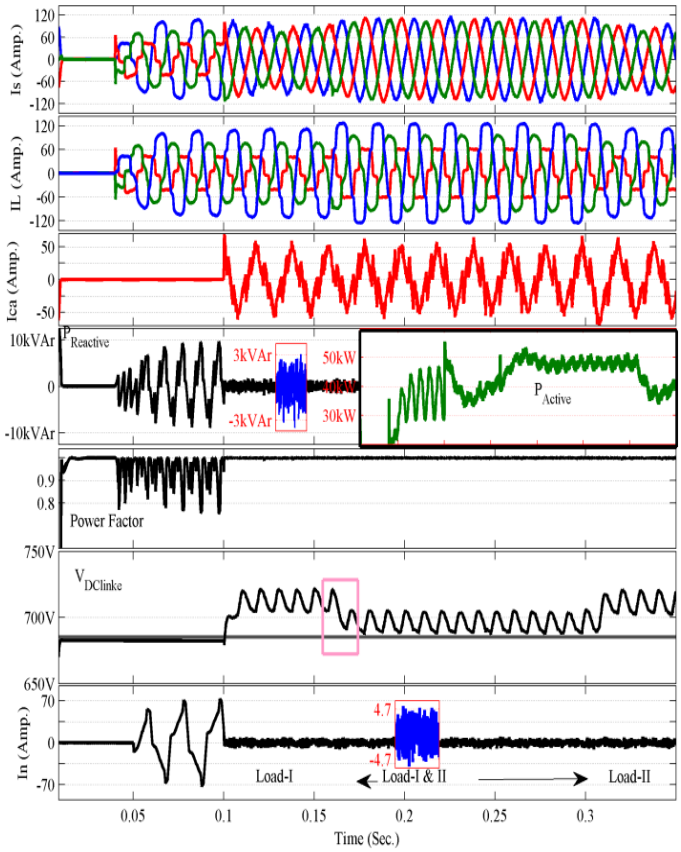


Figure 8: Indirect current control method shown as I_{sourse}, I_{load}, $I_{netural}$, $P_{react.}$, PF, V_{DC} and I_n with APC-MFLC theory.

4.4. Performance comparison of direct APC-PI, Indirect APC-PI and Proposed Indirect APC-MFLC

The comparison of overall performance of proposed system are discussed in Table 3.The performance of APC- PI theory with direct and indirect current control methods are compared and found that in both conditions, the neutral current compensation is almost same but the remarkable improvement is observed in power factor, reactive power compensation and load balancing with indirect control method. It is proved through proposed simulation that the indirect current control method performance is better than direct current control for the distribution grid integrated with wind generation system. So, proposed Indirect APC-MFLC methods are suitable to compensate harmonics and reactive power, neutral current during unbalanced loading conditions. The system is capable of maintaining unity power factor as well as sinusoidal source currents. Hence, the power quality of distribution system is improved.

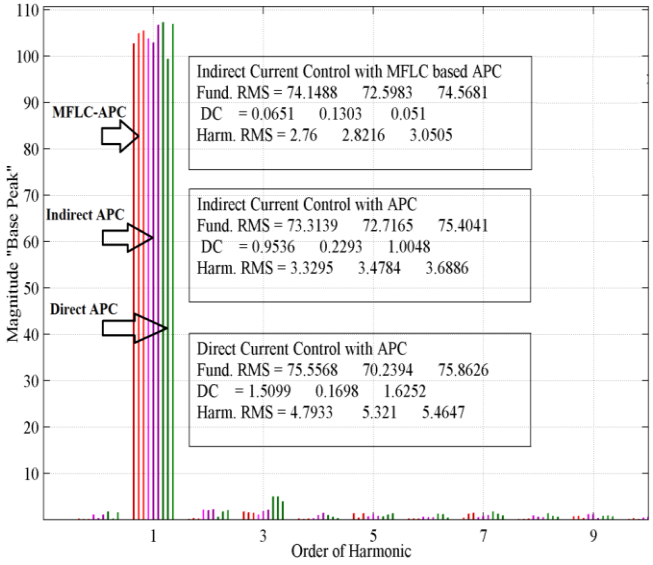


Figure 9: THD with Direct APCT-PI, Indirect APCT-PI & Indirect APCT-MFLC.

5. Conclusion

This paper presents a SAPF based on APC theory for reference current generation and proposed MFLC method for dc-link voltage control to enhance power quality of the distribution grid integrated with wind energy source under non-linear load. The performance of APC theory with direct and indirect current control methods is compared and found that in both conditions, the neutral current compensation is almost the same but the remarkable improvement shown in power factor, reactive power compensation, and load balancing with the indirect current control method.

Table 3: Comparison of direct APCT-PI, Indirec apct-pi and indirect Apct-MFLC

Parameter	**Direct control with APCT**			**Indirect direct control with APCT**			**Indirect direct control with MFLC- APCT**		
	I_{sa}	I_{sb}	I_{sc}	I_{sa}	I_{sb}	I_{sc}	I_{sa}	I_{sb}	I_{sc}
Balancing of source current I_s	108.4	98.06	105.1	103.6	102.8	106.6	110.8	1.8.1	108.1
Harmonic RMS after comp.	4.793	5.321	5.464	3.329	3.478	3.688	2.760	2.821	3.050
Neutral current I_n after amp.	7.0			6.8			4.7		
Reactive power after comp. (kVAr)	7			6			3.7		
Power Factor (PF)	0.94			0.996			0.999		

Hence, it has been proved through a simulation study that the indirect current control loop performance is better than the direct control loop for the unbalanced distribution system. The performance of the proposed method using APC-MFLC based realized SAPF is compared to that of SAPF designed based on reported strategies such as direct APC-PI controller and indirect APC-PI controller. The simulation results show the effectiveness of the proposed control algorithms to mitigate the harmonics, compensate the reactive power, and neutral current and to improve power factor with the balancing of source current using the MATLAB/Simulink.Finally, the proposed algorithms are capable of improving power quality problems in the distribution system with wind energy sources under non-linear load.

Conflict of Interest

All authors declare no conflicts of interests.

References

[1] N. F. Mohamed, "Wind energy conversion systems, classifications and trends in application," in 2019 5th International Renewable Energy Congress," 1-6, 2014, doi: 10.1109/IREC.2014.6826922.

[2] G. Shafiullah, M. Amanullah, A. Shawkat Ali, P. Wolfs, "Potential challenges of integrating large-scale wind energy into the power grid–A review," Renewable and Sustainable Energy Reviews, **20**, 306-321, 2013, doi: 10.1016/j.rser.2012.11.057.

[3] C. N. Bhende, S. Mishra, S. G. Malla, "Permanent magnet synchronous generator-based standalone wind energy supply system," IEEE Transastion Sustainable Energy, **2**(4), 361–373,2011, doi: 10.1109/TSTE.2011.2159253.

[4] W. Xueguang, W. Weisheng, D. Huizhu, C. Yunping, "Application of Models of the Wind Energy Conversion System to Wind Power Dynamic Analysis," in 1998 IEEE Power Systems Technology Conf., Beijing, 1406-1411, 1998, doi: 10.1109/ICPST.1998.729318.

[5] A. S. Bubshait, M. G. Sim, A. Mortezaei, T. D. C. Busarello, "Power quality achievement using grid connected converter of wind turbine system," in 2015 1st Annu. Meeting IEEE Ind. Appl. Soc.,1–8, 2015, doi: 10.1109/IAS.2015.7356817.

[6] M. Singh, V. Khadkikar, A. Chandra, "Grid synchronisation with harmonics and reactive power compensation capability of a permanent magnet synchronous generator-based variable speed wind energy conversion system," IET Power Electronics, **4**(1), 122-130, 2011, doi: 10.1049/iet-pel.2009.0132.

[7] Y. Oğuz, İ. Güney, H. Çalık, "Power quality control and design of power converter for variable-speed wind energy conversion system with permanent-magnet synchronous generator," The Scientific World Journal, **2013**(4), 122-132, 2013, doi: 10.1155/2013/783010.

[8] A. Bubshait, A. Mortezaei, M Simões, T. Busarello, "Power quality enhancement for a grid connected wind turbine energy system," IEEE Trans. on Industry Applications, **53**(3), 2495-2505, 2017, doi: 10.1109/TIA.2017.2657482.

[9] S. Kasa, P. Ramanathan, S. Amasamy, D. P. Kothari "Effective grid interfaced renewable sources with power quality improvement using dynamic active power filter," Electrical Power and Energy Systems, **82**, 150–160, 2016, doi:10.1016/j.ijepes.2016.03.002.

[10] H. Qazi, M. Wazi Mustafa "Review on active filters and its performance with grid connected fixed and variable speed wind turbine generator," Renewable and Sustainable Energy Reviews, 57, 420–438, 2016, doi: 10.1016/j.rser.2015.12.049.

[11] W. Ullah, M. Mekhilef, M. Seyedmahmoudian, "Active power filter for mitigation of power quality issues in grid integration of wind and photovoltaic energy conversion system," Renewable and Sustainable Energy Reviews, **70**, 635–655, 2017, doi: 10.1016/j.rser.2016.11.091.

[12] A. Hoseinpour, S. Masoud Barakati, R. Ghazi, "Harmonic reduction in wind turbine generators using a Shunt Active Filter based on the proposed modulation technique," International Journal of Electrical Power & Energy Systems, **43**(1), 1401-1412, 2012, doi: 10.1016/j.ijepes.2012.06.052.

[13] S. Kasa, P. Ramanathan, S. Ramasamy, "Effective grid interfaced renewable sources with power quality improvement using dynamic active power filter," International Journal of Electrical Power & Energy Systems, **82**(4), 150–160, 2016, doi:10.1016/j.ijepes.2016.03.002.

[14] D. Olivindo, J. A. Soa, I.R. Machado, "Shunt active power filter for energy quality improvement in distributed generation systems," in 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), 146-151, 2017, doi: 10.1109/ISIE.2017.8001238.

[15] H. Mojgan, Z. A. Abu, T. Arash, S. Mohammadsoroush "An Overview on Current Control Techniques for Grid Connected Renewable Energy Systems," 2012 2nd International Conference on Power and Energy Systems (ICPES 2012), Singapore, 119 – 126, 2012.

[16] O. F. Kececioglu, H. Acikgoz, C. Yildiz, A. Gani , M. Sekkeli, "Power Quality Improvement Using Hybrid Passive Filter Configuration for Wind Energy Systems," Journal of Electrical Engineering and Technology, **12**(1) , 207-216, 2017, doi: 10.5370/JEET.2017.12.1.207.

[17] M. Kouadria, T. Allaoui, M. Denai, "Power Quality improvement in isolated Wind-Diesel Power System," WEASE Trans. on Systems and Control, **12**(50), 490-498, 2018, doi: 10.5370/JEET.2017.12.1.207.

[18] J. Fei, K. Ma, S. Zhang, W. Yan, Z. Yuan "Adaptive current control with PI-fuzzy compound controller for shunt active power filter," Mathematical Problems in Engineering, **2013**(1), 150-158, 2013, doi.org/10.1155/2013/546842.

[19] P. Barik, G. Shankar, P. K. Sahoo "Power quality assessment of microgrid using fuzzy controller aided modified SRF based designed SAPF," International Transactions on Electrical Energy Systems, **29**(12), 1-24, 2019.

[20] V. N. Jayasankar**,** U Vinatha "Advanced Control Approach for Shunt Active Power Filter Interfacing Wind- Solar Hybrid Renewable System to Distribution Grid," Journal of Electrical System, **14**(2), 88-102, 2018, doi:10.1002/2050-7038.12289.

[21] S.H. Qazi, M.W. Mustafa, U. Sultana "Current harmonics mitigation from grid connected variable speed due to nonlinear loads using shunt active power filter," journal Teknnology, **79**(4), 2017, doi: 10.11113/jt.v79.4702.

[22] S. Ramasamy ,K. Sudheer "Enhancement of Power Quality in Renewable Source Integrated Three Phase System with ICC Controller," Journal of Green Engineering,, **8**(1), 71–88, 2018, doi: 10.13052/jge1904-4720.815.

[23] K. Ramesh, G.T. Arasu "Power Quality Improvement in Wind Energy System using Seven Level shunt active power filter," Asian Journal of Technology, **17**(3), 189-196, 2018, doi:10.36478/ajit.2018.189.196.

[24] G. Arunsankar, S. Srinath, "Optimal controller for mitigation of harmonics in hybrid shunt active power filter connected distribution system: An EGOANN technique," Journal of Renewable and Sustainable Energy, **11**(2), 1–16, 2019, doi: 10.1063/1.5085028.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 269-273 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Pre-University Students' Learning Styles and Attitude towards Mathematics Achievements

Nurhilyana Anuar, Nurashikin Abdullah*, Sharifah Norasikin Syed Hod

Universiti Teknologi MARA, Cawangan Selangor, Kampus Dengkil Dengkil,Selangor, 43800, Malaysia

ARTICLE INFO

Article history:
Received: 08 September, 2020
Accepted: 01 November, 2020
Online: 10 November, 2020

Keywords:
Learning styles
Attitude
Academic achievement

ABSTRACT

Learning styles studies have been discussed widely in academic field. It can be considered as a factor that contributes to the achievement of the students in their learning because everyone has a unique learning style when learning. In this paper, we aimed to examine the association of learning styles and attitude towards academic achievement. By identifying the student's learning styles preference, it will offer benefits to the student and instructor by improving learning achievement. Sample data were collected among 328 pre-university science and engineering students using cross-sectional survey. Attitudes Towards Mathematics Inventory (ATMI) and Index of Learning Style© (ILS) were used as the instruments to collect data from the respondents. Data were analysed using Stata statistical tool to examine the result. Result showed that there was no association of learning styles and attitude towards academic achievement in this study.

1. Introduction

Mathematics is one of the subjects in Sijil Pelajaran Malaysia (SPM), a high school national examination taken by secondary school students in Malaysia, as it is one of the subject requirements to continue study in pre-university program. However, the percentage of students who passed with distinction and credit in this subject has not improved from 2017 to 2019. Students who passed with distinction were 31.5% in 2017, 31.2% in 2018 and 31.2% in 2019 while those who passed with credit were 25.2%, 24.9% and 24.1% for 2017, 2018 and 2019 respectively [1, 2]. It is quite worrisome to see these statistics as mathematics does not only exist as individual subject, but it is also used as a tool in other subjects such as physics, chemistry, accountancy and economics to name a few. There are many aspects needed to be studied and improved in order to overcome this problem and the aspects of the students' learning styles and attitudes toward mathematics are among the aspects worth looked into.

In [3], the author mention that the learning style theory was originated from psychology research domain. However, this theory appears to have wide acceptance in many educational researches in exploring the issue of student achievement through learning style. There are several disciplines that have applied this theory in their studies such as nursing, engineering, mathematics [4-6] and online learning mode course [7].

In literatures, some studies on learning styles and academic achievement have been discussed extensively. For example, studies in [8], the studies found that there was no significant difference between learning style and academic achievement among Polytechnic undergraduate business students in Malaysia. The study concluded that by promoting student self-awareness and learning strategies, it will affect their learning style. In [9], the author studied the relationship of learning styles and academic achievement in mathematics among Grade 8 students. The study concluded that learning style is significantly related to academic achievement.

Based on the literature, there is a lack of study that discussed on pre-university students' learning style and attitude towards mathematic achievements. Moreover, the mixed results found in literature have motivated this study to be conducted. In addition, this study is an extension of the original research [10] that examined pre-university student's attitude towards mathematics achievement. Despite the fact that pre-university students from science program have always gained better achievement compared to students from engineering program, the original research showed that mathematical capability of the students was not influenced by the program they enrol in a local public university. Thus, the result showed that there was no significant findings on the attributes associated to academic achievement [10].

Therefore, it is significant to broaden this study on learning styles and attitudes towards mathematics achievement for pre-university students. The objectives of the studies are:

a. To study the student's preference learning style towards mathematics.

*Corresponding author: Nurashikin Abdullah., nurash2856@uitm.edu.my

b. To study the differences of student attitudes towards mathematics.

c. To measure the association of learning style and attitudes towards mathematics achievement.

In the next section, it discusses the related work on learning style. It is followed by research methodology while the final section is result and conclusion.

2. Related Works

What makes individual learning style distinct? Individual learning style is driven by some attributes such as characteristics and behaviour of the learner. According to literature, learning style has three basic elements which are cognitive, motivational and physiological. These elements indicate how the learner perceive, respond, and cooperate with the learning environment [11]. Every student has different level of cognitive, motivation and psychology responsive to the learning atmosphere [12]. Moreover, learning styles give an impact to the academic performance [13].

In an analysis done in [14], result showed there was a positive relationship between students' learning styles and the attitude towards mathematics. Many researches agreed that the effect of learning styles and positive attitude contributed to better achievement in mathematics. In [15], the author suggested that learning environment designed according to learning styles may increase students' success while the researcher in [16] said that individual learning style correlated the highest with mathematics performance. Likewise, studies conducted in [17], the author stated that positive attitudes towards mathematics and science are also associated with higher achievement in each subject.

There are several types of learning model discussed in literature on learning style which includes Kolb's Learning Style instrument (LSI), Learning Style Questionnaire (LSQ), Canfield Learning Style Inventory (CLSI), Cognitive Styles Analysis (CSA), Index of Learning Survey (ILS) and VARK (Visual, Aural, Reading or Write and Kinaesthetic) model [18].

Table 1: ILS dimension

Domains	Description
Sensing vs. Intuiting	Sensing – concrete and practical learner. Prefer facts, sound, sight, data, and physical sensation.
	Intuiting – abstract and theoretical learner. Prefer concept, theories, and insights.
Visual vs. Verbal	Visual – learn from what they have seen.
	Verbal – learning through the words whether from spoken or written.
Active vs. Reflective	Active – learn by performing practical activity and collaboration with others.
	Reflective – learn through thinking or pondering and working alone.
Sequential vs. Global	Sequential – linear thinking process. Prefer to do things step by step.
	Global – holistic thinking process. Prefer to study the whole picture to understand entire subject.

The most common models used in literatures are VARK model and ILS model [5-6],[19]. Different models of learning style have their own strengths and weaknesses. VARK model examines on sensory modality that includes visual, aural, read/write and kinaesthetic of learning styles. This model is used to identify learner preferred learning mode in which the learner may have multimodal learning styles [19]. Another model used among many researchers is Richard M. Felder and Linda K. Silverman's Learning and Teaching Styles based on its instrument of Index of Learning Styles (ILS). Initially, the model consists of five dimensions of Felder and Silverman's Learning Styles; Sensory-Intuitive, Visual-Auditory, Inductive-Deductive, Active-Reflective and Sequential-Global [12] but it has been updated to four dimensions which are Sensing-Intuiting, Visual-Verbal, Active-Reflective and Sequential-Global. In ILS model, information was perceived and processed by four domains [20, 21].

Furthermore, literature on learning style preferences among engineering student showed that students prefer active, sensing and visual learning styles in their problem-based learning (PBL) curriculum course [22]. Moreover, other studies also found that engineering students favoured sensing, visual, active and sequential learning styles [6]. In another perspective, learning styles preferences may differ between genders in which females have reflective, sensing, visual and sequential learning styles while males prefer active, sensing, visual and sequential [23]. In 2015, the author found that there was a significant relationship between learning styles and educational achievement such as in mathematical field [7]. In addition, there is a positive significant relationship between students learning style who used Active-Reflective and Visual-Verbal learning style and their score means.

ILS is more reliable, valid, and suitable for identifying learning styles [20]. It is a well-established tool and widely used to assess learner's learning styles. Moreover, various studies have been done on learning style using ILS [23]. Thus, this study adapted ILS model to fulfil the objectives of the study. Instruments used in this study are Attitudes Towards Mathematics Inventory (ATMI) and Index of Learning Style© (ILS). How does ATMI measure the attitude of the learner in mathematics? In ATMI instrument, it consists of four attributes; i) Self-confidence: learner's confidence and self-concept in comprehend mathematics; ii) Value: belief of usefulness, relevance and worth; iii) Enjoyment: learner has fun working with mathematics; and iv) Motivation: learner's interest and desire to study in mathematics [24].

3. Methodology

This study used cross-sectional survey among students attending pre-university studies in a local university and all students were invited to participate in this survey. A total of 328 students participated in this survey, giving the response rate of 13%.

A set of questionnaire consisted a range of information including socio-demographic as well as questions measuring their attitudes towards mathematics using Attitudes Towards Mathematics Inventory (ATMI) [25] and Index of Learning Style© (ILS) [26] was distributed among the participants. There were four self-reported domains focused in ATMI which included to measure self-confidence, value, enjoyment and motivation. ATMI consisted of 40 items on 5-point Likert scale ranging from strongly disagree to strongly agree. Its Cronbach's alpha indicated good reliability (α= 0.97). To find the result on the attitude of the student, each domain score was summed up. The higher scores showed positive attitudes towards mathematics.

The ILS consisted of 44-items which were divided into four subscales (active/reflective, sensing/intuitive, visual/verbal, and sequential/global). The Cronbach's alpha for ILS was considered

good with 0.74 [21]. The scores in ILS were computed by taking the difference on the number of items selected for each of the two styles in a scale. The total scoring was between -11 to 11. The scoring was categorised as strong (+/- 11-9), moderate (+/--5) and mild (+/-3-1) preferences.

Data analysis was done using Stata statistical software [27]. Prior to analysis, data were cleaned and verified for any errors, outliers or missing values. Continuous data were reported as means and standard deviations while categorical data were reported as frequencies and percentages. Normality test was conducted using Shapiro-Wilk's test. Bartlett's test for equal variances was conducted prior to mean comparisons between groups. P-value of more than 0.05 from Bartlertt's test confirmed the assumption of equality of variance. Bivariate analysis between continuous variables was run using t-test, while ANOVA was used for more than 2 comparison groups. Pairwise correlation was measured using Pearson correlation. The association between mathematics attitude and learning style on mathematics achievement was measured by linear regression. All statistical tests were conducted at 5% significant level.

4. Results

4.1. Participants

Table 2 shows a total of 328 pre-university students who participated in this study. It reflected the same participants who were used in both studies. The respondents were about 53% female and 47% male students. About 60% of the participants were from science program and 40% of the participants were from engineering program. There was a statistically significant difference between program and gender (Chi-square=49.233, p-value <0.001). More than half of the respondents reported that they attended boarding school previously (57%). More than 70% from both male and female students scored A+ in mathematics during their Higher School Certificate (HSC) exam. However, result showed that there was no statistically significant association between school type and gender (Chi-square=0.658, p-value=0.417).

Table 2: Distribution of participants in the survey

	All (N)	Male (n)	Female (n)	Chi-square test	p-value
Program					
Science	195	61	134	49.233	<0.001
Engineering	133	94	39		
Type of school					
Boarding school	187	92	95	0.658	0.417
Non-boarding school	141	63	78		
HSC Mathematics					
A+	235	115	120	0.939	0.333
A	93	40	53		

4.2. Finding from the original research

Result in Table 3 was taken from the original research which was presented in ICEED, 2019. The original research was more focused on the students' attitude towards mathematics final exam achievement for pre-university students. ANOVA test was conducted and the result showed there was no statistical difference between attitude and mathematics achievement (F=074, p-value = 0.658) of the students [10].

Table 3: Attitude and mathematics achievement

Attitude	Sum of squares	Df	Mean square	F	Sig.
Between groups	606.636	8	75.829	0.74	0.658
Within groups	32784.668	319	102.773		
Total	33391.304	327			

4.3. Learning styles preferences

Figure 1 shows the distribution of domain in learning style preferences. Female students were found to score more than 50% in most of the domains. Those who attended boarding school scored higher in all domains compared to non-boarding school attendees. Students who got A+ in their HSC mathematics exam were found to dominate all domains (more than 65%) while those who got between B+ and A for their HSC Additional Mathematics exam were found to score between 20% and 35% in all domains. Majority of the students from science program were found to prefer reflective (65%), sensing (67%), visual (60%) and global (62%) learning styles.

	Learning Style Preferences (Domain)							
	Active	Reflective	Sensing	Intuitive	Visual	Verbal	Sequential	Global
Gender								
Male	(47.1%)	(47.6%)	(40.4%)	(56.4%)	(47.7%)	(41.7%)	(47.2%)	(47.3%)
Female	(52.9%)	(52.4&)	(59.6%)	(43.6%)	(52.3%)	(58.3%)	(52.8%)	(52.7%)
School type								
Boarding	(60.3%)	(51.6%)	(59.0%)	(54.3%)	(57.6%)	(50.0%)	(53.9%)	(57.0%)
Non-boarding	(39.7%)	(48.4%)	(41.0%)	(45.7%)	(42.4%)	(50.0%)	(46.1%)	(43.0%)
HSC Mathematics								
A+	(70.6%)	(73.4%)	(67.6%)	(77.1%)	(72.0%)	(66.7%)	(70.6%)	(73.0%)
A	(29.4%)	(26.6%)	(32.4%)	(22.9%)	(28.0%)	(33.3%)	(29.4%)	(27.0%)
HSC Additional Mathematics								
A+	(4.4%)	(3.2%)	(2.7%)	(5.7%)	(4.3%)	(0%)	(5.0%)	(2.7%)
A	(30.4%)	(31.4%)	(27.7%)	(35.0%)	(30.6%)	(33.3%)	(31.1%)	(30.4%)
A-	(29.4%)	(29.8%)	(28.2%)	(31.4%)	(28.9%)	(37.5%)	(27.2%)	(32.4%)
B+	(23.0%)	(18.5%)	(24.5%)	(17.1%)	(21.7%)	(16.7%)	(21.1%)	(21.6%)
B	(12.2%)	(16.9%)	(16.5%)	(10.7%)	(14.1%)	(12.5%)	(15.0%)	(14.0%)
B-	(0.5%)	(0%)	(0.5%)	(0%)	(0.3%)	(0%)	(0.6%)	(0%)

Chi-square test: *<0.05, **<0.01, ***<0.001

Figure 1: Percentages of Learning Style Preferences (Domain)

Figure 2 shows the distribution of domain in learning style preferences for science and engineering students. Majority of the students from science program were found to prefer reflective (65%), sensing (67%), visual (61%) and global (62%) learning styles.

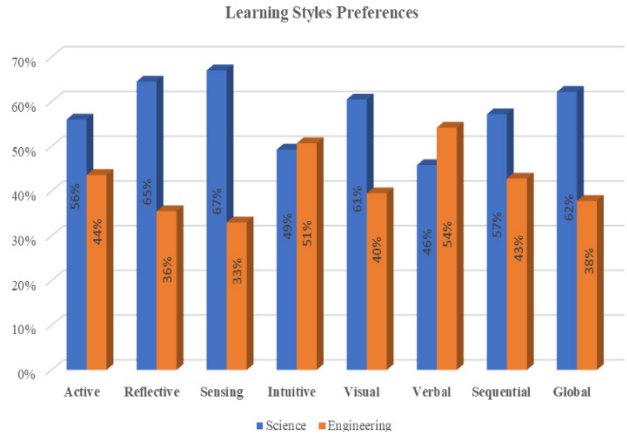


Figure 2: Learning Style Preferences (Domain) by program

4.4. Learning Style Preferences (Domain) during pre-university studies

Students who did very well (grade A) for both pre-university mathematics examinations during their studies had preference in all domains of learning styles. For semester 1 paper, 50.6% of those who scored A grade preferred sequential learning style while for semester 2 paper, students who scored A preferred both verbal (70.8%) and sequential (70%) learning style, respectively. In comparison, students who scored B grades (B- to B+) were found to prefer global learning style for both semester 1 and semester 2 papers as shown in Figure 3.

Learning Style Preferences (Domain)								
Grade	**Active**	**Reflective**	**Sensing**	**Intuitive**	**Visual**	**Verbal**	**Sequential**	**Global**
Semester 1								
A	48.0%	40.3%	44.7%	45.7%	45.1%	45.8%	50.6%	38.5%
A-	22.1	26.6%	22.3%	25.7%	23.0%	33.3%	23.3%	24.3%
B+	12.7%	12.9%	13.8%	11.4%	12.8%	12.5%	14.4%	10.8%
B	7.8%	12.9%	9.0%	10.7%	10.5%	0%	6.7%	13.5%
B-	5.9%	4.0%	6.4%	3.6%	5.3%	4.2%	2.2%	8.8%
C+	2.9%	1.6%	2.7%	2.1%	2.3%	4.2%	1.7%	3.4%
C	0.5%	0. %	1.1%	0%	0.7%	0%	0.6%	0.7%
D	0%	0.8%	0%	0.7%	0.3%	0%	0.6%	0%
Semester 2								
A	67.2%	60.5%	63.3%	66.4%	64.4%	70.8%	70.0%	58.1%
A-	10.3%	12.9%	11.7%	10.7%	10.5%	20.8%	9.4%	13.5%
B+	7.8%	11.3%	8.5%	10.0%	9.5%	4.2%	8.3%	10.1%
B	8.3%	7.3%	8.5%	7.1%	8.5%	0%	7.2%	8.8%
B-	3.4%	2.4%	3.2%	2.9%	3.0%	4.2%	1.1%	5.4%
C+	1.5%	2.4%	2.7%	0.7%	2.0%	0%	2.2%	1.3%
C	1.5%	1.6%	2.1%	0.7%	1.5%	0%	1.1%	2.0%
D+	0%	0.8%	0%)	0.7%	0.3%	0%	0%	0.7%
D	0%	0.8%	0%)	0.7%	0.3%	0%	0.6%	0%

Figure 3: Learning Style Preferences (Domain) in pre-university program

4.5. Students' attitude towards mathematics

The summary from ATMI survey that measured the students' attitude towards mathematics is presented in Table 4. The results indicate that male students have more positive attitudes compared to female students. Students from both types of school showed interest or positive attitudes towards mathematics. Those who did very well in both mathematics papers during HSC examinations have much higher positive attitudes compared to others. The same was true for those who were from engineering program. There were variations of attitude scores found across grades in mathematics subjects during pre-university studies. For semester 2 paper, students who scored B- and below were found to have higher positive attitudes' scores compared to the others. However, for semester 1 paper, their attitudes towards mathematics seemed to be equally distributed across all grades.

Table 4: Score on ATMI by gender, school type, HSC Mathematics and HSC Additional Mathematics

Characteristics	**N**	**Mean**	**SD**
Gender			
Male	155	137.941	10.237
Female	173	136.584	9.971
School			
Boarding	187	137.064	10.093
Non-boarding	141	137.410	10.153
HSC Mathematics			
A+	235	137.745	10.334
A	93	135.914	9.429
HSC Additional Mathematics			
A+	13	140.538	9.812
A	101	139.129	8.975
A-	97	137.505	10.586
B+	70	135.514	9.970
B	46	134.283	10.946
B-	1	130.0	0
Program			
Science	195	136.533	10.470
Engineering	133	138.241	9.492
Pre-university Mathematics Semester 1 grade			
A	148	138.412	9.926
A-	78	136.667	9.979
B+	42	134.048	12.791
B	32	137.375	8.454
B-	17	136.353	9.006
C+	8	138.500	6.655
C	2	139.500	2.121
D	1	134.000	0
Pre-university Mathematics Semester 2 grade			
A	212	137.439	10.209
A-	37	136.567	9.907
B+	30	136.867	10.274
B	26	134.154	10.376
B-	10	141.000	9.333
C+	6	138.667	10.875
C	5	140.400	5.771
D+	1	148.000	0
D	1	134.000	0

4.6. Association between learning style and attitudes towards mathematics achievement

Multiple linear analysis was conducted to measure the association between learning style and attitudes towards mathematics on the mathematic score while controlling for the effect of gender, type of school and HSC grade in mathematics. A significant regression model was observed ($F(5,322)=7.44$, Prob>F= <0.001 and Adjusted $R^2=0.089$). In this model, we found that learning style and attitude were not statistically associated with mathematics scores ($B_{learning\ style}=0.446$, $p=0.054$ and $B_{attitude}=0.019$, $p=0.708$, respectively) as presented in Table 5.

Table 5: Association between learning style and attitudes towards mathematics on the mathematics score in pre-university (semester 2 paper)

Predictors	**B**	**t**	**p-value**
Learning style scores	0.446	1.93	0.054
Attitude scores	0.019	0.37	0.708
Female	4.047	3.75	<0.001
Non-boarding school	3.682	3.39	0.001
HSC Mathematics, (A+)	3.221	2.690	<0.001

F (5,322)=7.44, Prob>F= <0.001 and Adjusted R2=0.089

5. Conclusion

This study showed that preference learning style was dominated by the female students from science program who had attended boarding school. Those students who got A+ in HSC mathematics dominated more than 65%. These students' preference learning styles were reflective, sensing, visual and global learning styles. Meanwhile, the result showed that when the students were in pre-university studies, those who scored

grade A during semester 1 preferred sequential learning style while students in semester 2 who scored A preferred verbal and sequential learning styles. Thus, it shows that there is no specific learning style that the students prefer to use because it depends on the environment of the studies. In terms of the student's attitude, we can conclude that there was a variation of attitude score found across grades in mathematics subject during pre-university studies. Those who had positive attitude scored a good grade in mathematics both in HSC and pre-university studies. Besides that, we have studied the factors that contributed towards academic achievement. The result showed no association was found between learning style and attitudes towards mathematics based on the mathematics scores. Thus, this study can help the instructors to plan their teaching through student's learning style to improve student achievement in mathematics. In addition, this study was conducted on pre-university students from science and engineering background which only focused on mathematics academic achievement. Therefore, this research can be expanded by investigating pre-university learning style for all subjects taken by the students and their academic achievement.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to thank Institute of Research Management and Innovation, Universiti Teknologi MARA for the financial support in this research under the grant 600- IRMI/Dana KCM 5/3/LESTARI (227/2017).

References

[1] Lembaga Peperiksaan Malaysia, "Pengumuman Analisis Keputusan Sijil Pelajaran Malaysia (SPM) Tahun 2018," 15, 2020.

[2] Lembaga Peperiksaan Malaysia, "Laporan Analisis Keputusan Sijil Pelajaran Malaysia (SPM) Tahun 2019," 17, 2020.

[3] F. Shahzad, S.A. Saif-ur-Rehman, "Learning Styles: History, Conceptualization and Continuum," Social Sciences Review (SSR), **2**(2), 15–31, 2014.

[4] N. Ojeh, N. Sobers-Grannum, U. Gaur, A. Udupa, M.A.A. Majumder, "Learning style preferences: A study of pre-clinical medical students in Barbados.," Journal of Advances in Medical Education & Professionalism, **5**(4), 185–194, 2017.

[5] C.K. Lee, M.S. Sidhu, "Engineering students learning preferences in UNITEN: Comparative study and patterns of learning styles," Educational Technology and Society, **18**(3), 266–281, 2015.

[6] M. Rezaeinejad, A. Azizifar, H. Gowhary, "The Study of Learning Styles and its Relationship with Educational Achievement Among Iranian High School Students," Procedia - Social and Behavioral Sciences, **199**, 218–224, 2015, doi:10.1016/j.sbspro.2015.07.509.

[7] M.M. El-Bishouty, A. Aldraiweesh, U. Alturki, R. Tortorella, J. Yang, T.W. Chang, S. Graf, Kinshuk, "Use of Felder and Silverman learning style model for online course design," Educational Technology Research and Development, **67**(1), 161–177, 2019, doi:10.1007/s11423-018-9634-6.

[8] H. Awang, N. Abd Samad, N.S. Mohd Faiz, R. Roddin, J.D. Kankia, "Relationship between the Learning Styles Preferences and Academic Achievement," IOP Conference Series: Materials Science and Engineering, **226**(1), 2017, doi:10.1088/1757-899X/226/1/012193.

[9] A.-M.L. Villajuan, "Relationship between Learning Styles & Academic Achievement in Mathematics of Grade 8 Students," International Journal of English Literature and Social Sciences, **4**(4), 1052–1055, 2019, doi:10.22161/ijels.4419.

[10] N. Abdullah, N. Anuar, N.F. Zolkipli, R.H. Shaharudin, S.N.S. Hod, "Assessing Pre-University Students' Attitude Towards Mathematics," Proceedings of the 2019 IEEE 11th International Conference on Engineering Education, ICEED 2019, 251–254, 2019, doi:10.1109/ICEED47294.2019.8994930.

[11] J. Keefe, "Assessment of Learning Style Variables: The NASSP Task Force Model, " Theory Into Practice, **24**(2), 138-144, 1985.

[12] R. Felder, L. Silverman, "Learning and Teaching Styles in Engineering Education.," Engineering Education, **78**(7), 674–81, 1988.

[13] M. İlçin, N. Tomruk, S.S. Yeşilyaprak, D. Karadibak, S. Savcı, "The relationship between learning styles and academic performance in Turkish physiotherapy students," Physiotherapy, **102**, e84–e85, 2016, doi:10.1016/j.physio.2016.10.084.

[14] N. Sirmaci, "The relationship between the attitudes towards mathematics and learning styles," Procedia - Social and Behavioral Sciences, **9**, 644–648, 2010, doi:10.1016/j.sbspro.2010.12.211.

[15] A. Ozerem, B. Akkoyunlu, "Öğrenme Stillerine Gore Tasarlanan Öğrenme Ortamları Ve Matematik Başarısına Etkisi," Egitim Arastirmalari - Eurasian Journal of Educational Research, **15**(61), 61–80, 2015, doi:10.14689/ejer.2015.61.4.

[16] A. Bosman, S. Schulze, "Learning style preferences and mathematics achievement of secondary school learners," South African Journal of Education, **38**(1), 1–8, 2018, doi:10.15700/saje.v38n1a1440.

[17] N. Berger, E. Mackenzie, K. Holmes, "Positive attiitudes towards mathematics and science are mutually beneficial for student achievement: a latent profile analysis of TIMSS 2015," The Australian Educational Reseacher, **47**, 409-444, 2020. doi: 10.1007/s13384-020-00379-8

[18] G. Balasubramaniam, I. K, "A Study of Learning Style Preferences among First Year Undergraduate Medical Students Using VARK Model," Education in Medicine Journal, **8**(4), 15–21, 2016, doi:10.5959/eimj.v8i4.440.

[19] H.R. Mozaffari, M. Janatolmakan, R. Sharifi, F. Ghandinejad, B. Andayeshgar, A. Khatony, "The Relationship Between the VARK Learning Styles and Academic Achievement in Dental Students," Advances in Medical Education and Practice, **11**, 15–19, 2020, doi:10.2147/amep.s235002.

[20] K. Middleton, E. Ricks, P. Wright, S. Grant, "Examining the Relationship Between Learning Style Preferences and Attitudes Toward Mathematics Among Students in Higher Education," Institute of Learning Styles Journal, 1(January), 1–15, 2013.

[21] S.R. Viola, S. Graf, Kinshuk, T. Leo, "Analysis of felder-silverman index of learning styles by a data-driven statistical approach," ISM 2006 - 8th IEEE International Symposium on Multimedia, 959–964, 2006, doi:10.1109/ISM.2006.30.

[22] A. Kolmos, J.E. Holgaard, "Learning styles of science and engineering students in problem and project based education," Proceedings of 36th European Society for Engineering Education, SEFI Conference on Quality Assessment, Employability and Innovation, 2008.

[23] C. SENELER, H. PETRIE, "Adaptation Of The Felder-Soloman Index Of Learning Styles (ILS) Into Turkish And An Assessment Of Its Measurement Quality," Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, (December), 711–736, 2018, doi:10.16953/deusosbil.332763.

[24] M. Tapia, G.E. Marsh, An Instrument to Measure Mathematics Attitudes, Academic Exchange Quarterly, **8**(2), 16–22, 2004.

[25] M. Tapia, "Attitudes toward Mathematics Inventory Redux," Academic Exchange Quarterly, **9**(3), 272, 2005.

[26] T.A. Litzinger, S.H. Lee, J.C. Wise, R.M. Felder, "A psychometric study of the index of learning styles©," Journal of Engineering Education, **96**(4), 309–319, 2007, doi:10.1002/j.2168-9830.2007.tb00941.x.

[27] StataCorp 2013. Stata Statistical Software: Release 13, 2013.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 274-279 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

A Comprehensive Review of Traditional Video Processing

Helen Kottarathil Joy*, Manjunath Ramachandra Kounte

School of ECE, REVA University, Bengaluru, 560064, India

ARTICLE INFO

Article history:
Received: 08 September, 2020
Accepted: 01 November, 2020
Online: 10 November, 2020

Keywords:
Video compression
MPEG
HEVC
Deep learning

ABSTRACT

Video and its processing are an interesting area as the increase in usage of internet videos, online streaming, CCTV, impact of internet on normal crowd increased. The need to know about video and its processing become an eminent area in research in current era. The paper tries to cover the traditional video processing, the advancement in video codec from the initial year, its origin, features, drawbacks and advancement lead to next stage. It provides an insight to need of video compression, steps involved in it, followed by overall review about video compression in various areas. The detailed explanation with reason of emergence, origin, characteristics are pointed. This information helps to add knowledge about the past and that helps to focus on the advancement and transitions that can be done to the video codecs. It summarizes the advancement in recent video processing using CNN, NN, deep learning too.

1. Introduction

'Video' the meaning has shifted far from a set of moving pictures from a traditional point of view to far extends. But the knowledge of traditional video processing is important to have a clear perspective towards the current video processing era. Video processing have a vast meaning from enhancement of various parameter in video, resolution, restoration of videos, denoising of videos, video compression etc. Each has its own impact and development in each stage of video processing development from video in a digital camera to HDTV to mobile camera to 4K,8K,10K videos. "The development of video compression started from the very started from the very start with a shadow of image compression such as Huffman coding, Golomb coding [1], arithmetic coding [2] etc. Later transform coding was introduced by encoding in spatial frequency including Fourier transform [3], Hadamard transform [4], Discreet Cosine Transform [5] etc followed by JPEG" [6]

The scope of video standardisation focus on video optimisation, allows complexity reduction for implementation and it no guarantee in terms of quality. Only decoder, bit stream and syntax of decoder is standardised. The scope if standardisation lies in the codec part of video processing. The basic knowledge about the fundamental terms in video processing helps to have strong view about any processing or standardisation done in video.

Video processing needs basic overview about the fundamentals like bitrate, display resolution, frame rate, frame type, interlacing, aspect ratio, video quality and compression techniques.

Bit rate is termed as the video data transferred in a frame of time or number of bits in the video transferred in a time durartion.it is related directly to the sharpness of the frame. This is a vital thing when we consider live broadcast into consideration. The term bitrate should be handled intelligently based on the encoder and decoder. Transfer speed also is a matter to be considered here. Display resolution in video is another area to be noted in video processing. Screen resolution /display resolution was not that important part in traditional era of video processing but as time moved and screen came with various sizes colour option, video

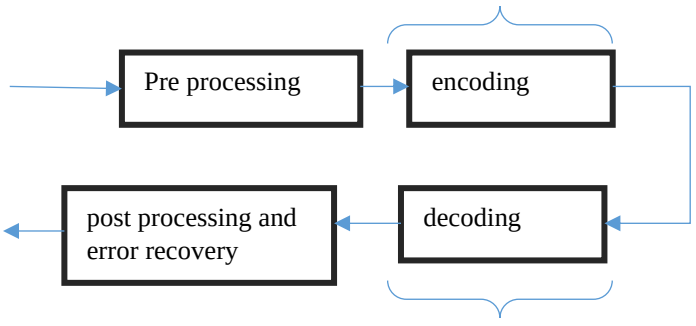


Figure 1: Scope of standardization

*Corresponding author: Helen Kottarathil Joy, helenjoy88@gmail.com

Screen resolution comes with an extension of 'p' and 'i' like 780p,1080p,1080i etc, it represents progressive /interlaced. The CRT monitors and TV followed interlaced scanning were the problem is flickering. Progressive scanning came into market as boom because the recent display screen responds fast. The refresh rate will be 60 normally to have better view effect. There are various frame resolution or frame type as in table

Table 1: Screen resolution Table

Sl.no	Resolution name	Horizontal pixels count	Vertical pixel count	Another Name
1	720p	1280	720	HD
2	1080p	1920	1080	Full HD
3	1440p	2560	1440	Quad HD
4	2160P	3840	2160	4K
5	4320P 8k	7680	4320	8K
6	4320p 10k	10240	4320	10K

Its noted that high resolution video can be viewed in low resolution screen. Video resolution doesn't matter on the screen it watching as it is taken care by the down sampling part in the decoder.

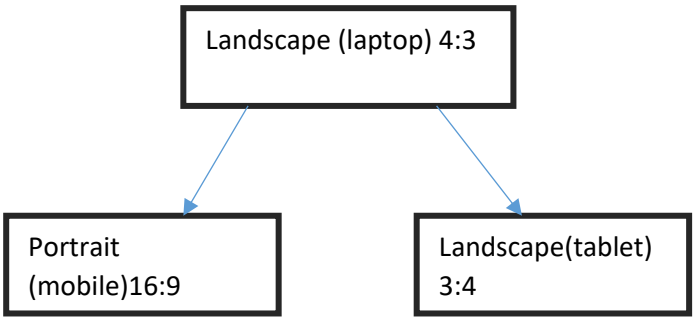


Figure 2: Aspect Ratio ideas

Aspect ratio says how the screen is with respect to the height of the display. In early ages 4:3 was common aspect ratio. In early 2010's by the boom of mobiles 16:9 become so prominent. Display orientation is another thing to be known. Size of the screen resolution should be considered properly for better display.

The paper is divided into three sections. The section 2 give brief idea about the steps in video compression, section 3 give a review of traditional video coding standards its origin, features and developmental stage towards next version. Comparison is done between various compressions followed by a overview about recent video compression with CNN and deep learning [7]. The paper ends with a conclusion with the need of learning traditional video compression for future research.

2. Video compression

2.1. Overview of video compression

Video compression is defined as a data reduction method ie used to encode the video. Video coding process helps that is the reduction of size of video file by making it compatible to store and need only less bandwidth for transfer. Video compression started as a succession of image compression. The compression mainly happens in a better way in video processing as the information carried by the frames are similar in most cases. A better percentage of compression can be achieved by taking this area into consideration. Video compression is essential as storage is taken into consideration. The disk space is important as we can more videos in fact more information can be loaded. Various transforms like DCT, DST etc can also be considered to reduce file size or in other words compression.

2.2. The stages of video compression

The basic steps in video processing are divided into six steps staring with partitioning the picture, predicting the similarity within the frame and between the frame. Majority of compression happens in this stage [8]. The predicted frame and other information's are transformed using DCT and quantized. All these were coded by using entropy coding and send as bit steams with proper bit rate.

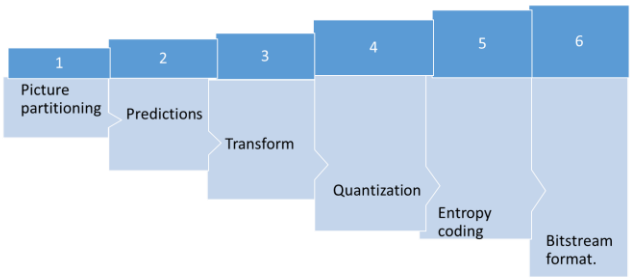


Figure 3: Basic Video Compression Steps

The input frame is transformed using DCT after picture partitioning (Splitting picture into macro blocks as apart of coding tree unit), then quantized. The quantized frame is inversed and compared with the previous frame in motion compensation and estimate is done by subtracting reconstructed frame from input frame and the residual with motion compensation. This is done by intra and inter prediction techniques basically comparing CTU within the frame and between frame to reduce redundancy and to gain maximum compression. The residual frame obtained will be 75 percentage compressed by this stage. The obtained frame 1&2 is transformed and quantized and the reconstructed frame is compared and subtracted. The process repeats and the output is encoded and framed and send in prescribed bit stream format.

3. Traditional video coding standards

Digital video technology covers the area of communication digital video telephony, digital TV, video storage, and a series of applications. This helps in the development of video codec in a faster way. The way of development of video codec are mainly focused on the organization ITU and ISO/IEC, joint venture of IFC and ISO are also available. ITU standards covers video codecs from H.261, H.263, H.263+, and this focus mainly in the video compression in real time communication for example video

conferencing whereas ISO/IEC mainly focus on the internal streaming, video storage etc. The bird view of video compression is shown in figure 4. Based on the purpose and application the compression features vary, upon that the segregation can be as the

one used in web or internet application, in cinemas, and for other medias. The ISO/IEC organization works on MPEG series as ITU/VCEG works on H.26x series. The ancestry of recent video compression followed this way.

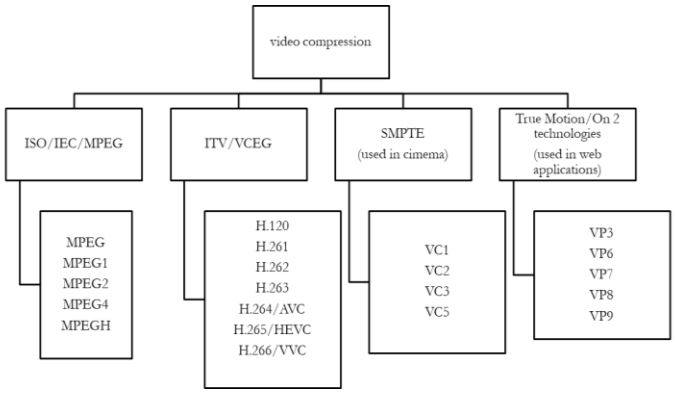


Figure 4: Video Compression Overview

Table 2: Specifications of H.262

1	Bit rate	1.5-2.0 Mbps CBR/VBR constant /variable bit rate
2	Packet/ cell loss rate (CLR)	<1 in 10 (-8)
3	BER	<1 in 10(-10)
4	Packet/ cell delay variation	<500ns

3.1. H.120 Video Compression Standard

The video coding standard starts with H.120 i.e. developed in the year 1984 by CCITT currently known as ITU-T. The primary application was video telephony and teleconferencing over ISDN. The bit rate available was 64kbits /sec.H.120 was not that good enough for real life applications, its spatial domain was good and temporal resolution was poor. The discussion came that encoding should have less than 1 bit as an average in a pixel. This leads to the idea of DCT block-based codec ie H.261.

3.2. H.261 Video Compression Standard

H.261 is the first member in the H.26x family developed in the year 1988-1990 in November by VGCE ie ITU-T9 video coding expert groups. The focus for the development of H.261 was similar to H.120 like video conferencing in ISDN and video telephony as it couldn't satisfy it [9]. The bit rate it was focusing was multiples of 64 kbits /sec. It was used as a backward compatibility mode in H.323 and some video conferencing system still it couldn't be perfect and need refinement that leads to next video codec. But still it remains as a milestone in the history of video codecs.

3.3. H.262/ Moving Picture Expert Group-2

H.262/MPEG-2 was proposed in 1995 by the joint venture of VCEG and MPEG. It is mainly for storage of video like SVCD, digital video, Blue ray, broadcasting, DVD video etc. Its bit rate is suitable for only less than 1Mbps, performs bad for high bit rate. MPEG -2 is a lossy compression method with motion estimation vector, DCT quantization an encoding by Huffman coding, the technique happens in this sequence [10]. The motion vector estimation compares the frame and approximate it into similar set of video frame in a translated form with changes between the frame. This helps to reduce the greater percentage of temporal redundancy in the video frame. This is followed by DCT that helps to convert the spatial information to frequency domain and discard all high frequency information that will not affect the visual experience of human eye. Quantization is applied to the coefficient of DCT. Huffman coding shortens the code and helps in better compression but its lossy with entropy coding.MPEG is good for high quality video with a particular set of transmission limit but not that better for internet transmission as the QoS of MPEG can't be satisfied always with intend.

3.4. H.263 Video Compression Standard

H.263 was designed for low bitrate mainly on video conferencing. This was evolved from the pros and cons of H.261,

MPEG-1, MPEG-2 standard developed in 1996 by ITU -T9. The main application was B-ISDN video on mobile phones (3Gp), video conferencing and video telephony adopted H.324 PSTN. the bit rate followed was 33.6 Kbps. It was developed as their levels that is 1,2,3 as H.263, H.263+, H.263++ with various extra features added on it. H263+ was focusing on improved compression performance and bit stream scalability. H.263++/H.263 2000 namely MPEG 4 part 2, it helps surveillance system its simple profile formal.

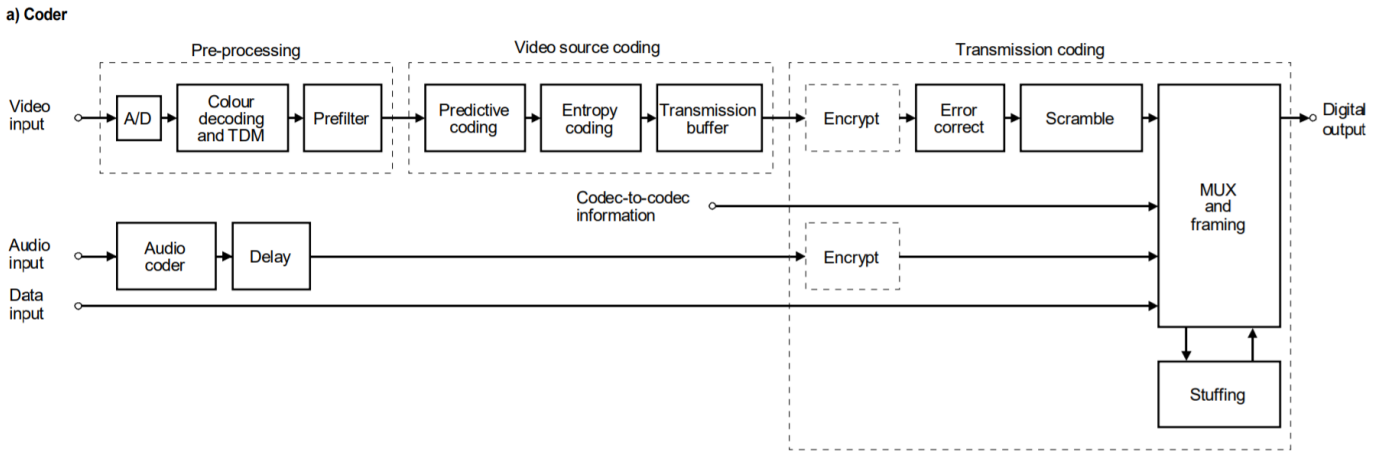


Figure 5: H.261 [7]

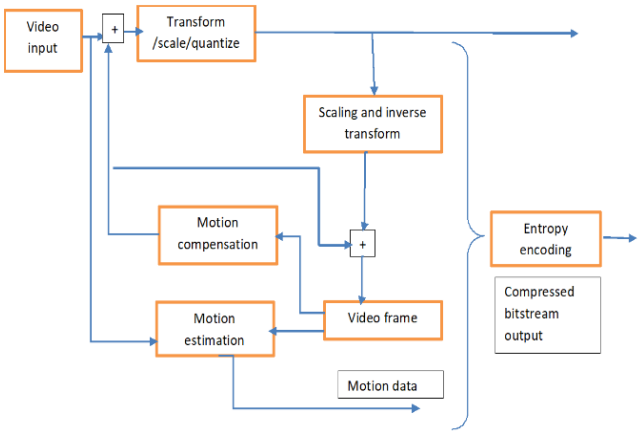


Figure 6: Hybrid video encoder [1]

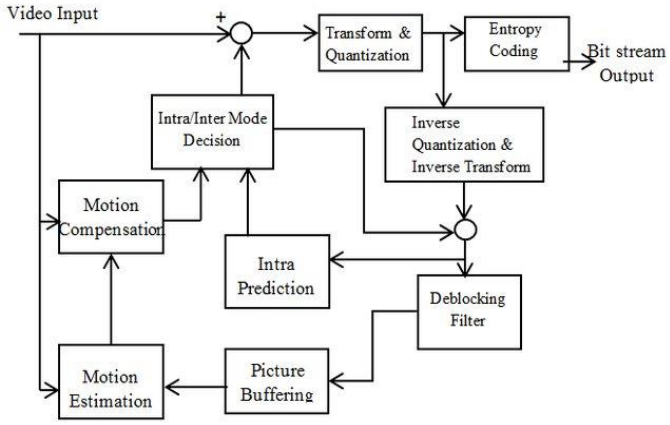


Figure 7: H.264 codec [12]

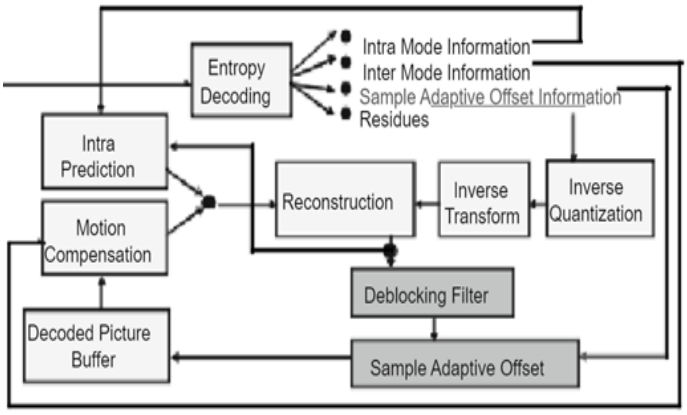


Figure 8: Video Compression with HEVC [14]

3.5. H.264 Video Compression Standard

H. 264/MPEG 4, AVC i.e. [10]an extended version of H.263, MPEG 4 visual released on 30th may 2003 is developed by ISO/IEC and ITU and H .264 and distribution of video content. The resolution it support is 4096X2304 that includes UHD and 4K videos. It uses variable bit rate. Advanced Video Coding /AVC is 25th version of AVC ie MPEG 4 part 10[11]. It is the best-known video coding standard for media storage like Blue ray disc and widely used in streaming in internet source that used variable bit rate. Scaling video coding is a domain feature here added with multi view video coding, 3D-AVC and multi resolution frame compatible (MFC). The application area includes DVBC (digital video broadcast project), HD-DVD, ATSC (advanced television system connector), CCTV (closed circuit TV), DSLR etc.

3.6. H.265 Video Compression Standard

H.265 is devoloped as the successor of AVC with reduced file size, reduced bandwidth[12]. There are CTUs in HEVC than macro blocks in H.264. The HEVC have variable CTU size from 4x4 to 64x64. The main features of HEVC includes flexibility in partitioning size, flexibility in trasform block size and predicting the modes, parallel processing architecture, sophistication for prediction, signalling mode and motion vector[13]. HEVC architecture has encoder and decoder. In encoder picture partioning is done. In the next process is the partioned picture is predicted by intra /inter prediction and the result is subtractred from the original unit[14]. The residue is transformed by DCT and quantised in next step. At the end transformed output, predicted information, mode information, and header are entropy encoded. In decoder the counterpart does the reverse operation to deliver the picture to other end of communication[15].

3.7. Society of Motion Picture and Television Engineers standards

It was introduced in founded in 1916 as the Society of Motion Picture Engineers or SMPE is a global professional association, of engineers, technologists, and executives working in the media and entertainment industry. An internationally recognized standards organization, SMPTE has more than 800 Standards, Recommended Practices, and Engineering Guidelines for broadcast, filmmaking, digital cinema, audio recording, information technology (IT), and medical imaging[16-19]. In addition to development and publication of technical standards documents, SMPTE publishes the SMPTE Motion Imaging Journal, provides networking opportunities for its members, produces academic conferences and exhibitions, and performs other industry-related functions"[20].Compression Systems in SMPTE has standardized five VC standards: VC-1 to VC-5 to provide well-reviewed documentation and enhanced interoperability. The latest of these is the VC-5 standard family that provides documentation and reference software for the video compression used in GoPro systems and workflows. SMPTE also has a new project to document the Apple ProRes codec.

3.8. Video Processing for Internet

On 2 codec designs are popularly known as VP3 then VP7.VP6 was used in flash 8 video codec. It was used in google, skype, YouTube etc. Microsoft and Windows media player version 9 properties similar to Real video, DivX and On 2 tech. AVC improves compression quality, H.264 becomes leading standard and used by many video applications such as in iPod PlayStation Portable, as well as in TV broadcasting standards like DVB-H and DMB[21].AVS audio video standards were initialized by China, the main reason china holds the key factors in development of consumer electronics. Coding efficiency is equivalent to H.264 while computation complexity is less. It was widely used standard [22].

Table 3: Comparison Table of video codecs

Parameter	HEVC	AVC	VP9
Compression Efficiency	Twice compared to AVC	Less than HEVC	Less than HEVC
Resolution &uploading speed	Less compared to AVC	Better compared to both	More
Bandwidth required for broadcast	15mbps	32mbps	More than HEVC
Compatibility	4k TVs, VOD	3Dvideo coding.	Better in chrome, opera Firefox…
Royalty	Not open source	Not open source	Open source
Computation cost	low	high	low
Encoding quality	Good for low bit rate videos	Use lower bit rates better than previous standards	Good for high bit rate videos
FPS	Support up to 300fps	Support up to 59.94fps	Support up to 60fps for video

4. Neural network and Deep learning-based video coding standards.

After the development of H.265/HEVC, the blocks in the codec were explored and improved by adding neural network/ deep learning techniques [22-27] into it. The development of video codec in the following year will be mainly on making a smart video codec by incorporating the deep learning features into it[28-30]. The trial of neural network /deep learning into HEVC standard by the application of trying to make smart modules in HEVC. By using multiple features, the modules are trying to be made smart. The future of codec lies in the hands of deep learning i.e. the future is expecting a smart video codec by absorbing the basic steps in codec and improvising it by adding deep learning techniques into it. Work have been done in HEVC modules intra-prediction, inter-prediction, quantization and encoding and loop filter with neural network methods [31] that can be revised and learned separately.

As an overview the neural network for video coding can be reviewed this way, for intra prediction all the method developed for still image can be used. As we know the applications are many. For neural network based intra-prediction the effective tool considered can be motion, with the better training network it can be effectively done. Some other research happened in this area using CNN are post processing, resolution up conversion sub pixel interpolation, loop filtering, intra-coding, encoding optimization etc.

5. Conclusion

Video and its processing are the area that is focusing on a smart era in the near future with high quality advancement. The paper points the video processing developmental stages from initial days till now. This helps to focus on the recent research in video processing and thus paves way for new research knowing about the origin, features and properties of early codecs, earns us to work on the future of video processing. The early era of video compression started with basic coding like Huffman coding, arithmetic coding etc and transforms like Fourier transform, DCT etc followed by the hybrid video processing era with many compression techniques from Motion JPEG to H 265. The detailed study of this helps to identify that in each stage the was trying to improve the compression ratio, reduce the complexity of circuit, and with high uploading speed. The research and advancement in this area can be done by having a strong base about the existing video processing. The drawbacks of the system can be reviewed and that helps in the research to make it smarter one. The future is focusing on smart video processing and deep learning, NN, CNN are the strong pillars for intelligent video processing.

Conflict of Interest

The authors declare no conflict of interest

References

[1] D.A. Huffman, “A Method for the Construction of Minimum-Redundancy Codes,” Proceedings of the IRE, ,**40**(9) , 1098-1101,1952,doi:10.1109/JRPROC.1952.273898.

[2] S.W. GOLOMB, “Run-Length Encodings,” IEEE Transactions on Information Theory, **12**(3), 399-401, 1966, doi:10.1109/TIT.1966.1053907

[3] I.H. Witten, R.M. Neal, J.G. Cleary, “Arithmetic coding for data compression,” Communications of the ACM, **30**(6), 520-540, 1987, doi:10.1145/214762.214771.

[4] H.C. Andrews , W.Pratt ,”Fourier Transform Applications”, Proc. Hawaii Int. Conf. System Sciences , 677-679, 1968, doi:10.5772/2658.

[5] W.K. Pratt, H.C. Andrews, J. Kane, “Hadamard Transform Image Coding,” Proceedings of the IEEE, **57**(1), 58-68, 1969, doi:10.1109/PROC.1969.6869.

[6] H. K. Joy and M. R. Kounte, "An Overview of Traditional and Recent Trends in Video Processing," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 848-851, 2019, doi: 10.1109/ICSSIT46314.2019.8987896.

[7] A.S. Coyner, J.P. Campbell, M.F. Chiang, Demystifying the Jargon: The Bridge between Ophthalmology and Artificial Intelligence, Ophthalmology Retina, 2019, doi:10.1016/j.oret.2018.12.008.

[8] A.N. Netravali, J.D. Robbins, “Motion-Compensated Television Coding: Part I,” Bell System Technical Journal, **58**(7), 1703-1718,1979, doi:10.1002/j.1538-7305.1979.tb02238.x.

[9] N. Ahmed, T. Natarajan, K.R. Rao, “Discrete Cosine Transform,” IEEE Transactions on Computers, **100**(1), 90-93, 1974, doi:10.1109/T-C.1974.223784.

[10] D.K. Kwon, M. Budagavi, V. Sze, W.S. Kim, Video compression, 2017, doi:10.1201/b12494.

[11] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, A. Luthra, “Overview of the H.264/AVC video coding standard,” IEEE Transactions on Circuits and Systems for Video Technology, **13**(7), 560-576, 2003, doi:10.1109/TCSVT.2003.815165.

[12] Wen Gao, "AVS standard - Audio Video Coding Standard Workgroup of China," 14th Annual International Conference on Wireless and Optical Communications, 2005. WOCC 2005, Newark, NJ, USA, 125–166, 2005, doi: 10.1109/WOCC.2005.1553738.

[13] M.A. Ansari, I.U. Khan, “Performance analysis and evaluation of proposed algorithm for advance options of H.263 and H.264 video codec,” in 2015 International Conference on Recent Developments in Control, Automation

and Power Engineering, RDCAPE 2015, 371-376, 2015, doi:10.1109/RDCAPE.2015.7281427.
[14] G.J. Sullivan, J.R. Ohm, W.J. Han, T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," IEEE Transactions on Circuits and Systems for Video Technology, **22**(12), 1649–1668, 2012, doi:10.1109/TCSVT.2012.2221191.
[15] H. Lv, R. Wang, X. Xie, H. Jia, W. Gao, "A comparison of fractional-pel interpolation filters in HEVC and H.264/AVC," in 2012 IEEE Visual Communications and Image Processing, VCIP 2012, 1-6, 2012, doi:10.1109/VCIP.2012.6410767.
[16] K.M. Khan, J. Arshad, M.M. Khan, "Secure digital voting system based on blockchain technology," International Journal of Electronic Government Research, **14**(1), 53–62, 2018, doi:10.4018/IJEGR.2018010103.
[17] V.M. Harshini, S. Danai, H.R. Usha, M.R. Kounte, "Health record management through blockchain technology," in Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019, 1411-1415, 2019, doi:10.1109/icoei.2019.8862594.
[18] H.K. Joy, S.L. Das, "A novel approach for biomedical web image super resolution," in Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2013, 876–879, 2013, doi:10.1109/ICCPCT.2013.6528858.
[19] S.J. Kamble, M.R. Kounte, "Machine Learning Approach on Traffic Congestion Monitoring System in Internet of Vehicles," in Procedia Computer Science, **171**(1), 2235–2241, 2020, doi:10.1016/j.procs.2020.04.241.
[20] S. Ma, X. Zhang, C. Jia, Z. Zhao, S. Wang, S. Wang, "Image and Video Compression with Neural Networks: A Review", IEEE Transactions on Circuits and Systems for Video Technology, **30**(6), 1683-1698, 2020, doi:10.1109/TCSVT.2019.2910119.
[21] S. Matsumoto, T. Yoshihisa, T. Kawakami, Y. Ishi and Y. Teranishi, "A Distributed Video Processing System for Internet Live Broadcasting Services," 2016 19th International Conference on Network-Based Information Systems (NBiS), Ostrava, 2016, 311-316, doi: 10.1109/NBiS.2016.28.
[22] M.R. Kounte, P.K. Tripathy, P. Pramod, H. Bajpai, "Implementation of Brain Machine Interface using Mind wave Sensor," in Procedia Computer Science, 244–252, 2020, doi:10.1016/j.procs.2020.04.026.
[23] B. Tian, L. Li, Y. Qu, L. Yan, "Video Object Detection for Tractability with Deep Learning Method," in Proceedings - 5th International Conference on Advanced Cloud and Big Data, CBD 2017, 397-401, 2017, doi:10.1109/CBD.2017.75.
[24] H. Wu, X. Zhang, B. Story, D. Rajan, "Accurate Vehicle Detection Using Multi-camera Data Fusion and Machine Learning," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 3767-3771, 2019, doi:10.1109/ICASSP.2019.8683350.
[25] M.R. Kounte, P.K. Tripathy, P. Pramod, H. Bajpai, "Analysis of Intelligent Machines using Deep learning and Natural Language Processing," in Proceedings of the 4th International Conference on Trends in Electronics and Informatics, ICOEI 2020, 956-960 ,2020, doi:10.1109/ICOEI48184.2020.9142886.
[26] W. Zhang, D. Zhao, L. Xu, Z. Li, W. Gong, J. Zhou, "Distributed embedded deep learning based real-time video processing," in 2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings, 001945-001950, 2017, doi:10.1109/SMC.2016.7844524.
[27] M.R. Kounte, P.K. Tripathy, P. Pramod, H. Bajpai, "Implementation of Brain Machine Interface using Mind wave Sensor," Procedia Computer Science, **171**, 244–252, 2020, doi:10.1016/j.procs.2020.04.026.
[28] J. Lainema, F. Bossen, W.J. Han, J. Min, K. Ugur, "Intra coding of the HEVC standard," IEEE Transactions on Circuits and Systems for Video Technology, **22**(12) ,1792-1801, 2012, doi:10.1109/TCSVT.2012.2221525.
[29] J.L. Lin, Y.W. Chen, Y.W. Huang, S.M. Lei, "Motion vector coding in the HEVC Standard," IEEE Journal on Selected Topics in Signal Processing, **7**(6), 957-968 , 2013, doi:10.1109/JSTSP.2013.2271975.
[30] X. Zhang, S. Wang, Y. Zhang, W. Lin, S. Ma, W. Gao, "High-Efficiency Image Coding via Near-Optimal Filtering," IEEE Signal Processing Letters, 2017, doi:10.1109/LSP.2017.2732680.
[31] L. Vinet, A. Zhedanov, "A 'missing' family of classical orthogonal polynomials," in Journal of Physics A: Mathematical and Theoretical, **44**(8), 085002, 2011, doi:10.1088/1751-8113/44/8/085201.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 280-289 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Overmind: A Collaborative Decentralized Machine Learning Framework

Puttakul Sakul-Ung[1,*], Amornvit Vatcharaphrueksadee[2], Pitiporn Ruchanawet[3], Kanin Kearpimy[1], Hathairat Ketmaneechairat[1], Maleerat Maliyaem[1]

[1]*King Mongkut's University of Technology North Bangkok Bangkok, 10800, Thailand*

[2]*North Bangkok University Bangkok, 12130, Thailand*

[3]*King Mongkut's University of Technology Thonburi Bangkok, 10140, Thailand*

ARTICLE INFO

Article history:
Received: 29 August, 2020
Accepted: 01 November, 2020
Online: 10 November, 2020

Keywords:
Decentralization
Machine Learning
Distributed Artificial Intelligence
Feature Transfer
PM2.5 Prediction
Collaborative Machine Learning

ABSTRACT

This paper is an extension of work originally presented in PM2.5 Prediction-based Weather Forecast Information and Missingness Challenges: A Case Study Industrial and Metropolis Areas, which focused on imputation algorithm to solve missingness challenge and demonstrated a basic prediction system to prove the proposed algorithm, II-MPA. Distributed and decentralized systems, recently, have been proven for their effectiveness in multiple perspectives. This paper introduces "Overmind", the solution that governs and builds the network of decentralized machine learning as a prediction framework named after its functionality: it aims to discover a set of data and associated attributes for assigning machine learnings in the collaborative decentralized manner. Overmind also empowers feature transfer learning with data preservation. It demonstrates how discovered features are transferred and shared among synergic agents in the network. This model is tested and evaluated the accuracy against the traditional single machine learning prediction model in the original work. The results are satisfactory in both prediction performance and transfer learning.

1. Introduction

This paper is an extension of work in [1]. Traditionally, a single machine learning is built to process upon a dataset; this can be served as a recommendation, prediction, classification, etc. Unfortunately, the drawback of the traditional model is realized where machine learning problems and algorithms are far enough that all scenarios can be resolved using a single model [2]. A traditional approach relies on the training process over a whole dataset; data exploration, perhaps, consumes time and resources. Also, the experience in data analytics does not guarantee the effectiveness of the built machine learning systems to specific business scenarios. Moreover, discovering new attributes in real-world scenarios may require a brand-new dataset and retraining process. As business prefers to preserve data as much as possible, this is supported by system design which deals with a new data pattern and the least trade-off and penalty in performance.

This paper innovates a framework under the "Data Blind" concept – it indicates the knowledge over dataset and scenarios, which are unknown and undiscovered. Hence, the proposed framework performs data analytics for building machine learning models that serve multiple purposes. Overmind, a system conceptual model in distributed artificial intelligence, is named after its theoretical design. Overmind aims for managing dataset, discovering data pattern, and creating collaborative agents, as a complete framework. This paper experiments a demonstration of how it works, and steps, which are taken to gather results and contributions. An interpretation of behaviours and outcomes will be studied and presented in future works.

2. Previous works

The concept of Overmind has been forged base on previous machine learning algorithms and the decentralized systems. We also explore related works on feature transfer and selection. Thus, this section illustrates in two aspects, which are Multi-agent and Decentralized System, and Feature Transfer.

*Corresponding Author: Puttakul Sakul-Ung, King Mongkut's University of Technology North Bangkok Bangkok, Thailand, puttakul.s@gmail.com

2.1. Multi-Agent and Decentralized System

A term of Distributed Artificial Intelligence (DAI) has been defined based on the concept of distributed intelligent agents. Many projects and methodologies have been reported and summarized into categories proposed in [3] as system conceptual and agent conceptual models. The fine examples of DAI contributions were presented in [4–6] in the area of Intelligent Manufacturing System (IMS) since 1990. Then, a decentralized and distributed machine learning works were published in 1993 [7, 8]. The Cooperative Distributed Problem Solving (CDPS) that involves active agents sharing cooperative behaviours and protocols was invented in 1995 [9]. However, the cooperation is working based on broadcasting negotiation between a system and active agents by receiving the message to perform solving collaboratively.

Undeniably, this paper is highly influenced by the concept proposed in years ago. Although, machine learning has been continuously improved its capability and intelligence to the level that can resolve complex problems in the real-world scenario. Today, a decentralized reinforcement learning [10], which is again inspired by the multiagent system, describes problem statements that require large-scale real-world machine learning whereas the completed work focuses on two dimensions - Networked Evolution Strategies (NES), and Bayesian Optimality in the Wisdom of the Crowd (BOWC). NES describes network structure of communication between parallel-processing agents, that speed up learning process. This network structure of communication inspires Overmind's concept to integrate network analysis components and allow communication among decentralized machine learnings. Lastly, BOWC draws attention to the modelling on how humans learn from the influence of each other; this also inspires the Overmind concept for incorporating the knowledge management framework to mimic the collaboration in humans.

Decentralized Clustering in Large Multi-Agent Systems [11] draws a concept of using multi-agents in decentralized clustering, which is one of the major machine learning objectives. To overcome data clustering real-world problems such as data is widely distributed, data sets are volatile, and data instances also cannot be compactly processed, decentralization of multi-agent has proven its better performance against the traditional k-mean algorithm. This reviewed work hints decentralized clustering of a single dataset, perhaps, gives significant contributions in pre-data processing.

Recently in 2019, Feature Distributed Machine Learning (FDML) is studied collaboratively called FDML: A Collaborative Machine Learning Framework for Distributed Features [12]. This work inspires privacy preservation in feature engineering when one machine learning model can acquire prediction power by using additional features from other machine learning without sharing actual features. This works fully introduces the collaborative machine learning framework with high capability in transfer learning. The experiments are conducted using a different algorithm, and the results yield significance in real-world application development.

Nowadays, a concept of distributed and decentralized artificial intelligence has been widely reported, which shows upcoming era in machine learning integrated fields of research such as artificial intelligence in distributed ledger technology [13], artificial intelligence as IoT [14, 15], distributed security solution [16, 17], etc. Surprisingly, these integrations are all based on a distributed and decentralized concept.

2.2. Transfer Learning

In this sense, transfer learning refers to a process where discovered knowledge is fed to existing machine learning for reinforcement and adaptation over modelling that intents to re-train for better performance. Furthermore, the communication among models in a network is established for transferring particular knowledge from one machine learning to others [18]

Similarly, work in [19] introduces the transferring knowledge over different classes. This work also emphasizes transfer learning in feature selection for machine learning models using a proposed framework. The results of this work influents possibility of transfer learning in real-world dataset across classes and domains.

In sum, this paper draws conceptual designs based on previous works in two dimensions, which are distributed and collaborative systems, and transfer learning. However, the algorithm and process behind our work are different from the origins due to its' concepts and purposes. Ultimate research questions under our contributions are following: (1) how to establish an end-to-end platform for demonstrating collaborative machine learning models in a dynamic network in order to power up the prediction and (2) how to create a flexible system for transfer learning between models in a network that reflects the reinforcement.

3. Methodology

This work follows a quantitative research method. Firstly, the literature review is performed by limiting the scope of previous works and using keyword-searching. The results of the review are compounded into the conceptual design of the proposed framework, Overmind, which will be described in the next section of this paper. The reviewing of the proposed framework is conducted by researchers and authors. Experiment and testing are conducted on our previous work dataset, Thailand Air Quality Index (AQI) and weather forecast information. The result of applying this proposed framework is studied by comparison against the traditional framework on the same dataset, as presented in sections 5 and 6 of this paper.

4. Overmind: A Framework for Decentralized Machine Learnings

Overmind, a collaborative decentralized machine learning framework, is proposed considering dealing with anonymous dataset before building a machine learning model to serve purposes such as prediction, recommendation and classification. Figure 1 shows the main conceptual design of deploying multiple collaborative machine learning models (“agents”) each of which is responsible for a specific cluster of data. Collaboration means the agents can communicate and exchange knowledge, and share the workload for data processing.

This paper proposes Overmind as a collection of functional processes which are:

1. Determine Data Clustering

2. Assign Agent
3. Establish Network
4. Transfer Learning

In the following section, the aforementioned functional processes are explained in the respective order.

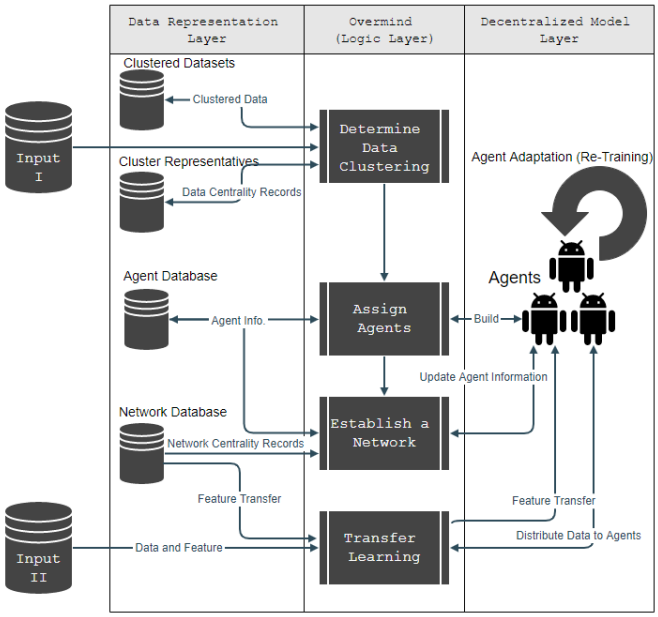


Figure 1: Overmind Architecture

4.1. Determine Data Clustering

Firstly, just the traditional approach, Overmind mimics the data cleansing process by using unsupervised learning algorithms to divide the dataset into clusters. This can be demonstrated as follows:

1. Let D be the training dataset where $D = \{d_1, d_2, d_3, \dots, d_n\}$, given that n is the number of instances.
2. Let F be the feature set, where $F = \{f_1, f_2, f_3, \dots, f_z\}$, given that z is the number of features in D.
3. Overmind deploys unsupervised learning method to divide D into clusters, given C is a set of clusters where $C = \{c_1, c_2, c_3, \dots, c_k\}$ and members in D are allocated and distributed over members of C, and k is the number of clusters.
4. The framework proposes s as centrality node of similarity graph of data under c_i. Hence, each cluster contains centrality node of data which is represented as s_i, where $i \leq k$

Figure 2 illustrates the output of determining data clustering and assigning agent.

4.2. Assign Agent

Agents are the supervised machine learning models trained using data in each cluster. Overmind determines the number of agents and builds a supervised machine learning algorithm to create agents, which can be explained as follows:

1. Let A be a set of agents trained by using C, given that a_i is trained by using data under c_i where $i \leq k$
2. Hence, a_i will be agent based on z features, given each feature contains weight/coefficient for particular a_i. Let W_i be a set of weight/coefficient of features for particular agent a_i

4.3. Establish Network

To build a collaborative framework, there shall be linkages between agents and rules for cooperation. This paper introduces a similarity network of agents based on the graph as below:

1. Let G be a graph of pairs $(V(G), E(G))$ where $V(G) = A$.
2. Similarity of agents in A is measured by W_i of particular a_i, given that y is a pre-determined threshold of similarity.
3. $E(G)$ will be created and linked between agent in $V(G)$ where the similarity measurement of a pair of agents satisfies y. This results in a network of agents.
4. Denote the centrality of G by H_G. If the centrality of G corresponds to multiple agents, Overmind selects one.

Eventually, Overmind creates decentralized machine learning as a collection of nodes in a network by using similarity measurement. Each node represents a supervised machine learning model.

assessed in term of similarity, which can be advanced for future works when one would like to measure affinity among trained models.

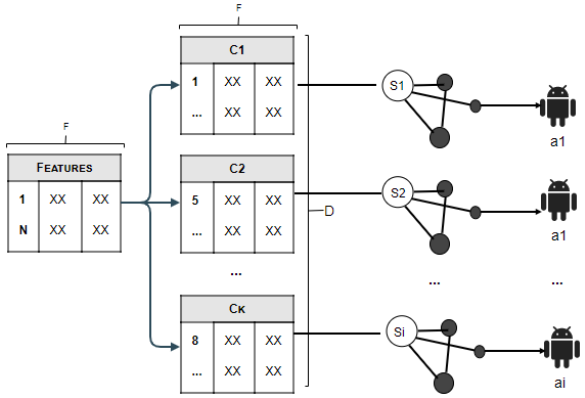


Figure 2: Illustration of Determining Data Clustering and Assigning Agent

4.4. Transfer Learning

This section describes two major transfer learnings of Overmind –data and feature transfers.

4.4.1 Data Transfer

Data transfer is involved when new data arrive for processing. Assumption on testing scenario when d_{n+1} is discovered and fed to framework, Overmind determines a suitable a_i for d_{n+1} by using s_i. This process implies similarity measurement on the centrality of clusters and new data instances. The steps are taken as below:

1. Let D' be a test dataset of this framework. Hence, $\{d'_1, d'_2, d'_3, \dots, d'_q\} = D'$.
2. Overmind determines each d'_x in D' and measures the similarity of d'_x to s_i.
3. Overmind assigns d'_x to a_i which has the highest similarity measured between s_i and d'_x. This step can be considered as

a batch process where D' is divided into subsets for feeding to particular a_i.

4. The accuracy of framework is measured as average accuracy of d'_x or average accuracy of subsets in D'.

4.4.2 Feature Transfer

Features and Feature Transfer are challenging components in designing the machine learning model. A traditional approach may require retraining of model when dealing with newly discovered features. A retaining process takes resources and time consuming for a big dataset. Yet, in the real-world scenario, some business may obtain a set of new information, both data and features, whether it contributes significances or not, but they have no willing to dispose of an existing dataset or pre-trained model.

In this sense, feature transfer refers to the ability that the existing trained model determines and deploy the discovered new features without retraining model; this means the transfer process will occur in every agent in the network. To elaborate flexibility in a dynamic machine learning model which accepts changes in both dataset and features, Overmind is conceptually designed to overcome the aforementioned challenges by described steps below:

1. When f_{z+1} is discovered within a new dataset D_{new}, Overmind predetermines a need of f_{z+1} by mimicking the supervised machine learning model of H_G and process upon D_{new} along with f_{z+1}. However, to avoid data volume bias, Overmind will proportionally random data in each c_i to form a dataset as equal as D_{new}.
2. Let accuracy of forming dataset is denoted by acc_{ext} and accuracy of model upon D_{new} is denoted by acc_{new}.
3. If $acc_{ext} > acc_{new}$, Overmind drops f_{z+1} from D_{new} and takes as data transfer process. This indicates that f_{z+1} cannot improve performance of existing system.
4. If $acc_{ext} \leq acc_{new}$, Overmind considers f_{z+1} as significant feature and process further steps for feature transfer process which is described below:
   4.1. Where the existing dataset lacks of f_{z+1} true values, Overmind creates imputation model by changing target label to f_{z+1} using forming dataset. This imputation model is denoted by a_{imp}.
   4.2. Overmind inserts f_{z+1} for each c_i in C, denoted by c'_i. This may result in extreme missing values in f_{z+1}. Hence, a_{imp} is used for imputation of missingness.
   4.3. Starting from H_G, agent a_i retrains its model along with f_{z+1}, the retrained agents are denoted by a'_i
   4.4. Retrained agents a'_i perform cross-validation and compare a result against a_i. There are two possible outcomes
      - a'_i gives a better performance, then f_{z+1} is accepted. a'_i adjusts W_i, denoted by W'_i, or else
      - a_i gives a better performance, then f_{z+1} is rejected. a'_i flags weight/coefficient of f_{z+1} as zero in W'_i
5. Overmind reshapes a network of agents, this creates $G' = (V(G), E(G'))$. Perhaps, centrality of network may be different from G, Overmind updates the new centrality, denoted by $H_{G'}$

Figure 3 shows the establishing of network of agents and transferring learning that impacts changes over a network.

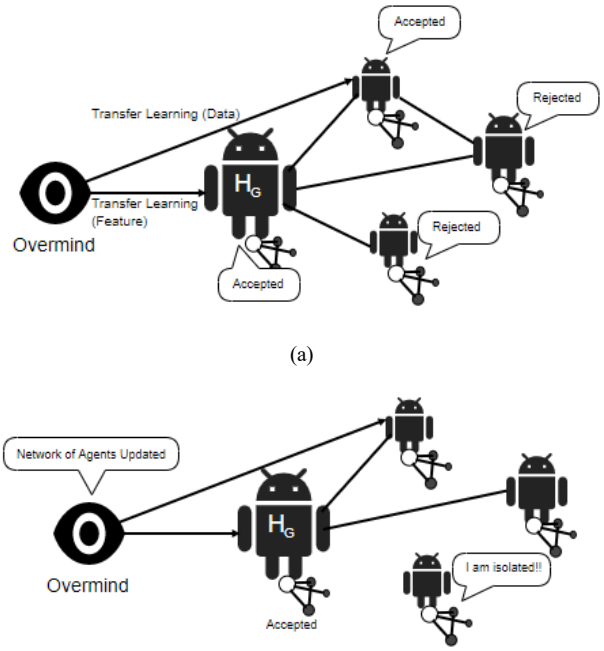


Figure 3: (a) Illustrating an established network with transfer learnings for both data and feature (b) Demonstrating a reshaped network resulted from transfer learning process.

Until now, a proposed framework creates a dynamic network which illustrates the transfer process among decentralized agents. Hence, this is called the collaborative decentralized machine learning framework.

5. Performance evaluation: a case study Thailand PM2.5 and weather forecast information database

To evaluate framework performance, this paper simulates the framework over the testing dataset, which is Thailand PM2.5 and weather forecast dataset presented in [1]. This dataset contains 2,383 instances with 21 features (including target). Overmind's simulation takes place in two functions, Framework Performance and Demonstration on Transfer Learning for Feature Transfer, to predict PM2.5 values from a given set of features.

5.1. Framework Performance

This paper deploys 80% of dataset for training and 20% for testing, for both traditional machining learning model –a single model –and Overmind.

In the training process, five uncorrelated features are removed from the dataset. Two algorithms, Gradient Boosted Trees (GBT) and Neural Net (NN), are used in this paper because the previous work has described Thailand PM.5 prediction using these algorithms. The set of tuning parameters is left by default using RapidMiner 9.6. For GBT, number of trees = 50 with learning rate = 0.01. For NN, training cycle = 200 with learning rate = 0.01. In addition, 10 folds cross-validation with automatic sampling type is used.

Throughout this paper, gradient color codes are used to indicate how good the results are. Green-colored fields indicate the higher accuracy and correlated values and red-colored fields indicate the opposite.

The clustering and agents training results show in Table 1

Table 1: Distributed Agents in Training Process

Agents	Train Dataset (0.8, 1906, F=15)		
	Data Instances (N)	GBT (RMSE)	NN (RMSE)
$D: a_s$	1906	13.50	7.21
$c_0: a_0$	1162	6.56	5.25
$c_1: a_1$	256	17.14	9.62
$c_2: a_2$	488	12.28	7.00
s: single traditional model $C = \{c_0, c_1, c_2\}$ $Distributed\ Agents, A = \{a_0, a_1, a_2\}$			

Table 1 proves a better performance, measured by Root Mean Square Error (RMSE), on agents, a_0 and a_2 than on a traditional single model, a_s, in both algorithms. It may indicate the abnormal data pattern in c_1 which causes variation of performance for a traditional model and assumption of removing these instances are not acceptable by business scenario. In other words, 13.43% of the instances drop performance of agent for from RMSE 6.56 to 13.50, and it seems to be unacceptable tradeoffs.

Establishing a network of agents, we explore the weights of features based on GBT variable importance percentage as shown in Table 2

Table 2: Variable Importance in agents based GBT

Variable Importance	$D: a_s$	$c_0: a_0$	$c_1: a_1$	$c_2: a_2$
AQI	0.762	0.689	0.123	0.282
CO	0.114	0.004	0.683	0.202
PM10	0.075	0.139	0.069	0.112
O3	0.017	0.003	0.041	0.018
PRESSURE	0.009	0.095	0.035	0.002
PRECIP	0.007	0.000	0.013	0.000
HUMIDITY	0.006	0.020	0.019	0.008
CLOUD	0.005	0.018	0.007	0.008
SO2	0.005	0.011	0.000	0.357
NO2	0.000	0.001	0.003	0.004
WIND	0.000	0.011	0.001	0.001
GUST	0.000	0.001	0.001	0.003
TEMP	0.000	0.006	0.005	0.003
TEMP_FEEL	0.000	0.002	0.000	0.000
TIME_POINT	0.000	0.000	0.000	0.000

By examining variable importance in Table 2, both of three highest importance features and lowest importance features were examined giving a clue that the criteria as feature weight on the prediction of PM2.5 differ for each agent. The data in each cluster should require a specifically trained agent.

Recall, an establishing network of agents based GBT algorithm, Overmind determines the similarity measurement of agents using variable importance. A Euclidean Distance Measurement is used to find the distance among agents. In Table 3, the similarity is measured by using the equation:

$$Similarity(s, d) = 1 - Distance(s, d) \quad (1)$$

where s is source node, d is destination node.

Table 3: Agent Similarity Measurement

Source	Destination	Distance	Similarity
$D: a_s$	$c_0: a_0$	0.17	0.83
$D: a_s$	$c_1: a_1$	0.86	0.14
$D: a_s$	$c_2: a_2$	0.60	0.40
$c_0: a_0$	$c_1: a_1$	0.89	0.11
$c_0: a_0$	$c_2: a_2$	0.58	0.42
$c_1: a_1$	$c_2: a_2$	0.62	0.38

Setting a threshold $y = 0.3$ means the edges will be created when the similarity between nodes is higher or equal to 0.3. However, this threshold is an adjustable parameter, the lower the threshold the higher chance for connected agents, the similarity of agents affects and shapes the transfer process. Where connection exists, a pair of the agent can communicate and transfer both data and feature, in another hand, isolated agent requires specially and specifically consideration. In this study, 0.3 is represented for the demonstration that creates linkages between agents unless the transfer process cannot be illustrated. Considering a tradition model, an agent in a network will results in Figure 4, however, this network will be examined again in the feature transfer section.

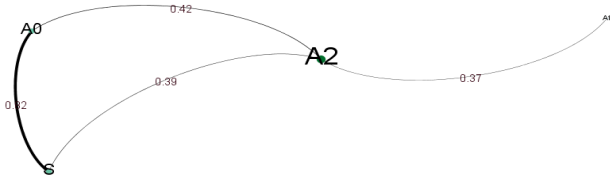


Figure 4: Network of Agents (|F|=15, y =0.3)

In the testing process, 477 instances of the dataset were distributed to specific agents based on the data similarity measurement. Moreover, the prediction assessments on both Gradient Boost Tree (GBT) and Neural Network (NN) were shown in Table 4.

The results show high accuracy when distributed agents perform a prediction over the associated clustered dataset. Undeniably, the traditional model is flexible and stable for input data. Perhaps, it should be considered as a tradeoff. We calculate average performance for distributed agents as:

$$Acc = \frac{\sum_0^i RMSE(a_i, c_i)}{i} \quad (2)$$

where $RMSE(a_i, c_i)$ is an accuracy of prediction by a_i on dataset c_i

Table 4: Test Results

Agent	Trained By	Test Dataset (0.2, n=477)					
		GBT (RMSE)			NN (RMSE)		
		c_0	c_1	c_2	c_0	c_1	c_2
		294	55	128	294	55	128
a_s	477	8.28	22.87	11.25	4.29	8.69	6.99
a_0	294	5.80	36.46	18.51	3.99	22.41	12.34
a_1	55	38.64	17.04	21.40	12.30	7.80	12.04
a_2	128	19.32	22.34	11.00	13.31	15.36	7.23

The GBT has achieved average performance RMSE of 11.28 whereas the traditional model has obtained RMSE of 14.13. On the other hand, the NN achieve the RMSE of 6.34, while the traditional model achieves at 6.65. These performances described in Figure 5.

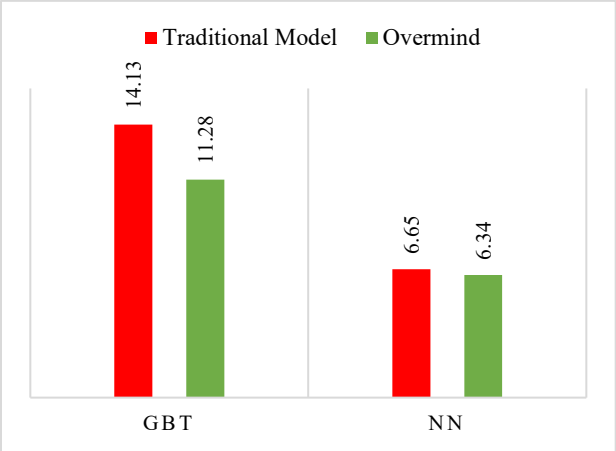


Figure 5: Average Performance, measured in RMSE.

5.2. Demonstration on Transfer Learning for Feature Transfer

In this demonstration, one correlated feature from a training dataset is removed, but a testing dataset remains with full features. Based on variable importance in Table 2, AQI is assumedly removed from the training dataset. The results of newly trained agents are shown in Table 5.

Table 5: Distributed Agents in Training Process, |F|=14

Agents	Train Dataset (0.8, 1906, |F|=14)		
	Data Instances (n)	GBT (RMSE)	NN (RMSE)
D: a_s	1906	14.28	9.97
c_0: a_0	553	15.72	11.96
c_1: a_1	234	16.20	12.45
c_2: a_2	1119	8.17	7.05

The performance of two minor clustered (n =553, and n = 234) significantly drops. However, it is not yet comparable results where |F|=15 because the clustering process may yield a different output. Base on this experiment, a same threshold value, $y = 0.3$ will result isolated a_2 which indicates that the data in c_0 and c_1 should be removed from the demonstration and only one agent required in this scenario, it is a_2; the demonstration will be then bias, hence, we lower a threshold to $y = 0.2$ and the result of networks of agents shows in Figure 6.

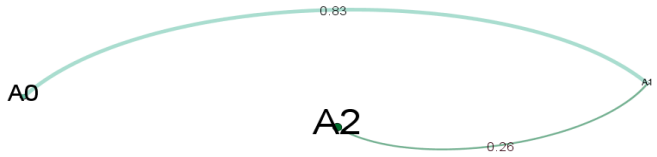


Figure 6: Network of Agents (|F|=14, y =0.2)

Figure 6 shows that a centrality of network is a_1, this means centrality of the network is determined neither by agent performance nor numbers of instances. The centrality of the network is responsible for sharing and transferring the discovered both data and feature.

Demonstration on a discovered new dataset and feature (n=477, |F|=14+1), assumedly, AQI is a new feature. Overmind firstly determines the need for AQI data if it can improve the performance of the overall agent or not. Overmind takes sampling proportional data from clusters in a network, then building a model to evaluate performance against a new dataset.

Table 6: Comparing Results for a Transfer Learning

Agent	No. of instance (n)	No. of feature (|F|)	GBT (RMSE)	NN (RMSE)
Existing Dataset: acc_{ext}	477	14	14.56	10.12
New Dataset: acc_{new}	477	14+1	12.43	7.90

Table 6 shows a better performance with a new feature, hence, a transfer learning should be accepted for the new feature, AQI.

An existing dataset in a network contains no value of AQI, and the business scenario does preserve existing dataset. Hence, an imputation model is built by using a new dataset for training process (n=477, |F|=14+1). Eventually, an existing dataset is imputed for a new feature, AQI; it is deployed to the centrality of the network, all agents accept AQI as a new feature because the comparison identifies AQI can improve performance as shown in Table 7.

Table 7: Distributed Agents in Training Process, |F|=14+1 (imputed AQI)

Agents	Train Dataset (0.8, 2383, |F|=14+1 imputed AQI)		
	No. of instances (n)	GBT (RMSE)	NN (RMSE)
D': a'_s	1906	13.39	6.85
c'_0: a'_0	553	14.19	7.10
c'_1: a'_1	234	14.55	9.55
c'_2: a'_2	1119	7.03	4.71

Overmind update a network of agents based on changing in similarity measurement among agents.

A re-connected nodes between a'_2 and a'_0 is formed, this results in a fully connected network, and the centrality of the network is shifted to a'_2 because it has the highest performance and controlling a majority of data in the network (N=1119). Overmind then continues data transfer process as described in (A). The data prediction results are shown in Figure 7.

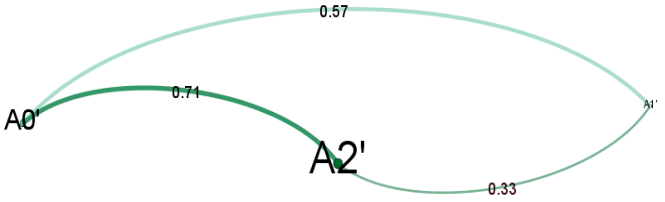


Figure 7: Network of Agents (F=14+1, Y=0.2)

Table 8: Test Result after feature transfer process

Agent	Test Dataset (0.2, 477, |F|=14+1)						
	No. of instance	GBT (RMSE)			NN (RMSE)		
		c_0	c_1	c_2	c_0	c_1	c_2
a'_s	477	14.44	17.02	11.92	7.11	13.11	6.36
a'_0	162	13.25	18.16	17.23	7.07	16.9	11.79
a'_1	37	19.73	13.61	25.73	19.81	10.38	15.64
a'_2	278	18.39	21.18	13.22	11.64	16.21	8.62

The comparison shows average performance in RMSE using GBT with 13.36 in this testing dataset, while a traditional model would give an accuracy, 14.46. For NN, the average performance in RMSE is 8.69, and 8.86 for a traditional model. This comparison is described as shown in Figure 8.

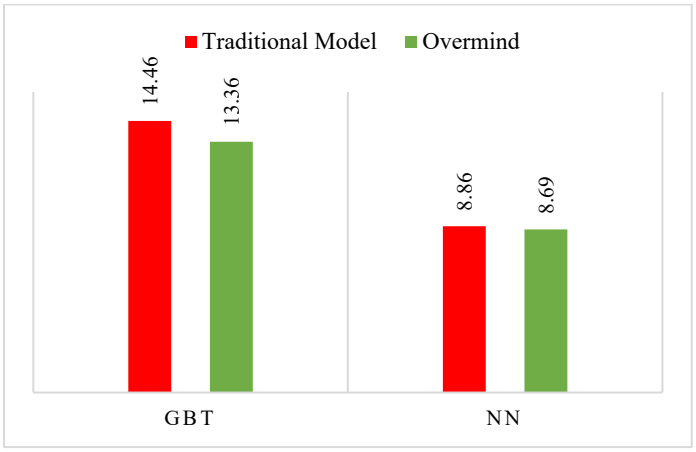


Figure 8: Average Performance, measured in RMSE after feature transfer process

Figure 8 shows penalty in transfer process caused by the imputation model that increases the RMSE after transfer process as shown in Figure 9. However, it should be reminded that demonstrating feature transfer is taken place on the important variable which is the AQI.

To verify the consistency of this approach, the progression of the framework functional processes, Determine Data Clustering, Assign Agent, and Establish Network has been performed. However, in the second round, the dataset is slightly different because (1) 1,906 instances (80%) of training dataset have imputed AQI values, and (2) 477 instances (20%) of the testing dataset have true values of AQI. To avoid confusing, the agents in the second execution is denoted by $a_{x:2}$ where x is a number of clusters produced by this execution.

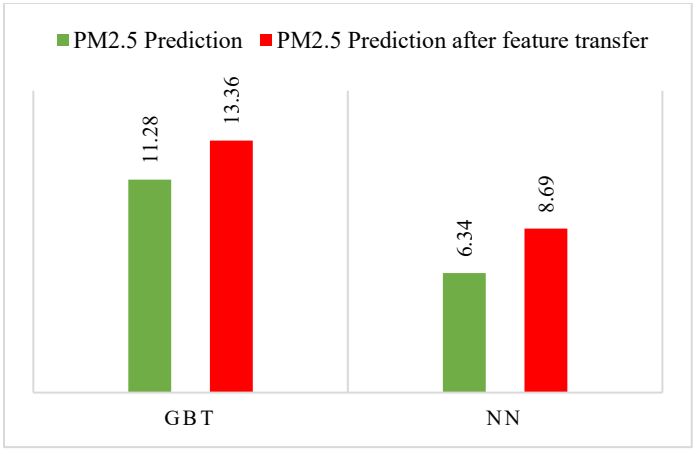


Figure 9: Penalty after feature transfer process

Table 9: Distributed Agents in Training Process by Second Execution

Agents	Train Dataset (1, 2383, |F|=15)		
	No. of instances (n)	GBT (RMSE)	NN (RMSE)
$D: a_{s:2}$	2383	13.05	6.95
$c_0: a_{0:2}$	1458	6.24	4.22
$c_1: a_{1:2}$	316	17.11	11.27
$c_2: a_{2:2}$	609	11.90	6.44
s: single traditional model $C = \{c_0, c_1, c_2\}$ $Distributed\ Agents, A = \{a_{0:2}, a_{1:2}, a_{2:2}\}$			

The result in Table 9 shows consistency over the clustering process which still produce three clusters with similar proportions. Moreover, the overall performance of agents is consistency and even slightly improved. Examining agent similarity using variable importance between the first and second executions are shown in Table 10 and Figure 10. The comparison in Figure 10 shows changes in variable importance, but the agents are keeping their characteristics over the same set of features.

Eventually, a network of agents is re-established as shown in Figure 11; a network of agents also preserves its structure where the centrality of network remains unchanged. However, if the similarity threshold is recalled in the very first demonstration of establishing a network where $y = 0.3$, the isolated agent $(a_{1:2})$ may occur which is assigned to the minority of data (n=316) that contribute poor performance.

Table 10: Variable Importance in agents based GBT by Second Execution

Variable Importance	$D: a_{s:2}$	$c_0: a_{0:2}$	$c_1: a_{1:2}$	$c_2: a_{2:2}$
AQI	0.762	0.689	0.123	0.282
CO	0.114	0.004	0.683	0.202
PM10	0.075	0.139	0.069	0.112
O3	0.017	0.003	0.041	0.018
PRESSURE	0.009	0.095	0.035	0.002
PRECIP	0.007	0.000	0.013	0.000
HUMIDITY	0.006	0.020	0.019	0.008
CLOUD	0.005	0.018	0.007	0.008

Variable Importance	$D: a_{s:2}$	$c_0: a_{0:2}$	$c_1: a_{1:2}$	$c_2: a_{2:2}$
SO2	0.005	0.011	0.000	0.357
NO2	0.000	0.001	0.003	0.004
WIND	0.000	0.011	0.001	0.001
GUST	0.000	0.001	0.001	0.003
TEMP	0.000	0.006	0.005	0.003
TEMP_FEEL	0.000	0.002	0.000	0.000
TIME_POINT	0.000	0.000	0.000	0.000

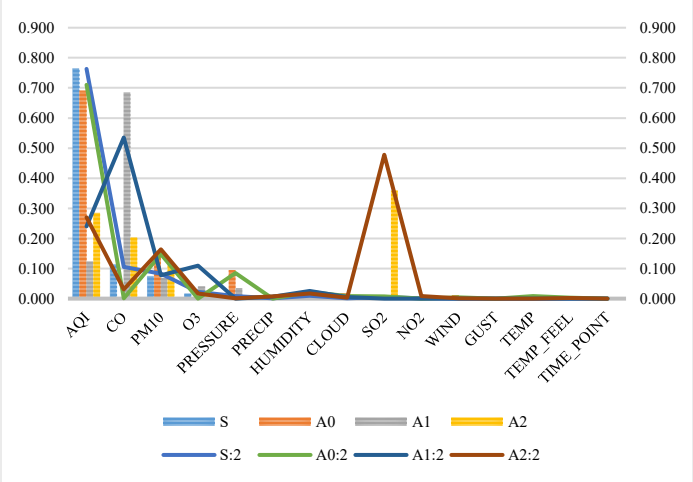


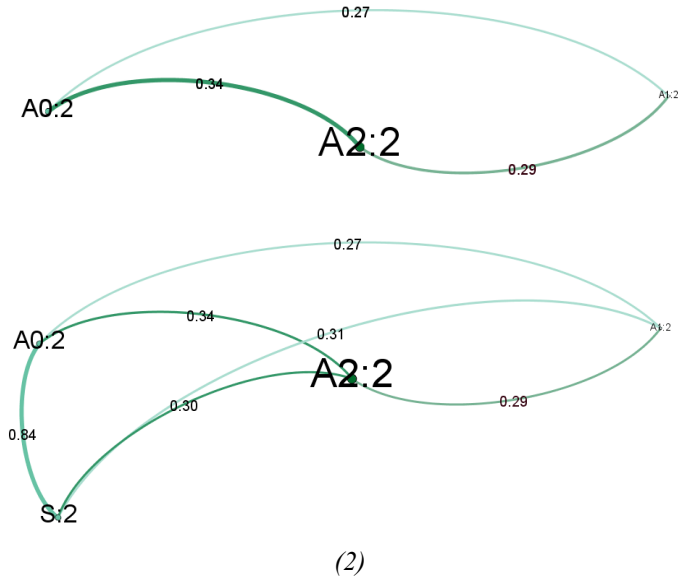
Figure 10: Comparison of Agent Variable Importance based GBT by first and second executions

This scenario implies that the business may consider, (1) a separate studying of agents in order to build a specific algorithm for prediction over the associated dataset, (2) removing associated data instances out of dataset in order to reduce inconsistency and improve prediction performance, (3) reducing a threshold (y) in order to keep particular agent connected to a network because this agent can be a specific dataset where a business foresees this data will be increasing over time in soon future.

6. Results and Discussion

Till now, demonstrations of both framework performance, and the transfer process are explained. The summary of the performance of Overmind with four comparing scenarios in cross-validation agents' average performance can be concluded as follows:

1. Scenario I: Overmind assigns distributed agents on a fully completed dataset.
2. Scenario II: Overmind assigns distributed agents on a dataset where importance attributes is removed.
3. Scenario III: Overmind assigns distributed agents to demonstrate a situation where a new feature is discovered and transferred, and
4. Scenario IV: Overmind assigns retrained distributed agents and restates its network



(2)

Figure 11: (1) Network of Agents in Second Execution (F=15, Y=0.2) without Traditional Model, (2) Network of Agents in Second Execution (F=15, Y=0.2) with Traditional Model,

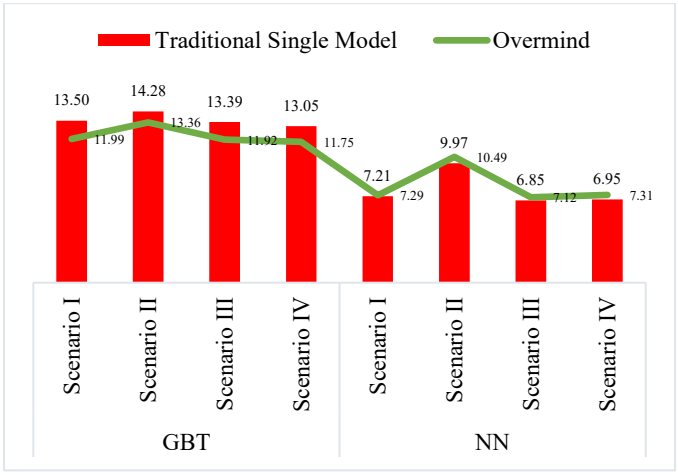


Figure 12: Comparison of Average Performance in Cross-Validation Agent Generated Models measured in RMSE

Figure 12 explains, for all scenarios, the cross-validation performance in Overmind generated models is impressive compared to a traditional single model under GBT algorithm. In the other side, the performance is not satisfied when compared to the traditional model under NN algorithm.

To summarize testing performance in testing data, we divide into two scenarios, (1) Scenario I: Overmind assigns distributed agents on a fully completed dataset, (2) Scenario II: Overmind assigns distributed agents to demonstrate a situation where a new feature is discovered and transferred. The results are shown in Figure 13.

Figure 13 indicates a better testing performance in all scenarios, but the performance is slightly better in the NN algorithm. In sum, Overmind can achieve its primary intended design and concept, this achievement can be demonstrated through our previous dataset for PM2.5 prediction in Thailand.

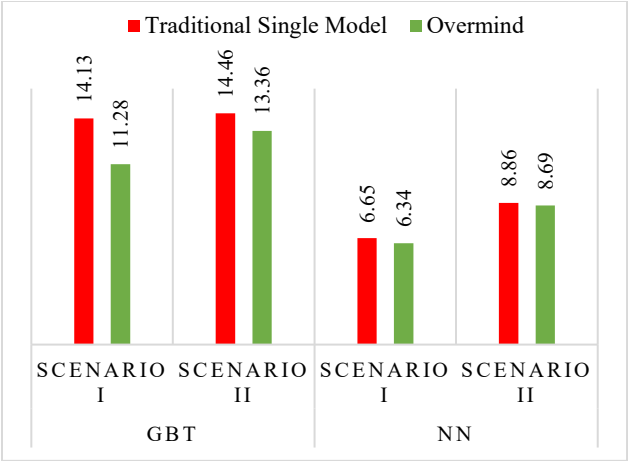


Figure 13: Summary and Comparison on Testing Performance measured in RMSE

7. Conclusion

Overmind is a framework aiming to demonstrate and facilitate a decentralized machine learning-based network analysis. This paper proves the concept and its performance through the mixed dataset between Thailand PM2.5 dataset and weather forecast dataset.

In data prediction, this framework can significantly reduce the data bias scenario, where input data is uncontrollable, variant, and clustered. Moreover, Overmind has proven its performance by comparing an average accuracy measured in RMSE on both traditional models (single machine learning) and distributed agents (multiple machine learning, agents). Overmind performs better performance than the traditional method in selected dataset using the GBT algorithm, reducing the RMSE from 14.13 to 11.28 (20.16% decrease). Furthermore, the RMSE is slightly reduced from 6.65 to 6.34 (4.6% decrease) in the NN algorithm.

A feature transfer process is demonstrated as a conceptual design for data preservation. However, the evaluated performance yields penalty due to imputation method.

In sum, Overmind has successfully formed a collaborative decentralized machine learning framework serving as an alternative in the design of a machine learning system. Possible contributions of Overmind are:

7.1. Higher Performance and Unaffected by Data Bias.

Overmind proves its effectiveness as a prediction system in the selected testing dataset. The design of Overmind is not affected by data bias because distributed agents are designed to function over associated sub-dataset.

7.2. Distributed Resource Utilization

Overmind mimics the concept of a distributed system so that it inherits the advantage where resources are allocated in a decentralized manner. Each agent utilizes computational resources for processing associated input instead of processing the whole dataset. This contribution also involves parallel processing and queue system where particular agents may transfer their queued inputs, as a bottleneck, to similar connected agents in a network for parallel processing where penalty inaccuracy can be traded off. This parallel processing for Overmind is already considered and planned for future work.

7.3. Dynamic Models in Collaborative Manner

Overmind demonstrates dynamic machine learning models for feature discovery and transferring across collaborative network analysis with data preservation. The concept of similarity measurement and network of agents has innovated methods in a dynamic framework. There are still opportunities in exploring the aforementioned methods to study optimization parameters and techniques.

8. Future Work

A study on optimization is recommended on Overmind's parameters. An interpretation over a network of agents formed by Overmind is an interesting topic to be explored and described how it contributes significances to business viewpoints.

In addition, the enterprise system architecture of Overmind can be improved and enhanced for higher performance and resilience such as queueing system and redundancy in agent level For example, when one agent is in abnormal stage, data can be continuously processed by neighbors or connected agents in the network; this will reduce the impact of disruption and error in artificial intelligence system.

Acknowledgment

We would like to thank the members of Faculty of Information Technology and Digital Innovation, King Mongkut's University of Technology North Bangkok, for their guidance and suggestions. Importantly, we would also like to thank Bangkok Air Quality for supporting data.

References

[1] P. Sakul-Ung, P. Ruchanawet, N. Thammabunwarit, A. Vatcharaphrueksadee, C. Triperm, M. Sodanil, "PM2.5 Prediction based Weather Forecast Information and Missingness Challenges: A Case Study Industrial and Metropolis Areas," in 2019 Research, Invention, and Innovation Congress (RI2C), 1–5, 2019,doi:10.1109/RI2C48728.2019.8999941.

[2] K.F. Li, "Engineering Distributed Intelligence: Issues and Challenges," in 2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems, 404–408, 2014, doi:10.1109/CISIS.2014.56.

[3] L. Eduardo, C. Hern, "On distributed artificial intelligence," The Knowledge Engineering Review, **3**(1), 21–57, 1988, doi:10.1017/S0269888900004367.

[4] G.M.P. O'Hare, A.W.J. Chisholm, "Distributed Artificial Intelligence: An Invaluable Technique for the Development of Intelligent Manufacturing Systems," CIRP Annals, **39**(1), 485–488, 1990, doi:https://doi.org/10.1016/S0007-8506(07)61102-0.

[5] B. Chaib-draa, "Industrial applications of distributed AI," **38**(11 %J Commun. ACM), 49–53, 1995, doi:10.1145/219717.219761.

[6] S. Kirn, "Organisational intelligence and distributed AI," No. 40. Arbeitsberichte des Instituts für Wirtschaftsinformatik, Westfälische Wilhelms-Universität Münster, 1995.

[7] Y. Feng-Chao, H. Yu-Kuen, "Design of a distributed artificial intelligence system shell," in Proceedings of TENCON '93. IEEE Region 10 International Conference on Computers, Communications and Automation, 192–195 vol.1, 1993, doi:10.1109/TENCON.1993.319961.

[8] S. Sen, "Multiagent systems: Milestones and new horizons," Trends in Cognitive Sciences, **1**(9), 334–340, 1997, doi:10.1016/S1364-6613(97)01100-5.

[9] G. Weiß, L. Steels, "Distributed Reinforcement Learning," Springer Berlin Heidelberg, Berlin, Heidelberg: 415–428, 1995.

[10] D. Adjodah, Decentralized Reinforcement Learning Inspired by Multiagent Systems, Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, 1729–1730, 2018.

[11] E. Ogston, B. Overeinder, M. van Steen, F. Brazier, "A method for decentralized clustering in large multi-agent systems," in Proceedings of the second international joint conference on Autonomous agents and multiagent systems - AAMAS '03, ACM Press, New York, New York, USA: 789, 2003, doi:10.1145/860575.860702.

[12] Y. Hu, D. Niu, J. Yang, S. Zhou, FDML: A Collaborative Machine Learning Framework for Distributed Features, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2232–2240, 2019, doi:10.1145/3292500.3330765.

[13] K.D. Pandl, S. Thiebes, M. Schmidt-Kraepelin, A. Sunyaev, "On the Convergence of Artificial Intelligence and Distributed Ledger Technology: A Scoping Review and Future Research Agenda," IEEE Access, **8**, 57075–57095, 2020, doi:10.1109/ACCESS.2020.2981447.

[14] Y. Lin, Y. Lin, C. Liu, "AItalk: a tutorial to implement AI as IoT devices," IET Networks, **8**(3), 195–202, 2019, doi:10.1049/iet-net.2018.5182.

[15] X. Lin, J. Li, J. Wu, H. Liang, W. Yang, "Making Knowledge Tradable in Edge-AI Enabled IoT: A Consortium Blockchain-Based Efficient and Incentive Approach," IEEE Transactions on Industrial Informatics, **15**(12), 6367–6378, 2019, doi:10.1109/TII.2019.2917307.

[16] H.F. Rashvand, K. Salah, J.M.A. Calero, L. Harn, "Distributed security for multi-agent systems - review and applications," IET Information Security, **4**(4), 188–201, 2010, doi:10.1049/iet-ifs.2010.0041.

[17] X. Zhao, R. Jiang, "Distributed Machine Learning Oriented Data Integrity Verification Scheme in Cloud Computing Environment," IEEE Access, **8**, 26372–26384, 2020, doi:10.1109/ACCESS.2020.2971519.

[18] S.J. Pan, Q. Yang, "A Survey on Transfer Learning," IEEE Transactions on Knowledge and Data Engineering, **22**(10), 1345–1359, 2010, doi:10.1109/TKDE.2009.191.

[19] Z. Wang, X. Ye, C. Wang, P.S. Yu, "Feature Selection via Transferring Knowledge Across Different Classes %J ACM Trans. Knowl. Discov. Data," **13**(2), 1–29, 2019, doi:10.1145/3314202.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 290-298 (2020)

www.astesj.com

Special Issue on Innovative Research in Applied Science, Engineering and Technology

ASTESJ

ISSN: 2415-6698

The Role of Promotion in Mobile Wallet Adoption – A Research in Vietnam

Ha Hoang[1,*], Tan Trinh Le[2]

[1]*University of Economics, The University of Danang, Da Nang 550000, Vietnam*

[2]*Business department, FPT University, Da Nang 550000, Vietnam*

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 16 October, 2020
Online: 10 November, 2020

Keywords:
Mobile wallet
Mobile payment
Digital economy

ABSTRACT

Mobile wallet is an application that allows users to make online payments using their mobile phones. In the context of an outbreak of COVID-19 epidemic in Vietnam and the world, mobile wallets are considered having many opportunities to change people's cash spending habits. The current study assesses the factors impact on the intention to use the mobile wallet, focusing on understanding the relationship of promotion with common factors in technology adoption research such as perceived risk, perceived usefulness, habits, social influences. The research results also show the direct and indirect effects of promotion on intent to use. From this result, 7 of 9 research hypotheses were accepted. Promotion, often overlooked by researchers, has been proven to be a critical factor when researching technology adoption because it significantly improves the level of intentional interpretation of the research model. This study will be the premise for future researches when studying adoption technology, researchers could integrate promotion to the model to achieve a significant improvement in the level of interpretation as showed in this study.

1. Introduction

PricewaterhouseCoopers (PwC) survey in 27 countries recorded Vietnam as the fastest growing mobile payments market in 2018, with the percentage of users increasing from 37% to 61% [1]. Facing increasingly complicated developments of the COVID-19 epidemic, the Vietnam's Government has also recommended people to limit cash payment and increase the use of online payment services to decrease the risk of infection. As a result, cashless payment methods, including mobile wallets, have a great opportunity to become more and more popular. In Vietnam, the ratio of non-cash payments to the total means of payment is only 14%, so the scope for growth of mobile wallets is huge. This context has created many favorable conditions to help mobile wallets develop and increasingly change the cash using habits of most Vietnamese people, actively contributing to bringing Vietnam towards the cashless economy [2].

Mobile payment means all payment services are made through mobile devices such as laptops, tablets or smartphones. Among them, mobile wallets are the latest e-commerce method that helps customers to make payment, shopping online, booking and share services. In terms of technology, a mobile wallet is an application installed in a smartphone or a tablet which allows consumers to input money and make online payments directly with the mobile wallet [3]. Mobile wallets can also be recognized as the revolution derived from e-wallets. Owning a wide ecosystem is a way of integrating into the digital transformation stream of users, especially young people who love technology. Because of the nature of mobile wallets, it is convenient, fast and secure in payment for daily activities, reducing the risks associated with spending and managing cash.

Although the application of technology has received long-standing attention from the government, it has not developed to achieve the desired scale, required for application in the entire systems payments in Vietnam. Vietnamese clients still prefer to use cash, reflected in the fact that there are still nearly 90% of transactions are paid by cash. Until now, almost no scholars doing research on mobile wallet in Vietnam whereas Vietnam is one of fastest growth in Asia area and is considered as a very attractive market for investors in all fields, especially technological payment. This study was carried out to fill the research gap in Vietnam related to mobile wallet.

This research aims to explore the factors affecting the intention to use of mobile wallet in Vietnam. The results are expected to provide the suggestions for mobile wallet suppliers to reach customers more effectively, and policymakers to drive a cashless economy in Viet Nam. In addition, the role of promotion demonstrated in this study will be the premise for future researches

*Corresponding Author: Ha Hoang, hahoang@due.edu.vn

when studying the adoption of technology, researchers could integrate promotion to the model to achieve a significant improvement in the level of interpretation as showed in this study.

We divided the content of this paper into five parts: The first part focuses on statement problem and meaning of this study comparing previous researches. Part two presents theories related study topic and constructing framework model. Besides that, authors have description research method and sampling characteristic in part three. To test the relationship between factors in the proposed model, we have showed analysis and results in part four and the last part is summary research results and implications.

2. Theoretical background and research framework

2.1. Theoretical background

Vietnam belongs to the group of countries with a high percentage of internet users (70.3%), equivalent to 68.5 million Internet users in 2019. With 43.7 million smartphone users (accounting for 45% of the population in 2019), Vietnam are at the regional average, higher than India, the Philippines, Indonesia and Thailand. Many advanced technologies are being applied such as biometric authentication such as fingerprint, face; quick response code (QR Code); Tokenization of information[4].

The development of technology, the popularity of the internet and smartphones has progressively increased the demand for cashless and digital transactions around the world. Customer attitudes and acceptance regarding mobile payments have also experienced a drastic change [5]. Many studies around the world have also confirmed that customers like technologies that can provide fast, useful services on a mobile platform. Mobile payment services, in which mobile wallets are one of the latest advances in the digital economy, have integrated these features [6, 7, 8].

Mobile wallet is one technology that have many growth opportunities in this condition [9]. Mobile wallets could replace physical wallets and even debit or credit cards in online and direct purchases. We see this technology as a major revolution in the digital economy because in the era of technological revolution, the issues of speed, interoperability, security and privacy of mobile technology have been resolved. In our study, we proposed and tested a research framework model which combined constructs in the TAM and UTAUT models, expanding with new variables for a transitional country like Vietnam such as promotion, a construct built from practice.

Promotion

Promotion is considered as a short-term marketing strategy and used to create awareness and interest in products or services. It helps the companies to attain sales and marketing goals [10]. Moreover, promotion could affect to the consumer's mind like a benefit to him/her, then creating the changes in consumer behavior [11]. Many studies have also confirmed the effect of sales promotions on consumer attitudes and consumer behavior [12, 13, 14]. As a result, promotion is a very useful strategic mean to improve profitability. Sellers are using various promotion strategies, for example coupons, product upgrades, price reduction, free samples and free gifts to increase intention to buy of customers. Promotion can be separated into non-financial and financial promotions [15]. The examples of financial promotions are discount of prices, coupons and vouchers while promotions related non-monetary include bonus products and gifts. The value of the promotion is also perceived relatively [16]. Customers could recognize a monetary promotion as a decline in losses because of this kind of promotion brings a reduction in the purchase cost that consumers have to pay. In contrast, promotion of non-monetary sales could be considered as a profit from the deal [17].

Related to the financial and payment sector, [18] showed the influence of sales promotion to persuade customers to open a bank account [18]. The results of this research showed that 50% of the increase accounts of a bank is the result followed a promotional campaign. Other researches showed the efficiency of promotion to influence the computer purchase and confirmed the positive linkages, promoting the behavior of purchase of financial services [19, 20].

Recent research on omnichannel shopping based on IT has included the promotion in the research model but has not found a meaningful impact of the promotion on intention to use [21].

In Vietnam, a recent study on the determinants of the choice of the customer mobile wallet, promotion in second place in the six elements outlined in the survey. As emphasized by several authors, there are academic and managerial deficiencies in the profound understanding of the relationship between promotion and consumer behavior [22, 23]. Especially in technology adoption, sales promotion still receives very little attention from researchers. That is why the construct of promotion is added to this study.

H1: The promotion of mobile wallet will positively affect the customers' perceived usefulness of this technology.

H2: The promotion of mobile wallet will negatively affect the customers' perceived risk of this technology.

H3: The promotion of mobile wallet will positively affect the social influences of this technology.

H4: The promotion of mobile wallet will positively affect the habit of customers for this technology.

Perceived usefulness

Perceived usefulness was mentioned for the first time by Davis in Technology Acceptance Model (TAM). This term refers to the ability of customers to improve performance by fully using the improved system, reflecting users' willingness to accept [24].

In technology adoption research, performance expectancy refers to an individual's perception that using a technology could provide benefits to customers in performing certain activities [25]. Reflecting a range of attributes that a technology could give benefits to clients, performance has been conceptualized by using system features that could enhance speed, productivity, and chances of task accomplishment and perceived usefulness [25, 26]. Unambiguously, in diverse task settings, performance expectancy was affirmed to affect intentions to use technological systems [27]. Consumers' perception that using mobile wallet would support them to achieve benefits in doing payment tasks could influence the behavioral intention of mobile wallet adoption. The following hypothesis was formulated:

H5: Perceived usefulness will positively affect the user's intention to use mobile wallet.

Perceived risk

Perceived risk was a popular construct used in extended models of UTAUT2 by several researchers. Perceived risk is considered negatively affect on intention to use [8]. Featherman and Pavlou also posited that individual's perceived risk of using a technology has negative influence on perceived usefulness (equivalent to performance expectancy) and behavioral intention toward that technology while his or her perceived ease of use (equivalent to effort expectancy) also has an adverse impact on the perceived risk [28]. Since mobile wallet is a new technology and includes complex and sensitive private information, perceived risks could be a difficulty in adopting this technology. Consumers could face monetary loss because of their improper operating system inaccuracy or potential fraud. As a result, their non-refundable paying money would be sent to wrong receivers or scammers or just be disappeared.

Perceived risk diminishes user's readiness and acceptance of new technology to more security dangers of technology like mobile wallets [7]. Perceived risk could make customers underestimate about the perceived usefulness of a technology and deny accepting it [29]. Some researches showed that perceived risk positively affect on user's intention ([30]; [31]). Two studies in 2015 further explored factors such as social influence, perceived risk and stressed their important role in mobile payment adoption [32, 33]. The following hypotheses were developed:

H6: Perceived risk will negatively affect the intention to use mobile wallet.

Social influence

Social influence is defined as the degree to which an individual perceives that essential referents such as family and friends believe he or she should adopt a specific technology [25].

Recently, in mobile payment adoption, many researches integrated social influences into their research frameworks and obtained some empirical support [34, 35, 36]. A study of Chong et al. (2010) showed that a customer considers the thoughts of their friends and their family before choosing a new technology [37]. If views of others are not favourable, they will resist adopting that technology. Social influence is also found as the most dominant on user's behavioural intention in mobile payments [38]. Based on the arguments above, the following hypothesis was developed:

H7: Social influences will positively affect the intention to use mobile wallet.

Habit

Habit is defined as the degree to which an individual performs behaviors automatically because of learning from prior experiences [25, 39]. The effects of habit on both behavioral intention and actual usage behavior have been proposed and validated by researchers [40, 41].

With the current development of mobile commerce, consumers are using mobile devices like smartphones for almost daily tasks such as internet surfing, social connecting, text and voice chatting, video calling, gaming, shopping, bill paying. Therefore, that development could make mobile-related habits relatively more relevant than before toward the influence on both behavioral intention and actual usage behavior of mobile payment. According to Venkatesh et al. (2012) the role of habits in technology use describes the various fundamental processes that affect technology use. In this study, habit is considered being a familiar behavior or an automatic behavior in using mobile wallets. The following proposition was developed:

H8: Habit will positively affect the user's intention to use mobile wallet.

Intention to use

According to Davis, behavioral intention is considered as the degree to which an individual believes that they will implement a particular behavior [42]. In technology adoption theories, the relationship between behavioral intention and usage behavior has been consistently confirmed [25, 26, 27, 43, 44]. In mobile payment context, there is still a gap between intention and actual use toward the technology. Although the development of mobile commerce has led consumers to use mobile devices in various aspects of their daily lives, the actual amount of mobile payment is still small compared to other forms of payment. Under the technological adoption literature, it can be hypothesized that:

H9: Promotion will positively affect the user's intention to use mobile wallet.

2.2. Research framework

Considered the relationships extracted by hypotheses from previous studies, Figure 1 shows the proposed research model. We included perceived usefulness, social norms, habit; integrated them with perceived risk and a novel construct named promotion. Intention to use and frequency of use applied to measure using behavior.

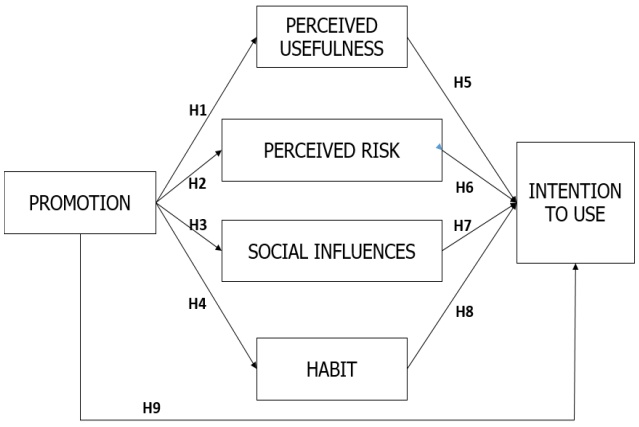


Figure 1: Conceptual framework (Source: own elaboration)

We proposed this research framework because predicting consumer's behavioural adoption and usage should be evaluated from a variety of perspectives. This integrated model can help us have a better recognition of the relationships between these research constructs. These relations are already confirmed in the literature review of previous researches, and the validity of model is confirmed in similar studies except the promotion factor that we added in this study [45, 46, 47]. This proposed model has showed relationship among factors affecting using mobile wallet in Vietnam so that previous studies have not focused on evaluating it.

3. Materials and Methods

Authors have used secondary data from previous studies to construct the proposed research model. We used a survey method with a questionnaire based on the 5-Point Likert Scale to test our research model and developed hypotheses. The potential participants of this research were chosen in convenience sampling method from Da Nang City. 400 young people in Da Nang were contacted by e-mail and social network during the period from January 2020 to April 2020. A link to the survey was included in the messages. 315 valid responses were received. The overall response rate was 78.75%.

3.1. Measurement instrument

The research had seven constructs: perceived risk, social influence, perceived usefulness, habit, promotion, intention to use. Constructs and measurement items in this research were modified from previous works on technology acceptance. Measurement items for constructs of perceived usefulness, social influence, perceived risk, habit were altered from Venkatesh et al. (2012). The scale of intention to use was employed from Davis (1989). All measurement items were measured on a five-point Likert scale, ranging from totally disagree (1) to totally agree (5). The frequency of use measurement was operationalized by one item that measure consumers' actual frequencies of mobile wallet usage.

3.2. Analytical procedures

First, we tested the scales by the Cronbach's alpha, the exploratory factor analysis (EFA) and the confirmatory factor analysis (CFA). In the next stage, we used the Structural Equation Modeling (SEM) method to evaluate the structural relationships between constructs in the theoretical model.

4. Results and Discussion

4.1. Reliability and validity

The reliability and validity of the constructs in the research were evaluated with SPSS 23. We use Cronbach's Alpha reliability test because it is a widely used measure that examines the scale reliability [48]. By measuring the reliability coefficient, the reliability test could assess the consistency of the entire scale. Baron & Kenny (1986) suggested that a scale would be high reliable level if the coefficient alpha is greater than 0.7 while the coefficient alpha is higher than 0.6 means the scale is reliable [49]. Table 2 showed that Cronbach's Alpha of all constructs are above the threshold of 0.7, satisfied internal consistency. All items kept will be included in the analysis Exploratory Factor Analysis (EFA). We used this technique to minimize and summarize data and it is very useful for identifying groups of variables. The relationship of mutually correlated groups of variables consider as several basic factors. Each observed variable will be calculated with a Factor Loading factor, which shows which factor each measurement variable belongs to. The KMO (Kaiser-Meyer-Olkin) coefficient must reach 0.5 or higher, showing that factor analysis is appropriate ($0.5 \leq KMO \leq 1$) [48]. The result of KMO coefficient of the study = 0.889 satisfied the requirements. In addition, Barlett's test is statistically significant (sig = 0.000 < 0.05) proving that the observed variables correlate with each other in the whole [50]. The results of the EFA analysis yielded results consistent with the structure of the six construct. Therefore, all constructs of this model ensure reliability and convergent validity. These constructs would be used in further analysis to test the proposed hypotheses.

In [51], the author recommended that CFA is rigorous to test the overall measurement model and to assess the reliability and validity of the constructs. We measure convergent validity by using composite reliability (CR) and average variance explained (AVE). By the rule of thumb, CR should be above 0.6 and AVE should be above 0.5 for all constructs. As the results shown in Table 1, CRs of all constructs ranges from 0.723 to 0.898 while the AVEs vary from 0.573 to 0.746. These results confirm that our model meets the requirement for convergent validity.

We used the procedure suggested by [52] to check discriminant validity. From this perspective, the AVE for each of the research constructs should be greater than the squared correlation between the construct and other construct. The Table 2 showed the measurement model satisfy the requirement for discriminant validity. The results indicated that the various AVE are lower than the diagonal variables suggesting all the constructs in this study have satisfactory discriminant validity.

Table 3 showed that the CFA model with six concepts also shows a good fit with the data. All the indices satisfied the recommended cut-off point. We conclude that the model fits data well, so this model can be used to test the research hypotheses.

Table 1: Reliability and validity of the tested model

Latent constructs	Cronbach's alpha	CR	AVE	MSV	FL range
SI	0.755	0.723	0.573	0.234	0.503-0.910
PR	0.897	0.898	0.746	0.021	0.810-0.905
PU	0.875	0.883	0.716	0.413	0.775-0.874
HB	0.866	0.870	0.771	0.196	0.836-0.909
PM	0.857	0.856	0.664	0.473	0.743-0.850
IU	0.891	0.892	0.674	0.473	0.692-0.917

Table 2: Factor correlation coefficients

	AVE	IU	PR	PU	PM	SI	HB
IU	0.573	**0.821**					
PR	0.746	-0.143*	**0.864**				
PU	0.716	0.642***	-0.086	**0.846**			
PM	0.771	0.688***	0.005	0.589***	**0.815**		
SI	0.664	0.484***	-0.019	0.422***	0.472***	**0.757**	
HB	0.674	0.443***	-0.116†	0.280***	0.312***	0.396***	**0.878**

Significance of Correlations: (The square root of AVE shown as bold at diagonal)
† $p < 0.100$, * $p < 0.050$, ** $p < 0.010$, *** $p < 0.001$

Table 3: Indices fit criteria

Fit indices	Measurement Model	Structural Model	Threshold limit value	Sources
CMIN/df	1.749	2.221	< 5	[53]
CFI	0.975	0.953	> .90	
GFI	0.937	0.912	> .90	[48]
AGFI	0.907	0.877	> .80	[54]
RMSEA	0.049	0.062	< .08	

Table 4: Results of the hypothesis testing from the structural model

H#	Paths			β	S.E.	C.R.	P-value	Conclusion
H1	PM	→	PU	.722	.075	9.635	***	Supported
H2	PM	→	PR	-.016	.086	-.186	.853	Rejected
H3	PM	→	SI	.387	.081	4.800	***	Supported
H4	PM	→	HB	.380	.084	4.534	***	Supported
H5	PU	→	IU	.282	.055	5.124	***	Supported
H6	PR	→	IU	-.079	.035	-2.255	.024	Supported
H7	SI	→	IU	.097	.071	1.361	.174	Rejected
H8	HB	→	IU	.153	.042	3.602	***	Supported
H9	PM	→	IU	.447	.080	5.556	***	Supported

Table 5: Results of evaluation of indirect effects

Indirect effects	Estimate	Lower	Upper	P	Conclusion
PM to PU to IU	.204	.119	.325	.000	Supported
PM to SI to IU	.038	-.010	.113	.194	Rejected
PM to PR to IU	.001	-.014	.023	.866	Rejected
PM to HB to IU	.058	.026	.106	.002	Supported

Evaluation of direct effects

After assessing reliability and validity of the measurement constructs, in this section we will test the research hypotheses based on review of literature and the framework of research model. To conclude, the significance of each hypothesis, S.E (standardized estimates) and p-value was assessed using SEM (structural equation modelling). As showed in the Table 4, seven of nine hypotheses were accepted.

The results in Table 5 showed that the direct path from PM to IU was significantly mediated by PU and HB. As discussed above, as PU and HB were found to have significant direct effects on intention to use the mobile wallet, the mediating effects of SI and PR in the relationship between PM and IU were partial in this study.

Evaluation of predictive capability

In the current study, 60% of the variance in intention to use mobile wallet was explained by five factors: PM, PU, SI, PR and HB. The R^2 values of intention to use in this attain a moderate level of predictive accuracy [55]. Hence, the predictive accuracy of the research model was satisfied. Comparing with the model not integrated the construct promotion, we found the R^2 value was only 51%. This result showed that promotion is an important factor when researching technology adoption significantly because it improves the level of intentional interpretation of the research model.

4.2. Discussion

In this study, we integrated and explored the promotional factor on users' intentions of mobile wallet. This is an extra factor and thus we have built 3 items to measure the effect of this factor. We combined with other constructs: social influence, perceived risk, perceived usefulness, habit and promotion to measure consumer's perception. The research model was successfully predicting a large variance in intention to use mobile wallet (60%). These results indicate the effectiveness of various constructs of this research model.

Taken as a whole, results show that promotion may affect intention to use and suitable for the study hypothesis. Seven hypotheses were accepted and two hypotheses were rejected (H2 and H7).

PM impact positively to PU

This is the first study to confirm the direct impact of PM on PU. This result also confirms the separation of PM and PU when in many studies that PU also implies the benefits from service use. Indeed, in the current fierce competition context, the perceived benefits measured by traditional items have become almost inevitable. PU-free technologies from customers will almost certainly be rejected. Therefore, PU is almost a must for all companies who want to enter the market need to prove the advantages of their products. However, in order for customers to choose their products, companies need to have appropriate PM strategies, ensuring long-term maintenance, avoiding batches, and ensuring financial efficiency. Implementing the above will ensure mobile wallet providers a long-term competitive advantage in an extremely fierce market with more than 30 mobile wallet providers like in Vietnam.

PM impact positively to SI

This result is also the first time that PM has been shown to have a direct effect on SI. Indeed, when there are attractive promotions, customers will recommend more mobile wallets to acquaintances because they believe that acquaintances will need to use and will bring similar positive experiences. as they experienced. PM has a strong impact on SI and thus mobile wallet providers can take advantage of this factor to offer promotions that can take advantage of even social influences such as referrals to acquaintances who will receive. attractive promotions, for example vouchers, gifts ... Thus, the effectiveness of the promotions in influencing customer behavior in using the mobile wallet will be improved.

PM impact positively to HB

As expected, more promotions will be an important factor in creating habits among users. Therefore, companies need to offer relevant and lasting promotions, striving to maintain the user's habit as a must-have routine. To do this, providers need to develop and integrate more features into the mobile wallet. Just using the mobile wallet can help users solve many problems in life. All of these will form the usage habits for customers, creating customer loyalty for the mobile wallet. The development of a payment ecosystem to provide a convenient and seamless payment experience for users is seen as a key factor for wallets to retain customers in the long run.

PM don't have a significant effect on PR

Usually, when faced with a decision like whether or not to use a mobile wallet, customers often compare gains and losses, risks and opportunities.

However, this result contradicts the original assumption that PM will negatively affect perceptions of risk. The promotions also do not reduce the risk perception of customers because in recent times in Vietnam, many new technology applications have relied on customers' trust and subjectivity to implement the fraudulent behavior like P2P lending applications. Despite being warned, many customers still have serious problems because of the use of these new technologies [56]. Hence the vigilance and precaution of the customer is very high.

PU have a significant and positive impact on IU.

Perceived usefulness in this research has a positive and significant impact on intention to use mobile wallet ($\beta = 0.274$; p-value < 0.001). This result is similar with many previous studies which found that perceived usefulness was one of the most important factors to determine intention to use [3, 57, 58]. Usefulness of a new technology is essential to increase intention to use. These studies also used UTAUT2 as their basis model and reinforced the impact of perceived usefulness on using behaviour [5, 8]. However, we found others study in which perceived usefulness is insignificant in the prediction of the user's intention of mobile wallet services [32].

PR have significant and negative impact on IU

PR has negative and significant effects on IU. In fact, security is always a major worry for the customer while performing payments via mobile wallet [3]. Public sharing of personal information, private information leaks while doing transactions

with mobile wallet are the most common concerns [59]. In contrast, we have several studies in literature revealed the positive relationship between perceived risk and intention to use [31, 60]. Marketing strategies that reach target customers must show customers that there are no risks involved in using mobile wallet. This would lessen perceived risk of customers and therefore improve their intention to use mobile wallet. A risk-free environment once created would result in a greater intention to use a mobile wallet app.

SI don't have a significant effect on IU

SI in the current research does not show a significant impact on IU. It is contrast with results of previous studies which confirmed that friends, family and relatives significantly influence the user's intention [8, 25, 33].

HB have a significant and positive impact on IU

HB was accepted in this research as a significant factor influencing intention to use mobile wallet. This result supports the perspective of Venkatesh et al (2012) who confirmed HB as a critical predictor of customer intention. Take into consideration, the mobile wallet providers need to make an effort in creating the condition where mobile wallet becomes a habit with consumers. For example, they can integrate the e-commerce ecosystem in using the mobile wallet as a payment method or think about give cashback promotion for payment in the retail market. This effort would generate a habit for the consumer to use a mobile wallet more frequent for daily life.

PM have significant and positive impact on IU

Among the constructs in the model which have a direct relationship with IU, PM was shown to have the strongest impact to IU ($\beta = 0.439$; p-value < 0.001). In a survey conducted in major cities in Vietnam about the key factors influencing consumers' choice of mobile wallets, diverse and regular promotions were assessed as one of the deciding factors to choose a mobile wallet. In fact, many people used a technology for the first time because of attractive promotions. As mentioned in the first part of this paper, this is a factor that has received little attention from academics when studying the consumer adoption of new technology. However, as showed in the results, mobile wallets' suppliers need to have many attractive promotions to gain consumer acceptance. Compared to other platforms, mobile wallets do not have many attractive promotions and that may be the reason that mobile wallets are still not popular in the Vietnamese market.

From practical perspectives, the results of this study provide empirical evidence on critical factors to be considered by not only suppliers but also marketers of the mobile wallet. It reveals that promotion and usefulness of use of mobile wallets are the most important factors in user adoption. These factors will improve intention of use of a mobile wallet. To get users to use a mobile wallet, application developers must emphasize the benefits of usefulness and attractive promotion related with this new financial service.

5. Conclusions and limitations

This study is successful in providing an integrated research model for academics to test the impact of social, psychological, and risk factors on technology adoption. In addition, it also helps companies providing mobile wallet services by identifying key factors that influence users' decisions and helping them to have the most comprehensive view of customers. Using SEM analysis, 7 of 9 hypothesis research were accepted. Results of this research are found significant and coherent with previous works on mobile payment and mobile wallet.

This research model explains 60% of intention to use mobile wallet in Vietnam, whereas the same model but without construct promotion explained only 50% in the variation in intention to use mobile wallet. This result demonstrated that putting the new promotion variable to the research model of the adoption and use of technology is necessary and that makes the theoretical contribution of the current study. This is also the premise for future researches when studying the adoption of technology, researchers could integrate promotion to the model to achieve a significant improvement in the level of interpretation as showed in this study.

Last but not least, in this study, we proposed the scale used to measure promotion based on the literature, expert panel and interviews with focus group. The items in this scale will provide the foundation for further research concerning technology adoption such as mobile wallet.

Beside of our study's major contribution that adds into the existing body of knowledge, we also recognize its limitations, mostly regarding the sampling with typically young, highly educated people as responders. In addition, our results have limitation regarding the specific context of Danang city. Danang is an average city with a relatively limited market as its population is approximately one million people. This sample only covers with young people in this city. In order to further enhance its generality, future research could extend the study to more cities or conduct in more countries in Southeast Asia to improve the generalization of the study. Although this research contributes more empirical results in this area, it also has some limitations and therefore further studies in this area will be needed.

Conflict of Interest

There is no conflict of interest regarding this study.

Acknowledgment

This study involves no supporting grants.

References

[1] PricewaterhouseCoopers, Mobile payments in Vietnam fastest growing globally, Thailand emerges second in Southeast Asia, PwC, Nov. 2020.

[2] H. Ha, "The Cashless Economy in Vietnam-The Situation and Policy Implications," Journal of Reviews on Global Economics, **9**, 216-223, 2020. doi: 10.6000/1929-7092.2020.09.20.

[3] K. Madan, R. Yadav, "Behavioural intention to adopt mobile wallet: a developing country perspective," Journal of Indian Business Research, 2016, doi: 10.1108/JIBR-10-2015-0112.

[4] B. An, (2020). The ecosystem makes a breakthrough for electronic wallets [Press release]

[5] A.A. Alalwan, Y.K. Dwivedi, N.P. Rana, "Factors influencing adoption of mobile banking by Jordanian bank customers: Extending UTAUT2 with trust," International Journal of Information Management, **37**(3), 99–110, 2017, doi: 10.1016/j.ijinfomgt.2017.01.002

[6] D.-H. Shin, "Towards an understanding of the consumer acceptance of mobile wallet," Computers in Human Behavior, **25**(6), 1343-1354, 2009, doi:10.1016/j.chb.2009.06.001.

[7] P.G. Schierz, O. Schilke, B.W. Wirtz, "Understanding consumer acceptance of mobile payment services: An empirical analysis," Electronic Commerce

Research and Applications, **9**(3), 209–216, 2010, doi: 10.1016/j.elerap.2009.07.005.
[8] R. Thakur, M. Srivastava, "Adoption readiness, personal innovativeness, perceived risk and usage intention across customer groups for mobile payment services in India," Internet Research, 2014, doi:10.1108/IntR-12-2012-0244.
[9] Kapoor, K. K., Dwivedi, Y. K., & Williams, M. D., "Examining the role of three sets of innovation attributes for determining adoption of the interbank mobile payment service," Information Systems Frontiers, **17**(5), 1039-1056, 2015, doi: 10.1007/s10796-014-9484-7.
[10] J. Dubey, "Personal care products: Sales promotion and brand loyalty," Journal Of Contemporary Management Research, **8**(1), 52, 2014.
[11] J.B. Yusuf, "Ethical implications of sales promotion in Ghana: Islamic perspective," Journal of Islamic Marketing, 2010, doi:10.1108/17590831011082400.
[12] F. de O. Santini, C.H. Sampaio, M.G. Perin, V.A. Vieira, "An analysis of the influence of discount sales promotion in consumer buying intent and the moderating effects of attractiveness," Revista de Administração (São Paulo), **50**(4), 416–431, 2015, doi:10.5700/rausp1210.
[13] S. Mussol, P. Aurier, G.S. de Lanauze, "Developing in-store brand strategies and relational expression through sales promotions," Journal of Retailing and Consumer Services, **47**, 241–250, 2019, doi: 10.1016/j.jretconser.2018.11.020.
[14] A. Mughal, A. Mehmood, A. Mohiuddeen, B. Ahmad, "The impact of promotional tools on consumer buying behavior: a study from Pakistan," Journal of Public Administration and Governance, **4**(3), 402–414, 2014, doi: 10.5296/jpag.v4i3.6680.
[15] P. Chandon, B. Wansink, G. Laurent, "A benefit congruency framework of sales promotion effectiveness," Journal of Marketing, **64**(4), 65–81, 2000, doi: 10.1509/jmkg.64.4.65.18071.
[16] J.Y. Kim, B.P. Min, "Comparison of effectiveness between various sales promotions," Korean Management Review, **34**(2), 445–469, 2005.
[17] M. Palazon, E. Delgado-Ballester, "Effectiveness of price discounts and premium promotions," Psychology & Marketing, **26**(12), 1108–1129, 2009, doi: 10.1002/mar.20315.
[18] R.H. Preston, F.R. Dwyer, W. Rudelius, "The Effectiveness of Bank Premiums: Stoneware… calculators… and Mercedes Sport Coupes–do they pay their way?," Journal of Marketing, **42**(3), 96–101, 1978, doi: 10.1177/002224297804200317.
[19] A. d'Astous, I. Jacob, "Understanding consumer reactions to premium-based promotional offers," European Journal of Marketing, 2002, doi: 10.1108/03090560210445173.
[20] R.B. Boschetti, Promoção de vendas não monetária e seus efeitos na intenção de compra e na escolha da marca de serviços financeiros, Pontifícia Universidade Católica do Rio Grande do Sul, 2012.
[21] Y. Lee, H. Cheon, "A Study on the Factors Affecting the User Intention of Omnichannel Shopping Based on Information Technology," in Proceedings of the 2019 5th International Conference on E-Business and Applications, 20–24, 2019, doi: 10.1145/3317614.3317623.
[22] A. d'Astous, V. Landreville, "An experimental investigation of factors affecting consumers' perceptions of sales promotions," European Journal of Marketing, 2003, doi: 10.1108/03090560310495447.
[23] H. Haans, E. Gijsbrechts, "'One-deal-fits-all?' On category sales promotion effectiveness in smaller versus larger supermarkets," Journal of Retailing, **87**(4), 427–443, 2011, doi: 10.1016/j.jretai.2011.05.001.
[24] F.D. Davis, R.P. Bagozzi, P.R. Warshaw, "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," Management Science, **35**(8), 982–1003, 1989, doi:10.1287/mnsc.35.8.982.
[25] Venkatesh, Thong, Xu, "Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology," MIS Quarterly, **36**(1), 157, 2012, doi:10.2307/41410412.
[26] Venkatesh, Morris, Davis, Davis, "User Acceptance of Information Technology: Toward a Unified View," MIS Quarterly, **27**(3), 425, 2003, doi:10.2307/30036540.
[27] G. Baptista, T. Oliveira, "Understanding mobile banking: The unified theory of acceptance and use of technology combined with cultural moderators," Computers in Human Behavior, **50**, 418–430, 2015, doi:10.1016/j.chb.2015.04.024.
[28] M.S. Featherman, P.A. Pavlou, "Predicting e-services adoption: a perceived risk facets perspective," International Journal of Human-Computer Studies, **59**(4), 451–474, 2003, doi: 10.1016/S1071-5819(03)00111-3.
[29] E. Swilley, "Technology rejection: the case of the wallet phone," Journal of Consumer Marketing, **27**(4), 304–312, 2010, doi:10.1108/07363761011052341.
[30] H. Lee, Y. Lee, D. Yoo, "The determinants of perceived service quality and its relationship with satisfaction," Journal of Services Marketing, 2000, doi: 10.1108/08876040010327220.
[31] J.-H. Wu, S.-C. Wang, "What drives mobile commerce?," Information & Management, **42**(5), 719–729, 2005, doi:10.1016/j.im.2004.07.001.
[32] E.L. Slade, Y.K. Dwivedi, N.C. Piercy, M.D. Williams, "Modeling consumers' adoption intentions of remote mobile payments in the United Kingdom: extending UTAUT with innovativeness, risk, and trust, " Psychology & Marketing, **32**(8), 860-873, 2015, doi:10.1002/mar.20823.
[33] Y. Yang, Y. Liu, H. Li, B. Yu, "Understanding perceived risks in mobile payment acceptance," Industrial Management & Data Systems, 2015, doi:10.1108/IMDS-08-2014-0243.
[34] J.-C. Gu, S.-C. Lee, Y.-H. Suh, "Determinants of behavioral intention to mobile banking," Expert Systems with Applications, **36**(9), 11605–11616, 2009, doi:10.1016/j.eswa.2009.03.024.
[35] U. Tandon, R. Kiran, A.N. Sah, "Understanding Online Shopping Adoption in India: Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) With Perceived Risk Application," Service Science, **8**(4), 420–437, 2016, doi:10.1287/serv.2016.0154.
[36] P. Duarte, S. Costa e Silva, M.B. Ferreira, "How convenient is it? Delivering online shopping convenience to enhance customer satisfaction and encourage e-WOM," Journal of Retailing and Consumer Services, **44**, 161–169, 2018, doi:10.1016/j.jretconser.2018.06.007.
[37] A.Y.L. Chong, N. Darmawan, K.B. Ooi, B. Lin, "Adoption of 3G services among Malaysian consumers: an empirical analysis," International Journal of Mobile Communications, **8**(2), 129, 2010, doi:10.1504/IJMC.2010.031444.
[38] G.W.-H. Tan, K.-B. Ooi, "Gender and age: Do they really moderate mobile tourism shopping behavior?," Telematics and Informatics, **35**(6), 1617–1642, 2018, doi:10.1016/j.tele.2018.04.009.
[39] Limayem, Hirt, Cheung, "How Habit Limits the Predictive Power of Intention: The Case of Information Systems Continuance," MIS Quarterly, **31**(4), 705, 2007, doi:10.2307/25148817.
[40] Pavlou, Fygenson, "Understanding and Predicting Electronic Commerce Adoption: An Extension of the Theory of Planned Behavior," MIS Quarterly, **30**(1), 115, 2006, doi:10.2307/25148720.
[41] N.K. Lankton, E.V. Wilson, E. Mao, "Antecedents and determinants of information technology habit," Information & Management, **47**(5), 300–307, 2010, doi:10.1016/j.im.2010.06.004.
[42]. F. D. Davis, A technology acceptance model for empirically testing new end-user information systems: Theory and results, Ph. D Thesis, Massachusetts Institute of Technology, 1985.
[43] F.D. Davis, R.P. Bagozzi, P.R. Warshaw, "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," Management Science, **35**(8), 982–1003, 1989, doi:10.1287/mnsc.35.8.982.
[44] V. Venkatesh, F.D. Davis, "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," Management Science, **46**(2), 186–204, 2000, doi:10.1287/mnsc.46.2.186.11926.
[45] C. Kim, M. Mirusmonov, I. Lee, "An empirical examination of factors influencing the intention to use mobile payment," Computers in Human Behavior, **26**(3), 310–322, 2010, doi: 10.1016/j.chb.2009.10.013.
[46] F. Liébana-Cabanillas, V. Marinković, Z. Kalinić, "A SEM-neural network approach for predicting antecedents of m-commerce acceptance," International Journal of Information Management, **37**(2), 14–24, 2017, doi:10.1016/j.ijinfomgt.2016.10.008.
[47] N. Shaw, "The mediating influence of trust in the adoption of the mobile wallet," Journal of Retailing and Consumer Services, **21**(4), 449–459, 2014, doi:10.1016/j.jretconser.2014.03.008.
[48] J.F. Hair, W.C. Black, B.J. Babin, R.E. Anderson, R.L. Tatham, Multivariate data analysis, Prentice hall Upper Saddle River, NJ, 2006.
[49] R.M. Baron, D.A. Kenny, "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations.," Journal of Personality and Social Psychology, **51**(6), 1173–1182, 1986, doi:10.1037/0022-3514.51.6.1173.
[50] M.S. Bartlett, "A Note on the Multiplying Factors for Various χ^2 Approximations," Journal of the Royal Statistical Society: Series B (Methodological), **16**(2), 296–298, 1954, doi:10.1111/j.2517-6161.1954.tb00174.x.
[51] M.S. Garver, J.T. Mentzer, "Logistics research methods: employing structural equation modeling to test for construct validity," Journal of Business Logistics, **20**(1), 33, 1999.
[52] C. Fornell, D.F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," Journal of Marketing Research, **18**(1), 39–50, 1981, doi: 10.2307/3151312.

[53] P.M. Bentler, P. Dudgeon, "COVARIANCE STRUCTURE ANALYSIS: Statistical Practice, Theory, and Directions," Annual Review of Psychology, **47**(1), 563–592, 1996, doi:10.1146/annurev.psych.47.1.563.

[54] L. Hu, P.M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," Structural Equation Modeling: A Multidisciplinary Journal, **6**(1), 1–55, 1999, doi:10.1080/10705519909540118.

[55] J. Henseler, C.M. Ringle, R.R. Sinkovics, The use of partial least squares path modeling in international marketing, Emerald Group Publishing Limited, 2009.

[56] T.D. Uyen, H. Ha, "The sharing economy and collaborative finance: The case of P2P lending in Vietnam," Jurnal Ilmiah Ekonomi Bisnis, **22**(2), 2017.

[57] F. Liébana-Cabanillas, J. Sánchez-Fernández, F. Muñoz-Leiva, "Antecedents of the adoption of the new mobile payment systems: The moderating effect of age," Computers in Human Behavior, **35**, 464–478, 2014, doi:10.1016/j.chb.2014.03.022.

[58] I.R. de Luna, F. Liébana-Cabanillas, J. Sánchez-Fernández, F. Muñoz-Leiva, "Mobile payment is not all the same: The adoption of mobile payment systems depending on the technology applied," Technological Forecasting and Social Change, **146**, 931–944, 2019, doi:10.1016/j.techfore.2018.09.018.

[59] A. Hossain, R. Quaresma, H. Rahman, "Investigating factors influencing the physicians' adoption of electronic health record (EHR) in healthcare system of Bangladesh: An empirical study," International Journal of Information Management, **44**, 76–87, 2019, doi:10.1016/j.ijinfomgt.2018.09.016.

[60] Y.-Y. Wang, Y.-S. Wang, T.-C. Lin, "Developing and validating a technology upgrade model," International Journal of Information Management, **38**(1), 7–26, 2018, doi:10.1016/j.ijinfomgt.2017.07.003.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 299-306 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Complex Order $PI^{\alpha+j\beta}D^{\gamma+j\theta}$ Design for Surface Roughness Control in Machining CNT Al-Mg Hybrid Composites

Ravi Sekhar [*,1], Tejinder Paul Singh[2], Pritesh Shah[1]

[1]*Symbiosis Institute of Technology (SIT), Symbiosis International (Deemed University) (SIU), Pune, India*

[2]*Department of Mechanical Engineering, Thapar Institute of Engineering & Technology, Patiala, India*

ARTICLE INFO

Article history:
Received: 28 August, 2020
Accepted: 01 November, 2020
Online: 08 November, 2020

Keywords:
Complex order controller
Fractional calculus
Fractional order controller
Nano composites
Carbon nanotubes
Fractional modeling
Bode diagram

ABSTRACT

Accurate machining control is indispensable for the smart factories of tomorrow. Variations in controller responses may cause unacceptable process deviations during machining leading to productivity losses and possible damage. In the present work, a complex order $PI^{\alpha+j\beta}D^{\gamma+j\theta}$ (COPID) controller was designed to effectively control surface roughness generation while machining CNT Al-Mg hybrid composites. Performance of the designed complex order controller was compared against the conventional PID and fractional order PID (FOPID) controllers for the machined surface roughness system. Output signal responses indicate that the complex order controller attains the desired surface roughness set point with zero percent overshoot in almost same settling time (46 sec) as PID (41 sec) and FOPID (46 sec) controllers. The PID and FOPID output signals registered overshoots of 96.8 % and 36.7 % respectively. Similarly, in case of control signals (feed rate) the COPID controller successfully minimised peak overshoot to 4.6 %; as compared to 64.2 % and 96.5 % in case of the PID and FOPID controllers respectively. The COPID controller was also effective in reducing its peak time response metric (2.001 % for control signal and no peak time for output signal due to zero overshoot). In comparison, the PID and FOPID controllers recorded higher response peak times (7 sec / 26 sec for the PID ouput/control and 5 sec / 7.84 sec for FOPID output/control signal responses). Overshoot elimination in output signal (surface roughness) is crucial for consistency of the machined surface quality. Similarly, overshoot minimisation in control signal (feed rate) is critical because excessive feed rate can damage the cutting tool, work piece, machinery and is a potential safety hazard for the machine operator as well. Hence, the COPID controller can be safely and extensively applied in smart industrial control systems of the future.

1 Introduction

This paper is an extension of the work originally reported in the 2019 IEEE International Conference on Mechatronics, Robotics and Systems Engineering (MoRSE) [1]. In that work, firstly the multi input single output system of surface roughness generation during machining of the CNT Al-Mg composites was identified using the ARX and ARMAX model structures. Thereafter, the fractional order PID (FOPID) controller was designed to control the identified system at the desired set point (surface roughness). The authors reported that the ARMAX model based FOPID control exhibited better time domain characteristics as compared to the ARX model based FOPID controller design. The current work extends this work further to include complex order controller design to improve the time domain performance even more. PID controller has also been implemented for better comparisons. Moreover, the output signal time domain specifications corresponding to the machining variables not discussed in the previous work have also been included in the present study.

Metal matrix composites are new age materials finding new and varied applications due to their superior mechanical properties over conventional alloys [2]. Carbon nanotubes (CNT) in particular have attracted attention of researchers worldwide owing to the superior strengthening characteristics imparted by these nano sized particles to the base metal alloys [3]. Aluminum based alloy systems have almost a universal presence across industry sectors and products [4]. Combination of carbon nantubes and aluminum-magnesium alloys results in composites with superior mechanical characteristics.

[*]Corresponding Author: Ravi Sekhar, Symbiosis Institute of Technology, Pune, India Contact No. +91 20 28116400 & Email ravi.sekhar@sitpune.edu.in

Modeling and designing effective controllers is an important step towards automating the machining of these modern materials.

Generally, PID controllers find extensive applications due to their ease of implementation and maintenance in industrial systems. However, specialized machinery may require more sophisticated controllers having better design characteristics. Fractional order controllers offer a better design alternative because they have two more parameters to tune as compared to the traditional PID controllers [5, 6]. These controllers are based on fractional calculus, a branch of integral calculus that enables fractional orders of the model structure parameters. Fractional order PID controllers (FOPID) have been demonstrated to exhibit better performance characteristics in various application areas such as control systems, signal processing, medical research, material science and many more [7]–[10]. FOPID controllers prove to be more robust against system variations against their PID counterparts. Therefore, FOPID controllers are being increasingly preferred in more dynamic systems across industries [11]–[14]. Machining of complex materials like the metal matrix composites is one such example of a dynamic system; wherein the material composition is anisotropic and calls for a more robust controller architecture [6].

Complex order PID (COPID) controllers are the next step in the controller evolution [15]. They are an extension of the FOPID controllers just as the FOPID are extended versions of the PID controllers. The COPID have further two parameters to tune over the FOPID controllers. In the recent years, COPID controllers have been gradually attracting increasing investigations into their design and application aspects [16]–[18]. Khandani et al [19] designed a complex order controller for DC motors. Pinto and Carvalho [20] derived a complex order model for HIV infection drug resistance. Machado [18] optimized complex order controller for linear and non linear systems using genetic algorithm. Ayadi and Amairi [21] applied numerical optimization to tune complex order controller for a second order time delay resonant system. Hanif et al [22] implemented genetic algorithm for tuning of complex order controller design. Silva et al [23] applied complex order structure to model hexapod robot locomotion based on foot-ground transfer function. Jacob et al [15] reviewed frequency domain, time domain and stability aspects in complex order modeling of linear systems. Tare et al [24] compared integral, fractional and complex order controller performances for fractional order systems. Adams et al [25] determined the transfer function solution to the complex order differential equation. Shahiri et al [26] investigated robust control of a non linear fuel cell system using complex order architecture. From the literature review, it appears that the application of COPID controller in machining systems needs attention. The current work aims to bridge this gap.

The CNT Al-Mg composite experimental details may be referred to in the previous part [1] of the current work. It involved variation of the input machining parameters (cutting speed, feed and depth of cut) and data collection of the corresponding values of the machined surface roughness. Machining experiments were performed on a CNC turning center (Simple Turn 5057 Haas Automation) using chemical vapour deposition (CVD) coated tool inserts (Seco CNMG120408). Surface roughness was measured by a Handy Surf E-DTS570 tester. The best performing model structure (ARMAX) describing this machined surface roughness system in the previous work was adopted in the current study for the COPID and PID controller design. The FOPID controller time domain characteristics from the previous work [1] were referred to in the current work for comparisons with the COPID and PID controller responses. In the previous work, high peak overshoots were detected in output / control signal responses of the FOPID controller. The current study aimed to minimise / eliminate peak overshoots by employing COPID controller design. The following section details upon the design methodologies followed for the implementation of the PID, FOPID and COPID controllers in composite material machined surface roughness system.

2 Controller Design Methodology

This section describes the design methodologies of the PID, FOPID and COPID controllers for the machined surface roughness system. All controllers were designed to attain output set point of 1 micron surface roughness. In machining, cutting speed is inversely related to surface roughness whereas depth of cut directly increases machined surface roughness levels. However, feed rate has the most significant effect on surface roughness generation [2]. Therefore, feed rate parameter was selected as the manipulated variable in the current study whereas cutting speed and depth of cut were maintained at their mid levels as per the experimental design [1]. Thus, all controllers were designed to control this multi input (cutting speed, feed, depth of cut) and single output (surface roughness) system as shwon in Figure 1.

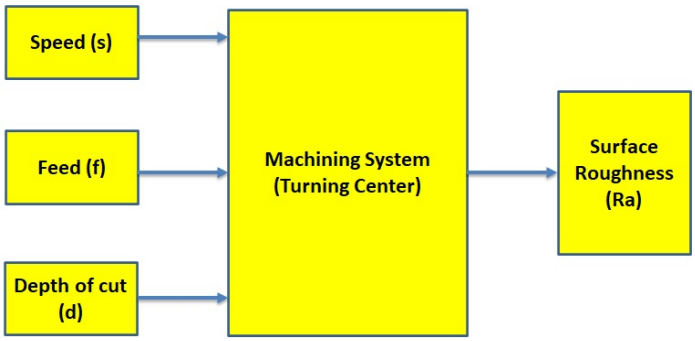


Figure 1: Multi input single output system schematic for the nano composite machined surface roughness system

2.1 PID controller

The PID controller has been widely implemented in industries since many years because of its simple structure and compatibility with the industrial automation software [27]. In the current work, Simulink block was utilised to implement PID controller for the machined surface roughness plant. The PID controller structure is given as follows:

$$C(s) = K_P + K_I\left(\frac{1}{s}\right) + K_D\left(\frac{N}{1 + N\frac{1}{s}}\right) \quad (1)$$

where, K_P is the proportional gain, K_I is the integral gain, K_D is derivative gain and N is the filter coefficient of derivative. The PID block was auto tuned in Simulink using the time domain approach.

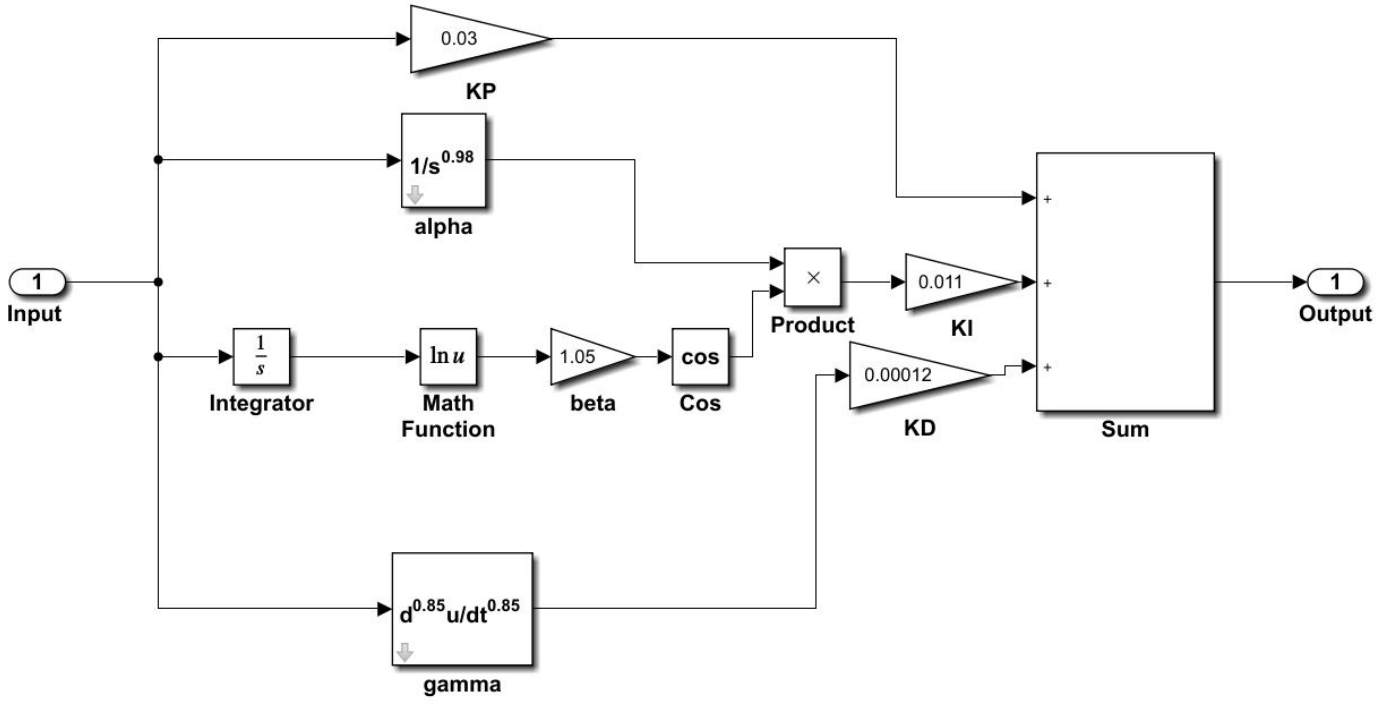


Figure 2: Complex $PI^{\alpha+j\beta}D^{\gamma+j\theta}$ Controller Structure in Simulink

2.2 FOPID controller

Fractional PID (FOPID) controller is an extension of the classical PID controller and is designed based on fractional calculus [28]. In this controller, the orders of integration and derivation are real numbers [29, 14]. The FOPID controller has the following controller structure:

$$C(s) = K_P + K_I\left(\frac{1}{s}\right)^{\lambda} + K_D s^{\mu} \tag{2}$$

where μ is the order of derivative and λ is the order of integration. There are five parameters in the fractional PID controller. This controller has the property of isodamping making it more robust against variations in the system parameters. The values of λ and μ were taken between 0 to 2 for the closed loop system to be stable.

Following equations were employed to synthesise the fractional calculus for the fractional PID controller application:

$$\omega_u = \sqrt{\omega_h \omega_b}$$

$$\omega_0^{'} = \alpha^{-0.5}\omega_u; \omega_0 = \alpha^{0.5}\omega_u;$$

$$\frac{\omega_{k+1}^{'}}{\omega_k^{'}} = \frac{\omega_{k+1}}{\omega_k} = \alpha\eta > 1$$

$$\frac{\omega_{k+1}^{'}}{\omega_k} = \eta > 0; \frac{\omega_k}{\omega_k^{'}} = \alpha > 0$$

$$N = \frac{\log(\omega_N/\omega_0)}{\log(\alpha\eta)}$$

where, N is the order of approximation, ω_h and ω_b are the frequency bounds for approximation.

In the present work, FOMCON toolbox was used for implementation of FOPID and COPID controllers [12]. In this toolbox, Simulink blocks are available for various applications of fractional calculus. It can be used for designing controllers in continuous/discrete time scales, fractional order state space and many more. In FOPID design, the derivative action was considered zero ($K_D = 0$ and $\mu = 0$). The other gain parameters are taken from PID controller settings. Value of λ was selected based on the approach given by Shah and Agashe [29].

2.3 COPID controller

Complex order PID (COPID) controller idea originated from the 3rd generation of CRONE controller [30, 11]. COPID controller is similar to FOPID controller except that the orders of integration and derivation are complex instead of real numbers. The COPID controller has the following structure [16, 26].

$$C(s) = K_P + K_I\left(\frac{1}{s}\right)^{\alpha+j\beta} + K_D\,(s)^{\gamma+j\theta} \tag{3}$$

where, K_P is the proportional gain, K_I is the integral gain, K_D is derivative gain and $\alpha, \beta, \gamma, \theta$ are the orders of the complex order controller.

To implement the above equation, a simplification procedure was carried out in two steps. Firstly, the complex integration was written as:

$$K_I\left(\frac{1}{s}\right)^{\alpha+j\beta} = K_I\left(\frac{1}{s}\right)^{\alpha}\left(\frac{1}{s}\right)^{j\beta}$$
$$K_I\left(\frac{1}{s}\right)^{\alpha+j\beta} = K_I\left(\frac{1}{s}\right)^{\alpha} * e^{\ln\left(\frac{1}{s}\right)^{j\beta}}$$
$$K_I\left(\frac{1}{s}\right)^{\alpha+j\beta} = K_I\left(\frac{1}{s}\right)^{\alpha} * e^{j\beta\ln\left(\frac{1}{s}\right)}$$
$$K_I\left(\frac{1}{s}\right)^{\alpha+j\beta} = K_I\left(\frac{1}{s}\right)^{\alpha} * \left[\cos\left(\beta\ln\left(\frac{1}{s}\right)\right) + j\sin\left(\beta\ln\left(\frac{1}{s}\right)\right)\right]$$

The imaginary part of the above equation cannot be synthesised for time domain implementation of the system. Hence, for implementing the COPID controller in Simulink, imaginary part of the above equation was omitted [16].

$$K_I\left(\frac{1}{s}\right)^{\alpha+j\beta} = K_I\left(\frac{1}{s}\right)^{\alpha} * \left[\cos\left(\beta\ln\left(\frac{1}{s}\right)\right)\right] \quad (4)$$

Similarly, derivative component of the Eq. (3) was simplified for implementation as follows -

$$K_D * s^{\gamma+j\theta} = K_D * s^{\gamma} * s^{j\theta}$$
$$K_D * s^{\gamma+j\theta} = K_D * s^{\gamma} * e^{\ln s^{j\theta}}$$
$$K_D * s^{\gamma+j\theta} = K_D * s^{\gamma} * e^{j\theta\ln s}$$
$$K_D * s^{\gamma+j\theta} = K_D * s^{\gamma} * [\cos(\theta\ln s) + j\sin(\theta\ln s)]$$

The imaginary part of the above equation cannot be synthesised for time domain implementation of the system. Hence, for implementing the COPID controller in Simulink, imaginary part of the above equation was omitted [16].

$$K_D * s^{\gamma+j\theta} = K_D * s^{\gamma} * [\cos(\theta\ln(s))] \quad (5)$$

Combining Eq. (4) and (5), the COPID controller structure may be rewritten as

$$C(s) = K_P + K_I\left(\frac{1}{s}\right)^{\alpha} * \left[\cos\left(\beta\ln\left(\frac{1}{s}\right)\right)\right] + K_D * s^{\gamma} * [\cos(\theta\ln(s))] \quad (6)$$

The COPID controller given in Eq. (6) was initially tuned by incorporating the reference values of FOPID controller and was further fine tuned based on the principles given in literature [31, 29]. The fine tuned COPID controller was implemented in Simulink (Matlab) as shown in figure 2. The initial value of integer order integrator was assumed to be a small number during the controller implementation.

3 Results and Discussions

This section describes the different controller structures, output / control signal responses, time domain characteristics and Bode plot for the COPID controller. The discrete-time ARMAX model (order 3331) with 1 sec sample time from the previous work [1] was obtained as follows:

$$A(z)y(t) = B(z)u(t) + C(z)e(t)$$
$$A(z) = 1 - 0.3675z^{-1} + 0.4876z^{-2} - 0.3772z^{-3}$$
$$B_1(z) = -0.001019z^{-1} - 0.003656z^{-2} + 0.01446z^{-3}$$
$$B_2(z) = 0.6741z^{-1} + 0.2097z^{-2} + 13.82z^{-3}$$
$$B_3(z) = -1.569z^{-1} + 0.4574z^{-2} - 1.532z^{-3}$$
$$C(z) = 1 + 0.4142z^{-1} + 0.554z^{-2} - 0.4004z^{-3}$$

This model depicts the CNT Al-Mg composite machined surface roughness system considered in the current study. The PID controller applied on this model structure has the following structure:

$$C(s) = 0.0407 + 0.0143\frac{1}{s} - 0.0633\frac{0.4367}{1 + 0.4367\frac{1}{s}} \quad (7)$$

The output and control signal responses of this PID controller for a set point of 1 micron surface roughness are shown in Fig. 3 and Fig. 4 respectively.

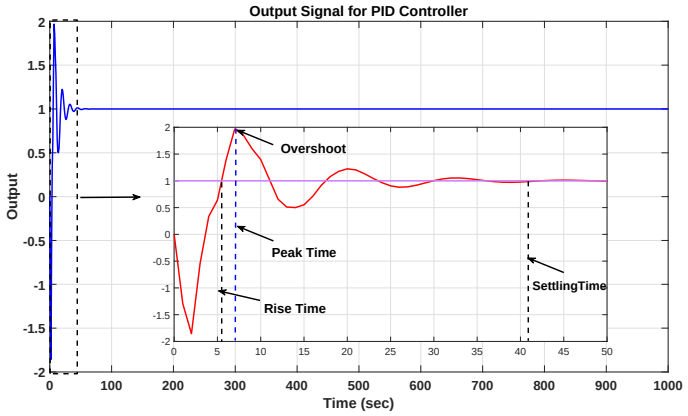


Figure 3: PID controller output (surface roughness, μ) for the machined surface roughness system

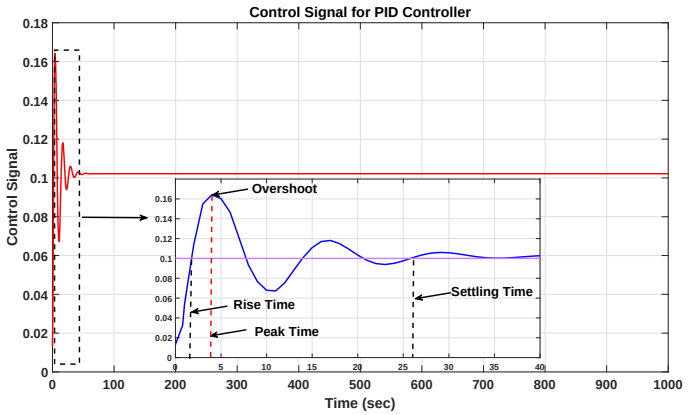


Figure 4: PID control signal (feed rate, mm / rev) for the machined surface roughness system

Following is the structure of the FOPID controller applied to the

composite machined surface roughness system -

$$C(s) = 0.0407 + 0.0143\frac{1}{s^{0.952}} \tag{8}$$

Fig. 5 and Fig. 6 depict the output and control signal plots for the FOPID controller set at 1 micron surface roughness output.

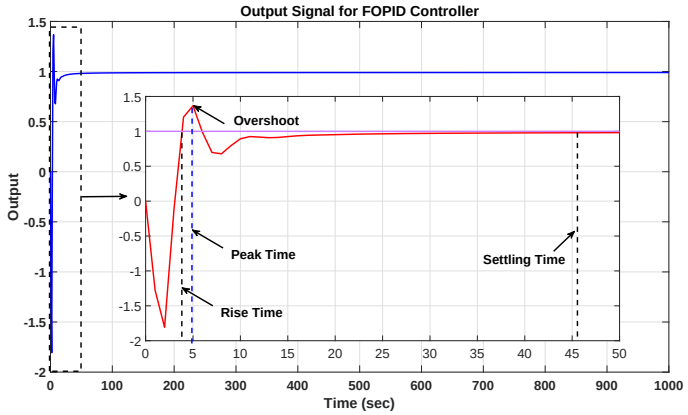


Figure 5: Fractional PID controller output (surface roughness, μ) for the machined surface roughness system

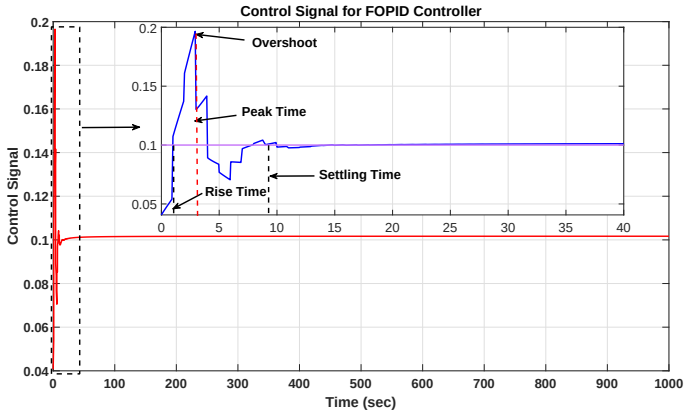


Figure 6: Fractional PID control signal (feed rate, mm / rev) for the machined surface roughness system

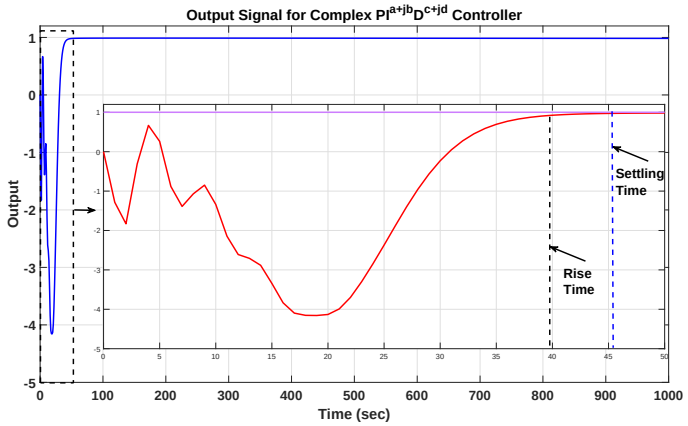


Figure 7: Complex $\mathrm{PI}^{\alpha+j\beta}\mathrm{D}^{\gamma+j\theta}$ controller output (surface roughness, μ) for the machined surface roughness system

The complex order controller structure for the same system is shown below -

$$C(s) = 0.03 + 0.011\frac{1}{s^{0.98+j1.05}} + 0.00012s^{0.85+j0} \tag{9}$$

The output and control signal trends for the COPID controller are shown in Fig. 7 and Fig. 8 respectively.

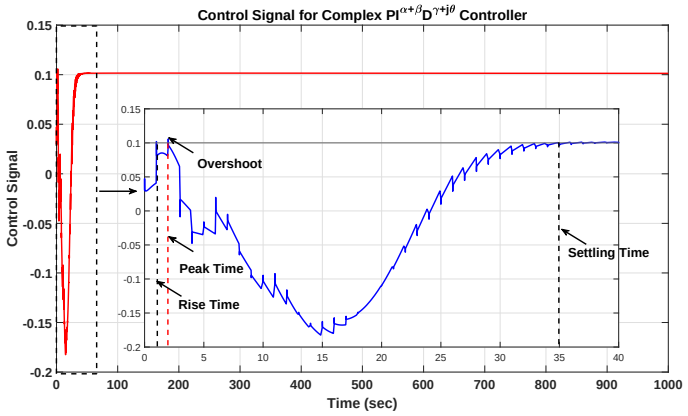


Figure 8: Complex $\mathrm{PI}^{\alpha+j\beta}\mathrm{D}^{\gamma+j\theta}$ control signal (feed rate, mm / rev) for the machined surface roughness system

Table 1: Time domain specifications of output signals

	PID	FOPID	COPID
Rise Time (sec)	5	3.8	39
Peak Time (sec)	7	5	NA
Settling Time (sec)	41	46	46
Overshoot (%)	96.8	36.7	0

Table 2: Time domain specifications of control signals

	PID	FOPID	COPID
Rise Time (sec)	2	1	1
Peak Time (sec)	4	2.924	2.001
Settling Time (sec)	26	7.844	34.38
Overshoot (%)	64.2	96.5	4.6

Table 1 shows the output signal (surface roughness) time domain specifications of the PID, FOPID and COPID controllers for the machined composite surface roughness system. The PID controller scores slightly better than the FOPID and COPID controllers in case of response settling time. PID also scores significantly better than the COPID controller in terms of the rise time. FOPID controller attains the least rise time. However, the COPID controller positively removes peak overshoot from the output signal. The other two controllers have high overshoots in their respective output signal responses. High overshoot in the surface roughness output is undesirable from quality perspective. Generally, critical industrial components require tight dimensional and surface roughness tolerances. Hence, for such cases of precision manufacturing, complex

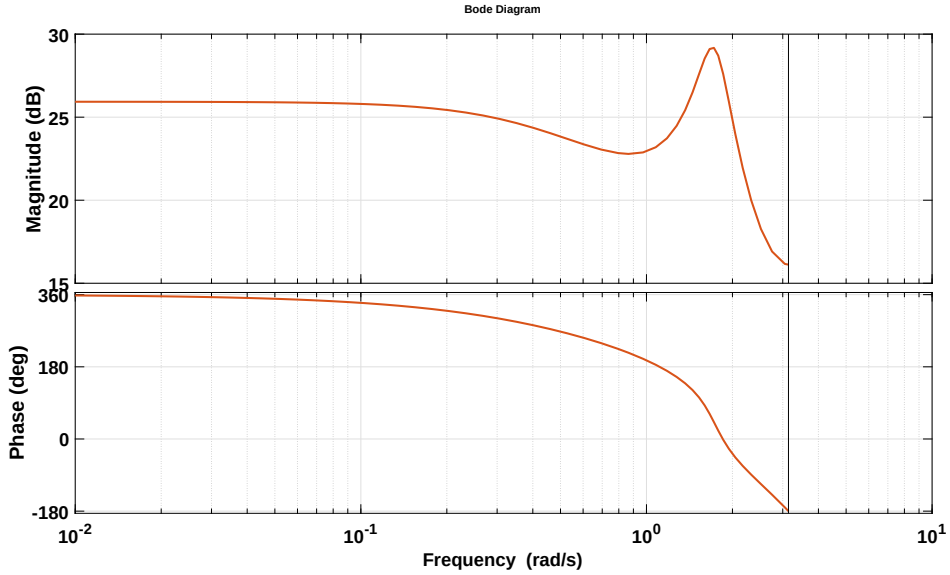


Figure 9: Bode plot for complex order controller

order controller promises to be a better alternative over PID and FOPID controllers.

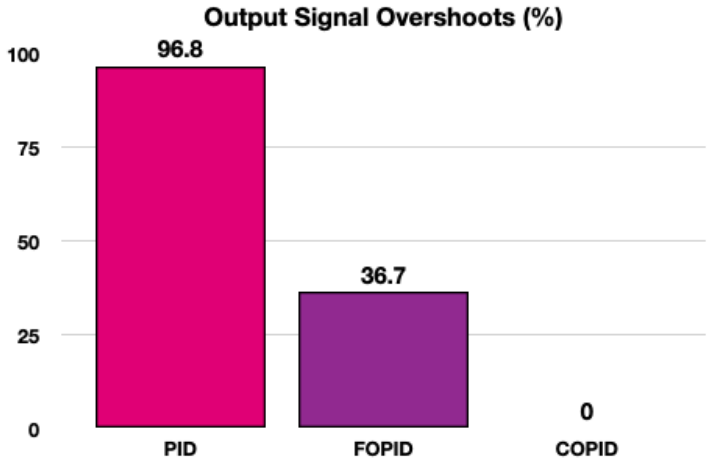


Figure 10: Output Signal Overshoots (%)

Table 2 shows the time domain characteristics of the control signal (feed rate) responses. In this case, the FOPID controller performs well in terms of the rise and settling times. The COPID outperforms the rest for the peak time and overshoot percentages. FOPID has the worst overshoot percentage, whereas the PID controller has the worst peak time characteristic. However, PID controller does perform better than the COPID for settling time. In the current study, control signals correspond to feed rate manipulation in the actual machining systems. High overshoots in feed rate settings may lead to cutting tool damage / breakage, machine tool chatter and possible safety hazard for the machinist. Therefore, minimisation of peak overshoots in feed rate manipulation is a critical machining requirement being fulfilled by the complex order PID controller. Fig. 9 and 10 give visual representations of the relative differences among the output and control signal overshoot percentages attained by the three controllers considered in the current study. Thus, the superiority of the COPID controller over others may be appreciated in light of the importance of the peak overshoot minimisation. Fig. 11 shows the Bode diagram for the complex order controller. This plot indicates that the gain and phase margin for the COPID controller is ∞, implying that the machined composite surface roughness system is robust against variations in the system parameters. This result is crucial from the point of view of the anisoptropic structure of composite materials; that lead to system variations during machining.

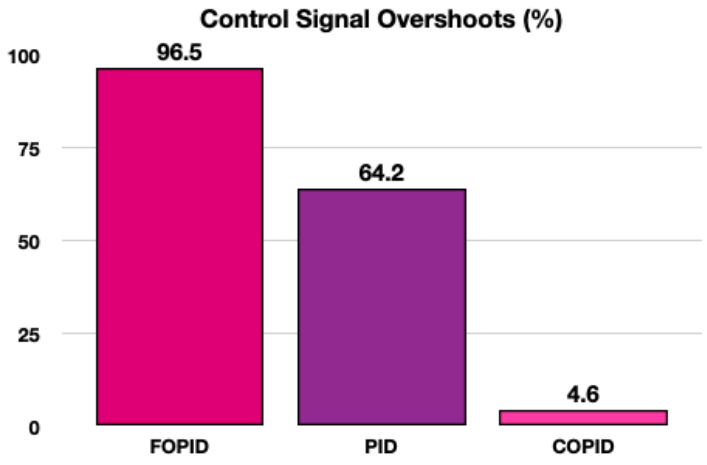


Figure 11: Control Signal Overshoots (%)

4 Conclusions

The current work is an extension of the preliminary study reported in the 2019 IEEE International Conference on Mechatronics, Robotics

and Systems Engineering (MoRSE) [1]. The preliminary study involved system identification and FOPID control of surface roughness evolution in machining of CNT Al-Mg composite materials. The ARMAX model structure controlled by the FOPID controller produced best output signal responses as reported in the maiden paper [1]. The current study expanded that work by including PID and COPID controller designs on the best performing ARMAX model structure determined in the previous work. All three controllers' responses were compared in terms of time domain specifications of the respective output and control signal responses. The current work also included analysis of the output signal trends for the said controllers (which was not considered in the preliminary study). The COPID controller conclusively demonstrated its ability to minimise and eliminate peak overshoot percentages in the control signal (feed rate) and output signal (surface roughness) responses respectively. The other two controllers (PID and FOPID) proved effective in reducing other time domain metrics viz. rise, peak and settling times. However these controllers were unable to stem the peak overshoots to minuscule levels. It is absolutely important to minimise and possibly eliminate the control/output signal overshoots to avoid the adverse affects of such overshoots on the machining process and product quality. This work establishes that the complex order PID controllers can be safely and widely adopted towards effective control of nano composite manufacturing systems for high productivity.

References

[1] R. Sekhar, T. Singh, P. Shah, "ARX/ARMAX Modeling and Fractional Order Control of Surface Roughness in Turning Nano-Composites," in 2019 International Conference on Mechatronics, Robotics and Systems Engineering (MoRSE), 97–102, IEEE, 2019, doi:10.1109/MoRSE48060.2019.8998654.

[2] R. Sekhar, T. Singh, "Mechanisms in turning of metal matrix composites: a review," Journal of Materials Research and Technology, **4**(2), 197–207, 2015, doi:https://doi.org/10.1016/j.jmrt.2014.10.013.

[3] R. Sekhar, T. Singh, "Determination of Johnson Cook Parameters in Turning of Micro and Nano Reinforced Aluminum Composites using Trust Region Reflective Algorithm," International Journal of Innovative Technology and Exploring Engineering, **8**(12), 1712–1716, doi:10.35940/ijitee.L3183.1081219.

[4] V. S. Jatti, R. Sekhar, R. Patil, "Study of ball nose end milling of LM6 al alloy: Surface roughness optimisation using genetic algorithm," Int. J. Eng. Technol, **5**, 2859–65, 2013, doi:http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.411.1573&rep=rep1&type=pdf.

[5] I. Podlubny, "Fractional-order systems and fractional-order controllers," Institute of Experimental Physics, Slovak Academy of Sciences, Kosice, **12**(3), 1–18, 1994, doi:http://people.tuke.sk/igor.podlubny/pspdf/uef0394.pdf.

[6] R. Sekhar, T. Singh, P. Shah, "Micro and Nano Particle Composite Machining: Fractional Order Control of Surface Roughness," in Third International Conference on Powder, Granule and Bulk Solids: Innovations and Applications PGBSIA 2020 February 26-28, 2020, 35, 2020, doi:https://www.pgbsia.com/wp-content/uploads/Conference-Proceedings-PGBSIA-2020.pdf#page=40.

[7] P. Shah, R. Sekhar, S. Agashe, "Application of Fractional PID Controller to Single and Multi-Variable Non-Minimum Phase Systems," International Journal of Recent Technology and Engineering, **8**(2), 2801–2811, 2019, doi:10.35940/ijrte.b2805.078219.

[8] P. Shah, R. Sekhar, "Closed Loop System Identification of a DC Motor using Fractional Order Model," in 2019 International Conference on Mechatronics, Robotics and Systems Engineering (MoRSE), 69–74, IEEE, 2019, doi:https://doi.org/10.1109/MoRSE48060.2019.8998744.

[9] P. Shah, S. Agashe, "Design and optimization of fractional PID controller for higher order control system," in International conference of IEEE ICART, 588–592, 2013, doi:https://www.semanticscholar.org/paper/Design-and-Optimization-of-Fractional-PID-for-Order-Shah-Agashe/6e32b76ffe9609e0423e062b8aedd87361dedd4c.

[10] R. Bhimte, K. Bhole, P. Shah, "Fractional Order Fuzzy PID Controller for a Rotary Servo System," in 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 538–542, IEEE, 2018, doi:https://doi.org/10.1109/ICOEI.2018.8553867.

[11] S. Das, Functional fractional calculus, Springer Science & Business Media, 2011, doi:https://doi.org/10.1007/978-3-642-20545-3.

[12] A. Tepljakov, "FOMCON: fractional-order modeling and control toolbox," in Fractional-order Modeling and Control of Dynamic Systems, 107–129, Springer, 2017, doi:https://doi.org/10.1007/978-3-319-52950-9_6.

[13] P. Shah, S. Agashe, "Design of controller for a higher order system without using model reduction methods," Progress in Fractional Differentation and Applications, **3**, 289–3004, 2017, doi:http://dx.doi.org/10.18576/pfda/030405.

[14] S. Pritesh, S. AGASHE, A. J. KULKARNI, "Design of a fractional $PI^{\lambda}D^{\mu}$ controller using the cohort intelligence method," Frontiers of Information Technology & Electronic Engineering, **19**(3), 437–445, 2018, doi:https://doi.org/10.1631/FITEE.1601495.

[15] J. A. Jacob, A. V. Tare, V. A. Vyawahare, V. N. Pande, "A review of time domain, frequency domain and stability analysis of linear complex-order systems," in 2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), 164–169, IEEE, 2016, doi:https://doi.org/10.1109/WIECON-ECE.2016.8009110.

[16] M. Shahiri, A. Ranjbar, M. R. Karami, R. Ghaderi, "New tuning design schemes of fractional complex-order PI controller," Nonlinear Dynamics, **84**(3), 1813–1835, 2016, doi:https://doi.org/10.1007/s11071-016-2608-5.

[17] C. Li, X. Dao, P. Guo, "Fractional derivatives in complex planes," Nonlinear Analysis: Theory, Methods & Applications, **71**(5-6), 1857–1869, 2009, doi:https://doi.org/10.1016/j.na.2009.01.021.

[18] J. T. Machado, "Optimal controllers with complex order derivatives," Journal of Optimization Theory and Applications, **156**(1), 2–12, 2013, doi:https://doi.org/10.1007/s10957-012-0169-4.

[19] K. Khandani, A. A. Jalali, M. R. R. Mehdiabadi, "Robust complex order controller design for DC motors," in 20th Iranian Conference on Electrical Engineering (ICEE2012), 900–903, IEEE, 2012, doi:https://doi.org/10.1109/IranianCEE.2012.6292481.

[20] C. M. Pinto, A. R. Carvalho, "Effect of drug-resistance in a fractional complex-order model for HIV infection," IFAC-PapersOnLine, **48**(1), 188–189, 2015, doi:https://doi.org/10.1016/j.ifacol.2015.05.162.

[21] A. Guefrachi, S. Najar, M. Amairi, M. Aoun, "Tuning of fractional complex order PID controller," IFAC-PapersOnLine, **50**(1), 14563–14568, 2017, doi:https://doi.org/10.1016/j.ifacol.2017.08.2093.

[22] O. Hanif, G. B. Babu, S. Sharma, "Performance Improvement of PI^{x+iy} D Fractional Complex Order Controller using Genetic Algorithm," in 2018 Fourth International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 1–5, IEEE, 2018, doi:https://doi.org/10.1109/AEEICB.2018.8480981.

[23] M. F. Silva, J. T. Machado, R. S. Barbosa, "Complex-order dynamics in hexapod locomotion," Signal processing, **86**(10), 2785–2793, 2006, doi:https://doi.org/10.1016/j.sigpro.2006.02.024.

[24] A. V. Tare, J. A. Jacob, V. A. Vyawahare, V. N. Pande, "Design of novel optimal complex-order controllers for systems with fractional-order dynamics," International Journal of Dynamics and Control, **7**(1), 355–367, 2019, doi:https://doi.org/10.1007/s40435-018-0448-5.

[25] J. L. Adams, T. T. Hartley, L. I. Adams, "A solution to the fundamental linear complex-order differential equation," Advances in Engineering Software, **41**(1), 70–74, 2010, doi:https://doi.org/10.1016/j.advengsoft.2008.12.014.

[26] M. Shahiri, A. Ranjbar, M. R. Karami, R. Ghaderi, "Robust control of nonlinear PEMFC against uncertainty using fractional complex order control," Nonlinear Dynamics, **80**(4), 1785–1800, 2015, doi:https://doi.org/10.1007/s11071-014-1718-1.

[27] K. J. Åström, P. Eykhoff, "System identification—a survey," Automatica, **7**(2), 123–162, 1971, doi:https://doi.org/10.1016/0005-1098(71)90059-8.

[28] P. Shah, S. Agashe, "Review of fractional PID controller," Mechatronics, **38**, 29–41, 2016, doi:https://doi.org/10.1016/j.mechatronics.2016.06.005.

[29] P. Shah, S. Agashe, "Experimental Analysis of Fractional PID Controller Parameters on Time Domain Specifications," Progress in Fractional Differentiation and Applications, **3**, 141–154, 2017, doi:http://dx.doi.org/10.18576/pfda/030205.

[30] P. Lanusse, A. Oustaloup, B. Mathieu, "Third generation CRONE control," in Proceedings of IEEE Systems Man and Cybernetics Conference-SMC, volume 2, 149–155, IEEE, 1993, doi:https://doi.org/10.1109/ICSMC.1993.384864.

[31] K. H. Ang, G. Chong, Y. Li, "PID control system analysis, design, and technology," IEEE transactions on control systems technology, **13**(4), 559–576, 2005, doi:https://doi.org/10.1109/TCST.2005.847331.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 307-314 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Dense Deep Neural Network Architecture for Keystroke Dynamics Authentication in Mobile Phone

Lubna Abdelkareim Gabralla*

Department of Computer Science and Information Technology, Community College, Princess Nourah bint Abdulrahman University, 84428, Riyadh, Saudi Arabia

ARTICLE INFO

Article history:
Received: 18 July, 2020
Accepted: 29 October, 2020
Online: 10 November, 2020

Keywords:
Smartphones
Deep learning
Dense neural network

ABSTRACT

The ever-growing technology in mobile smartphones has enabled users to store sensitive and private information; as a result, it required the need for an improved security system. Previous approaches heavily relied on shallow machine learning algorithms that require feature extraction which is time consuming, laborious and can cause, resulting in poor authentication. In this paper, we propose a deep learning - dense neural network to avoid the limitation of the classical algorithms and build a mobile smartphone touch screen authentication scheme based on keystroke dynamics. A deep learning – dense neural network classifier was trained using keystroke dynamics features extracted from users. A comparative analysis was made between our proposed DNN classifier and some selected classical machine learning algorithms on the keystroke dynamics data. The data is split into five different data partition of training and testing. Results clearly indicated that the deep learning – dense neural network has eliminated the feature extraction steps required by the classical algorithms and improved the overall authentication accuracy, as such, improved the security of the smartphone device. In addition, it is found that the propose deep learning – dense neural network authentication scheme is more robust than the classical algorithms and has the potential to be fully implemented on smartphone to improve the security system of the mobile smartphone touch screen devices.

1. Introduction

The rapid advancement of technology in mobile smartphone devices makes them an important part of human life [1–4] . Also, the reduction in cost makes it easier for everyone to possess [5] including Those with disability,[6, 7] and their popularity is increasing [8, 9] They are now equipped with tremendous functionalities and a flexible operating system, providing the opportunity to store sensitive information such as official documents as well as installing official applications to improve the efficiency of office work [10]. Mobile phone users have become addict due to the social media applications and other applications being made available through google play store and apple store, [10] which carry numerous private information. Hence, the security of smartphone devices is becoming sacrosanct [11]. Protecting sensitive information requires a strong security scheme to protect the smartphone from intrusion, many mobile phone authentication schemes have proposed including person number (PIN),password and authentication [2] which are all vulnerable to shoulder surfing [2, 12]and smudge attacks [13, 14].Recently, biometrics have been incorporated into smartphone devices [15] in an effort to improve their security system . These include fingerprint auth entication and facial recognition. Fingerprint authentication is also vulnerable to attacks as a user's fingerprint can be copied from a touched object and used to gain access to the device [16]. Similarly, Facial recognition is also vulnerable to attack as the user's photo or video can be used to gain unauthorized access to the device and this attack is made easier as the pictures of users can be found on social media platforms [17].

The vulnerability to attack of the current security systems of mobile devices has led many researchers to propose different machine learning algorithms to improve their authentication scheme. For example, [18] they used support vector machine (SVM) to build a behavioral authentication scheme to improve the traditional authentication system. The scheme authenticates users

*Corresponding Author: Lubna Abdelkareim Gabralla, Email: lubnagabralla@gmail.com

Based on their behaviors when interacting with social media applications. It is found to perform better than other classical methods. As [19] presented a continuous behavioral authentication model to improve the security of mobile application. In the study, k-nearest neighbor (KNN), random forest (RF) and gradient boosting (GB) were used to build authentication classifiers. In [20], author built an authentication scheme using SVM based on physical activity performed by different mobile phone users. The proposed SVM model showed a promising result when compared with decision tree (DT) and KNN. In [21], the author hybridized SVM and hidden Markova model (HMM) to develop an authentication model using EEG signal while a user is drawing a pattern. The SVM-HMM model provides the best result when compared with naïve Bayes and cosine similarity. In [22], the author trained SVM classifier on facial attributes extracted from real images to build a continuous classification scheme. Results indicated that the proposed scheme shows a significant improvement on the existing continuous facial recognition system. Similarly, in [23] the author used a one-class SVM to build a behavioral biometrics authentication that automatically adapt to human behavior change over time while considering memory constraint. Results show that there is a possibility of using online machine learning to adapt to recent human behavior. In [24], the author trained a two-class SVM classifier using the user's SCG signal which captures heartbeat signal to improve the security scheme of smartphone devices. The heartbeat signal can be used as a unique feature to authenticate smartphone users.

In [25], the author explored four mobile phone non assisted sensors; transmitted data, noise, battery and ambient light to develop a continuous user authentication based on KNN. The KNN classifier achieved a reasonable accuracy. In [26], it exploits user's hand geometry and behavioral biometrics to build a one class classification model based on KNN. The model was compared with SVM and experimental results show that KNN outperforms the SVM in all the different positions. In [27] It creates an authentication scheme using artificial neural network (ANN) based on thumb stroke dynamics. The scheme was evaluated and compared with other .machine learning algorithms and the results of the experiment indicated that the ANN provided the best result. It was found that researchers in this domain heavily relied on SVM [28–33]; [18, 20–23] for mobile smartphone authentication. The previously used algorithms such as SVM, KNN, ANN etc. proved to be good in improving smartphone security especially when the data size is small. However, these algorithms require independent feature extraction before the extracted features are fed to the algorithms [19]. The performance of the algorithms depend on how well the features are extracted before modeling the classifier. Therefore, inefficient feature extraction may lead to poor classification accuracy [34] unlike deep learning which does not require independent feature extraction as it is done automatically and can work on large data size [35, 36]. Finally, a [37] convolutional neural network is applied to create an authentication scheme based on tap sequence and usage behavior of users.

In this paper, we proposed deep learning (dense neural network) based keystroke dynamics authentication for mobile smartphones touch screen device. The closest work to our proposal is [38] the one that employed deep learning technique to build an identification system that identify smartphone users based on keystroke behavior captured via a special keyboard or a web browser. This differs from our work in the sense that it focuses on a user's identification while our proposed model focuses on the user's authorization and authentication. Similarly, it explores gated recurrent unit and bidirectional recurrent neural network (GRU-BRNN) to build the identification model while our work explored the dense neural network. Also, [39] the deep neural network is explored to develop an authentication scheme based on user keystroke dynamics on mobile phone. Furthermore, our work explored supervise learning dense neural network on 71 different features of keystroke dynamics to build an authentication model while [39] applied deep neural network unsupervised learning on timing, tapping and inertial attributes of keystroke dynamics to develop an authentication scheme.

We choose the DNN in view of the fact that the learning features are provided from all the combination of the features in the layers while convolutional layer depends on minor repetitive field with features that are consistent [40].

The rest of the paper is structured as follows: Section 2 describes the concept of deep learning. Section 3 describes the methodology of the study. Section 4 presents the results and discussion before presenting Section five that comprises conclusions and recommendation for future work.

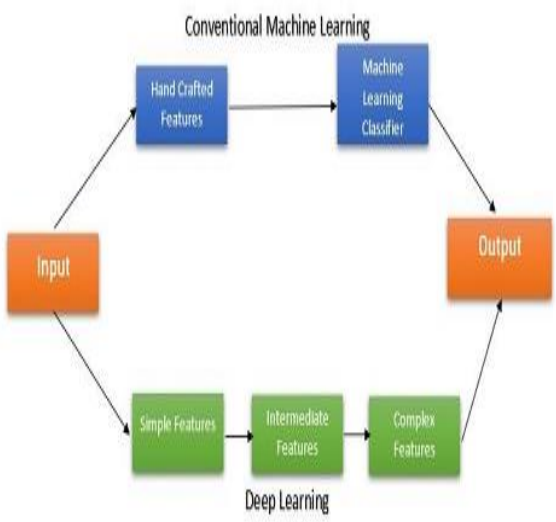


Figure 1: Conventional machine learning VS deep learning

2. Basic Concept of Deep Learning – Dense Neural Network

Deep Learning is an aspect of machine learning which represents multiple hidden layers that can learn on multiple attributes to produce better results [36]. Traditional machine learning algorithms have limitation on processing real data. Building a classification model with conventional machine learning algorithms, requiring reasonable amount of human expertise and efficient feature engineering to transform the raw data to desirable features that can be fed into any learning system or classifier. Deep learning permits inputting raw data into the algorithm without being transformed to feature vector or feature engineering as the features are learnt automatically during training the network [36].

Problems are solved in a hierarchical manner in deep learning, where the lower layers depict the basic representation of the problem and the upper layers are created based on the lower layers to build a more complex model. Deep learning is a hierarchical process as each layer in the deep learning network uses the output of the previous layer and the output of the current layer serves as the input to the successive layer to continuously build a complex concept [36]. The number of layers in a network determines the depth of the network. Cconventional machine learning only focuses on one or two layers, whereas in deep learning, the network contains at least three or more hidden layers. The unique aspect of deep learning is that the feature layers are automatically learned on the raw data not extracted by human expertise unlike the conventional machine learning[36]. The major advantages of deep learning over traditional machine learning are; automatic feature extraction and efficient handling of massive amount of data as a result, it keeps improving as the data gets larger [36]. Based on the concept presented in[36], we created Figure 1 to show deep learning and the conventional machine learning processes.

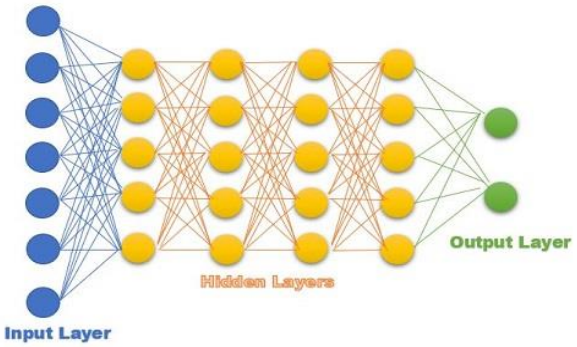


Figure 2: A simple example of dense neural network

In DNN, the layers of the network are fully connected as the name (dense) suggest, each neuron in a layer is connected to all the neurons in the previous layer which means all the nodes are fully connected to the nodes in the next layer. A densely connected layer learns features from all the combined features of the previous layer. The goal of the classifier is to adequately categorize the input data [40]. Figure 2 presents the simple example of DNN.

3. The proposed deep learning - dense neural network authentication model

This section presents the methodology of the study, where the data collection procedure, propose authentication model, propose work flow and performance metrics are outlined. The proposed model authenticates smartphone users is based on user's password typing behavior on a touch screen. The framework consists 4 steps namely; accusation of data, training, user authentication and performance evaluation.

3.1. Dataset

The dataset used for this study contains keystroke dynamics (user typing behavior) data of users collected from nexus 7 touch screen smartphone device when typing the "tie5Roanl" password. The data was collected from 56 subjects and each subject was asked to perform the task 51 times. Therefore, the dataset contains 2856 records. The dataset contains 71 attributes (features), the main features include: hold, up-down, pressure, finger area, average of hold, average pressure and average finger area. Each feature has some feature elements which correspond to the typed characters. The dataset is obtained from the UCI machine learning repository [41]. It was normalized between 0.00 – 1.00, consisting of the user's keystroke dynamics. The data extracted from touch screen smartphone is used to conduct the experiment. The dataset was partitioned several times and the keystroke attributes were used to train the deep learning – DNN classifier model. During the authentication phase, user's tested data for all the 56 subjects. As such, the classes are 56 to authenticate each of the 56 subjects. Each of the 56 subject is authenticated. If the similarities between the two models for each of the 56 users matches, the user is authenticated into the smartphone else the user is rejected. To evaluate the performance of the proposed deep learning model - DNN, the model is compared with previously used machines learning algorithms such as SVM, ANN and KNN. These three algorithms were chosen because the algorithms are found to be highly relied on especially SVM in creating classifiers for improving the security of mobile smartphones device [42].

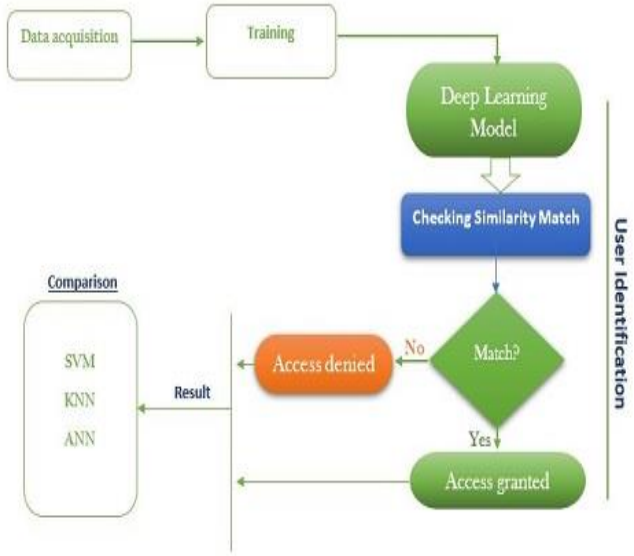


Figure 3: proposed model workflow

The SVM, KNN and ANN were trained to build classifier based on the keystroke dynamics data for the purpose of validation. The performance of the proposed deep learning - DNN classifier is compared with that of the SVM, KNN and ANN. The process involved assessing how well the algorithms perform on the test dataset. The performance of the algorithms on the keystroke dynamic dataset is measured based on performance evaluation metrics. The complete work flow of the methodology is presented in Figure 3.

3.2. Performance Evaluation Metrics

The performance of the proposed touch screen authentication model is evaluated using the true positive (these are instances that are correctly identified), true negative (instances that are correctly rejected), false positive (these are instances that are incorrectly identified) and false negative (these are instances that are incorrectly rejected). The performance metrics were computed

from the results generated from the validation of the models. The performance of the deep learning - DNN algorithm and the classical algorithms are being evaluated using the performance metrics [43].

Table 1: Summary of evaluation metrics

Evaluation metrics	Description
False Positive Rate (FPR) = $\frac{FP}{(FP+TN)}$	Used to determine how many instances are wrongly classified.
F1-Score = $\frac{2TP}{(2TP+FP+FN)}$	Determines the harmonic average between precision and recall.
Recall $(r) = \frac{TP}{(TP+FN)}$	Used to determine how many instances are been classified correctly.
Accuracy (A) = $\frac{TP+TN}{(TP+TN+FP+FN)}$	This is used to measure the accuracy of the technique.

4. Results and Discussion

In this section, the results generated from the experiment were presented including the discussion. The hardware platform, parameter settings of the proposed deep learning - DNN, comparative analysis of the DNN performance compared to the state of the art methods: SVM, ANN, and KNN were presented as well as the implication of the result in theory and practice.

4.1. Parameter Setting

The implementation was conducted on a platform equipped with Microsoft Window 10 with the following specifications:

- System Type: x64-based processor, 64-bit Operating System
- Memory installed on system (RAM): 8.00 GB
- Processor: Intel(R) Core (TM) i3-4000M @ 2.40 GHz 2.40 GHz

The platform for DNN is TensorFlow using Python. a preliminary experiment was conducted to obtain the best parameters of the propose deep learning - DNN. The number of hidden layers and the nodes on each hidden layer are selected after different combination of layers and nodes were run during the preliminary experiment. It was found that the deep learning - DNN setting with the best performance is presented in Table 2. Different learning rates were tested ranging from (0.1 - 0.9), the best value was adopted. Adam optimizer provided the best performance when compared with Gradient Descent Optimization and Adagrad Optimizer. Relu activation function was selected after been compared with other activation functions such as softmax, tanh, etc. The summary of the preliminary experiment result for setting the deep learning – DNN is presented in Table 2.

Table 2: Optimal parameter setting for the proposed DNN

Parameter	Setting
Input layer neuron(s)	71
Hidden layer 1 neuron(s)	20
Hidden layer 2 neuron(s)	20
Hidden layer 3 neuron(s)	20
Hidden layer 3 neuron(s)	20
Output layer neuron(s)	56
Learning rate	0.1
Optimizer	Adam optimizer
DNN structure	71-20-20-20-20-56
Epoch	1000
Activation Function	Relu Activation Function

The deep learning – DNN with the parameter setting in Table 2 is applied to run on keystroke dynamic datasets to authenticate access to mobile smartphone touch screen device. The DNN was run 10 times to ensure consistency of the result produced by the DNN. Data partition ratio affects the performance of the DNN, as such, several partition ratios were used to ascertain the robustness of the DNN and it is performance. An algorithm intend for real world application should be robust in addition to performance. The performance metrics in Table 1 were used to measure the performance of the DNN in user authentication for the mobile touch screen device. To evaluate the performance of the DNN, the classical algorithms SVM, KNN and ANN were also applied on the same dataset to authenticate user access to the touch screen mobile device based on keystroke dynamic authentication. The results of the experiments are presented in Tables 3 – 6 and Figure 4. The first column indicates the different data partition ratio used to evaluate the algorithms. The second column shows the algorithm for the experiment while the third, fourth and fifth columns indicate the mean, best and the worst performance respectively. The bold values in each Table indicates the best result obtained.

Table 3: Performance comparison of the propose DNN with the SVM, KNN and ANN based on Recall

Partition	Algorithms (%)	Mean (%)	Best (%)	Worst (%)
90-10	DNN	89.7	**96.0**	89.0
	SVM	68.7	73.7	51.7
	KNN	44.6	47.7	34.5
	ANN	2.57	3.5	1.8
80-20	DNN	91.2	**94.0**	86.0

	SVM	70.8	72.2	69.7
	KNN	44.6	46.1	43.7
	ANN	2.02	2.3	1.8
70-30	DNN	95.0	**99.0**	92.0
	SVM	71.3	72.7	70.1
	KNN	42.7	45.2	41.2
	ANN	2.1	2.3	1.8
60-40	DNN	93.1	**97.0**	89.0
	SVM	70.4	71.5	68.7
	KNN	41.7	42.4	41.1
	ANN	2.0	2.3	1.7
50-50	DNN	93.0	**98.0**	91.0
	SVM	68.8	69.3	68.2
	KNN	40.4	41.2	40.2
	ANN	1.9	2.2	1.7

The result presented in Table 3 represent the performance of the algorithms based on the Recall for the five different data partitions. It clearly shows that the proposed DNN model produce the best results compared to the classical algorithms. The DNN result indicates that the propose DNN classifier was able to authenticate user access to the mobile touch screen device based on keystroke dynamics with very high level of accuracy. Meaning that the DNN determines the number of instances that were correctly classified as the legitimate user of the mobile phone touch screen device. ANN produced the worst result in all the different data partitions. This is not surprising because the shallow ANN produce poor result as data size increases.

Table 4 clearly indicated that the propose DNN outperforms the classical algorithms in terms of the F1-measure. The performance of the DNN is far more than that of the classical algorithms recording more than 90% in each case. The ANN has maintained consistency in producing the worst performance in all the data partition ratio. This result means that the DNN is a good algorithm with potential to determine the harmonic average between the precision and the recall better than the classical algorithms.

Table 4: Performance comparison of DNN with SVM, KNN and ANN based on F1-Score

Partition	Algorithms (%)	Mean (%)	Best (%)	Worst (%)
90-10	DNN	89.6	**96.0**	84.0
	SVM	69.1	74.5	51.0
	KNN	45.2	48.0	36.0
	ANN	0.1	0.2	0.1
80-20	DNN	91.2	**94.0**	86.0
	SVM	71.1	72.4	69.7
	KNN	45.1	47.0	44.0
	ANN	0.5	0.2	0.1
70-30	DNN	95.0	**99.0**	92.0
	SVM	71.6	73.0	70.3
	KNN	44.0	46.1	42.0
	ANN	0.1	0.2	0.1
60-40	DNN	93.1	**97.0**	89.0
	SVM	70.5	71.8	68.1
	KNN	41.8	42.6	41.3
	ANN	0.1	0.2	0.1
50-50	DNN	93.0	**98.0**	91.0
	SVM	69.1	75.0	53.0
	KNN	40.4	41.2	40.3
	ANN	0.1	0.2	0.1

Table 5: Performance comparison of the propose DNN with SVM, KNN and ANN based on accuracy

Partition	Algorithms (%)	Mean (%)	Best (%)	Worst (%)
90-10	DNN	89.7	**96.0**	89.0
	SVM	84.1	86.6	75.4
	KNN	71.7	73.3	66.2
	ANN	5.0	5.0	5.0
80-20	DNN	91.3	**94.1**	86.2
	SVM	85.1	85.8	84.6
	KNN	72.0	72.6	71.4
	ANN	5.0	5.0	5.0
70-30	DNN	94.8	**98.7**	92.2
	SVM	85.4	86.1	84.8
	KNN	70.8	72.1	70.1
	ANN	5.0	5.0	5.0
60-40	DNN	93.2	**97.0**	89.2
	SVM	84.9	85.5	84.1
	KNN	70.3	70.7	70.1
	ANN	5.0	5.0	5.0
50-50	DNN	92.8	**97.6**	90.6
	SVM	84.1	84.4	83.8
	KNN	69.7	70.1	69.6
	ANN	5.0	5.0	5.0

Table 5 presents the accuracy of the DNN in authenticating the user of a mobile touch screen device in terms of keystroke dynamic. It is clearly indicated that the propose DNN has better accuracy than the compared algorithms.

Table 6: Performances comparison of the DNN with SVM, KNN and ANN based on FPR

Partition	Algorithms (%)	Mean (%)	Best (%)	Worst (%)
90-10	DNN	9.7	3.6	15.8
	SVM	0.5	**0.4**	0.9
	KNN	1.0	1.0	2.0
	ANN	2.5	1.8	3.5
80-20	DNN	8.5	5.8	13.6
	SVM	0.5	**0.5**	0.5
	KNN	1.0	1.0	1.0
	ANN	2.0	1.8.	2.3
70-30	DNN	5.3	1.6	7.6

	SVM	0.5	**0.5**	0.5
	KNN	1.0	1.0	1.0
	ANN	2.0	1.8.	2.3
60-40	DNN	6.7	3.0	10.6
	SVM	0.5	**0.5**	0.6
	KNN	1.0	1.0	1.0
	ANN	1.9	1.7	2.3
50-50	DNN	7.2	2.8	11.0
	SVM	0.6	**0.5**	0.6
	KNN	1.0	1.0	1.0
	ANN	1.8	1.7	2.2

Table 6 clearly indicates that in terms of the FPR, the SVM has the lowest FPR compared to the propose DNN. In this parameter, the DNN fails to outperform the classical algorithms.

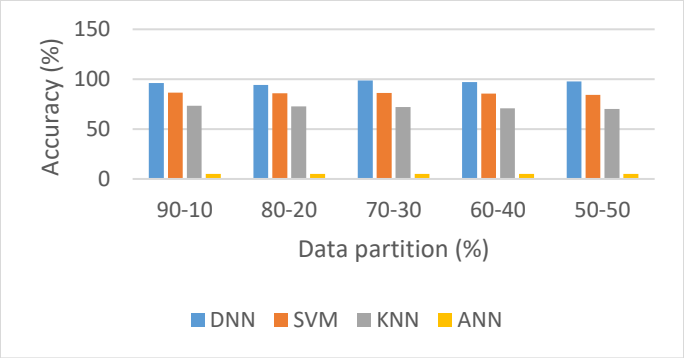


Figure 4: Overall accuracy rate of the proposed deep learning – DNN compared to the SVM, KNN and ANN

Figure 4 indicates that the DNN model has the best classification accuracy than the compared algorithms: SVM, KNN and ANN with ANN producing the worst classification accuracy based on the keystroke dynamics. Our proposed DNN model proves to be the overall best performing algorithm for classification based on the keystroke dynamic data providing the best accuracy rate. It indicates that the DNN model provides a high rate of correctly classified instances and maintaining a low rate of classifying incorrect instances.

Our research work provided a description of the application of DNN in authenticating users of mobile touch screen device based on keystroke dynamics. The authentication security system is to identify authentic users based on keystroke dynamics on mobile touch screen device. As a result, keystroke dynamic features were used for the training of the propose DNN without feature extraction with varying data partition ratios. The DNN authentication based on the keystroke dynamics was applied to classy users based on evaluation measures: f1-measure, recall, accuracy and FPR.

The feature extraction method that the classical algorithms heavily rely on for their performance in the modelling process is a multiple work and human intervention is required significantly. Our propose DNN approach for authentication based on keystroke dynamics is able to eliminate the double work of the feature extraction mostly practiced by researchers in the mobile touch screen device domain.

From the results presented in Tables 3 – 6 and Figure 4. It is clear that the propose DNN can perform better than the classical algorithms in terms of authenticating users of mobile touch screen devices based on keystroke dynamic. The recall and F1-measure values of the propose DNN for authentication based on keystroke dynamics ranges from 94 to 99% while accuracy ranges from 94 – 98.7%. The possible reason why the DNN performs better than the classical algorithms is because of the ability of the DNN to process large size of dataset without requiring feature extraction to discover intricate structures. The propose DNN has proven to provide better authentication of mobile touch screen device based on keystroke dynamic without feature extraction typically required by the classical algorithms. As such, the extra steps of feature extraction that is tedious and incurring extra computational processes can be eliminated with the propose DNN. The DNN has proven that it is a good algorithm with the required robustness to improve the security of the mobile touch screen devices. Therefore, mobile touch screen authentication system can be developed with the propose DNN to improve security of the mobile devices. The experimental result obtained from the study shows that the DNN is a choice algorithm for building classifier for the future research work on keystroke dynamic based user authentication.

The task of authenticating mobile touch screen device considered accuracy, F1-measure and recall as critical measures for evaluating the effectiveness of the authentication systems. The higher the value of these measures the better is the authentication system. The propose DNN authentication based on the keystroke dynamic can be considered to be successful because of its performance on the three performance measures. Though, the propose DNN has inferior performance on FPR compared to the classical algorithms regarding rejections - FPR. This means that the propose DNN can wrongly classify some few instances as the percentage is not much. It is argued that a classification technique is said to be good if the TPR is high while maintaining a low FPR [42]. Therefore, the propose deep learning - DNN is considered the best because the SVM does not maintain high TPR and low FPR like the deep learning – DNN classifier.

5. Conclusions and Future Work

In this paper, we proposed a deep learning approach to build a touch screen mobile phone authentication scheme based on keystroke dynamics. An experiment was carried out and the deep learning model was evaluated by comparing it with conventional machine learning algorithms: the SVM, KNN and ANN. Results of the experiment show that the propose deep learning model outperforms the compared algorithms. The results indicated the feasibility of using deep learning to improve the security system of the mobile smartphone touch screen devices. The deep learning approach has succeeded in eliminating the tedious feature extraction step required by the conventional algorithms: SVM, KNN and ANN. The propose dense DNN low performance on FPR which is considered to be the next future work to improve the value of the FPR. In the future, we plan to use behavioral biometric data extracted from users when drawing a pattern password.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the Fast-track Research Funding Program.

References

[1] M. Shahzad, A.X. Liu, A. Samuel, 'Secure unlocking of mobile touch screen devices by simple gestures - You can see it but you can not do it', in Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM, 39–50, 2013, doi:10.1145/2500423.2500434.

[2] P. Studies, C. Science, 'ECG Authentication for Mobile Devices by Juan Sebastian Arteaga Falconi', **65(3)**, 591–600, 2013.

[3] M. Shahzad, A.X. Liu, A. Samuel, 'Behavior Based Human Authentication on Touch Screen Devices Using Gestures and Signatures', IEEE Transactions on Mobile Computing, **16**(10), 2726–2741, 2017, doi:10.1109/TMC.2016.2635643.

[4] T. Vu, A. Baid, S. Gao, M. Gruteser, R. Howard, J. Lindqvist, P. Spasojevic, J. Walling, 'Capacitive touch communication: A technique to input data through Devices' Touch Screen', IEEE Transactions on Mobile Computing, **13**(1), 4–19, 2014, doi:10.1109/TMC.2013.116.

[5] G. Li, P. Bours, 'Studying WiFi and accelerometer data based authentication method on mobile phones', in ACM International Conference Proceeding Series, 43–49, 2018, doi:10.1145/3230820.3230824.

[6] M. Ernst, T. Swan, V. Cheung, A. Girouard, 'Typhlex: Exploring Deformable Input for Blind Users Controlling a Mobile Screen Reader', IEEE Pervasive Computing, **16**(4), 28–35, 2017, doi:10.1109/MPRV.2017.3971123.

[7] J. Lai, D. Zhang, 'ExtendedThumb: A target acquisition approach for one-handed interaction with touch-screen mobile phones', IEEE Transactions on Human-Machine Systems, **45**(3), 362–370, 2015, doi:10.1109/THMS.2014.2377205.

[8] R. Francese, M. Risi, G. Tortora, M. Tucci, 'Visual Mobile Computing for Mobile End-Users', IEEE Transactions on Mobile Computing, **15**(4), 1033–1046, 2016, doi:10.1109/TMC.2015.2422295.

[9] J. Yu, H. Han, H. Zhu, Y. Chen, J. Yang, Y. Zhu, G. Xue, M. Li, 'Sensing human-screen interaction for energy-efficient frame rate adaptation on smartphones', IEEE Transactions on Mobile Computing, **14**(8), 1698–1711, 2015, doi:10.1109/TMC.2014.2352862.

[10] D. Kunda, M. Chishimba, 'A Survey of Android Mobile Phone Authentication Schemes', Mobile Networks and Applications, 1–9, 2018, doi:10.1007/s11036-018-1099-7.

[11] X. Zhao, T. Feng, W. Shi, I.A. Kakadiaris, 'Mobile user authentication using statistical touch dynamics images', IEEE Transactions on Information Forensics and Security, **9**(11), 1780–1789, 2014, doi:10.1109/TIFS.2014.2350916.

[12] N. Wakabayashi, M. Kuriyama, A. Kanai, 'Personal authentication method against shoulder-surfing attacks for smartphone', in 2017 IEEE International Conference on Consumer Electronics, ICCE 2017, 153–155, 2017, doi:10.1109/ICCE.2017.7889266.

[13] W. Meng, W. Li, D.S. Wong, J. Zhou, 'TMGuard: A touch movement-based security mechanism for screen unlock patterns on smartphones', in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 629–647, 2016, doi:10.1007/978-3-319-39555-5_34.

[14] W. Meng, Y. Wang, D.S. Wong, S. Wen, Y. Xiang, 'TouchWB: Touch behavioral user authentication based on web browsing on smartphones', Journal of Network and Computer Applications, **117**, 1–9, 2018, doi:10.1016/j.jnca.2018.05.010.

[15] M. Gao, X. Hu, B. Cao, D. Li, 'Fingerprint sensors in mobile devices', in Proceedings of the 2014 9th IEEE Conference on Industrial Electronics and Applications, ICIEA 2014, 1437–1440, 2014, doi:10.1109/ICIEA.2014.6931394.

[16] K. Cao, A.K. Jain, Hacking Mobile Phones Using 2D Printed Fingerprints, 2016.

[17] E. Vazquez-Fernandez, D. Gonzalez-Jimenez, 'Face recognition for authentication on mobile devices', Image and Vision Computing, **55**, 31–33, 2016, doi:10.1016/j.imavis.2016.03.018.

[18] W. Meng, W. Li, L. Jiang, J. Zhou, 'SocialAuth: Designing touch behavioral smartphone user authentication based on social networking applications', in IFIP Advances in Information and Communication Technology, 180–193, 2019, doi:10.1007/978-3-030-22312-0_13.

[19] S. Alotaibi, A. Alruban, S. Furnell, N. Clarke, 'A novel behaviour profiling approach to continuous authentication for mobile applications', in ICISSP 2019 - Proceedings of the 5th International Conference on Information Systems Security and Privacy, 246–251, 2019, doi:10.5220/0007313302460251.

[20] M. Ehatisham-ul-Haq, M. Awais Azam, U. Naeem, Y. Amin, J. Loo, 'Continuous authentication of smartphone users based on activity pattern recognition using passive mobile sensing', Journal of Network and Computer Applications, **109**, 24–35, 2018, doi:10.1016/j.jnca.2018.02.020.

[21] P. Kumar, R. Saini, P. Pratim Roy, D. Prosad Dogra, 'A bio-signal based framework to secure mobile devices', Journal of Network and Computer Applications, **89**, 62–71, 2017, doi:10.1016/j.jnca.2017.02.011.

[22] M. Smith-Creasey, F.A. Albalooshi, M. Rajarajan, 'Continuous face authentication scheme for mobile devices with tracking and liveness detection', Microprocessors and Microsystems, **63**, 147–157, 2018, doi:10.1016/j.micpro.2018.07.008.

[23] B. Kolosnjaji, A. Hufner, C. Eckert, A. Zarras, 'Learning on a Budget for User Authentication on Mobile Devices', in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2042–2046, 2018, doi:10.1109/ICASSP.2018.8461898.

[24] L. Wang, K. Huang, K. Sun, W. Wang, C. Tian, L. Xie, Q. Gu, 'Unlock with Your Heart', Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, **2**(3), 1–22, 2018, doi:10.1145/3264950.

[25] J.M. de Fuentes, L. Gonzalez-Manzano, A. Ribagorda, 'Secure and usable user-in-a-context continuous authentication in smartphones leveraging non-assisted sensors', Sensors (Switzerland), **18**(4), 2018, doi:10.3390/s18041219.

[26] Y. Song, Z. Cai, Z.L. Zhang, 'Multi-touch Authentication Using Hand Geometry and Behavioral Information', in Proceedings - IEEE Symposium on Security and Privacy, 357–372, 2017, doi:10.1109/SP.2017.54.

[27] L. Zhou, Y. Kang, D. Zhang, J. Lai, 'Harmonized authentication based on ThumbStroke dynamics on touch screen mobile phones', Decision Support Systems, **92**, 14–24, 2016, doi:10.1016/j.dss.2016.09.007.

[28] C. Bo, L. Zhang, X.Y. Li, Q. Huang, Y. Wang, 'SilentSense: Silent user identification via touch and movement behavioral biometrics', in Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM, ACM Press, New York, New York, USA: 187–189, 2013, doi:10.1145/2500423.2504572.

[29] U. Burgbacher, K. Hinrichs, 'An implicit author verification system for text messages based on gesture typing biometrics', in Conference on Human Factors in Computing Systems - Proceedings, 2951–2954, 2014, doi:10.1145/2556288.2557346.

[30] Y. Chen, J. Sun, R. Zhang, Y. Zhang, 'Your song your way: Rhythm-based two-factor authentication for multi-touch mobile devices', in Proceedings - IEEE INFOCOM, 2686–2694, 2015, doi:10.1109/INFOCOM.2015.7218660.

[31] H. Xu, Y. Zhou, M.R. Lyu, 'Towards Continuous and Passive Authentication via Touch Biometrics: An Experimental Study on Smartphones', 187–198, 2014.

[32] M. Frank, R. Biedert, E. Ma, I. Martinovic, D. Song, 'Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication', IEEE Transactions on Information Forensics and Security, **8**(1), 136–148, 2013, doi:10.1109/TIFS.2012.2225048.

[33] P. Saravanan, S. Clarke, D.H. Chau, H. Zha, 'LatentGesture: Active user authentication through background touch analysis', in ACM International Conference Proceeding Series, 110–113, 2014, doi:10.1145/2592235.2592252.

[34] F. He, L. Bao, R. Wang, J. Li, D. Xu, X. Zhao, 'A multimodal deep architecture for large-scale protein ubiquitylation site prediction', in Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017, 108–113, 2017.

[35] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature, **521**(7553), 436–444, 2015, doi:10.1038/nature14539.

[36] M.A. Wani, F.A. Bhat, S. Afzal, A.I. Khan, Introduction to Deep Learning, 1–11, 2020, doi:10.1007/978-981-13-6794-6_1.

[37] Y. Liang, Z. Cai, J. Yu, Q. Han, Y. Li, 'Deep Learning Based Inference of Private Information Using Embedded Sensors in Smart Devices', IEEE Network, **32**(4), 8–14, 2018, doi:10.1109/MNET.2018.1700349.

[38] L. Sun, Y. Wang, B. Cao, P.S. Yu, W. Srisa-An, A.D. Leow, 'Sequential Keystroke Behavioral Biometrics for Mobile User Identification via Multi-view Deep Learning', in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 228–240, 2017, doi:10.1007/978-3-319-71273-4_19.

[39] Y. Deng, Y. Zhong, Keystroke Dynamics Advances for Mobile Devices Using Deep Neural Network, 59–70, 2015, doi:10.15579/gcsr.vol2.ch4.

[40] M. Rampurawala, Classification with TensorFlow and Dense Neural Networks, Heart Beat, 2019.

[41] B. Zheng, S.W. Yoon, S.S. Lam, Y. Xiang, Y. Zhou, H. Liu, W.-L. Xiang, M.-Q. An, B. Subanya, P.G. Scholar, K. Srinivas, G.R. Rao, A.Govardhan, S. Shilaskar, A. Ghatol, E. Rashedi, H.N. Saeid, H. Nezamabadipour, K. Polat, G. Salih, D. Pal, K.M. Mandana, S. Pal, D. Sarkar, C. Chakraborty, H.S.N. Murthy, M.Meenakshi, R.A. Mohammadpour, S.M. Abedi, et al.,

Machine Learning Repository, Computers and Operations Research,Elsevier, **40**(November), 21–22, 2015.

[42] A.A. Bello, H. Chiroma, A.Y. Gital, L.A. Gabralla, S.M. Abdulhamid, L. Shuib, 'Machine learning algorithms for improving security on touch screen devices: a survey, challenges and new perspectives', Neural Computing and Applications, 2020, doi:10.1007/s00521-020-04775-0.

[43] A. De Luca, A. Hang, F. Brudy, C. Lindner, H. Hussmann, 'Touch me once and i know it's you! Implicit authentication based on touch screen patterns', in Conference on Human Factors in Computing Systems - Proceedings, 987–996, 2012, doi:10.1145/2207676.2208544.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 315-322 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Emotion Recognition on FER-2013 Face Images Using Fine-Tuned VGG-16

Gede Putra Kusuma*, Jonathan, Andreas Pangestu Lim

Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia

ARTICLE INFO

Article history:
Received: 25 August, 2020
Accepted: 21 October, 2020
Online: 10 November, 2020

Keywords:
Facial Emotion Recognition
FER-2013 Dataset
Convolutional Neural Network
VGG-16 Model
Unconstrained Setting

ABSTRACT

Facial emotion recognition is one among many popular and challenging tasks in the field of computer vision. Numerous researches have been conducted on this task and each proposed either standalone- or ensemble-based processing technique. While many researches strive for better accuracy, this research also attempts to increase the processing efficiency of computer correctly classifying human emotions based on human face by utilizing a single standalone-based neural network. This research proposes the use of standalone-based modified Convolutional Neural Network (CNN) based on Visual Geometry Group – 16 (VGG-16) classification model which was pretrained on ImageNet dataset and fine-tuned for emotion classification. The classification is performed on the publicly available FER-2013 dataset of over 35,000 face images with in-the-wild setting for 7 distinct emotions with the provided 80% training, 10% validation, and 10% testing data distributions. The proposed approach outperforms most standalone-based model results with 69.40% accuracy.

1. Introduction

Recognizing facial emotions is a challenging and interesting task, proven with numerous previously held competitions [1–8], available datasets [1, 9–14], and conducted researches on the subject. The progresses made for computer vision in recognizing emotion by face have been described in many publications [15–19]. A notable mention is the extensive paper by Li and Deng [19] which gives explanatory and thorough review about existing and commonly used datasets for facial emotion recognition, along with state-of-the-art (SOTA) approaches and their respective results.

The reviews from [15–19] described many available facial emotion recognition datasets, such as Facial Emotion Recognition 2013 (FER-2013) [1, 9], Static Facial Emotion in the Wild (SFEW) [10], Japanese Association of Female Facial Expression (JAFFE) [11], Cohn-Kanade (CK) [12], Extended Cohn-Kanade (CK+) [13], and Expression in the Wild (ExpW) [14] among many others. These datasets vary in many aspects, commonly described by one or combination of the following: amount of data, number of emotion classes, image- or sequential-based, and in lab-like or in-the-wild condition. The amount of data available in each dataset varies from hundreds to tens of thousands, with some datasets having pre-set training, validation, and/or testing data distribution. These datasets also vary in number of emotion classes, commonly ranging from six to eight emotion classes which include anger, disgust, fear, happiness, sadness, surprise, contempt, and is often appended with neutral emotion as well. Although some datasets do not include neutral and/or contempt expression. Image- or sequential-based (video-based or image-sequence-based) datasets also give variance in the conducted researches, as each dataset type's processing require different approach. Another difference is the condition of the dataset, where lab-like dataset differ from in-the-wild dataset. The former is recorded in an ideal condition (proper lighting, proper face alignment, and/or minimal to no usages of face accessories such as eyeglasses), whereas the latter depicts face expressions captured in a non-ideal condition from real-world scenarios.

These datasets are utilized in many works related to solving facial emotion recognition task, some of which uses conventional methods, deep learning, pretrained model, ensemble neural networks, combination of deep learning with hand-crafted feature selection technique(s), and other related works that will be described in following sections. These approaches achieve various results, which could be summarized in most researches' test accuracy for lab-like datasets often achieving above 90%, while test accuracy for in-the-wild datasets barely surpasses 75% [19].

With such low accuracy result from previous researches for facial emotion recognition with in-the-wild dataset (as compared

*Corresponding Author: Gede Putra Kusuma, inegara@binus.edu

with lab-like datasets), this paper attempts to improve facial emotion recognition accuracy for image-based in-the-wild dataset FER-2013. The proposed model achieved test accuracy of 69.40%, which is better than most existing researches with standalone-based neural network architecture, with the advantage of being simpler in terms of network depth and topology and having the capability of end-to-end training.

Continuing this introduction is a review of core concepts, brief description of FER-2013 as the used dataset in this paper's experiments, existing approaches for facial emotion recognition task using FER-2013 dataset, and the works that inspire and are the base of this paper's model. Afterwards, the proposed model is elaborated in detail. Then, the results from the proposed model are discussed and compared with other related works to understand how well the model performs compared to other taken approaches. Finally, this paper concludes with final thoughts about the proposed model and future works to improve facial emotion recognition task.

2. Literature Review

2.1. Studies and Developments

The study of emotion has gained interest since early 1970s, with one remarkable and frequently used research result being the work of Ekman and Friesen. They introduced six basic emotions: anger, disgust, fear, happiness, sadness, and surprise [20]. Many publications would later improve the known basic emotions, such as the work of Matsumoto which described contempt as another universal basic emotion [21]. Other improvements include the introduction of tools for supporting study of emotion recognition, such as the development of Facial Action Coding System (FACS) [22] with the updated version introduced in 2002 [23].

The interest in emotion study gained attention from computer science researchers as well, resulting in published researches on how to create automated emotion recognition system since the late 1980s. Even a survey on the developments made during those times [24]. These researches were conducted with broad and separate focus, such as establishing dataset [11, 12, 25], mapping features or feature selection process(es) [26, 27], designing architecture [27, 28], and adopting approaches from other fields of study [25, 27]. More recent studies adopted CNN for face emotion recognition, such as the approaches proposed in [29–31]

One, if not the most, attractive feature of emotion recognition is the human face itself. The main reason behind this is because mostly human face expresses emotions one feels [32]. The expression(s) is/are shown due to the work of face muscles' coordination which produces many expression configurations [33]. Note that the human face can express more than one emotion at the same time or at an approximately same time [34].

2.2. Facial Emotion Recognition Dataset (FER-2013)

For such an attractive feature, many datasets have been developed for facial emotion recognition task. As this paper only utilizes one dataset (FER-2013 [1,9]), it is the only dataset discussed in the paper.

Created in 2013, FER-2013 dataset contains 35,887 grayscale 48x48-pixel images, with the images being stored in a spreadsheet where each image's pixel values are stored in cells per row. The images were obtained utilizing Google search and are later grouped per emotion classes, which are anger, disgust, fear, happiness, neutral, sadness, and surprise. As the dataset was built utilizing Google search, the images are in in-the-wild condition, with even very few images being animated characters.

The dataset originally had data distribution of 28,709 images for training and 3,589 images for public test, but after the competition ended, another 3,589 images which were used for private test were added to the dataset. The usages of FER-2013's data distribution vary among published researches, with each using public test images for different purposes, either as part of training set, validation set, or test set. Several images contained in FER-2013 dataset are given in Figure 1.

Figure 1: FER-2013 sample images

2.3. Standalone-Based Neural Network (SBNN)

An architecture was proposed using CNN and batch normalization, in which the batch size was set constant to 256. The training process ran for 8 epochs as higher epochs was reported to result in overfitting [35]. Such architecture resulted in 60.12% test accuracy on FER-2013's private test data.

Other architectures using both CNN and batch normalization, varying in number of filters, were proposed in [36]. The uniqueness of both architectures is that both do not have any fully connected layer and maintain kernel size of 8, and with this design, both architectures achieved test accuracy above 65%.

Global average pooling (GAP) [37] was also used as an addition to both CNN and batch normalization [38]. The architectures vary in terms of depth-wise separable convolution layer, and although both were designed for multi-task purposes (emotion and gender recognition), both architectures achieved 66% accuracy.

Another new architecture inspired by network-in-network architecture and Inception was proposed with addition of utilizing a polynomial-based learning rate decreasing method for faster convergence and better performance [39]. The data used as input were pre-processed first for face feature extraction while still maintaining the image form, which resulted in a maximum of 10% increase in accuracy. With the proposed architecture and data pre-processing, the test result for FER-2013 dataset reached 66.4% after training for 200 epochs with batch size of 250, noting that the test data used is unknown to be either public or private test data.

Two SBNN architectures, one using custom CNN and another using One-vs-All (OVA) Support Vector Machine (SVM) with Histogram of Oriented Gradients (HOG) which resulted in 66.67%

and 45.95% accuracy on test data of FER-2013 dataset respectively were also introduced [40].

A new SBNN architecture based on attentional convolutional neural networks was proposed in [41]. The proposed model was able to focus on only face features that are important to emotion recognition, achieving 70.02% test accuracy on FER-2013's private test data.

The winner of FER-2013 challenge, Yichuan Tang, proposed SBNN architecture with L2-SVMs (DLSVM) as replacement for softmax layer at the end of the neural network architecture. Such modification resulted in the proposed model reaching 71.2% test accuracy on private test dataset [42].

The current SOTA SBNN architecture was proposed in [43], in which they theorized that features at mid-level and high-level of a CNN would have impact on the final prediction. They based their model, termed as Multi-Level Convolutional Neural Network (MLCNN), on VGG architecture with 18 layers and achieved 73.03% test accuracy on FER-2013's private test data.

2.4. Ensemble-Based Neural Network (EBNN)

An EBNN architecture with 3 subnets, containing 3 and increasing convolution layers respectively for each subnet was proposed [44]. With batch size of 100, momentum value of 0.9, learning rate value of 0.01 decreasing to 0.001 by validation error condition, and running from varying epochs of 20 to 100, and 80% and 20% data distribution for training and test set respectively, they reached 65.03% test accuracy.

Referred in their paper as aggregator or hybrid CNN–SIFT aggregator, [45] combined CNN with Scale Invariant Feature Transform (SIFT), specifically Dense SIFT and regular SIFT. The EBNN, constructed and trained from scratch with combining CNN, Dense SIFT, and regular SIFT, resulted in 73.4% test accuracy on FER-2013 private test data.

Another researcher proposed an EBNN coined as hierarchical committee for facial emotion recognition task with 216 retrained deep convolutional neural network (DCN) models using transfer learning and 108 DCN models trained on aligned faces [46]. This method achieved 70.58% test accuracy on FER-2013's private test dataset, with randomly splitting training data for training and validation purposes after removing 11 images from the training data. The work is further developed in [47] which consisted of 9 DCNs, with each having 3 convolution-max pooling stages and ended with 2 fully connected layers. By using both aligned and non-aligned face structure dataset in training each DCN and using Alignment-Mapping Network (AMN) with the ensemble, they achieved 73.73% accuracy on FER-2013 private test data with the same training and validation data distribution as used in [46].

Another proposed EBNN architecture was offered in Pramerdorfer and Kampel's work [48]. The ensemble consisted of 8 DCN models based on VGG, Inception, and ResNet, where each base model underwent convolution and/or pooling layer(s) removal. They managed to achieve 75.2% test accuracy on FER-2013 private test data.

Currently known SOTA EBNN architecture was proposed in [49], which combined 3 VGG models, 2 pre-trained models which were retrained, and 1 model trained from scratch, combined with SIFT and k-means clustering, and flattened extracted feature vector as input to local SVM. This approach achieved 75.42% test accuracy on FER-2013's private test data.

2.5. Other Researches

There are other notable researches that inspire the proposed model which impacts the proposed model's architecture and learning and/or generalization capability. These researches are reviewed in this section in the order of previously stated impact descriptions, with the purpose of giving brief and proper information regarding the researches' results prior to elaborating the proposed model.

The architecture of the model is based on VGG-16 architecture [50], which contains 13 convolution layers with 3 additional fully connected layers at the end of the network. After each convolution layer, ReLU [51] layer is used for non-linearity rectification. Table 1 describes VGG-16 architecture.

Table 1: VGG-16 architecture

Type	Layer
Feature Extractor	Convolution
	Convolution
	Max Pooling
	Convolution
	Convolution
	Max Pooling
	Convolution
	Convolution
	Convolution
	Max Pooling
	Convolution
	Convolution
	Convolution
	Max Pooling
	Convolution
	Convolution
	Convolution
	Max Pooling
Classifier	Dense/Fully Connected (4096)
	Dense/Fully Connected (4096)
	Dense/Fully Connected (1000)
	Softmax

However, the existence of fully connected layers in a convolutional neural network architecture may lead to higher overfitting probability [37]. To overcome this, the authors in [37] proposed GAP which works by averaging each of final feature map results and directly using the averaged results for classification. This pooling method regularizes feature maps as confidence map for categories with no parameter requiring optimization, hence avoids overfitting.

Another attempt to reduce overfitting possibility is the use of early stopping [52]. Such implementation affects the training mechanism as the training may stop before fully reaching the specified number of epochs due to no further increase in accuracy or decrease in loss in a specific period. Other than using early stopping to avoid overfitting, a proper optimizer is required for better model learning and/or generalization capability.

One most used and popular optimizer is Stochastic Gradient Descent (SGD) [53, 54], which takes constant parameter learning rate and traverses the gradient by the specified learning rate to find global minima. With SGD, it was reported that the generalization capability of a model improves [55, 56], hence less overfitting probability. Another development for SGD was using momentum as additional parameter which accelerates model learning process, which has been proven in practice in [57]. However, SGD is often slow for learning process since the traversing process is done uniformly by the learning rate value.

An alternative to SGD was proposed in [58] and coined as Adam to overcome SGD's limitation. With Adam, the learning process adapts the learning rate parameter as the learning process continues, resulting in better learning performance compared to SGD. The major drawback of Adam is the generalization capability drops significantly as reported in [56].

In an effort to overcome Adam optimizer's drawback, Keskar and Socher proposed SWATS (Switching from Adam to SGD) [59], which at first uses Adam optimizer to have better initial learning process then switches to SGD on specified trigger condition. They empirically argued that SWATS optimizer gives better performance on most tasks, especially visual- and text-based tasks, by combining both Adam and SGD.

3. Proposed Model

Our proposed model was defined to be based on VGG-16 as the base model and using GAP as final pooling layer prior to VGG-16's classifier. The best model was obtained after experimenting on multiple combinations of SGD/Adam/SWATS as optimizer and inclusion/non-inclusion of early stopping. The researchers made some modifications for VGG-16's classifier due to the use of GAP.

VGG-16 was chosen as the base model for several reasons. The first reason is because VGG-16 uses convolution layer with 3x3 kernel instead of 5x5 or 7x7, hence lesser parameters to train [50]. The second reason is by using two or three stacked convolution layers with 3x3 kernel as an equal for one convolution layer with 5x5 or 7x7 kernel respectively, VGG-16's architecture adds more rectification layers, which results in a decision function that is more discriminative [50]. The third reason is based on empirical findings as the researchers reviewed previous publications, in which the researchers found that several SBNN and EBNN researches used VGG based models and achieved good, with some achieving SOTA, results on FER-2013 dataset among other tested datasets. While VGG-13, VGG-16, and VGG-19, either in its original architecture or modified, are present in the found literature reviews, the researchers decided to go with VGG-16 since it is not too deep and complex as VGG-19, noting that both architecture demonstrated similar results as reported in [50]. The researchers also agreed to not use VGG-13 as Simonyan and Zisserman stated in their paper that increase in depth in networks results in better accuracy [50]. In addition to the architecture selection, the researchers theorized that the use of pretrained model may improve accuracy and reduce training time. The researchers used VGG-16's pretrained model which was trained on ImageNet [60] dataset rather than initializing the network's weights randomly.

The researchers also modified VGG-16's architecture by using GAP as final pooling layer before entering VGG-16's classifier, and therefore the researchers replaced all of VGG-16's classifier layers with a single fully connected layer, taking 512 averaged input neurons and outputting 7 neurons as the equivalent number of FER-2013's emotion classes. Our proposed model's architecture can be seen in Table 2.

With the proposed model's architecture, the researchers experimented on multiple combinations of training mechanism, optimizer choices, data distribution, frozen layers, and batch normalization usage. Our best model was trained using SGD optimizer and used early stopping in its training process. Such combination proved to be superior as compared with the researchers other experiments. The results of the experiments are discussed in the next section.

Table 2: Modified VGG-16 architecture

Type	Layer
Feature Extractor	Convolution
	Convolution
	Max Pooling
	Convolution
	Convolution
	Max Pooling
	Convolution
	Convolution
	Convolution
	Max Pooling
	Convolution
	Convolution
	Convolution
	Max Pooling
	Convolution
	Convolution
	Convolution
	Global Average Pooling
Classifier	Dense/Fully Connected (7)

4. Experiments

4.1. Experimental Design

The researchers experimented on combining VGG-16 with several optimizers and training method. The used optimizers were SGD, Adam, and SWATS, and the training method was customized to use early stopping. The experiment was conducted using PyTorch and ran on Google Colab with GPU acceleration support. Due to the training data's size and Google Colab's lifetime, the researchers had to perform run-and-pause per 10 epochs as part of the training mechanism to prevent sudden disconnect from the machine, since in such an event the training process may stop prematurely. Note that the training and validation were done in one epoch. After each run-and-pause, the test was performed per 10 epochs on the model and the results are recorded.

Initially, the researchers had several options to experiment on. The options include the use/disuse of batch normalization, freezing/not freezing selected layers, imbalanced (original)/balanced training data distribution, use/disuse of GAP, and SGD/Adam/SWATS optimizer selection. All in all, the researchers initially estimated the need to experiment on approximately 12 combinations of VGG-16 model. Due to the

quite large number of possible combinations and since each epoch takes about 7-10 minutes, the researchers had to estimate which combination merits further investigation per specified run epochs. After each combination has passed the specified run epochs, the combination that provided the highest test accuracy was used to undergo further training and/or further combination. At the end of the experiments, the researchers experimented on 23 combinations, with all combinations having equivalent hyperparameter values. The researchers used learning rate value of 0.001, 0.9 for momentum on SGD optimizer, and batch size of 32.

The results of the experiments are given in Table 3. For better readability, the following acronyms are used in Table 3:

- **EG**: Experiment group
  The researchers group the experiments to select the best model in each experiment group, and what the next experiment group's models' possible combinations are.
- **D**: Data distribution
  FER-2013 dataset is very imbalanced, with some classes having more than 5,000 images and some having less than 1,000 images. The researchers tried using both balanced and imbalanced FER-2013 dataset to see which would yield better results. The acronyms for the experiments' FER-2013 data distribution are as follows:
  - *B: Balanced*
  - *I: Imbalanced*
- **BN**: Use batch normalization
  The use of batch normalization should help the model learn in a more stable manner, yet some of the best performing SBNN models do not use batch normalization. To this extent the researchers included the combination of using and discarding batch normalization in the model.
- **GAP**: Use global average pooling (GAP)
  GAP layer helps aggregate feature map information and helps prevent overfitting caused by fully connected layers in the model. For this reason, the researchers tried the combination of using and not using GAP in the experiments. Note that although GAP can be used as the classifying layer, the researchers appended a fully connected layer instead as the classifying layer for GAP's aggregated feature map.
- **LL**: Learning layers
  Finetuning a pretrained model can be done with freezing some layers except for the new classifier layer(s). Our experiments freeze some layers and keep others to be able to learn (learning layers). The following acronyms show possible combinations of learning layers in the experiments:
  - *F447: Fully connected layers with neurons structure 4096-4096-7*
  - *F7: Fully connected layer with 7 neurons*
  - *1C: 1-Last convolution layer and all following fully connected layers*
  - *2C: 2-Last convolution layers and all following fully connected layers*
  - *3C: 3-Last convolution layers and all following fully connected layers*
  - *A: All layers*
- **OPT**: Optimizer
  The researchers selected three optimizers to be used in the experiments, taking into consideration the generalization capability and convergence speed of each optimizer. The three optimizers used in the experiments are:
  - *SGD*
  - *Adam*
  - *SWATS*
- **EP**: Benchmarking epoch
  Measuring the performance of each experiment group must be done using the same settings to have a fair comparison, which in this case being the number of epochs of each model's training.
- **TR**: Test result
  The obtained accuracy in percent on the test set of FER-2013 dataset.
- **M**: Model series number
  To easily refer to a specific model, the researchers number each model. Note that italicized series number means the model is used in multiple experiment groups.

4.2. Experimental Results

EG1 denotes the first experiments. All combinations in EG1 underwent training for 10 epochs due to the large number of combinations. The best result from this experiment turned out to be model with model series number 7 (M7), using balanced training data distribution with no batch normalization, GAP layer, and SGD optimizer with 64.11% test accuracy result.

From the best result of EG1, the researchers decided to perform the next experiment (EG2) on balanced training data with unfreezing 1- and 2-last convolution layer (denoted 1C and 2C in Table 3, respectively) since the researchers hypothesized that with the use of GAP, training the output (fully connected) layer does not do much on improving the model's generalization capability. In addition to the experiment on balanced training data, the researchers tried running an experiment for imbalanced data with only the output layer being unfrozen. The researchers also extended the number of epochs run from 10 to 30 in EG2. The result was out of what the researchers had thought, with the combination utilizing imbalanced training data having the highest test accuracy of 67.26% (M15), and the combination for balanced data and unfrozen 2-last convolution layer (M14) coming in second with 67.20% test accuracy.

The researchers noticed how M14 had a close accuracy gap with M15 (EG2's best model). The researchers then slightly deviated the experiment in EG3 with the thought of using both balanced and imbalanced dataset on no frozen layers. The researchers unfroze all layers in the proposed model and trained the networks for 30 epochs. The results show that using SGD with imbalanced training data and no frozen layer gave the best result of all the experiments, with 69.46% test accuracy, which is M17. The researchers also noted how the combination that uses balanced training data could not compete with the combination utilizing imbalanced training data.

Considering our last observation on EG3, EG4 consisted of experiments on imbalanced dataset with SGD optimizer, and the researchers tried to reduce the number of unfrozen layers by freezing only 3-last convolution layers (3C in Table 3) in hopes of speeding training time by reducing trainable parameter count. Then the researchers compared the combination's result with M15

and M17. Once again, M17 which was built by unfreezing all layers in the network and using imbalanced dataset produced the best accuracy result.

With the best combination now obtained after four experiment groups, EG5 established the comparison for such combination (imbalanced training data, no batch normalization, using GAP, and no frozen layer) but with different optimizers. The researchers also tried to use early stopping in this experiment group, with the early stopping being effective per epoch 30. The researchers set the minimum passing point of the training epoch to be 30 to have the models learn at first and not getting cut out by the early stopping mechanism. The early stopping mechanism was also executed if the validation accuracy does not improve for last 10 epochs since the last best recorded validation accuracy. To this extent, note that benchmarking epoch (EP in Table 3) refers to last epoch for each model after early stopping was effective. What was unique from this experiment group is that M17 had gotten better validation accuracy but lower test accuracy, proven with the decreasing test accuracy in EG5 as compared to EG3 and EG4's results.

After EG5, the researchers attempted to do training without early stopping, to see whether the model may be caught in local minima when early stopping was enabled. The experiments which did not use early stopping are grouped as EG6. EG6 proved our assumption wrong about our model being caught in local minima with the use of early stopping, as all three models in EG6 did not show much improvement after passing 30 epochs, nor after the benchmarking epoch in EG5. The results of EG6 also show the same findings as reported in EG5, in which the model using SGD optimizer with the specified hyperparameters value in Table 3 reigned triumphant compared to other models in EG6.

With the obtained best model from the experiments, the researchers provide their best model's test accuracy comparison with other works' test accuracy on Table 4. Note that the compared models are those from SBNN group as the proposed model is categorized as SBNN model.

Our proposed model outperforms most proposed approaches by a significant margin of 3-24%. The researchers argue that the capability of VGG-16 is adequate as feature extractor for facial emotion recognition and is further enhanced by the addition of GAP layer prior to classifier (fully connected) layer. GAP's addition is observed to greatly increase test accuracy, as the experiments have shown in Table 3 and shown by [38]. Unfreezing all layers in pretrained VGG-16 model also results in the best performance. The researchers believe that the model learns to better extract features as convolutional layers are learning as well during training instead of just letting the fully connected layers learn the preferred mapping. In addition to the proposed model's capability, the use of SGD optimizer benefits the proposed model as reported in [56] where SGD is reported to yield better performance on unseen data. What is unique is the use of Adam and SWATS optimizer without batch normalization leads to worst performances as shown in Table 3, and yet using batch normalization with SGD does not yield the best results.

Our proposed model also achieves similar, albeit lower, test accuracy as compared to the approaches from [40,41] with less model complexity. This would be beneficial since utilizing a simpler model may result in fewer required resources and faster performance in real-life scenarios. Using simpler model would also open further research on the possibilities of using simpler models when pursuing other objectives.

Table 3: Experiment results

EG	D	BN	GAP	LL	OPT	EP	TR	M
1	B	✓	✓	F7	SGD	10	51.85	1
			×	F447			60.29	2
			✓	F7	Adam		58.01	3
			×	F447			56.25	4
			✓	F7	SWATS		57.70	5
			×	F447			57.81	6
		×	✓	F7	SGD		**64.11**	***7***
			×	F447			63.02	8
			✓	F7	Adam		18.19	9
			×	F447			24.93	10
			✓	F7	SWATS		24.93	11
			×	F447			24.93	12
2	B	×	✓	F7	SGD	30	64.92	***7***
				1C			64.27	13
				2C			67.20	14
	I			F7			**67.26**	***15***
3	B	×	✓	N	SGD	30	66.48	16
	I						**69.46**	***17***
4	I	×	✓	F7	SGD	30	67.26	***15***
				3C			68.12	18
				N			**69.46**	***17***
5	I	×	✓	N	SGD	46	**69.40**	***17***
					Adam	30	60.57	19
					SWATS		57.09	20
6	I	×	✓	N	SGD	100	**69.15**	21
					Adam		59.87	22
					SWATS		57.75	23

Table 4: SBNN results comparison

Related Works	Proposed Method	Test Accuracy
[35]	CNN + Batch Normalization	60.12%
[36]	CNN + Batch Normalization + Varying number of filters	65%

[38]	CNN + Batch Normalization + GAP	66%
[39]	New architecture + polynomial learning rate	66.4%
[40]	Custom CNN	66.67%
	One-vs-All (OVA) SVM	45.95%
Our model	**VGG-16 + GAP**	**69.40%**
[41]	New architecture based on attentional CNN	70.02%
[42]	SBNN + L2-SVMs (DLSVM)	71.2%
[43]	Multi-Level CNN (MLCNN)	73.03%

5. Conclusions

In this paper, the researchers report their experiments using FER-2013 dataset as the data source. Several researches using FER-2013 are also reported and they used different kinds of approaches, which the researchers group as standalone-based neural network (SBNN) and ensemble-based neural network (EBNN) approaches. The proposed network is classified as SBNN with VGG-16 as the base model, which was modified into using 13 convolutional layers and GAP as the last pooling layer. The network is then experimented on by varying several variables, such as data distribution, use/disuse of batch normalization, GAP, optimizers choice, and freezing some layers. The researchers experimented on 23 models and found out that using an imbalanced dataset, GAP, non-frozen layer, and SGD optimizer results in the highest accuracy throughout the experiments, which is 69.40%. With this result, the network has surpassed most reported networks defined in Table 4 by quite a large margin. In addition, the model supports end-to-end training and is much simpler as compared to other three best SBNN models [41–43], which in return results in lower time and memory consumption.

Further investigations could be done regarding hyperparameter tuning as the researchers experimented on the variation of data distribution, pooling layer, and optimizer selection only. Other improvement includes using other existing models or developing new architecture with similar topology as the proposed model for better classifying capability. With the proposed model's result, the researchers believe that further modifications may lead to better performance in classifying emotions with in-the-wild settings.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] I.J. Goodfellow, D. Erhan, P. Luc Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," Neural Networks, **64**(2015), 59–63, 2015, doi:10.1016/j.neunet.2014.09.005.

[2] A. Dhall, R. Goecke, J. Joshi, M. Wagner, T. Gedeon, "Emotion Recognition in the Wild Challenge 2013," in Proceedings of the 15th ACM on International Conference on Multimodal Interaction, 2013.

[3] A. Dhall, R. Goecke, J. Joshi, K. Sikka, T. Gedeon, "Emotion recognition in the wild challenge 2014: Baseline, data and protocol," in ICMI 2014 - Proceedings of the 2014 International Conference on Multimodal Interaction, 461–466, 2014, doi:10.1145/2663204.2666275.

[4] A. Dhall, O. V. Ramana Murthy, R. Goecke, J. Joshi, T. Gedeon, "Video and image based Emotion recognition challenges in the wild: EmotiW 2015," in ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction, 423–426, 2015, doi:10.1145/2818346.2829994.

[5] A. Dhall, R. Goecke, J. Joshi, T. Gedeon, "Emotion recognition in the wild challenge 2016," in ICMI 2016 - Proceedings of the 18th ACM International Conference on Multimodal Interaction, 587–588, 2016, doi:10.1145/2993148.3007626.

[6] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, T. Gedeon, "From individual to group-level emotion recognition: EmotiW 5.0," in Proceedings of the 19th International Conference on Multimodal Interaction, 2017.

[7] A. Dhall, A. Kaur, R. Goecke, T. Gedeon, "EmotiW 2018: Audio-Video, Student Engagement and Group-Level Affect Prediction," in Proceedings of the 20th ACM International Conference on Multimodal Interaction, 2018, doi:10.1145/3242969.3264993.

[8] A. Dhall, S. Ghosh, R. Goecke, T. Gedeon, "EmotiW 2019: Automatic emotion, engagement and cohesion prediction tasks," in ICMI 2019 - Proceedings of the 2019 International Conference on Multimodal Interaction, 546–550, 2019, doi:10.1145/3340555.3355710.

[9] P.-L. Carrier, A. Courville, Challenges in representation learning: Facial expression recognition challenge, 2013.

[10] A. Dhall, R. Goecke, S. Lucey, T. Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," in Proceedings of the IEEE International Conference on Computer Vision, 2106–2112, 2011, doi:10.1109/ICCVW.2011.6130508.

[11] M. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, "Coding facial expressions with Gabor wavelets," in Proceedings - 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998, 200–205, 1998, doi:10.1109/AFGR.1998.670949.

[12] T. Kanade, J.F. Cohn, Y. Tian, "Comprehensive database for facial expression analysis," in Proceedings - 4th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2000, 46–53, 2000, doi:10.1109/AFGR.2000.840611.

[13] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010, 94–101, 2010, doi:10.1109/CVPRW.2010.5543262.

[14] Z. Zhang, P. Luo, C.C. Loy, X. Tang, "From Facial Expression Recognition to Interpersonal Relation Prediction," International Journal of Computer Vision, **126**(5), 550–569, 2018, doi:10.1007/s11263-017-1055-1.

[15] B.C. Ko, "A brief review of facial emotion recognition based on visual information," Sensors (Switzerland), **18**(2), 2018, doi:10.3390/s18020401.

[16] A.G. Sanjay Kumar, "Facial Expression Recognition: A Review," in National Conference on Cloud Computing and Big Data, 159–162., 2015.

[17] E. Owusu, E.K. Gavua, Z. Yong-Zhao, "Facial Expression Recognition – A Comprehensive Review," International Journal of Technology and Management Research, **1**(4), 29–46, 2020, doi:10.47127/ijtmr.v1i4.36.

[18] N. Samadiani, G. Huang, B. Cai, W. Luo, C.H. Chi, Y. Xiang, J. He, "A review on automatic facial expression recognition systems assisted by multimodal sensor data," Sensors (Switzerland), **19**(8), 2019, doi:10.3390/s19081863.

[19] S. Li, W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, 2020, doi:10.1109/TAFFC.2020.2981446.

[20] P. Ekman, W. V. Friesen, "Constants across cultures in the face and emotion," Journal of Personality and Social Psychology, **17**(2), 124–129, 1971, doi:10.1037/h0030377.

[21] D. Matsumoto, "More evidence for the universality of a contempt expression," Motivation and Emotion, **16**(4), 363–368, 1992, doi:10.1007/BF00992972.

[22] P. Ekman, W. V. Friesen, Facial Action Coding System: Manual, 1978.

[23] P. Ekman, W. V. Friesen, J.C. Hager, Facial Action Coding System. Manual and Investigator's Guide, 2002.

[24] A. Samal, P.A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," Pattern Recognition, **25**(1), 65–77, 1992.

[25] I.A. Essa, A.P. Pentland, "Facial expression recognition using a dynamic

model and motion energy," in IEEE International Conference on Computer Vision, 360–367, 1995, doi:10.1109/iccv.1995.466916.
[26] G.W. Cottrell, J. Metcalfe, "EMPATH: face, emotion, and gender recognition using holons," Nips, **1**(1987), 564–571, 1990.
[27] Z. Zhang, M. Lyons, M. Schuster, S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," in Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998.
[28] M. Pantic, L.J.M. Rothkrantz, "Expert system for automatic analysis of facial expressions," Image and Vision Computing, **18**(11), 881–905, 2000, doi:10.1016/S0262-8856(00)00034-2.
[29] P. Tarnowski, M. Kołodziej, A. Majkowski, R.J. Rak, "Emotion recognition using facial expressions," Procedia Computer Science, **108C**, 1175–1184, 2017.
[30] J. Lee, S. Kim, S. Kim, J. Park, K. Sohn, "Context-aware emotion recognition networks," in Proceedings of the IEEE International Conference on Computer Vision, 10142–10151, 2019, doi:10.1109/ICCV.2019.01024.
[31] M.S. Hossain, G. Muhammad, "Emotion recognition using deep learning approach from audio–visual emotional big data," Information Fusion, **49**(October), 69–78, 2019, doi:10.1016/j.inffus.2018.09.008.
[32] I. Cohen, N. Sebe, A. Garg, L.S. Chen, T.S. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," Computer Vision and Image Understanding, **91**(1–2), 160–187, 2003, doi:10.1016/S1077-3142(03)00081-X.
[33] A. Martinez, S. Du, "A model of the perception of facial expressions of emotion by humans: Research overview and perspectives," Journal of Machine Learning Research, **13**(2012), 1589–1608, 2012, doi:10.1007/978-3-319-57021-1_6.
[34] S. Du, A.M. Martinez, "Compound facial expressions of emotion: From basic research to clinical applications," Dialogues in Clinical Neuroscience, **17**(4), 443–455, 2015, doi:10.31887/dcns.2015.17.4/sdu.
[35] I. Talegaonkar, K. Joshi, S. Valunj, R. Kohok, A. Kulkarni, "Real Time Facial Expression Recognition using Deep Learning," SSRN Electronic Journal, 2019, doi:10.2139/ssrn.3421486.
[36] A. Agrawal, N. Mittal, "Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy," Visual Computer, **36**(2), 405–412, 2020, doi:10.1007/s00371-019-01630-9.
[37] M. Lin, Q. Chen, S. Yan, "Network in network," 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings, 1–10, 2014.
[38] O. Arriaga, M. Valdenegro-Toro, P.G. Plöger, "Real-time convolutional neural networks for emotion and gender classification," ESANN 2019 - Proceedings, 27th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 221–226, 2019.
[39] A. Mollahosseini, D. Chan, M.H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," 2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016, 2016, doi:10.1109/WACV.2016.7477450.
[40] M. Quinn, G. Sivesind, G. Reis, "Real-time Emotion Recognition From Facial Expressions," 2017.
[41] S. Minaee, A. Abdolrashidi, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network," 2019.
[42] Y. Tang, "Deep Learning using Linear Support Vector Machines," 2013.
[43] H.-D. Nguyen, S. Yeom, I.-S. Oh, K.-M. Kim, S.-H. Kim, "Facial expression recognition using a multi-level convolutional neural network," Proceedings from the International Conference on Pattern Recognition and Artificial Intelligence, 217–221, 2018.
[44] K. Liu, M. Zhang, Z. Pan, "Facial Expression Recognition with CNN Ensemble," Proceedings - 2016 International Conference on Cyberworlds, CW 2016, 163–166, 2016, doi:10.1109/CW.2016.34.
[45] T. Connie, M. Al-Shabi, W.P. Cheah, M. Goh, "Facial expression recognition using a hybrid CNN-SIFT aggregator," Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **10607 LNAI**, 139–149, 2017, doi:10.1007/978-3-319-69456-6_12.
[46] B.K. Kim, J. Roh, S.Y. Dong, S.Y. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," Journal on Multimodal User Interfaces, **10**(2), 173–189, 2016, doi:10.1007/s12193-015-0209-0.
[47] B.K. Kim, S.Y. Dong, J. Roh, G. Kim, S.Y. Lee, "Fusing Aligned and Non-aligned Face Information for Automatic Affect Recognition in the Wild: A Deep Learning Approach," IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 1499–1508, 2016, doi:10.1109/CVPRW.2016.187.
[48] C. Pramerdorfer, M. Kampel, "Facial Expression Recognition using Convolutional Neural Networks: State of the Art," 2016.
[49] M.I. Georgescu, R.T. Ionescu, M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition," IEEE Access, **7**(May), 64827–64836, 2019, doi:10.1109/ACCESS.2019.2917266.
[50] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 1–14, 2015.
[51] V. Nair, G.E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," Conference: Proceedings of the 27th International Conference on Machine Learning, 2010.
[52] L. Prechelt, "Early Stopping — But When?," **0**(1998), 53–67, 2012, doi:10.1007/978-3-642-35289-8_5.
[53] H. Robbins, S. Monro, "A Stochastic Approximation Method," The Annals of Mathematical Statistics, **22**(3), 400–407, 1951, doi:10.1214/aoms/1177729586.
[54] J. Kiefer, J. Wolfowitz, "Stochastic Estimation of The Maximum of a Regression Function," The Annals of Mathematical Statistics, **23**, 1952.
[55] M. Hardt, B. Recht, Y. Singer, "Train faster, generalize better: Stability of stochastic gradient descent," 33rd International Conference on Machine Learning, ICML 2016, **3**, 1868–1877, 2016.
[56] A.C. Wilson, R. Roelofs, M. Stern, N. Srebro, B. Recht, "The marginal value of adaptive gradient methods in machine learning," Advances in Neural Information Processing Systems, **2017-Decem**(Nips), 4149–4159, 2017.
[57] I. Sutskever, J. Martens, G. Dahl, G. Hinton, "On the importance of initialization and momentum in deep learning," ICML'13: Proceedings of the 30th International Conference on International Conference on Machine Learning, **28**, 1139–1147, 2013.
[58] D.P. Kingma, J.L. Ba, "Adam: A method for stochastic optimization," 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 1–15, 2015.
[59] N.S. Keskar, R. Socher, "Improving Generalization Performance by Switching from Adam to SGD," (1), 2017.
[60] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," (June), 248–255, 2010, doi:10.1109/cvpr.2009.5206848.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 323-330 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Multi-Model Security and Social Media Analytics of the Digital Twin

Jim Scheibmeir[1], Yashwant Malaiya[2]

[1]*Colorado State University, Systems Engineering, 80523, United States*

[2]*Colorado State University, Computer Science, 80523, United States*

ARTICLE INFO

Article history:
Received: 28 August, 2020
Accepted: 28 October, 2020
Online: 10 November, 2020

Keywords:
Internet of Things
Digital Twin
Cybersecurity
Contact Tracing
Social Media Analytics

ABSTRACT

Digital twins act through application programming interfaces to their physical counterparts to monitor, model, and control them. Beyond these traditional functions of digital twins, they must also act to secure their physical counterparts. A multi-model scheme is presented to help digital twins towards the task of securing the physical system. Additionally, this work includes an analysis of more than four hundred thousand tweets each relating to digital twin technology and cybersecurity which were collected during June and July in 2020. Of the first corpus of tweets collected by searching for #digitaltwin during the research period, only a small population of 10% reference security concepts. In the second and larger corpus of collected tweets, the top mentioned industries were health, education, and public. A naïve Bayes model reached a 70.3% accuracy at differentiating tweets that were either related to cybersecurity or the internet of things. The study also indicates that cybersecurity tweets are consistently more negative in many areas of sentiment when compared to tweets about the internet of things. The sentiment findings of cybersecurity tweets will reinforce the need to address culture in cybersecurity posture while the security multi-model schema contributes to the state of the art.

1. Introduction

An API development model for digital twins has been proposed recently [1]. This paper augments that work to address the cybersecurity concerns of the digital twin within the context of the various implementations and industries where the twin may be utilized. Furthermore, the adoption of current security models, such as state machine, lattice, non-interference, and information flow models are proposed for digital twins.

In the original paper, a development model for the APIs of the digital twin was created. The model was proposed to establish the context of the environment, the system, relationships between the two, and the functional and non-functional requirements.

This research includes the mining of Twitter tweets using appropriate search criteria for digital twins, the internet of things (IoT), and cybersecurity. The analysis includes mention analysis of industries, exploration of the sentiment, natural language processing (NLP) implementation towards the classification of tweets, and modeling to predict the twitter user behavior.

*Corresponding Author: Jim Scheibmeir, Email: jimscheibmeir@gmail.com

Three sets of tweets were collected for this study with the latter two sets being joined into a combined corpus. Each set was collected through the execution of scripts written in R. The first collection is smaller and has 3,102 tweets about digital twins. The second and third collections are larger. In total, 422,963 industrial internet of things (IIoT) tweets and 497,174 cybersecurity tweets were collected. These larger sets were combined into one corpus for analysis. The search criteria for the tweet collections are listed in Table 1.

Table 1: Tweet collection search criteria

Script Topic	Criteria used in actual R code
Digital Twin	"#digitaltwin"
IIoT	"#industry40 OR #IIOT OR #IOT"
Cybersecurity	"#cybersecurity OR #infosec OR #hacking"

This study is organized as follows. Section 2 reviews the literature and the background information on the subject. Section 3 describes our research method, including the collection of tweets as well as the processing. Sections 4 and 5 present the results from

social media analysis. Section 6 proposes a multi-model security scheme for digital twins. Finally, section 7 presents the conclusion of this study.

2. Background and related work

Digital twins are a logical conduit between physical entities and humans, geared towards monitoring and controlling a system [2]. The design of the interfaces themselves may utilize structured standards such as OpenAPI to list the various objects, paths, and operations [3].

Digital twins can be used during system design for vulnerability detection or during physical system operations to assist with intrusion detection [4]. Digital twins can utilize and manage the Internet of Things (IoT) devices and IoT systems of systems [5]. Wireless Sensor Networks (WSN) and attached IoT sensor components can be energy and cost-effective monitoring approaches [6]. WSNs are also frequently constrained to limited computing and power resources and thus security in WSNs can be challenged [7]. To maintain the integrity of the sensor data and overall system state awareness, secure links with proper authentication must be established [8].

Vulnerable digital twins offer hackers a blueprint of the physical counterpart as well as other backend systems that may be a part of system integration [9]. The digital twin can help secure the physical systems or be yet another vulnerability as IoT growth has created a large cyber-attack surface [10]. IoT has revealed security vulnerability ranging from trust management, authentication, privacy, to access control of embedded systems [11]. While digital twins may present such risk, they are still getting popular. The smart factory market, an economy of integrated automation solutions adopted to streamline manufacturing, is predicted to be worth approximately $205 billion-dollar by 2022 [9]. Hearn and Rix also cite another 2019 study whereof 220 security leaders in industrial and manufacturing who were a part of the research survey, 79% indicated they had experienced an IoT cyberattack within that past year. IoT and digital twins-based systems can be attractive targets for malicious actors.

Hearn and Rix identify advantages, such as the prevention of downtime and monitoring attacks against cyber-physical systems, as potential benefits of digital twins [9]. Risks of digital twins in cyber-physical systems (CPS) include the intellectual property incorporated in the digital twins as well as critical information about the CPS itself. For such reasons, the digital twin itself must go through rigorous software hardening routines, such as fuzz and penetration testing [9]. Security of the digital twin must start from the ground-up, inception to retirement. Security concerns must be incorporated with the organization's culture [9].

IoT systems and CPS are being deployed in the energy industry, and there have been instances of cyber-attacks on them. Some known attacks include the US electrical grid in 2009, Ukraine power outage in 2015, and Stuxnet attacks in 2005 and 2010 [10]. Common problems can facilitate such attacks, such as low power IoT devices have minimal resources and thus they are not resilient to denial of service attacks. A digital twin may create yet another vulnerability of an industrial or energy system.

Atalay & Angin in their 2020 study proposes that the digital twin can help to secure a cyber-physical system by establishing a security framework. The framework should contain an extensible digital twin that represents the smart grid, a cyber-threat database for applicable attack vectors, attack simulation tools, and an analysis and reporting module [10].

The framework suggested by Atalay & Angin starts with the creation of the digital twin through specification and reuse of other models. The framework separates the modeling of the CPSs into two phases, identification of the elements and their relationships, and building the network model using simulation. By creating the digital twin and using models, it is possible to achieve standardization through model reuse, utilization of a central threat intelligence database of smart grid attacks, and to achieve continuity of security evaluation [10].

Digital twins may secure physical systems through rule-based controls as well as intrusion detection as proposed by Eckhart and Ekelhart [4]. They identify a wide range of definitions for digital twins. Some definitions omit simulation and focus on visualization, such as implementing augmented reality (AR). AR applications have their security and quality concerns. Eckhart and Ekelhart also explain how a digital twin be used to implement intrusion detection for an industrial control system (ICS) or CPS. They identify concerns beyond intrusion detection, such as:

- the origin of digital twins was in the simulation of systems, although some authors omit simulation as a benefit of digital twins
- digital twins may have fidelity issues which may limit their use
- digital twins have a lifecycle that follows the lifecycle of their physical counterpart
- proper retirement of a digital twin system is imperative as it may contain private and confidential data
- digital twins may be used for security personnel training or in a cyber range as the attack target
- cost of digital twin creation can still be a prohibitive factor

Eckhart and Ekelhart refer to their 2018 publication, where, it was suggested that previously established specifications of the physical system could be modeled with AutoML, virtualized within Mininet and other tools, which can reduce the cost of creating digital twins [12]. They further proposed that digital twins can assist in intrusion detection by offering a baseline of normal behavior that can be analyzed and compared to the behavior of the physical system to detect outlying and potentially malicious activity [12].

3. Methodology

3.1. Collection

This paper focuses on the analysis of collections of Twitter tweets, as well as academic papers. Scripts have been written in R and executed throughout the research period. More R code has been utilized to clean the collected tweets, transform new classification data fields, and analyze them for n-grams, mention,

and sentiment analysis. The scripts use various libraries, such as the rtweet library as an API to the Twitter social media platform. Figure 1 below omits various keys and passwords yet illustrates the basics of library import and the implementation of a search and data frame store of tweets.

```
library(rtweet)
library(ggplot2)
library(dplyr)
library(tidyr)

c_key <- ''
c_secret<- ''
access_token <- ''
access_secret <- ''
appdt <- ''

twitter_token <- create_token(
  app = appdt,
  consumer_key = c_key,
  consumer_secret = c_secret,
  access_token = access_token,
  access_secret = access_secret)

setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)

digitaltwin_tweets <- search_tweets(q = "#digitaltwin",since = '2020-06-08', n = 50000)
```

Figure 1: Screenshot of example usage of rtweet library within R Studio

Three scripts were written to collect tweets for this research. The first used a single search criterion, "#digitaltwin", which has collected 3,102 tweets within 21 days of June 2020. The second and third scripts ran for the entire month of July 2020. 422,963 industry internet of things (IIoT) tweets and 497,174 cybersecurity tweets were collected. The tweets are stored with metadata made available via the rtweet API, but not all data has been retained. Additional data has been generated and stored such as the industry a tweet may mention and the sentiment score (positive or negative) of the text within the tweet. Table 2 lists the acquired and generated data.

Table 2: Tweet data fields

Field Name	Short Description
created_at	The date the tweet was originally posted to the Twitter platform
text	Character contents of the tweet
favorite_count	An integer representing the number of times a tweet has been marked as a favorite by Twitter users
retweet_count	An integer representing the number of times a tweet has been retweeted by Twitter users
country	Frequently NA within the dataset and sometimes providing the country of origin of the Twitter user posting the tweet
retweet_location	Frequently NA within the dataset and sometimes providing the country of origin of the Twitter user retweeting a tweet
Industry and count	A list of industries and the count of mentions from the gathered tweets
Sentiment	The individual sentiment score of a tweet's text
Retweet	Binomial value indicating if retweet_count > 0
Favorite	Binomial value indicating if favorite_count is > 0

3.2. Data Preparation and Processing

Data preparation activities may account for 80 percent of the time invested in data science effort [13]. In this study, the collection took many days, but the Twitter data was usable practically from the beginning. This is due to the limited search criteria used and the nature of the required data (such as we did not need the country or retweet_location fields for this research, both having many empty or not applicable entries). The longest time in preparation was determining industry mentions and labeling the tweets. Industry and identification keywords for cross-reference and classification efforts were collected from the International Labor Organization's web site [14]. Preprocessing was done cross-referencing the ILO keywords to the tweet's text to determine the tweet's topics. This enabled mention analysis by the labor industry.

In the second corpus of tweets, the labeling of tweets allowed the naïve Bayes models to be trained and then tested. Tweets were labeled based upon the search criteria that was used to collect the tweets. This label would consist of either IoT or cybersecurity. Eighty percent of the data was used to train the models and twenty percent was utilized to test the model accuracy. Various models and parameters were used to predict the correct tweet label. Other preprocessing for the model input parameters included sentiment analysis and word lemmatization. The package textstem was utilized for word lemmatization and the packages tidytext and sentimentr were utilized for the sentiment analysis.

4. Results of First Corpus Analysis

The word cloud illustrated in Figure 2 highlights common bigrams found within the first corpus of collected tweets, those 3,102 found with search criterion #digitaltwin, after removing common stop words. While concerning the word 'security' is not paired with another common word frequently enough to make this bi-gram chart, we do see that conversations on Twitter around defense, risk, and the industrial use of IoT are occurring, each having relevance towards this research.

Figure 2: Word Cloud of Bigrams within the Tweet Collection Text

The word cloud used in figure 2 omits counts of the actual instances of the bigrams. Table 3 outlines a few common security terms and the number of instances they occur within our first

corpus, those 3,102 tweets collected by searching #digitaltwin. Such terms may occur in one tweet multiple times. Overall, terms such as "encrypt", or "vulnerability" have almost no mention in such tweets and do not appear as a part of the social media conversation which is a concerning observation.

Table 3: Common security terms and their occurrence

Terms	Count of occurrences
"security"|"secure"	88
"risk"	212
"defense"|"defensive"	179
"iiot"|"industrial iot"	750
"encrypt"|"decrypt"|"cipher"	1
"vulnerability"|"threat"	1

Looking at this first corpus of tweets, the percentage of tweets containing common security terms make up 10% of the total volume of tweets collected using the #digitaltwin search criteria within the specified time range, as illustrated in Figure 3. We observe a dearth of conversation within the social media analysis towards securing digital twins, which could be used to implement the security models needed.

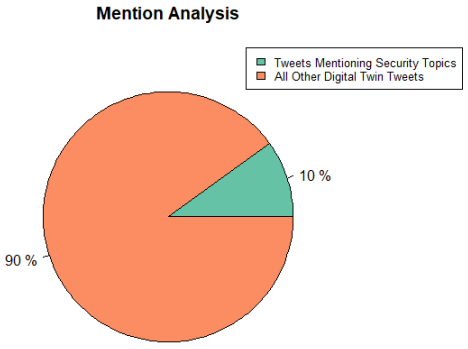


Figure 3: Most IoT tweets do not mention security topics

4.1. Most Favored and Most Retweeted Tweets about Digital Twins

Tweets can be favored as well as re-tweeted. The number of favorites indicates the count of unique user accounts that like or agree with the content of the tweet [15]. In the first corpus, the most favored tweet is also the most retweeted. It will be discussed later in this section. To provide more breadth of the findings, the second most favored tweet is illustrated in Figure 4 [16]. This tweet refers to Intel technology and links to a 2015 white paper where Intel published best practices to help with IoT sensor technology implementation against challenges such as cost of ownership, maintaining security, and designing for scalability [17].

Within the scope of this first corpus, the most retweeted tweet is about the use of digital twin used in contact tracing to help prevent the spread of Covid-19. The digital twin was specifically built and used to help contain the spread of the virus in Dharavi, a location within Mumbai that houses nearly one million residents. Many of these residents rely on daily wages. Living and working conditions may reduce or prevent social distancing. During the time of the tweet creation, epidemiologists had credited the Dharavi community for their efforts towards containment [18]. Figure 5 is a screen capture of this most favorited tweet within the first corpus [19]. While we cannot confirm the accuracy of the tweet, we can recognize from the tweet contents the broad application of digital twin usage to monitor, model, and control a pandemic. The nature of such a system and the associated data also provide support for why such the system requires a mature security posture, to protect individuals and their privacy rights.

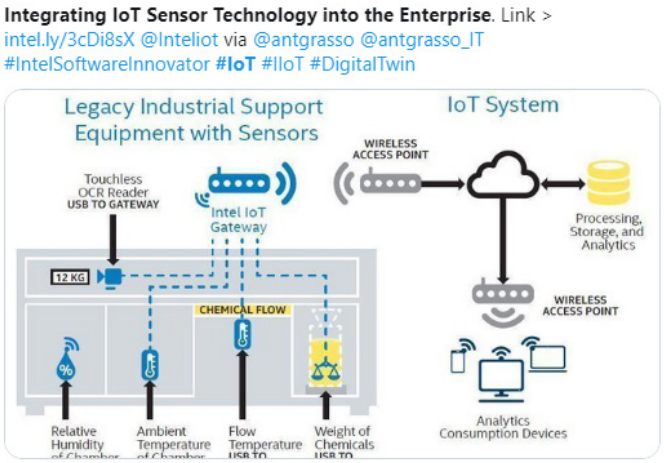


Figure 4: Second Most Favored Tweet within the Scope of this Research

Figure 5: Most retweeted tweet within the research scope

4.2. Mention and Sentiment Analysis

Studies show digital twins being utilized across industries; mention analysis from social media provides an additional perspective. To analyze industry mention among the first tweet collection, we crossed referenced the tweets' text with a list of known industry categories [14]. The simple algorithm reads the tweet texts and counts industry mentions, including when multiple

industries are mentioned within a single tweet's text. Figure 6 illustrates mention analysis results with health, education, and public as the top three topics with most mentions within the tweets' text.

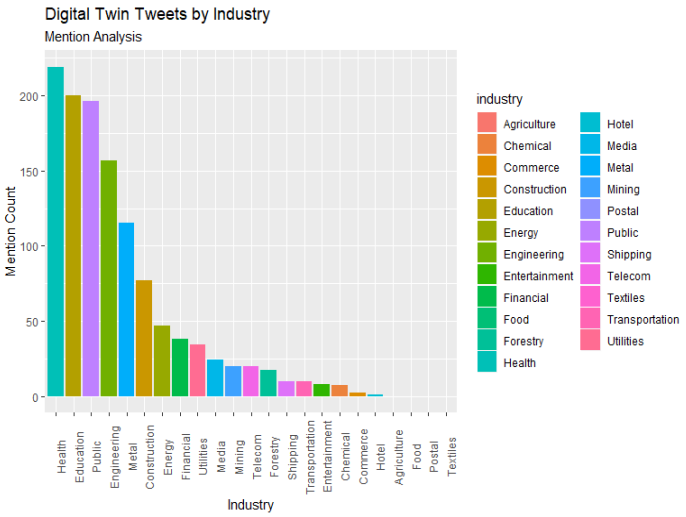


Figure 6: Industry mention analysis: health, education, and public are the top three mentioned industries

Utilizing the sentimentr package, each industry group has the individual tweet text sentiment scores averaged and placed on the chart found in Figure 7. The sentiment scores from the sentimentr package may range from a negative 2.0 to a positive 2.0. While the health and public industries were high in the mention-analysis, the sentiment scores for those industry groups are negative. The three industry groups having the highest average sentiment scores are engineering, commerce, and energy.

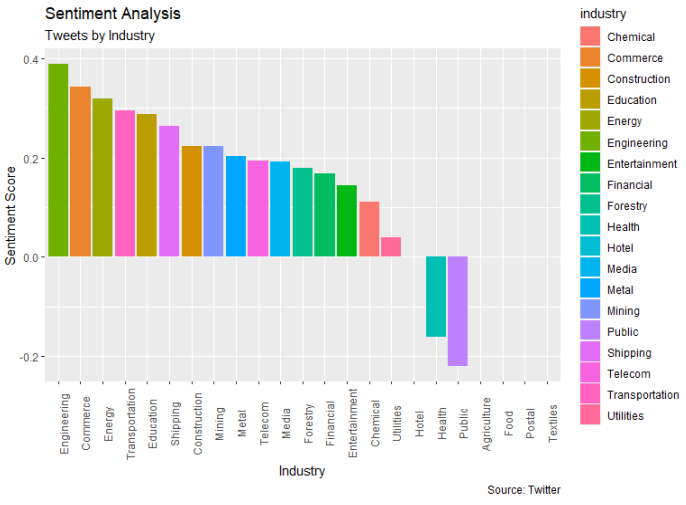


Figure 7: Aggregated sentiment scores grouped by industry

The tweet in the engineering group having the highest sentiment score is given in Figure 8 [20]. This tweet indicates the usage of digital twins to assist in the simulation of electric vehicles for optimal energy management.

Today's electric vehicles market is high stakes. The current pandemic threatens the automotive supply chain [21], a Tesla truck recently marketed towing a competitor's vehicle (F-150) up a hill [22], and Tesla competitor Nikola Motors saw a one-day stock jump of 104% [23]. Security posture is required around the intellectual property in this dynamic and competitive market, and therefore must be applied to such a digital twin.

Figure 8: Tweet having highest sentiment score within the engineering group

5. Results of Social Media Analytics of IIoT and Cybersecurity Tweets

Multiple prediction models were created to further inspect the tweets. This included attempts to predict whether a given tweet was towards the topic of cybersecurity or IIoT. Methods such as word stems, lemmatization, term frequency-inverse document frequency (TFIDF) were utilized in naïve Bayes and general linear models.

A naïve Bayes model that attempted to predict the label, cybersecurity or IIoT, of a given tweet using word lemmatization and sentiment analysis could predict the label correctly with a 70.3% accuracy. While this prediction accuracy is not high, perhaps the sentiment coefficient correlations are more telling. As previously stated, security must start with the organization's culture [9]. An organization's culture consists of the beliefs and values of the employees. Communication incorporating the sentiment of the topic may be more influential towards influencing behavior or belief. The sentiment level of the cybersecurity tweets, when compared to the IIoT tweets, was higher in the areas of anger, disgust, fear, negative, and sadness, whereas the IIoT tweets sentiment level, when compared to cybersecurity tweets, was higher in anticipation, joy, positive, surprise, and trust. Table 4 lists the average sentiment scores for each classification of sentiment by the cybersecurity and IIoT labels.

Given a tweet is about Cybersecurity, there is a 14.5% probability that the tweet is negative, an 11.6% probability that the tweet provokes anticipation, and an 11.3% probability that the tweet has the sentiment of fear. This information may be key for communicating cybersecurity strategies and plans within an organization. We should consider the sentiment of our security posture messaging, and how it fits into the culture, behaviors, and values of the communication's target audience.

The Term Frequency-Inverse Document Frequency (TFIDF) score utilization within a naïve Bayes model had less accuracy,

56.3%, compared with the model using both word lemmatization and sentiment analysis. TFIDF does allow us to see important words in a corpus, we observe that "bugbountytip" had the highest TFIDF score within the cybersecurity corpus at 0.00127.

Table 4: Sentiment coefficient probabilities on tweet labels

Conditional Probabilities of IIoT and Cybersecurity Tweet Sentiment		
Sentiment	Cybersecurity	IIoT
Anger	0.07145	0.03765
Anticipation	0.11663	0.14300
Disgust	0.01868	0.01400
Fear	0.11321	0.06652
Joy	0.05369	0.07048
Negative	0.14541	0.07757
Positive	0.25503	0.35884
Sadness	0.04404	0.02833
Surprise	0.03586	0.03683
Trust	0.14597	0.16677

6. Application of a Digital Twin as a Multi-Model Security Architecture

Digital twins are commonly used to help monitor, model, and control cyber-physical systems. Engineering a digital twin should be done using a systematic approach by performing a needs analysis from the physical device and context of the operating environment [1]. Needs analysis and environmental context must include non-functional requirements, including cybersecurity. Securing information confidentiality of these cyber-physical systems is a growing concern [24]. Beyond the common monitor, model, and control mechanisms, digital twins may also be utilized for incorporating multiple security models.

Several security control models exist. The information flow model is a common mechanism to maintain information confidentiality [25]. We outline in Table 5 how the information flow model may also work to maintain the integrity of a digital twin and its physical counterpart. Multiple models will need to be implemented in the system engineering effort to establish a secure posture. Each model has a strength, such as Bell-LaPadula working towards the confidentiality of information [26]. Furthermore, in Table 5, we indicate how the lattice model, incorporating Biba, can work to ensure information integrity.

The digital twin may also incorporate concepts and techniques such as automated policy exchange (APEX) to prevent data leakage. APEX technology works to keep honest people working within the guidance of security best practices and policies, as large systems and enterprises may have complex policies that not everyone is aware of [27]. Using APEX, the digital twin would monitor labeled files throughout their lifecycle and use control mechanisms, such as deployed agents within the cyber-physical system, to prevent or warn on file activities by users.

Another possible avenue would be for the monitoring portion of the digital twin to utilize a System Call Intercept (SCI) framework. SCI works to reduce data leakage by reviewing policies of subjects and objects, trapping some specific activities before an operating system completing the execution of the activity and perhaps suspending the activity completely [28]. Using the SCI framework, the monitor and modeling functions of the digital twin would use policies to review specific subjects or objects (perhaps all users and system components within a zero-trust approach) and trap some actions for review before the OS or another system component completing the action. An inventory of objects, subjects, and system calls that could be trapped would be required to implement such a framework, which will improve data protection. This inventory could be complemented with behavior analysis by machine learning models to identify malicious activity. Machine learning can be used for countering cybersecurity threats and vulnerability mitigation using regression, prediction, and classification techniques.

Table 5 provides high-level examples of how a digital twin implementation may satisfy or utilize common security models such as State Machine, Lattice, Interference, and Information Flow models.

Table 5: Application examples of a digital twin as a multi-model security schema

Security Model	**Example Digital Twin Application**
State Machine	Digital twin control input is first executed within the model. Validation occurs within the model for an appropriate output state. The input may be rejected before entering the digital twin control mechanism for actual processing on the cyber-physical system.
Non-Interference Model	Monitoring and control mechanisms must obfuscate users reading (monitoring) and writing (controlling) across domains. This reduces the probability of users being influenced by the actions of subjects with greater/lesser clearance. This approach may utilize common discretionary access control (DAC) across file systems, mandatory access control (MAC) through rule-based systems, and role-based control. DAC and MAC limit the availability of information objects to certain subjects to retain the confidentiality of information [29].
Lattice Model	Biba protects integrity by preventing a subject from reading lower integrity or lower-level access data [29]. Using the Biba Model approach, the digital twin modeling process accesses only known secured sensors (having a level of integrity through controls). While we may monitor many sensors, decision dashboards and predictive analytics should utilize only trusted and secured sensors. A matrix of such processes and secure sensors would instruct inputs for such high integrity algorithms.
Information Flow Model	Sensor information flow is one way. The Digital twin may read sensor data for monitoring or modeling purposes. No data is transmitted to a sensor to be later sent back, as sensor-originated data. This prevents levels of spoofing sensor data, by not allowing writes to the sensor memory. Any PUT type of action to sensors APIs would be rejected and logged as malicious intent.

Figure 8 illustrates a situation where a population, Bluetooth sensors from mobile devices, network paths, and system logic are used for the creation of a population's digital twin. In this case, the digital twin of the population would be used by public health officials and the application subscribers for contact tracing. The diagram's purpose is to provide examples illustrating the concepts given in Table 5 applied to a hypothetical contact tracing system.

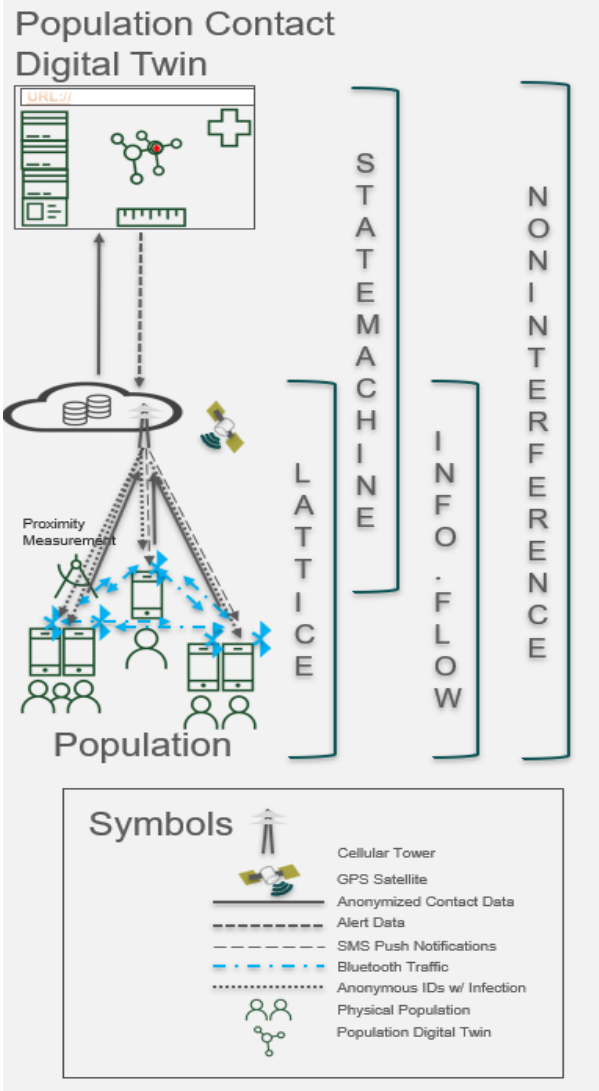


Figure 8: Applying the multi-model in digital twin security

If contact tracing systems were injected with malicious data, it could alter the behaviors of a population. Applying the lattice model helps assure that a system having integrity does not read from a system or sensor with low integrity. The lattice model can prevent a malicious actor's device farm from spoofing the contact system. To validate the integrity of a sensor or mobile device in this case, the system may validate a history of GPS or use human interaction to assure the device is not a simulator or otherwise driven by a malicious bot.

Public health officials may instruct the system to deploy warning notifications to participants who have opted in to use the feature. However, before deploying the notification, the digital twin should utilize the state machine model. Again, a notification could be a malicious message being sent to alarm a population. The logic should first model that notification within the digital twin. The modeling should evaluate the message accuracy and may even predict how the population may behave to help public health officials craft response plans. If the notification messages were not related by a contact tracing scheme, the message may be malicious and should be stopped.

The information flow model is utilized to protect the population and their privacy. The mobile device's application logic allows Bluetooth connections to write to ledgers indicating contacts. This data is anonymized and does not indicate a person's identity. Writing to this ledger is only allowed by the local mobile device which is in the possession of the user. The centralized system does not allow writing to this contact ledger. The mobile device application will pull anonymized alert data from the central system. This data is written to a separate ledger. The only mechanism for the centralized system to send data to the host devices is through the SMS push notification mechanism, which has levels of security such as users being required to opt-in and the state machine model.

Users of the Population Contact Digital Twin dashboard will utilize roles, authentication, and authorization to access the digital twin of the population. Low entitlement users would not have the ability to interfere with the population by spying on them using inference techniques. Entitled users, such as medical doctors, may be granted access to information having more fidelity to help protect the population's health. This protects a population using the non-interference model with access controls.

7. Conclusions and Future Work

7.1. Conclusions

Digital twins operate from and create new cyber-physical system information. A proper design approach for a digital twin should model its cyber-physical system, including its future states. In addition to monitoring, modeling, and controlling the cyber-physical system, a digital twin must also provide security. This fourth element as applied to the prior three digital twin requirements allows for a layered approach to securing cyber-physical systems by implementing multiple security models. While past researchers have focused on intrusion detection, this article proposes that common security control models such as state machine, non-interference, lattice, and information flow models can be utilized in digital twin implementation to secure itself the associated cyber-physical system.

In the first corpus of tweets collected in this study by searching for #digitaltwin, those referencing security concepts represented a small population of 10%. In the second and larger corpus of the collected tweets, the top mentioned industries or business sectors were health, education, and public in that order. The top three industries having the highest tweet sentiment were engineering, commerce, and energy. While naïve Bayes models reached only a 70.3% accuracy at differentiating a tweet that was about cybersecurity versus a tweet about digital twins, the sentiment findings of the messages leaned towards negative messaging from the cybersecurity tweets.

7.2. Future Work

Most of our research utilized social media, which can be time-consuming to prepare such data and has limits in what it can

contribute. Future research should include detailed case studies of direct implementation which would strengthen the approaches proposed in this work. The sentiment findings of cybersecurity tweets deserve further research as other researchers have stated that security must start with an organization's culture. The implication of negative messaging on an organization's cybersecurity posture and culture needs to be studied.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] J. Scheibmeir, Y. Malaiya, "An API Development Model for Digital Twins," in 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C), Sofia, Bulgaria, 518-519, 2019, doi: 10.1109/QRS-C.2019.00103.

[2] E. A. Saddik, "Digital Twins: The Convergence of Multimedia Technologies," IEEE Multimedia Mag., 25, 87-92, Apr. 2018.

[3] B. Martino. M. Di, M. Rak, A. Ficco, S. Esposito, S. Nacchia, "Internet of Things Reference Architectures, Security and Interoperability: A Survey." Internet of Things, 1-2, 99-112, 2018, doi:10.1016/j.iot.2018.08.008.

[4] M. Eckhart, A. Ekelhart, "Digital Twins for Cyber-Physical Systems Security: State of the Art and Outlook." Security and Quality in Cyber-Physical Systems Engineering. Springer, 2019, doi:10.1007/978-3-030-25312-7_14.

[5] A. Canedo. "Industrial IoT Lifecycle via Digital Twins." Proceedings of the Eleventh IEEE/ACM/IFIP International Conference on Hardware/software Codesign and System Synthesis. ACM, 1-1, 2016, doi: 10.1145/2968456.2974007

[6] D. Yadav, S. Jayanthu, S. Das, S. Chinara, P. Mishra, "Critical Review on Slope Monitoring Systems with a Vision of Unifying WSN and IoT," IET Wireless Sensor Systems, 2019, doi:10.1049/iet-wss.2018.5197.

[7] Y. Wang, A. Garhan, R. Byrav. "A Survey of Security Issues in Wireless Sensor Networks." IEEE Communications Surveys & Tutorials, **8**(2), 2-23, 2006, doi:10.1109/comst.2006.315852.

[8] P. Porambage, C. Schmitt, P. Kumar, A. Gurtov, M. Ylianttila, "Two-phase Authentication Protocol for Wireless Sensor Networks in Distributed IoT Applications," IEEE Wireless Communications and Networking Conference, 2014, doi:10.1109/WCNC.2014.6952860.

[9] M. Hearn, S. Rix, "Cybersecurity Considerations for Digital Twin Implementations report". IIC Journal of Innovation, 2019

[10] M. Atalay, P. Angin, "A Digital Twins Approach to Smart Grid Security Testing and Standardization," 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT" 435-440, 2020, doi: 10.1109/MetroInd4.0IoT48571.2020.9138264.

[11] A. Girma, "Analysis of Security Vulnerability and Analytics of Internet of Things (IoT) Platform," in Information Technology - New Generations. Advances in Intelligent Systems and Computing, 738, 2018, doi:10.1007/978-3-319-77028-4_16

[12] M. Eckhart, A. Ekelhart, "Towards Security-Aware Virtual Environments for Digital Twins," in Proceedings of the 4th ACM Workshop on Cyber-Physical System Security, ACM, 61–72, 2018, doi:https://doi.org/10.1145/3198458.3198464

[13] A. Cirillo, RStudio for R Statistical Computing Cookbook. Birmingham Packt Publishing, 2016.

[14] International Labor Organization. "Industries and Sectors." Industries and sectors, https://www.ilo.org/global/industries-and-sectors/lang--en/index.htm.

[15] R.S. Perdana, A. Pinandito, " Combining Likes-Retweet Analysis and Naive Bayes Classifier within Twitter for Sentiment Analysis", Journal of Telecommunication, Electronic, and Computer Engineering, **10**(1-8), 41-46, 2018.

[16] A. Grasso [@antgrasso]. (2020, June 8). Integrating IoT Sensor Technology into the Enterprise. [Twitter moment]. Retrieved from: https://twitter.com/antgrasso/status/1269962528530071552

[17] K. H. Gaiser, V. P. Harrington, G. T. Loughrin, S. J. Meyer, J. M. Sartini. "Integrating IoT Sensor Technology into the Enterprise." December 2015. https://www.intel.com/content/www/us/en/it-management/intel-it-best-practices/integrating-iot-sensor-technology-into-the-enterprise-paper.html.

[18] P. M.N., S. Bengali. "How an Indian Slum Became an Unlikely COVID-19 Success Story." Los Angeles Times. June 24, 2020. https://www.latimes.com/world-nation/story/2020-06-24/dharavi-slum-in-mumbai-india-contained-covid-19.

[19] Zuppa [@ZuppaGeoNavTech]. (2020, June 17). "Offence Is The Best Defense" [Twitter moment]. Retrieved from: https://twitter.com/ZuppaGeoNavTech/status/1273449554797408257

[20] Simcenter [@Simcenter_] (2020, June 12). Learn more about #simulation for optimal vehicle energy management in our new on-demand webinar series. [Twitter moment] Retrieved from: https://twitter.com/Simcenter_/status/1271442013586624512

[21] L. Harbour, "The Coronavirus' Impact On The Global Automotive Supply Chain." Forbes. March 14, 2020. https://www.forbes.com/sites/laurieharbour1/2020/03/13/the-coronavirus-impact-on-the-global-automotive-supply-chain/#17fd6b9f444e.

[22] M. Wayland, L. Kolodny, "Tesla Cybertruck Tug-of-war with a Ford F-150 Is High Stakes and Highly Unlikely." CNBC. November 27, 2019. https://www.cnbc.com/2019/11/26/tesla-cybertruck-tug-of-war-with-a-ford-f-150-is-high-stakes-unlikely.html.

[23] M. Fox, "Tesla Competitor Nikola Motors Skyrockets 104% after Setting Reservation Date for Its Electric Truck (NKLA, TSLA) | Markets Insider." Business Insider. June 8, 2020. https://markets.businessinsider.com/news/stocks/nikola-motors-stock-price-soars-setting-truck-reservation-date-competitor-2020-6-1029290291.

[24] A. Humayed, J. Lin, F. Li, B. Luo, "Cyber-physical systems security—A survey", IEEE Internet Things J., **4**, 1802-1831, 2017, doi: 10.1109/JIOT.2017.2703172.

[25] A. Sabelfeld, A. C. Myers, "Language-based information-flow security", IEEE J. Sel. Areas Commun., **21**(1), 5-19, 2003, doi: 10.1109/JSAC.2002.806121.

[26] X. Ma, Y. Huang, D. Li, "A Security Model Based on Lattice," International Conference on Electrical and Control Engineering, 4958-4961, 2010, doi: 10.1109/iCECE.2010.1199.

[27] S.J., Simske, and H. Balinsky. "Apex." Proceedings of the 10th ACM Symposium on Document Engineering, 10, 2010, doi:10.1145/1860559.1860587.

[28] H. Balinsky, D. S. Perez, and S. J. Simske. "System Call Interception Framework for Data Leak Prevention." 2011 IEEE 15th International Enterprise Distributed Object Computing Conference, 2011. doi:10.1109/edoc.2011.19.

[29] C. Schou, & S. Hernandez, Information assurance handbook:effective computer security and risk management strategies. New York: McGraw-Hill Education, 2015.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 331-340 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

The Contribution of Wind Energy Capacity on Generation Systems Adequacy Reliability using Differential Evolution Optimization Algorithm

Athraa Ali Kadhem[*,1], Noor Izzri Abdul Wahab[1], Ahmed Abdalla[2]

[1]Department of Electrical and Electronic Engineering, University Putra Malaysia, Selangor 43400, Malaysia

[2]Faculty of Electronics Information Engineering, Huaiyin Institute of Technology, Huai'an 223003, China

ARTICLE INFO

Article history:
Received: 07 May, 2020
Accepted: 30 October, 2020
Online: 20 November, 2020

Keywords:
Optimization Algorithms
Intelligent Search
Adequacy Systems

ABSTRACT

Currently, increasing penetration of the Wind Energy Conversion System (WECs) in Power generation systems has influenced the supply of electrical power reliability for the consumer in comparison with other traditional sources. In this paper, the performance and efficiency of a new optimization approach referred to as the "Differential Evolution Optimization Algorithm" (DEOA) to measure the reliability of power adequacy systems (RPAs). The proposed intelligent algorithm which relies on the Population-based Intelligent Search (PBIs) technique is viewed as a feasible alternative for the Monte Carlo simulation (MCS) method in the assessment of the non-chronological system. The benefit of utilizing this algorithm is apparent in the manner it expedites the calculation and achieves greater precision with less calculation effort. Additionally, there is a deeper understanding of the effect of the increasing levels of wind energy on generation adequacy from the WECs sources to satisfy future power electricity demands. In addition, the effectiveness of the suggested algorithm in assessing the RPAs was compared to the analytical and MCS method.

1. Introduction

The inconsistent nature of wind energy presents challenges to the planners and operation personnel of power plants. Therefore, it is now crucially important to assess the extent that power impacts the dependability and sufficiency of the power being generated by existing plants. In assessing the penetration of the wind power from the WECs based on the nature of the wind can be complicated as it involves a number of critical issues that have to be taken into account when designing the WECs, such as the speed and power of the wind. To design a method for estimating wind production necessitates a massive amount of historical data of wind velocity so that the stochastic nature of the wind at a particular location can be accurately simulated. Another possible option could be to use a dependable stochastic simulation technique for the generation of the time series artificial of the wind speed.

Analyzing how adequate the PS is possible through the identification and evaluation of failure states, which are situations when customer demand for power cannot be met by the power system, hence, making it necessary to practice load shedding to ensure the maintenance of the system integrity. Due to the fact that there are numerous potential failure situations [1], [2], it may not be practicable to directly count and estimate it all even in cases of medium-size power systems.

Currently, RE sources usually combined with the conventional electrical grid for generating less polluting energy; as for example WTGs [3]. As a result, the estimation of the reliability index for unconcentrated resources such as wind energy is complicated [4]. As such, new tools need to be used to assess PS with the wide presence of wind energy sources [5].

Evaluating the dependability of PGS is typically performed by simulation or analysis approaches. The majority of authors in [6] focused on assessing the dependability of PGS in wind farms' basis of analytic approaches. The MCS method offers an accurate estimate of the dependability indices. Therefore, the MCS would be beneficial for such a purpose, but it requires considerable computational work which is time-consuming and should be avoided if convergence efficiency is critical [7]. This research

[*]Corresponding Author: Athraa Ali Kadhem, University Putra Malaysia Email: athraaonoz2007@yahoo.com

offers a replacement for the MCS and PBIs method [8],[9]. This optional approach method is used to identify a num ber of possible failure situations to facilitate the adequacy indices of the system.

Numerous PBIs approaches are possible for intelligent research, epitome for these algorithms like the BPSO, ISSPs, OGA, EC, BPSO, and BCA [10]–[13]. Although some research has been done on the dependability of the generating systems, there is still a need for approaches that are more appropriate and which offer greater computational flexibility and are more realistic in representing the desired PGS adequacy [14]–[17].

Models are needed for the simulation of the stochastic characteristic of PGS behavior, which is not a new issue for population-based intelligent research approaches. However, it has emerged as an important issue when wind sources and the PS are combined and taken into account. This paper suggested a novel algorithm known as "DEOA" to lower the time amount cost of the computation to simulation pattern in the non-SMCS and other Optimization Algorithms by enhancing the checking efficiency in a given area state-space employing a PBIs approach. DEOA is based on the meta- heuristics searching employed in simulating the generating system's operation and taking into account the not expecting failure of current systems and the instability of "WECs" energy. DEOA exhibited benefits in the form of higher computational speed, utilizing a relatively uncomplicated structure, providing ease-of-use, and was shown to be highly robust. Besides, the numerically simulated solution was compared with the MCS-related Work. It also explains the impacts of wind energy capacity from the WECs on power systems to attain at the level of the incremental peak load capability.

Nomenclature

WECs	Wind Energy Conversion System
RPAs	Reliability of Power Adequacy Systems
PBIs	Population-Based Intelligent Search
DEOA	Differential Evolution Optimization Algorithm
MCS	Monte Carlo simulation
SMCS	Sequential Monte Carlo Simulation
PS	Power System
RE	Renewable Energy
WTGs	Wind Turbine Generators
PGS	Power Generation Systems
ISSP	Intelligent State Space Pruning
OGA	Optimization Genetic Algorithm
EC	Evolutionary Computation
PSO	Particle Swarm Optimization
BCA	Bee Colony Algorithm
LOLE	Loss of Load Expectation
LOEE	Loss of Energy Expectation
LOLF	Loss of Load Frequency
LDC	Load Duration Curve
RI	Reliability Indices
LH	load levels
PNS	Power Not Supplied
LOLP	Loss of Load Probability
WDM	Weibull distribution model
WPG	Wind Power Generation
BPSO	Binary Particle Swarm Optimizations
MSGA	Modified Simple Genetic Algorithm
RTS	Reliability Test System

2. Assistant Theoretical Materials

2.1. Assessment of Conventional Generation Systems Adequacy

Analyzing the reliability of electrical energy systems is important to determine PGS adequacy and for possible future increase in the system capacity. This is to ensure adequate availability of power to satisfy consumer requirements for particular installed system capacity. As a result, the could be utilized as an element of the quantitative conditions in a power generation system in planning for the system's risk paradigm. In this study, the essential evaluated were utilized in facilitating the approximation of the PS dependability level of their "Loss of Load Frequency (LOLF)", "Loss of Energy Expectation (LOEE)", and "Loss of Load Expectation (LOLE)".

Load designs system are normally non-sequential or sequential paradigms that can be used with various Optimization Algorithms approaches. The sequential Load Duration Curve scheme (" LDC") can, for epitome, can create values of load demand hourly states, the implication being that around "8736 Hours" separate states can be logged per annum [18]. The annual LDC approach for IEEE-RTS-79 - 96 is presented in Figure 1.

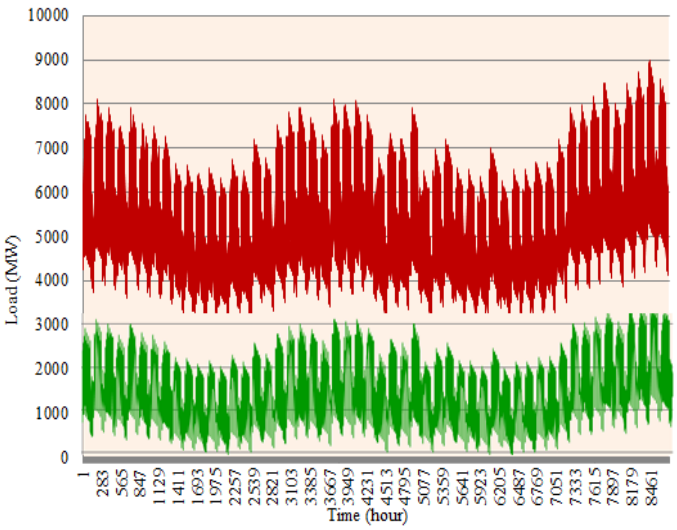


Figure 1: Annual LDC model for the RTS

To attain elevated levels of RPAs and the PGS valuation, the best design of RI has an important role in designing and developing a successful generation system [19]. Estimates of the LOLE, LOEE, and LOLF are made based on the attained capacity states' collections and load characteristics. In this study, denoted the discrete states for the load levels (LH) at a time (t). The "Loss of Load Probability" (LOLP) is estimated as equation (1):

$$LOLP(LH_i) = \sum_{j=1}^{sa} S_j \cdot P_j \cdot Copy_j \quad (1)$$

where:

sa = total number of state arrays, and

Sj = status of the system state.

In a failure state, the status value will=1; while in a successful state, the status will=0, this refers to the status of the load demand satisfaction [20]. equation (2) displayed the values of the LOLE annually:

$$LOLE = \sum_{j=1}^{8736} LOLP(LH_j) \quad (2)$$

Estimated the megawatts of the PNS by used equation (3):

$$PNS(LH_i) = \sum_{j=1}^{sa} S_j \cdot P_j \cdot Copy_j \cdot (LH_i - Cap_j) \quad (3)$$

$$LOEE = \sum_{j=1}^{8736} PNS(LH_j) \quad (4)$$

Equations employed to calculate the components of the LOLF index [19] as:

$$LOLF(LH_i) = \sum_{j=1}^{sa} S_j \cdot FG_j \cdot Copy_j \quad (5)$$

$$FG = \sum_{j=1}^{8736} LOLF(LH_j) \quad (6)$$

$$FL = \sum_{j=1}^{8736} V_j \cdot \left[LOLP(LH_j) - LOLP(LH_{j-1})\right] \quad (7)$$

$$LOLE = [FG_{(LH)}] + [FL_{(LH)}] \quad (8)$$

2.2. Wind Speed Paradigm

Simulating of wind speed/velocity to study WECs performance and dependability was done and reported by [21]. The sequential WECs simulation model involves the generation of hourly wind speed data over a given perio and specified site. The main challenge in maintaining the reliability of PGS is the fluctuation nature of wind power. It is therefore necessary to develop an appropriate wind velocity model before proceeding to analyze the system reliability.

The majority of research recently done dependent on the Weibull distribution model (WDM) to signify the hourly fluctuations in the annual mean wind velocity. The level of wind turbine generated power at a given wind site depends on the mean and standard variation of the wind velocity. Based on the assumption that the WDM is able to change particular variables such as the shape (b) and the scale (a), it is normally used for simulating wind speed fluctuations. Thus, the inverse transform function for the WDM in [22] in equation (9) can be employed to simulate the wind speed s.

$$s = a\left[-Ln(H)^{1/b}\right] \quad (9)$$

where; a and b = two parameters for the WDM;

H = random variable;

s = wind speed.

By applying the wind speed that generated by using the equation (9) to the wind turbine units output model, the turbine units power output is as shown in equation (10).

$$P_{WTG} = \begin{cases} 0 & ws < V_{ci} \\ P_r \times (A + B_x + Cx^2) & V_{ci} \le ws < V_r \\ P_r & V_r \le ws < V_{co} \\ 0 & ws > V_{co} \end{cases} \quad (10)$$

where

ws = wind speed (m/s),

Vci = WTG cut-in speed (m/s),

Vco = WTG cut-out speed (m/s),

Vr = WTG rated speed (m/s),

Pr = WTG rated power output (MW).

The constants A, Bx, and Cx have previously been calculated by [22].

3. Methodology

Two programs based on the simulation method have been developed using MATLAB. The first simulation program was developed to simulate the variations of wind speed using WDM, which was utilized for an estimation of the parameters “a” and “b”, that were deuced form wind speed data. Consequently, during the simulation process, the two parameters (a, b) will generate wind speed variations at any time. Additionally, the program can estimate a wind power result from the wind turbine. Figure 2 shows models for wind speed generation and output power simulation. The second simulation program is based on the DEOA technique to evaluate the dependability of the PGS and, wind power penetration consideration is incorporated in the generating capacity adequacy evaluation. According to the steps taken as stated in reference [22], the RI was calculated utilizing DEOA algorithm as shown in Figure 3.

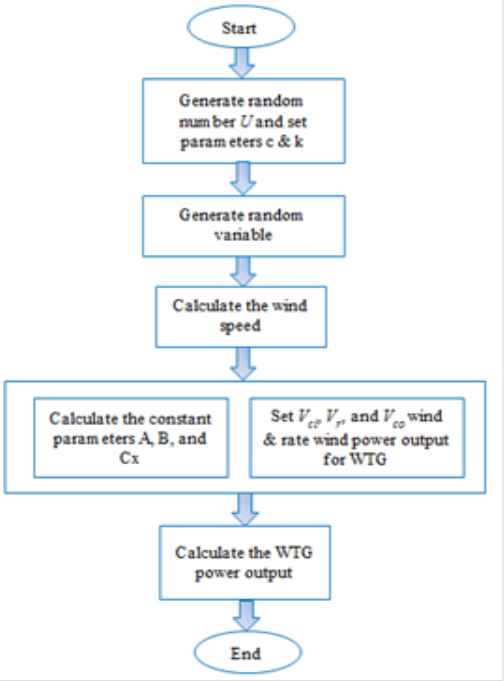


Figure 2: Diagram of speed and wind power for the turbine unit

4. Outcomes Analysis

4.1. Simulation Outcomes of Speed & Power Wind

This study explains the possibility to generate wind energy at different levels which parallel the probability of wind speed fluctuation, therefore, the wind power model can generate wind energy with different abilities states as instant at hourly. One of the features of the wind speed is that the turbine output power is determined subsequently when the hourly wind speed is obtained. The simulation of the profiles of the speed of the wind was employed to simulate ability states from the simulated generation of wind output power.

There is a need to examine the power output from the wind farm, to obtain a better picture of the selected site so that the WDM parameter, a, is set as the mean wind speed and b=2, is the usual wind characteristic [23]. The particular WTG characteristics employed here are the 4, 25, and 19 m/s (cut-in-speed, cut-out-speed, and rate speed) and at a rated power of 2 MW [22]. Figure 4 shows the forecast for hourly wind speed over a year, and also indicates the simulation wind power output for WTG (with 85 WTG) with a power rate of 2 MW for a similar period.

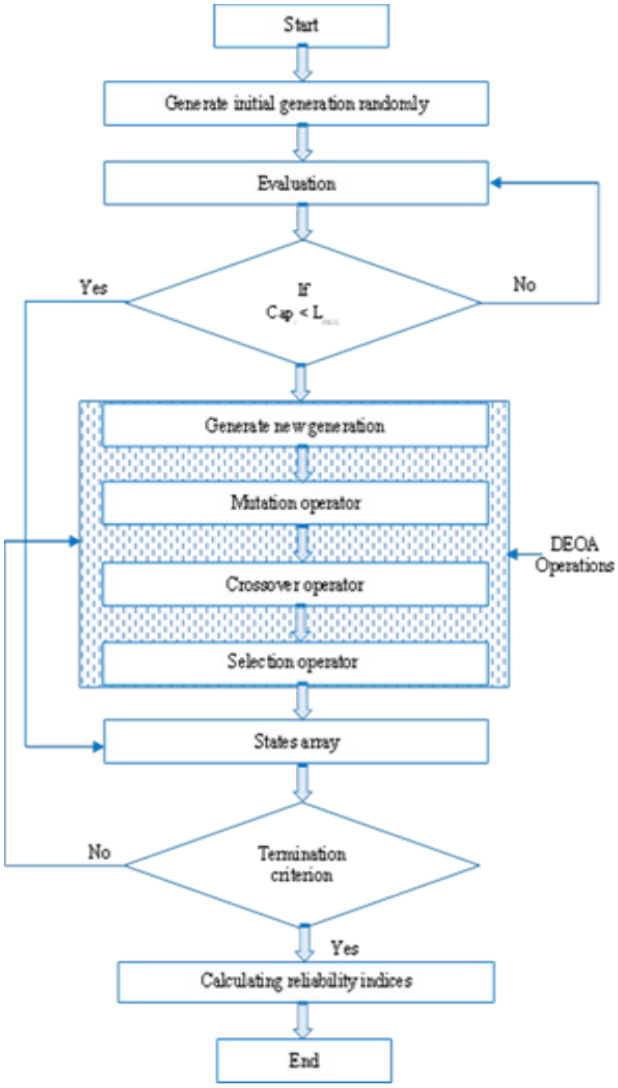


Figure 3: Diagram of reliability indices assessment using DEGA

4.2. Cases Running the Algorithm

This study develops and tests a method for evaluating the reliability of PGS using WECs with two dissimilar test systems ("IEEE-RTS-79" & "IEEE-RTS-96"), to validate how the DEOA performs to assess the dependability and sufficiency of the PGS. Meanwhile, it tests the efficiency of the proposed algorithm to evaluate multiple penetration-levels of wind energy.

The test systems "IEEE-RTS-79 and RTS -96" have 32 and 96 generating units respectively, with the capability to generate power ranging approximately from (12- 400 MW), with a total power supply of 3405 and 10215 MW respectively, at approximately (2850 MW) and (9000 MW) peak load [24], [25]. Figure 5 showing the diagram of the single area of the system RTS-79, and Figure 6 showing the outline of the RTS-96 of three-areas interconnected through merging three single areas from the RTS-79 system.

Run I: *Valuation of Algorithm Performance for the "RTS-79" System*

The designs of the variable control values for the proposed DEOA algorithm was documented as displayed in Table 1 [26], during applied at the "Test System- 79 & 96"

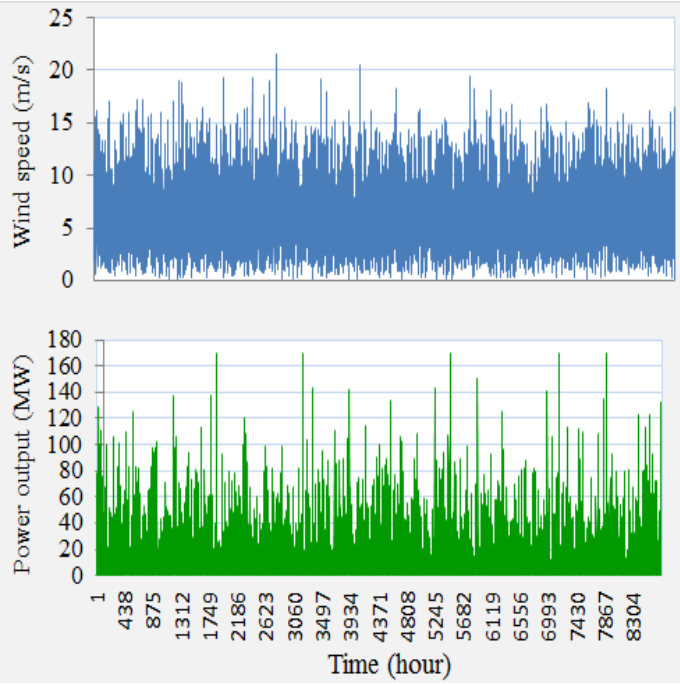


Figure 4: Wind forecasting and simulation of the wind MW generation for one year

The DEOA algorithm was applied for 100 iterations run consecutively in matching conditions as previously stated, to verify the robustness and confidence of this algorithm. The results of the various implementations are presented in Table 2. The results achieved from the DEOA were compared with those of [27]. Figures 7 and 8 show that following runs of 100 iterations, the values were wavering in the region of the real value. Repetitions of the runs of the experiment applying the algorithm displayed that this algorithm was highly accurate when determining the RI.

Table 1: The control values for the proposed algorithm

Parameter	DEOA (RTS-79)	DEOA (RTS-96)
Population size	60	100
Generation limit	100	150
Crossover probability	0.6	0.6
System Capacity (MW)	3405	10215
Mutation probability High	0. 1	0.6
Mutation probability Low	0.6	0.05
Load max	2850 MW	9000 MW
Threshold probability	1e-15	1e-20
Stopping Repetitions	100	150

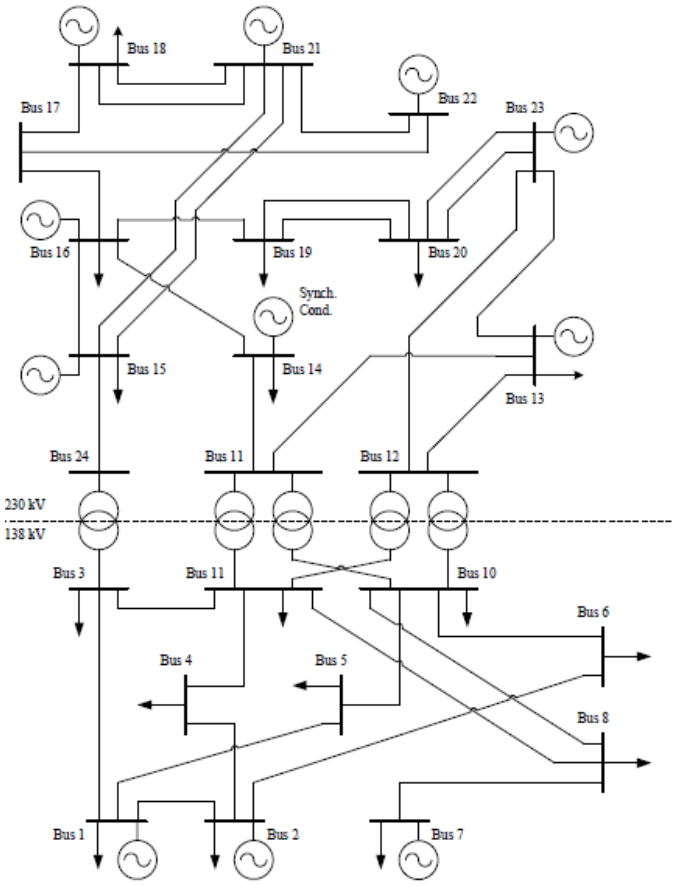


Figure 5: Diagram of single area for the system RTS-79

Table 2: The DEOA algorithm compare with analytical method through runs of the algorithm (100) times

Method	Reliability Indices		Error (%)
	LOLE (hour/year)	LOEE (MWh/year)	
(Accepted value) Analytical method [30]	9.39418	9.3670	0.28
(Experimental value) DEOA method	1176.00	1120.00	4.76

The consequences outcomes of the DEOA algorithm were compared with those of other optimization genetic algorithms earlier mentioned in [28], [29], as shown in Table 3. Comparisons of these outcomes were made. The absolute values utilized for these comparisons were from a single turn one mploying "MSGA", & "DEOA".

The computation effort of the proposed method was also assessed and a comparison was made with those of "MSGA", "MCS", and "DEOA" as shown in Table 4. On the other hand, comparison of the simulated proposed method with MCS was halted when the coefficient of variation level was approximately 5% [29]. This comparison was performed with absolute values obtained from a single turn on of the DEGA and MSGA. Evaluating the computation time of the DEOA, as presented in Table 4 was shorter compared to other methods. Hence, the suggested DEOA exhibits benefits such as enhanced computation speed, by employing a simple structure.

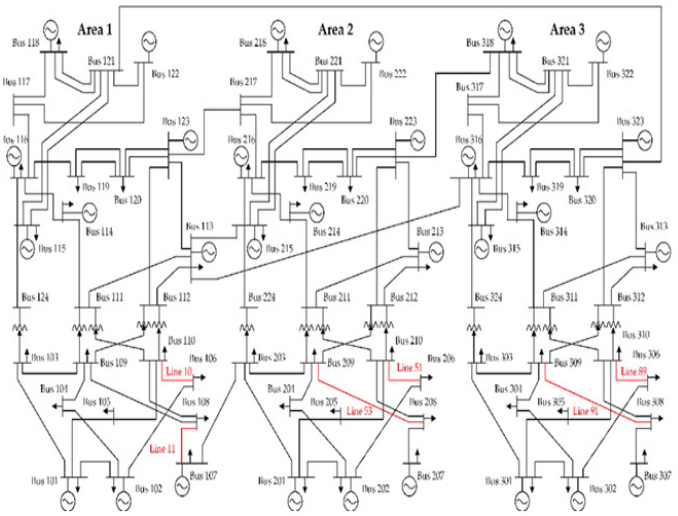


Figure 6: Diagram of three-areas interconnected for the system RTS-96

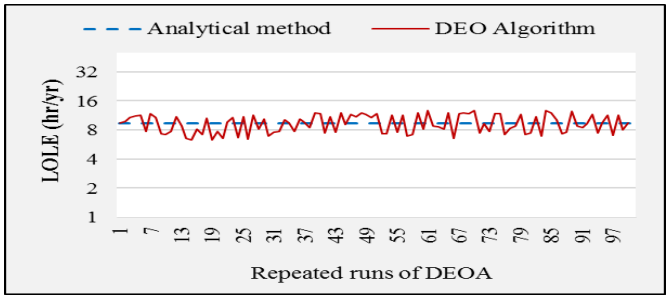


Figure 7: Calculation of estimated LOLE with 100 repeated runs for the RTS-79 System

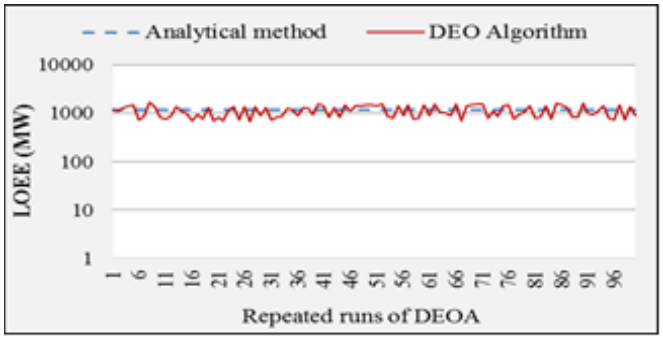


Figure 8: Calculation of estimated LOEE with 100 repeated runs for the system 'RTS-79'

Table 3: Compare the DEOA algorithm RI results for RTS-79 with accepted value

Reliability Indices	Experimental value		
	MCS [26]	MSGA [28]	DEOA
LOLE (hour/year)	9.371	9.324	9.370
Error (%)	0.244	0.745	0.255
LOEE (MWh/year)	1197	1163	1178
Error (%)	-1.785	1.105	-0.170
LOLF (occ/year)	1.9190	2.0030	2.300
Error (%)	0.00	4.37	19.85

Table 4: Comparison of the runs time between MSGA, DEOA, and MCS methods

Techniques	Experimental value		
	MCS	MSGA	DEOA
	Runs= 200	Runs= 750	Runs= 100
LOLE (hour/year)	9.541	9.324	9.37
Time (sec)	372	177	4

Run II: *Valuation of Algorithm Performance for the "RTS-96"*

In this run, the control parameters values had been setting beforehand as shown in Table 1. The tuning for the failures data for Generating stations was in line with previous data provided by [30]. The result of the DEOA was compared to those of Optimization algorithms declared by [31], as shown in Table 5.

In this paper, the percentage of error was considered, where the percentage error for individual values of the proposed DEOA algorithm was compared with individual values of the other techniques, as displayed in Tables 2, 3, and 5. On the other hand, the error percentage shows the effectiveness of the proposed DEOA algorithm in presenting separate values, large positive values of the error indicate a considerable deviation in the predicted value from the real one.

Table 5: Results of the RI for the system RTS-96 utilizing the DEOA method

Techniques	Reliability Indices			
	LOLE (hour/year)	Error (%)	LOEE (MWh/year)	Error (%)
Analytical method	0.1380	-	24.26	-
MCS Algorithm	0.1400	-1.44	23.97	1.195
DEOA Algorithm	0.140	-1.44	26.21	- 8.037

Run III: *Valuation of Performance SMCS with WPG for "RTS-79" System*

In this study, the SMCS method was employed to recompute the RI (for RTS-79 System) for comparison with the DEOA algorithm. From the simulation, Figure 9 shows what the system is capable of based on the conventional and the unconventional units. Figure 10 however presents a whole PGS and what it is capacity of, based on the simulation of the process overlaid together with the sequential load model and Figure 11 shows the LOLE, LOEE, and LOLF for 200 selected SMCS sampling years.

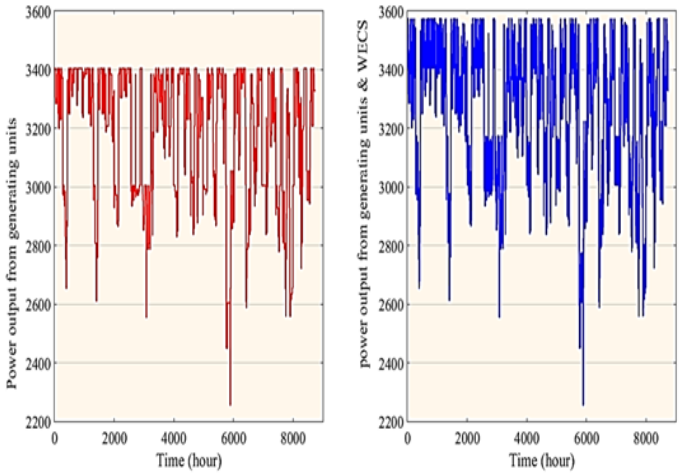


Figure 9: Available capacities obtained from conventional and unconventional units

Some intersections observable in Figure 10, which means that the available electrical power generation from the PGS insufficient for meeting the required load.

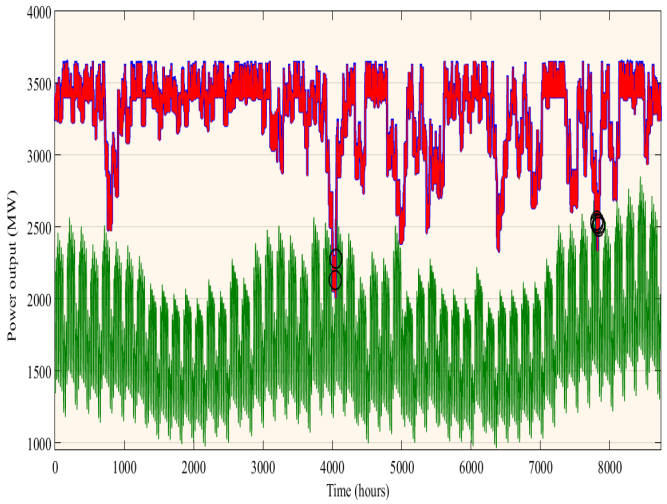


Figure 10: Simulation process shows the available capacities of the system, superimposition with the load paradigm

Run IV: *Valuation of Reliability Indices for DEOA with WPG for "RTS-79" System*

For peak demand load (2850MW) and 170 MW combined with wind energy, the RI the for "RTS-79" derived was from the Optimization Algorithm and the new proposed DEOA approach

as shown in Table 6. The MCS might be utilized for choosing failure states for the systems. However, due to its reliance on proportionate sampling, its effectiveness in the detection of failure states could be diminished. According to the results indicated in Table 6, the SMCS method could lead to extremely poor convergence time. Additionally, the DEOA provides the most efficient computation as well as easy usage in comparison with other algorithms.

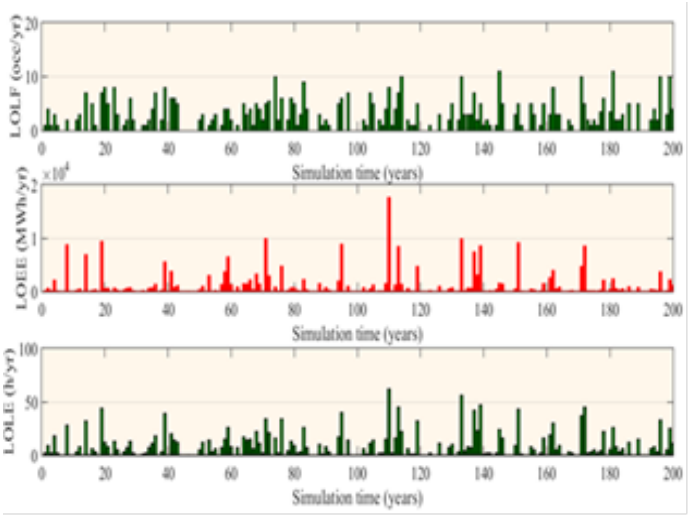


Figure 11: The reliability indices LOLE, LOEE, and LOLF for 200 samplings

Table 7 presents the outcomes of five models previously discussed [32] in comparison with those of DEOA. These techniques, computed the RI employing the MCS approach with variable wind Speed models (the actual wind speed recording data, Markov method, ARMA distribution model, Weibull distribution model, and Normal distribution). It was obvious that the outcomes derived from the use of the DEOA algorithm had values which were about the same as those obtained from using real wind speed data in the MCS approach.

Table 6: Comparison of the runs time between DEOA, BPSO, and SMCS Method

Proposed Methods	LOLE (hour/year)	LOEE (MWh/year)	Time (Sec)
SMCS Algorithm	7.550	941.070	64.641
BPSO Algorithm	7.430	823.780	8.979
DEOA Algorithm	7.450	1318.340	4.00

Table 7: Comparison between suggested methods in literature and proposed DEOA method

Techniques	LOLE (hour/year)	LOEE (MWh/year)
MCS technique with Actual wind data	7.450	908.70
MCS technique with Markov model	7.470	918.10
MCS technique with ARMA method	7.120	884.90
MCS technique with Weibull model	7.780	976.70
MCS technique with Normal model	6.950	858.50
Proposed DEOA with Actual wind data	7.450	1318.34

In real PGSs the LOLE index is used more often for the generation of adequacy analysis to determine sufficiency. On the other hand, the LOEE indicator is used for complex generating/transmitting of the adequacy analysis [33]. This means that outcomes presented in the tables are for assessing the PGS's dependability and sufficiency, which took into account the level of precision of LOLE from the suggested DEOA approach with shorter computation time in comparison with other approaches.

The proposal of an intelligent algorithm to assess the PGS is shown as a feasible replacement for the Non-SMCS in the evaluation of the non-sequential systems. Thus, this algorithm lowered the computation time needed to calculate the Non-MCS approach, which is efficient but does not take into account time-based phenomena, including the frequent computation interruptions of the LOLF and the duration indicatots (LOLD). This means that outcomes in tables that assess the LOLF indicate poor precision from the suggested PBIs approaches with shorter computation time in comparison with other approaches.

Run V: *Valuation of Reliability Indices with Wind Power Penetration for the "RTS-79" System Using DEOA*

In this section, the DEO algorithm is applied to examine the effects of wind power capabilities generated from WTGs on power systems to attain dependability indicators at a similar level to additional peak load capacity. The Weibull distribution approach was employed to simulate hourly repeating wind speed. Figure 12 shows the the simulation results of the forecast forecast for hourly wind speed and wind power output for several installed WTG units with a power rate of 2 MW for a year. The number of additional wind turbines that have been attached to a PGS is commensurate with the proportions of the wind power incorporation. Where the rate of penetration of wind energy ranges from 5% - 30% generates the output of wind energy ranges from 170 - 1020 MW from number WTG units (from 85 – 510).

This paper also examines the choice an appropriate penetration level of wind energy to attain a level similar to that of traditional PGSs. The DEO algorithm was used to analyze the RTS-79 system cases with changes in the number of WTG units, and changes in the penetration level of wind energy.

The effect of the wind energy penetration rate as a result of the number of available WTG units was tested on the values of reliability indicators for generation systems. Table 8 shows the different wind energy penetration levels (%), the number of WTG units and the output power of the WTG units, as well as the values of reliability indicators LOLE, LOEE.

Table 8: The wind power penetration beside on reliability indexes and percentage of installed capacity

Wind Pen. Rate %	No. of WTG	Product Power (MW)	LOLE	LOEE
5%	85	170	7.45	1318
10%	170	340	6.38	1364
15%	255	510	5.44	1214
20%	340	680	4.78	1256
25%	425	850	3.93	1079
30%	510	1020	3.60	1035

It is seen that the values of LOLE and LOEE are decreased for the RTS-79 system with additional WTG units. Figures 13 and 14 show LOLE and LOEE indicayors and the role of some of the WTG units added to RTS-79, where each WTG unit has a rated capacity of 2 MW.

Besides, it is noted that the LOLE value is lowered considerably in the event the wind energy penetration rate ranges from 5% to 20%; after which the rate of decline is relatively insignificant, implying that increasing the wind energy penetration to higher rates will have less effect in reducing the values of reliability indicators.

5. Conclusions

Both MCS and PBIs approaches have been used to assess the reliability of generating system combined with wind energy.

This paper presented a comprehensive performance clarification of the suggested technique for assessing the electric power systems dependability, such as; DEOA. The suggested intelligent algorithm which depends on the PBIs approach is viewed as a feasible alternative for the MCS in the assessment of non- sequential systems.

Furthermore, this study introduced the novel DEOA algorithm with WDM to assess the dependability of the PGSs combined with WTGs. It was shown that with the application of the proposed approach, it was likely that there would be simulation of the wind speed at a given hour for assessing the dependability of the power generation system. The main improvements achieved in this study are as presented below:

- Using the DEOA as an alternative to MCS for guided intelligent searches.
- The repeated runs of the experiment with the algorithm displayed the algorithm's elevated precision level when determining the RI.
- Incorporating WTGs into the power generating system, to revise the dependability evaluation process to adapt to this change.

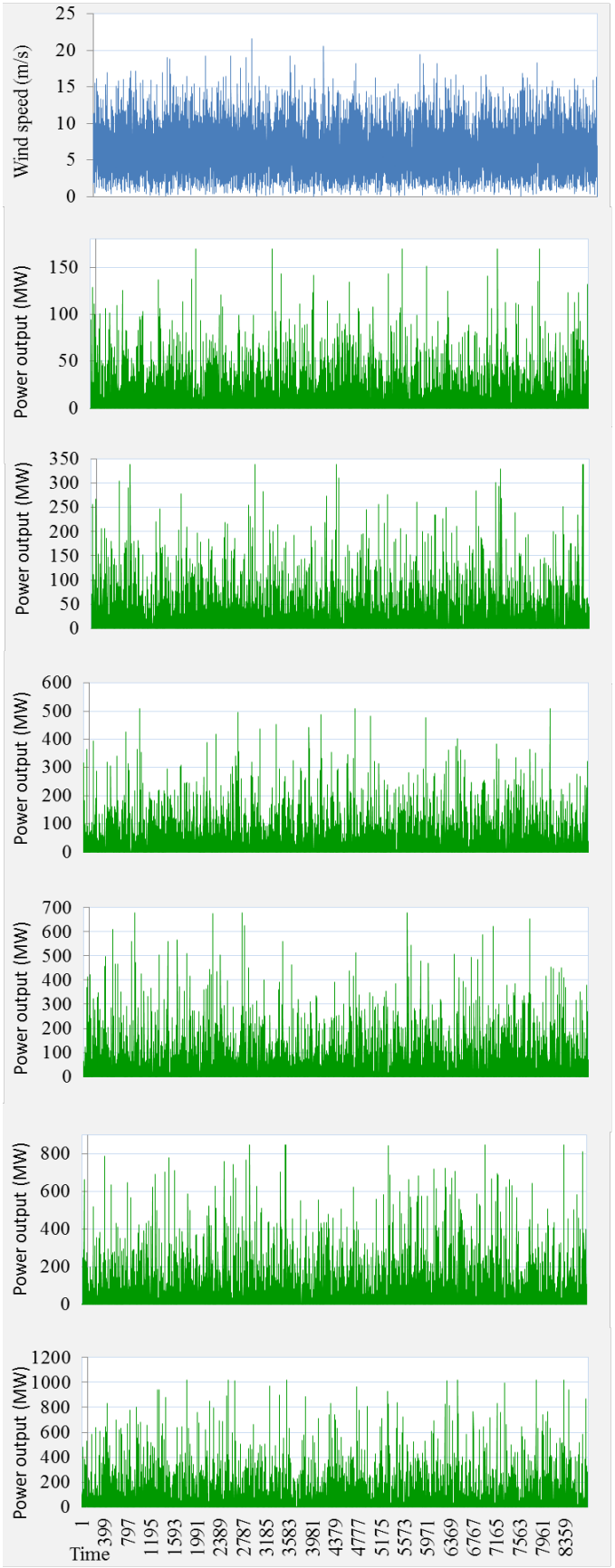


Figure 12: Wind forecasting and simulation of the generated wind energy for a year

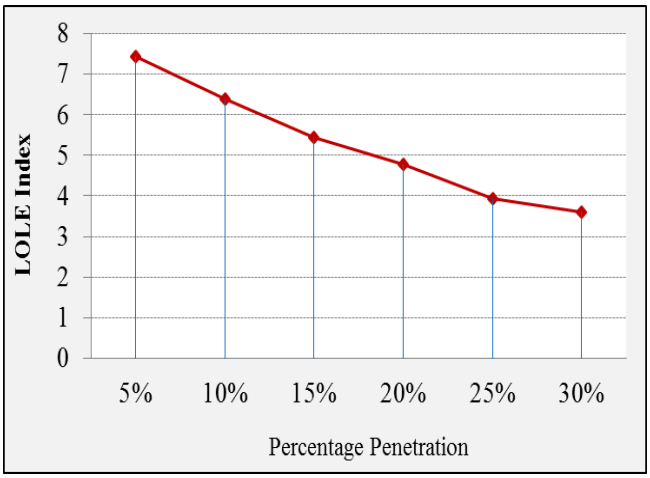


Figure 13: Shows the number of WTG units added to RTS-79 as a function for an estimate of the LOLE indices

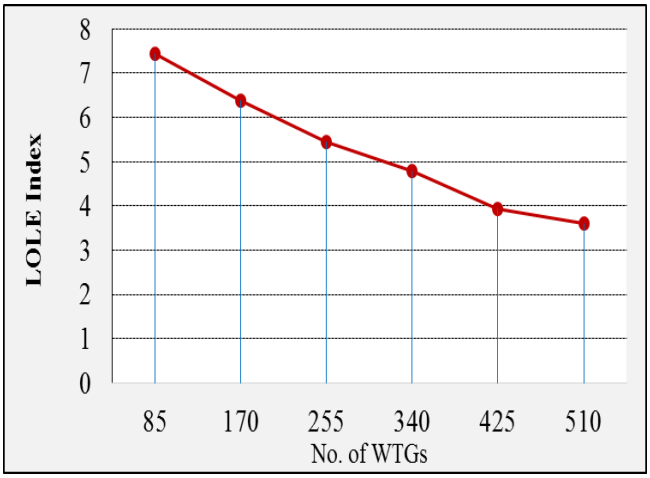


Figure 14: Shows the number of WTG units added to RTS-79 as a function for an estimate of the LOEE indices

On the other hand, RTS test system was used to make a comparison of the achievement of the DEOA with the conventional MCS to indicate the effectiveness of the proposed algorithm.

Besides, selecting an appropriate penetration level of wind power and determining its effect on the PGS and the essential element of investigating the potential of the wind energy source as a replacement for sustainable energy generation. Also, increasing the penetration rates of wind energy to higher rates will have less effect on reducing the values of RI.

References

[1] R. A. González-Fernández, A. M. Leite Da Silva, L. C. Resende, and M. T. Schilling, "Composite systems reliability evaluation based on Monte Carlo simulation and cross-entropy methods," IEEE Transactions on Power Systems,**28**(4), 4598–4606, 2013, doi: 10.1109/TPWRS.2013.2267154.

[2] A. M. L. Da Silva, R. A. Gonzalez-Fernandez, S. A. Flavio, and L. A. F. Manso, "Composite reliability evaluation with renewable sources based on quasi-sequential Monte Carlo and cross entropy methods," International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 1-6. 2014, doi: 10.1109/PMAPS.2014.6960618.

[3] M. H. Moradi and M. Abedini, "A combination of genetic algorithm and particle swarm optimization for optimal DG location and sizing in distribution systems," Electrical Power & Energy Systems, **34**(1), 66-74, 2012. doi: 10.1016/j.ijepes.2011.08.023.

[4] Y. Zhang, A. A. Chowdhury, and D. O. Koval, "Probabilistic wind energy modeling in electric generation system reliability assessment," IREP-Bulk Power System Dynamics and Control-VII Conference, 1–8, 2010, doi: 10.1109/TIA.2011.2127435.

[5] R. Karki, D. Dhungana, and R. Billinton, "An Appropriate Wind Model for Wind Integrated Power Systems Reliability Evaluation Considering Wind Speed Correlations," Appled Scinces, **3**, 107–121, 2013. doi: 10.3390/app3010107.

[6] S. H. Karaki, B. A. Salim, and R. B. Chedid, "Probabilistic model of a two-site wind energy conversion system," IEEE Transactions on Energy Conversion, **17**(4), 530–536, 2002. doi: 10.1109/TEC.2002.805215.

[7] H.J.Jayatheertha, "Evaluation of Composite Electric System Performance Indices Using Sequential Monte Carlo Simulation," International Journal of Advanced Engineering Research and Studies, **21**(2), 1–4, 2012.

[8] L. Wang, C. Singh, and K. C. Tan, "Reliability evaluation of power-generating systems including time-dependent sources based on binary particle swarm optimization," in 2007 IEEE Congress on Evolutionary Computation (CEC), 3346–3352, 2007. doi: 10.1109/CEC.2007.4424904.

[9] L. Wang and C. Singh, "Population-based intelligent search in reliability evaluation of generation systems with wind power penetration," IEEE Transaction on Power System, **23**(3), 1336–1345, 2008. doi: 10.1109/TPWRS.2008.922642.

[10] M. Benidris and J. Mitra, "Use of intelligent search methods in performing sensitivity analysis of power system reliability indices," The PES General Meeting, Conference & Exposition, National Harbor, MD, 2014, doi: 10.1109/PESGM.2014.6939915.

[11] R.C. Green II, L. Wang, and M. Alam, "Intelligent State Space Pruning with Local Search for Power System Reliability Evaluatio," in 2012 IEEE PES Innovative Smart Grid Technologies Europe (ISGT), 1-8, 2012.

[12] S. J. Huang and X. Z. Liu, "Application of artificial bee colony-based optimization for fault section estimation in power systems," Electrical Power & Energy Systems, **44**(1), 210–218, 2012. doi: 10.1016/j.ijepes.2012.07.012.

[13] R. C. Green, Z. Wang, L. Wang, M. Alam, and C. Singh, "Evaluation of loss of load probability for power systems using intelligent search based state space pruning," in 2010 IEEE the Probabilistic Method Applied to Power Systems (PMAPS), 319–324, 2010, doi: 10.1109/PMAPS.2010.5528892.

[14] A. Ali Kadhem, N. I. Abdul Wahab, I. Aris, J. Jasni, and A. N. Abdalla, "Computational techniques for assessing the reliability and sustainability of electrical power systems: A review," Renewable and Sustainable Energy Reviews, **80**, 1175–1186, 2017. doi: 10.1016/j.rser.2017.05.276.

[15] A. Almutairi, M. H. Ahmed, and M. M. A. Salama, "Probabilistic generating capacity adequacy evaluation: Research roadmap," Electric Power Systems Research, **129**, 83–93, 2015. doi: 10.1016/j.epsr.2015.07.013.

[16] J. Lin, L. Cheng, Y. Chang, K. Zhang, B. Shu, and G. Liu, "Reliability based power systems planning and operation with wind power integration: A review to models, algorithms and applications," Renewable and Sustainable Energy Reviews, **31**, 921–934, 2014. doi: 10.1016/j.rser.2013.12.034.

[17] D. Saxena, S. N. Singh, and K. S. Verma, "Application of computational intelligence in emerging power systems," International Journal Engineering Science Technology, **2**(3), 1–7, 2010. doi: 10.4314/ijest.v2i3.59166.

[18] A. A. Kadhem, N. I. Abdul Wahab, and A. N. Abdalla, "Wind energy generation assessment at specific sites in a Peninsula in Malaysia based on reliability indices," Processes, **7**(7), 1-23, 2019. doi: 10.3390/pr7070399.

[19] R. A. Bakkiyaraj and N. Kumarappan, "Optimal reliability planning for a composite electric power system based on Monte Carlo simulation using particle swarm optimization," Electrical Power & Energy Systems, **47**(1), 109–116, 2013. doi: 10.1016/j.ijepes.2012.10.055.

[20] Ž. B. Rejc and M. Čepin, "Estimating the additional operating reserve in power systems with installed renewable energy sources," Electrical Power & Energy Systems, **62**, 654–664, 2014. doi: 10.1016/j.ijepes.2014.05.019.

[21] M. Mosadeghy, T. K. Saha, R. Yan, and S. Bartlett, "Reliability evaluation of wind farms considering generation and transmission systems," the IEEE PES General Meeting Conference & Exposition, 1-5, 2014, doi: 10.1109/PESGM.2014.6939156.

[22] A. A. Kadhem, N. I. Abdul Wahab, I. Aris, J. Jasni, A. N. Abdalla, and Y. Matsukawa, "Reliability assessment of generating systems with wind power penetration via BPSO," International Journal on Advanced Science, Engineering and Information Technology, **7**(4), 1248–1254, 2017. doi: 10.18517/ijaseit.7.4.2311.

[23] A. K. Azad, M. G. Rasul, M. M. Alam, S. M. Ameer Uddin, and S. K. Mondal, "Analysis of wind energy conversion system using Weibull

distribution," International Conference on Mechanical Engineering (ICME), **90**, 725–732, 2014, doi: 10.1016/j.proeng.2014.11.803.

[24] A. A. Kadhem, N. I. Abdul Wahab, I. Aris, J. Jasni, and A. N. Abdalla, "Reliability assessment of power generation systems using intelligent search based on disparity theory," Energies, **10**(3), 1-13, 2017. doi: 10.3390/en10030343.

[25] C. Grigg and P. Wong, "The IEEE reliability test system -1996 a report prepared by the reliability test system task force of the application of probability methods subcommittee," IEEE Transactiona on Power Systems, **14**(3), 1010–1020, 1999. doi: 10.1109/59.780914.

[26] A. A. Kadhem, N. Izzri Abdul Wahab, and A. N. AbdAlla, "Differential evolution optimization algorithm based on generation systems reliability assessment integrated with wind energy," 5th International Conference on Power Generation Systems and Renewable Enegy Technologies (PGSRET), 1-6, 2019, doi: 10.1109/PGSRET.2019.8882705.

[27] C. Singh and V. Gubbala, "An alternative approach to rounding off generation models in power system reliability evaluation," Electric Power Systems Research, **7796**(95), 37–44. 1996.

[28] N. Samaan and C. Singh, "Adequacy assessment of power system generation using a modified simple genetic algorithm," IEEE Transactions on Power Systems, **17**(4), 974–981, 2002. doi: 10.1109/TPWRS.2002.804994.

[29] R. Billinton and W. Li, Reliability Assessment of Electric Power Systems Using Monte Carlo Methods. Plenum Press, New York, 1994.

[30] R. N. Allan, R. Billinton, and N.M.K. Abdel-Gawad "The IEEE Reliability Test System—Extensions to and Evaluation of the Generating System," IEEE Transactions on Power Systems, **PWRS(4)**, 1–7. 1986.

[31] V. Miranda, L. de Magalhaes Carvalho, M. A. da Rosa, A. M. L. da Silva, and C. Singh, "Improving power system reliability calculation efficiency with EPSO variants," IEEE Transactions on Power Systems, **24**(4), 1772–1779, 2009. doi: 10.1109/TPWRS.2009.2030397.

[32] S. Soleymani, M. E. Mosayebian, S. Mohammadi, and D. G. Gomes, "A combination method for modeling wind power plants in power systems reliability evaluation," Computer and Electrical Engineering, **41**, 28–39, 2015. doi: 10.1016/j.compeleceng.2014.12.005.

[33] P. Henneaux, F. X. Bouchez, and L. Rese, "Confidence intervals for adequacy assessment using Monte Carlo sequential simulation," in 2016 IEEE International Energy Conference, 1-6, 2016, doi: 10.1109/ENERGYCON.2016.7514142.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 341-346 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Enhanced Power Utilization for Grid Resource Providers

Tariq Alwada'n[1,*], Thair Khdour[2], Abdulsalam Alarabeyyat[3], Ali Rodan[4,5]

[1]*School of Computing, Engineering & Digital Technologies, Teesside University, Middleborough, UK*

[2]*Computer Information Systems Department, Al-Balqa' Applied University, Amman, Jordan*

[3]*Software Engineering Department, Al-Balqa' Applied University, Amman, Jordan*

[4]*Computer Information System Department, the Higher Colleges of Technology, AL Ain, UAE*

[5]*King Abdullah II School for Information Technology, the University of Jordan, Amman, Jordan*

ARTICLE INFO

Article history:
Received: 26 August, 2020
Accepted: 31 October, 2020
Online: 20 November, 2020

Keywords:
Grid Computing
Multi-Agent system
FIPA
Grid Resource Broker
Jade

ABSTRACT

A grid is a system that can manage and organize services and resources that spread amongst different control domains, employ interfaces and protocols, and offer a high quality of services. The integration of Multi-Agent Systems (MAS) with a grid environment significantly affects grid performance. MAS is considered a suitable solution for open systems that modify frequently. Grid offers a wide range of resources for its users, and some of these resources might not be used or utilized for some time before any new jobs come to the grid for processing. Usually, these resources are accumulated in massive data centers to cater to the grid users' growing demand. This accumulated will lead to consuming a considerable amount of electricity for their operation. The ever-increasing usage of grid computing has led to an increase in electrical energy usage by massive servers in their data centers. In this paper, we have proposed an automated system composed of modern agents that can be used the power wastage resulted from inactive servers inside the data centers for a specific amount of time. The proposed technique depends on switching inactive virtual and physical machines to lower power positions (Sleep/Wakeup or switched off) while still preserving customers' performance requirements. The Automated system has been tested and evaluated using the Jade tools. The results show that the newly proposed method can reduce power wastage for inactive grid resources.

1. Introduction

This paper is an extension of work presented initially in the ITT2019 conference [1]. Grid computing is an improved technology that involves a pool of linked machines that be owned by numerous organizations in diverse locations to construct a distributed system. The grid system can be accustomed to working with complicated scientific and business problems. It is created to assist in the sharing of distributed and diverse resources and to simplify resolving a considerable volume of computing issues [2]. The essential purpose of grid computing is to improve reliability, cut the cost of computing, and enhance flexibility by changing computers from an object that we purchase and operate privately to an object managed by a third organization [3]. However, grid computing is encountering various challenges, such as finding the right resources and decreasing the number of discarded jobs [4], [5]. Figure 1 displays an example of a grid computing environment [6], where the grid may contain different types of resources which need a resource management system in order to coordinates them.

The grid structural design shown in Figure 2 contains the following components [7]: a grid portal, which is also named a grid interface. It is used by the authorized grid clients to access the grid providers (resources). The portal should hide the complexity of the grid from clients by employing an uncomplicated interface. In this manner, it enables the sorting of grid job requisites.

The second element is the grid broker, where this element is described as the core of grid computing. It performs critical jobs in creating an effective grid system by connecting the user demands to the grid providers to fulfill the clients' tasks. The broker is also responsible for reducing transmission delays and increasing the

*Corresponding Author: Tariq Alwada'n, Email: t.alwadan@tees.ac.ae

resource assigning jobs through resources by avoiding the depending on a certain one. The grid broker's primary role is to locate and reserve the most appropriate support for the users' tasks. The broker receives the requests from the clients and sends the job's description files to the clients' requirements is responsible for monitoring the running tasks and hands over the outputs to their related users. To process the grid users' requests, the broker asks the Replica Catalogue Information Service to locate the available resources and data in the grid.

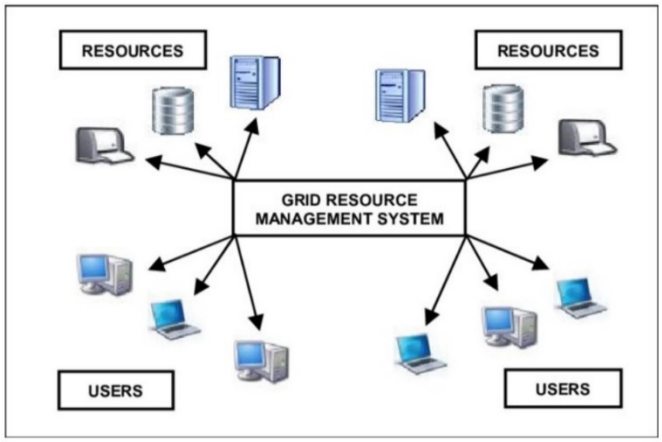


Figure 1: Grid Computing Environment

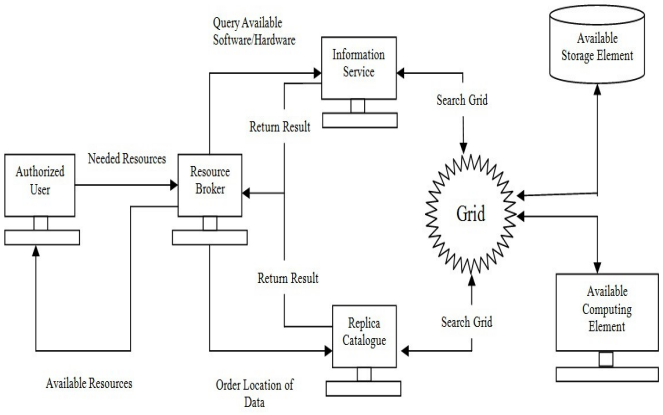


Figure 2: Grid Architecture. [8]

The third element is the Information Service. It is the directory service of the grid, where the grid keeps the information about all grid resources (nodes) and all the running jobs on those nodes. The grid resources have to register and update their information with the Information Services and their available resources. The grid broker retrieves all the information about the available and free resources from the Information Services. The information stored in the Information Services can be dynamic or static. The static information is employed to specify the operating system specifications and hardware requirements. In contrast, the dynamic information is associated with the current job running on the grid resources, the type of software specification, the available time in resources, disk space, and policies.

Replica Catalogue (RC) is the fourth element. RC is the directory that presents information, which helps in locating the saved data in the grid. To simplify employing the grid's data, the grid broker requests the RC for information about locations of the data in the grid and the way to identify the access mechanism needed to utilize this data.

To locate the best resources for the grid clients' requests, the grid broker gets the job specifications from the grid portal and looks for the suitable resources that can fit these specifications. For this purpose, the broker inquires for resources information and the data stored in the grid from the Information Service and the Replica Catalogue. Later it chooses the resource or resources that able to fit the job specifications.

Processing and modeling jobs have grown to be more complicated as the system scopes increased. Accordingly, it is difficult to carry out tasks with such an extensive system using centralized methods. Multi-Agent System (MAS) presents an excellent manner for delivering control. It also can be used to explore and change the local communications amongst entities [9]. Furthermore, artificial intelligence mechanisms can be combined with it. Intelligent agents are incorporated to operate autonomously in a changing environment. The intelligent agent also has the following main features: [10].

- Autonomy: the power to behave dynamically to a particular point in support of the clients and programs and adjust how agents accomplish their jobs.
- Cooperation and Communication: the power to cooperate and communicate with other agents to swap information, receives orders, and respond.
- Learning: the power to enhance the performance steadily while behaving with the external world.
- Re-activity: the ability to react to exterior calls and moderate the agents behave concerning these calls.
- Pro-activity: the ability to create choices as a consequence of internal evaluations.

In this introduction, we have introduced some background about Grid and MAS as a suitable solution for open systems, such as Grid, that modify frequently. This paper is organized as follows. Section 2 introduces related works to this research. Section 3 gives the system model. The experiment and results are presented in Section 4. Lastly, the conclusion of this research is given in Section 5.

2. Related Works

In recent times, combining the MAS with grid computing has taken the interest of lots of researchers. Most topics that have been concerned in the area of grid computing are how to achieve better use of grid data storage, service distribution, infrastructures, cost, and energy efficiency. Nevertheless, the worry in the agent environment's area is on the intelligent side of agents and the procedures that can be exploited to solve complicated problems.

The authors in [11] suggested a multi agent system that can support the growth of the IDAPS micro-grid where this proposed system includes a DER agent, database agent, and control and a user agent. The Agents transmit their requirements using a TCP/IP protocol embedded in the IEEE FIPA standard to assure the system capability. In addition, authors in [12] suggested a model for agent-based grid computing (AGEGC). Based on their model, AGEGC

was created by using the MAGE (Multi-Agent Environment) platform. This research included various levels, such as; information tier, knowledge tier, service-oriented tier, and operating system tier. AGEGC focuses on service-oriented tier provided that current processing system.

In recent times, there are many paradigms concerning the growth of agent systems, such as; MOLE [13], DMARS [14], RETSINA [15]. Lots of companies are contributing to building characteristics for standard agent models; for example, OMG [16] KSE [17] and FIPA [18].

Garimella in [19] employs a new preserve server that works in assistance with the Dynamic Soft Real-Time system [20]. The purpose of the Garimella system is to reserve the resource of the CPU beforehand. By using this system, the user request for some quality of service conditions. E.g. of that the percentage of CPU needed and the beginning time and the duration. Once the request for the preservation is begun, the reserved resources are ready to be reachable for the users. One more proposed model is the Resource Broker (RB) presented in [21] merged with a new preserve server shown in [19]. Both system have continual and quick response times and numerous negotiation probability for the users. Nevertheless, the volume of re-negotiation introduces an extra-large cost to the system, in the event of an unpredicted lack of resources.

The authors in [22] presented several algorithms for keeping an improved preservation approach for supercomputer scheduling systems. Those algorithms improve traditional organizing techniques by combing the reservation demand mechanism with scheduling regular tasks mechanism. Using this way of reservations, let clients ask for more than one resource in parallel from the scheduling systems at the same time. However, the mechanisms used in [22] preserve the whole time slot, which means the resources are not carried on in a distributed way by multiple users for the same amount of time.

3. System Model

Our suggested architecture is adopted from the grid architecture in [23], as shown in Figure 3. It employs Client/Server architecture since it is the most famous architecture in the heterogeneous system [24]. Our proposed architecture uses multi-tier architecture, considering that there is a necessity for the parted database to save the grid resources' location. Our grid architecture includes the following entities; the grid portal, the grid broker, and the grid resource providers.

A grid portal is an interface for an authorized client to send their jobs to the grid. A portal should hide the complexity of the grid from clients by employing an uncomplicated interface. In this manner, it enables the sorting of grid job requisites. The grid broker can connect with agents that operate, control, and support both the grid clients and grid resource providers. The broker is an agent handler who organizes and monitors the running jobs and maintains the grid clients updated with their running jobs.

The grid broker in this paper is moved from operating as a database warehouse that saves the available resources and conditions and matches the clients' jobs with these resources into an agent manager with the ability to control event creations and create and eliminate several agents in the system. The grid broker in our proposed architecture, is a MAS that can interconnect with the grid system on account of grid users to find the information regarding grid resource provider whom they can fit the users' requirements and choose the best one among them. Furthermore, our advanced system can provide and monitor the implementation of the jobs and support the grid users with live feedback concerning the job processes.

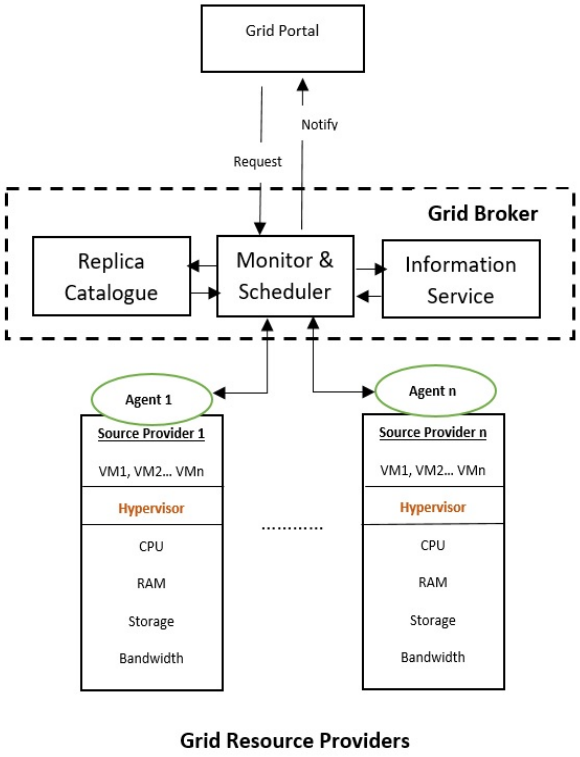


Figure 3: System Architecture

The grid broker is split into three primary entities. The first one is the monitor & scheduler. This entity is in charge of getting the grid clients' tasks and deciding the best available resource providers. The scheduler requests all data about the available resources from the Information Service entity and the data information saved in the Replica Catalogue (RC) entity. The monitor has the authority to monitor and provision the currently running jobs in the grid resource providers. The grid broker might have the improvements and ability to split clients' tasks between multiple resource providers to cut the cost and speed up the operations.

After the Monitor & Scheduler transmit the jobs to the available resource(s) for processing, it receives frequent feedbacks from the agents at the resource providers to verify the standing of the processing jobs and the percentage of the job completed to the grid broker. This point helps the grid broker to provision the running jobs' accomplishment, and by its turn, it will help determine an alternate resource provider(s) in case of failure at any point. The local agent at the resource providers is similarly in charge of updating the grid broker as the job is finished.

The resource provider can include several physicals and virtual machines. The Hypervisor in every machine is in charge of organizing and generating virtual machines on the upper of every physical machine. To react to the fast development of clients'

requests for data processing ability and storage, grid resource providers create a massive amount of data centers throughout the world. The grid datacenters typically include a large number of interconnected resources from physical machines and virtual machines, which will consume a considerable amount of electricity for their operations.

The proposed system depends on switching inactive virtual and physical machines to lower power positions (Sleep/Wakeup or switched off) while still preserving customers' performance requirements. The local agents at the resource providers are responsible for this operation. Each virtual machine (VM) has a local agent that can report the VM utilization to the data center's primary agent. The utilization here means the CPU utilization. Suppose the utilization for a specific VM has a zero amount for 300 seconds. In that case, the data center's primary agent will automatically shut down this VM to reduce the power consumption of the physical machine containing that turned off VM.

4. Implementation and Testing

The Jade platform was chosen to execute our proposed topology. Jade stands for the Java Agent Development framework, developed to help build out and enhance agent software as ruled by the FIPA conditions for smart MAS [25].

We introduced a simulation that deals with 200 physical hosts. The specifications of physical and virtual machines are shown in Table 1 and Table 2, respectively.

Table 1: Specification of Physical Machines

Type	Name	MIPS	RAM (MB)	Bandwidth (Gbps)	Number of Hosts
1	HP ProLiant ML110 G4 servers	2000	4096	2	100
2	HP ProLiant ML110 G5 servers	3000	4096	2	100

Table 2: Specification of Virtual Machines

Type	MIPS	RAM (MB)	Bandwidth (Gbps)	Number of VM
1	2000	1024	100	200
2	1500	2048	100	200
3	500	512	100	200

To create real power-consuming data, we took advantage of the actual data of power consumption offered by the SPECpower benchmark [26] and [27], as shown in Table 3. The power consumption by physical machines can be precisely expressed by a linear relationship between CPU utilization and power consumption. It can be noticed from Table 3 that even at small utilization, the machine exhausts a large volume of power. Therefore it is necessary to shut down such machines or reduce the number of VMs in it when not in use to reduce the power consumption of the physical machine. Figure 4 shows part of our simulation.

Table 3: Power Consumption (Watts per hour) at different load levels.

Machine type	0%	20%	50%	70%	100%
1	86	92.6	102	108	117
2	93.7	101	116	125	135

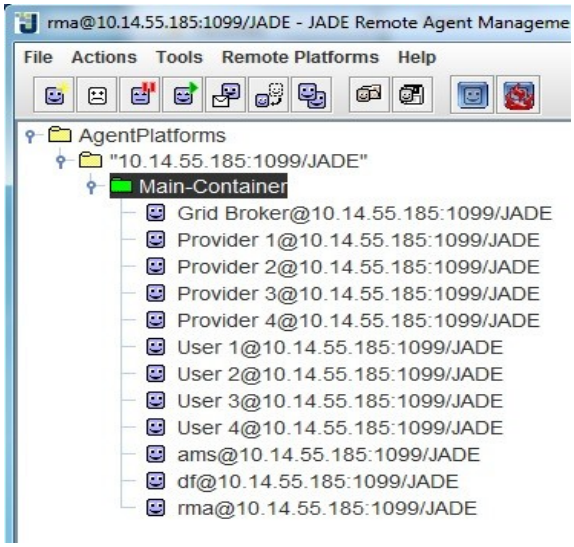


Figure 4: Part of the Summation Stage

The following stages explain the main steps of our simulation:

- Grid users state their tasks by choosing the required hardware and software specification to run their jobs through the grid portal.
- The grid broker receives the grid users' jobs and looks for the best provider(s) for those jobs. This step is done by helping the Information Services and the Replica Catalogue.
- Later, the grid broker sends the jobs to the suitable resource provider(s) and connects to the primary agent attached to that provider.
- The grid broker can divide the job between multiple resource providers to cut down the cost and speed up the operations.
- As the job is running at the resource provider, the provider's agent (primary agent) sends periodic feedbacks about the percentage of completion of the running job to the grid broker. This stage helps the grid broker to supervise the completion of the running job, and that will help find an alternative resource provider in case of failure.
- At the same time, the primary agent at the resource provider periodically checks the CPU utilization of its VMs at the physical machines.
- If the utilization reaches the zero point for 300 seconds, the primary agent will automatically shut down that VM.
- Finally, the grid broker sends the job outputs to grid users.

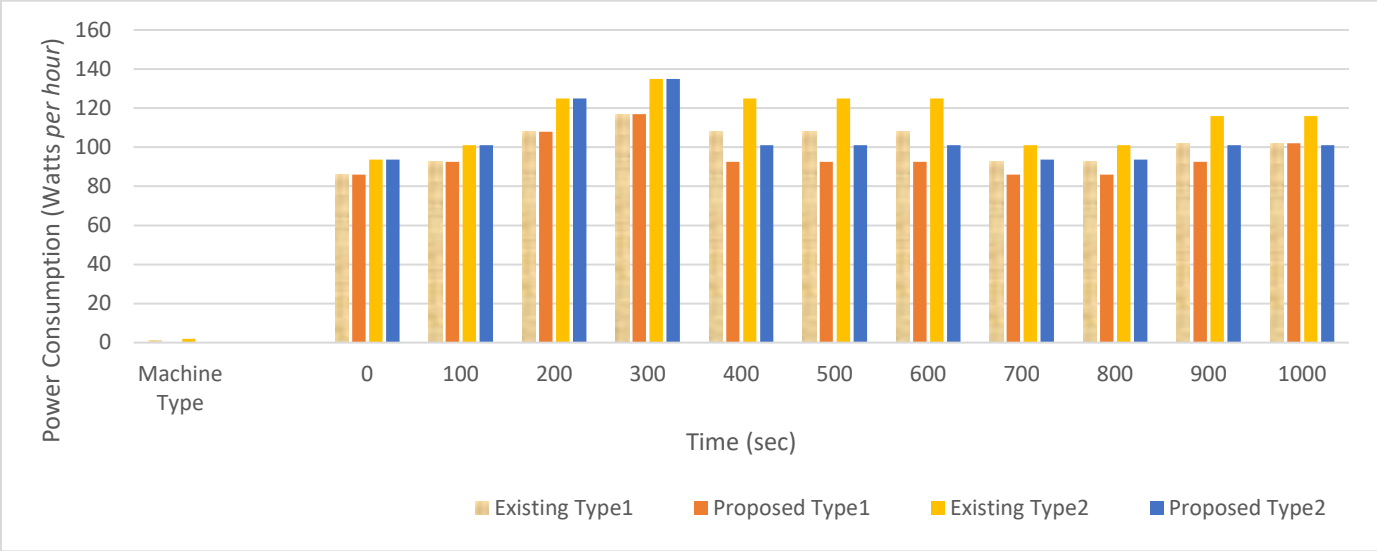


Figure 5: Simulation Results

From the simulation results in Figure 5. It can be seen that by using our proposed system, the power consumption has been slightly reduced for the two types of physical hosts (HP ProLiant ML110 G4 servers and HP ProLiant ML110 G5 servers), especially after the first 300s. This reduction is because the utilization of the physical machines has been reduced as the primary agents in both physical machines automatically shut down that VMs if the its utilization reaches the zero point for 300 seconds. As a result, the power consumption has been reduced in total for machines.

5. Conclusion

The purpose of the grid is to produce a broad scale, and heterogeneous systems intend to solve industrial or scientific problems. The integration of the Multi-Agent Systems (MAS) with the grid environment has a significant impact on grid performance. Grid offers a wide range of resources for its users, and some of these resources might not be used or utilized for some time before any new jobs coming to the grid for processing. As a result, the grid servers will consume a considerable amount of electricity for their operation. The growing usage of grid computing has led to an increase in electrical energy usage by the massive amount of servers in their data centers. In this paper, we have proposed an automated system composed of modern agents that can be used to reduce the power wastage resulted from inactive servers inside the data centers if the utilization of the VM reaches the zero point for 300 seconds, In this case the primary agent in the physical machine will automatically shut down that VM. The Automated system has been tested and evaluated, and the results show that the newly proposed method can reduce the amount of power wastage for the inactive grid resources.

References

[1] T. Alwada'n, A. Alarabeyyat, T. Khdour, and A. Rodan, "Utilizing Multi-Agent Systems in Grid Environments," Sixth Intentional Conference on HCT Information Technology Trends (ITT), Ras Al Khaimah, United Arab Emirates, 2019, doi: 10.1109/ITT48889.2019.9075099.

[2] T. Alwada'n, H. Aldabbas, H. Janicke, T. Khdour, and O. Aldabba. Dynamic "policy management in mobile grid environments", International Journal of Computer Networks & Communications (IJCNC), **4**(2), 2012.

[3] H. Alhakami, H. Aldabbas, and T. Alwadan. "Comparison between cloud and grid computing: Review paper", International Journal on Cloud Computing: Services and Architecture (IJCCSA), 2, 2012.

[4] T. Alwadan, H. Janicke, O. Aldabbas, and M. Alfawair. "New framework for policy support for mobile grid services", The 6th International Conference on Risks and Security of Internet and Systems (CRISIS2011), Romania 2011.

[5] T. Alwadan, H. Janicke, O. Aldabbas, and H. Aldabbas. "New framework for dynamic policy management in grid environments", In Recent Trends in Wireless and Mobile Networks, Third International Conferences, Springer, 2011.

[6] Ku. Ku-Mahamud and H. Abul Nasir. "Ant Colony Algorithm for Job Scheduling in Grid Computing", Asia International Conference on Modeling and Simulation, 2010, 10.1109/AMS.

[7] T. Alwadan, H. Janicke, A. Alarabeyyat, A. Alkharabsh, and T. Khdour. "Policy-based support for mobile grid services", International Journal of Computer Science Issues (IJCSI), 10, 2013.

[8] T. alwada'n, H. Aldabbas, and H. Janicke, editors. "Mobile Grid System: Policy and Privacy", LAP LAMBERT Academic Publishing, 2014.

[9] A. S. Grimshaw, M. A. Humphrey, and A. Natrajan. "A philosophical and technical comparison of legion and globus", IBM J. Res. Dev., March 2004.

[10] D. Talia. "Cloud computing and software agents: Towards cloud intelligent services", 2011.

[11] M. Pipattanasomporn, H. Feroze, and S. Rahman. "Multi- agent systems in a distributed smart grid: Design and implementation", 2009.

[12] Z. Shi, He Huang, J. Luo, Fen Lin, and H. Zhang. "Agent-based grid computing. Applied Mathematical Modelling", 2006, 30:629640, 07.

[13] J. Baumann, F. Hohl, K. Rothermel, and M. Straßer. "Mole – concepts of a mobile agent system", World Wide Web, Sep 1998.

[14] A. S Rao and M. P George. "Bdi agents: From theory to practice", In Proceedings of the First International Conference on Multi- Agent Systems (ICMAS-95), 1995.

[15] K. Sycara, A.e Pannu, M. Williamson, D. Zeng, and K. Decker. "Distributed intelligent agents", IEEE Expert: Intelligent Systems and Their Applications, December 1996.

[16] D. Milojicic, M. Breugst, I. Busse, J. Campbell, S. Covaci, B. Friedman, K. Kosaka, D. Lange, K. Ono, M. Oshima, C. Tham, S. Virdhagriswaran, and J. White. "Masif: The omg mobile agent system interoperability facility". Personal Technologies, Jun 1998.

[17] P. Patel-Schneider, D. P McKay, T. Finin Thomas, G. Ramesh Patil, R. Fikes, and R. Neches. "The DARPA Knowledge Sharing Effort: Progress Report". Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning. Morgan Kaufman, August 1992.

[18] fipa.org. "The foundation for intelligent physical agents", Website, 2019.

[19] G. Garimella, "Advance CPU Reservations with the Dynamic Soft Real-Time Scheduler", Master's Thesis, University of Illinois at Urbana-Champaign, 1999.

[20] H. Chu and K. Nahrstedt, "A Soft Real Time Scheduling Server in UNIX Operating System," European Workshop on Interactive Distributed Multimedia Systems and Telecommunication Services (IDMS '97), Sep. 1997.

[21] K. Kim and K. Nahrstedt, "A Resource Broker Model with Integrated Reservation Scheme," IEEE International Conference on Multimedia and Expo 2000 (ICME '00), Aug. 2000.
[22] W. Smith, I. Foster, and V. Taylor, "Scheduling with Advanced Reservations," International Parallel and Distributed Processing Symposium (IPDPS '00), May 2000.
[23] T. Alwada'n, S. Alghyaline., A. Alazzam. "High Priority Requests in Grid Environment", International Journal of Advanced Computer Science and Applications, No. 10, 2020.
[24] C. Jean Dollimore, and T. Kindberg. "Distributed Systems: Concepts and Design", fourth Edition, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.
[25] F. Bellifemine, A Poggi, and G. Rimassa. "Jade - A FIPA- compliant agent framework", The Practical Application Company Ltd., 1999.
[26] A. Verma, P. Ahuja, A. Neogi. "pMapper: power and migration cost aware application placement in virtualized systems", Proceedings of ACM/IFIP/USENIX 9th International Middleware Conference, Leuven, Belgium, 2008.
[27] N. Patel, H. Patel. "Energy efficient strategy for placement of virtual machines selected from underloaded servers in compute Cloud", Journal of King Saud University - Computer and Information Sciences, 6, 2020.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 347-352 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

DC-DC Buck Converter Driver with Variable Off-Time Peak Current Mode Control

Osvaldo Gasparri[*,1], Paolo Del Croce[2], Andrea Baschirotto[1]

[1]*Department of Physics "G.Occhialini", University of Milano, Bicocca, 20126, Italy*

[2]*IFAT DC ATV BP PD ADL, Infineon Technologies, Villach, 9500, Austria*

ARTICLE INFO

Article history:
Received: 29 August, 2020
Accepted: 01 November, 2020
Online: 20 November, 2020

Keywords:
DC-DC Converter
Peak Current Mode Control
Simulink
dSpace

ABSTRACT

Power converters in automotive industries are needed to source electronic devices. DC-DC Buck converter allows to drive loads ensuring their safety operation taking the car battery voltage as an input to generate a sub-voltage or current. By means of a control loop, the system makes sure that the output variable has the desired value in each operation condition. The Peak Current Mode Control (PCMC) is used to control the output current. Whenever the current reaches the reference maximum value (by load specs), the system acts to lower it. Intrinsically over-current protection is guaranteed. The paper presents an improvement of the basic PCMC, where, instead of just controlling the peak value, the loop controls also the average, more delicate feature, resulting in a more reliable driver system. The circuit is able to source 3A average-4.5V from a 13.5V nominal battery voltage, with a peak current of 3.3A and a 0.6A maximum ripple (3±10%A). The concept has been designed and simulated on Simulink and then tested on hardware using dSpace, a Rapid Control Prototyping set-up.

1 Introduction

Electronic devices in Automotive industries need for power converters to be correctly sourced. Indeed, battery voltage could range between 4.5 to > 27V, making it unsuitable for driving any load without provoking damages or malfunctions. DC-DC Converters allow to take as an input the car battery voltage and return as an output a controlled voltage or current required by the load. What is more, the DC-DC protects the load against over-current and over-voltage, ensuring its correct operation. Meanwhile, integrated circuits are required to occupy small space and reduce external components to minimize the associated cost. The core of a Buck DC-DC converter is a switch, or Power-DMOS, which is repeatedly turned on and off, allowing the voltage/current to increase/decrease quickly in the neighborhood of a desired value. The circuit that turns on and off the switch is the control circuit [1]. The control could be either made on the output current or voltage. This paper presents a current control loop able to control all the current features (peak, average and ripple) at once. This paper is an extension of work originally presented in the 26th IEEE International Conference on Electronics Circuits and Systems (ICECS) [2].

The solution arises from the well-known Peak-Current-Mode Control (PCMC), [3] operating in Continuous Conduction Mode (CCM). Instead of having a constant switching period, as it happens in the typical PCMC, the presented control loop has a variable period. In particular, the time in which the power-DMOS stays off, changes accordingly to the calculated current average value, to make it closer to a target desired value. The loop will be now on referred to it as Variable T_{off}. Moreover, the typical PCMC intrinsically suffers of instability issues directly attributable to the fixed switching frequency [4]. Therefore, by making it variable, such issues are consequently solved. The use of stabilization techniques such as Slope Compensation [4] can be avoided, which increase the complexity of the system without improving the control precision (indeed current features will change accordingly to the input voltage). Also protection against over-current is already guaranteed: indeed the maximum current, namely the peak, is fixed. The proposed system operates with 500kHz maximum switching frequency, delivering an average output current of 3 ± 10%A with a load-fixed 4.5V output voltage. The current peak will be limited to 3.3A. The application case could be, for instance, a LED driver, where the LED itself, as soon as the correct current is provided, fixes the output voltage. The circuit has been simulated and then tested in laboratory by using an emulator called dSpace. Instead of a LED, in first place has been used as a load a linear component: a 1.5Ω resistor.

Finally, the current sensing is integrated, with no need for

[*]Corresponding Author: Santa Maria Hoè (LC) Italy, o.gasparri@campus.unimib.it

an external pad to sense the current, resulting in a lighter circuit. Compared to the most common solutions [5], the here used fully-integrated Current Sensing technique requires less external components, with consequent lower cost in terms of BoM and board fabrication.

The paper is organized as follows. The algorithm principle is presented in Section II, together with the formula useful for its implementation, described in Section III. The simulation results are shown in Section IV and the conclusions are drawn in Section V.

2 Variable-T_{off} PCMC Algorithm

A Buck DC-DC converter is a device which in principle allows to convert a given input voltage to a lower output voltage by turning on and off a Power-DMOS, able to withstand to high power. In the application at stake from [4.5, 27]V are obtained 4.5V in output by controlling the current, I_L, flowing into the load. A simple scheme of a Buck converter is given in Figure 1.

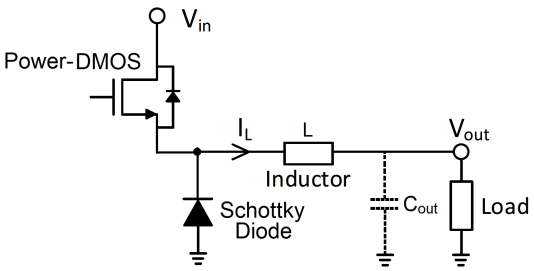


Figure 1: Buck converter scheme

Rather than other diodes, a Schottky diode has been used for its lower threshold (0.5V) and higher speed. The presence of the output capacitor, which minimizes the voltage ripple, can be neglected when controlling the current (and the output voltage is fixed by the load) leading to a lighter circuit. The switching period is divided into two phases: the ON-phase when the Power-DMOS conducts (T_{on}) and the OFF-phase when the diode conducts (T_{off}). Referring to the Duty-Cycle (D):

$$\begin{cases} T_{on} = D * T \\ T_{off} = (1 - D) * T \quad with \quad T = T_{on} + T_{off} \end{cases} \tag{1}$$

The Buck converter should sink an optimum average current I_{avg} to make the load work properly and, to avoid damages, the current must fluctuate limitedly. The current ripple should be indeed limited to the average value±10%, namely 0.3A. Consequently, the maximum current, the peak value, should be 3.15A. The rule of the control loop is to ensure that these specifications are met. The relation between the current features can be expressed as follows:

$$\Delta I_L = 2 * (I_{max} - I_{avg}) \tag{2}$$

In standard PCMC, the period T is fixed. Conversely, in Variable T_{off}, the period is variable. In particular, T_{on} changes automatically according to the input voltage, while T_{off} is forced to vary every cycle by the control loop, based on the vicinity between the calculated output average value of that cycle and the desired value.

The load current during the ON-phase increases with positive slope given by:

$$\left(\frac{dI_L}{dt}\right)_{ON} = \frac{\Delta I_L}{T_{on}} = \frac{V_{in} - V_{out}}{L} \tag{3}$$

Once I_L reaches I_{max}, the power-DMOS turns OFF and the output current decreases with a negative slope:

$$\left(\frac{dI_L}{dt}\right)_{OFF} = \frac{\Delta I_L}{T_{off}} = -\frac{V_{out}}{L} \tag{4}$$

After a time T_{off}, the DMOS is turned ON again and so on. The output current waveform (equal to the inductor current) has the typical triangular CCM shape as in Figure 2.

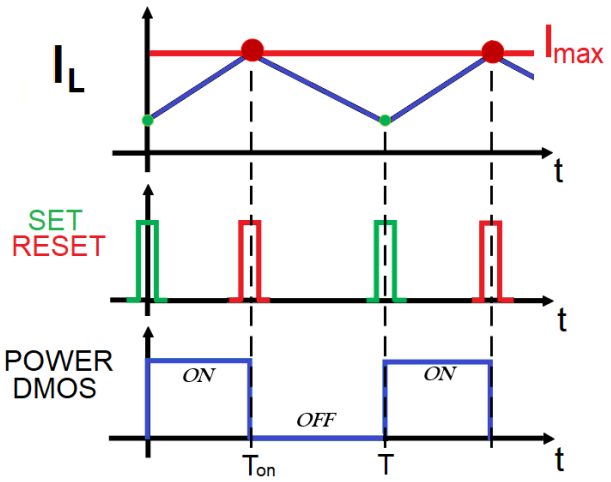


Figure 2: Output current

While I_{max} is fixed, the I_{min} is free to change. The loop, by means of a sensing technique, is able to read I_{min}. Indeed, the current sensing happens during the ON phase to insure the current doesn't grow beyond I_{max}. Therefore, to get I_{min}, it just needed a register to store the first current value when the ON phase begins. Once I_{min} is known, the loop calculates:

$$I_{avg} = \frac{I_{min} + I_{max}}{2} \tag{5}$$

Then, the difference between I_{avg} and $I_{mean,ref}$ = 3A gives an error with sign that, once properly amplified by a factor K, is summed to the previous T_{off} value. The following T_{off} value will be:

$$T_{off,next} = T_{off,prev} + k \cdot (I_{avg} - 3A) \tag{6}$$

In this way, if e.g. I_{avg} is higher than 3A, the T_{off} value will be increased (causing a stretching of the switching period) and allowing the current to decrease more during the OFF phase, reaching a lower I_{min} with the aim of reducing the average value. If otherwise I_{avg} is lower than 3A, the error will be negative, meaning that the T_{off} of the next switching cycle must decrease to make the average value higher and matching with the target. The loop operation is resumed in the flow-chart in Figure 3. The total number of regressions in the algorithm depends on the current rising slope, on the sensing technique sampling frequency and on the counter frequency. Every time a switching cycle ends, the operation restarts from the beginning giving rise to the next switching cycle.

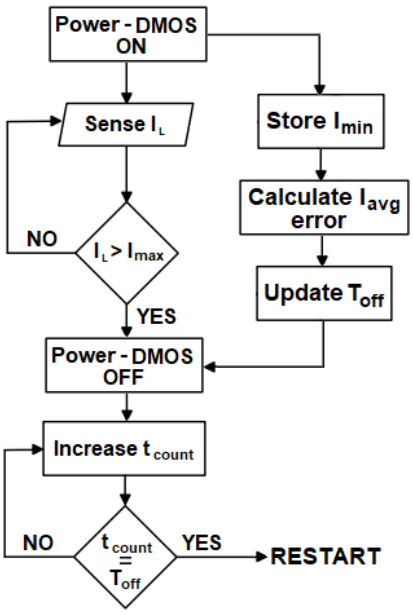


Figure 3: Variable-T_{off} PCMC algorithm flowchart

While T_{off} is decided by the loop, the T_{on}, from (3), changes not only with the the ΔI_L -loop dependent- but also with the battery voltage:

$$T_{on} \propto \frac{1}{(V_{in} - V_{out}) \cdot \Delta I_L} \tag{7}$$

It should be noticed that a first pre-memorized T_{off} value guess must exist (to be later overwritten by the cycle) to avoid malfunction at the cranking of the system during the very first switching cycle. This first guess is based on the application: if for instance the desired average is 3A, the peak is 3.15A, the inductance 30μH and $V_{out} = 4.5V$, it follows that the T_{off} value, from (2) and (4), could be grossly chosen as:

$$T_{off} = \frac{L \cdot \Delta I_L}{V_{out}} = 4\mu \tag{8}$$

Then, given some V_{out} variations due to the load or some uncertainties related to the inductance value, the references etc., the loop will act as described above, varying the T_{off} value to better adapt the system to the load specifications.

Moreover, the T_{off} value has an inferior limit, due to the maximum power-DMOS frequency (f_{max}) allowed. Ideally, the f_{max} can be evaluated for T_{on} tending to zero, namely:

$$f_{max} \simeq \frac{1}{T_{off}} \tag{9}$$

In realty, the minimum T_{on} value won't be zero, but, from (3), for a $V_{in} = 27$V it will be about 0.8μ s.

3 Instability digression

The typical PCMC loop is just able to control the maximum load current, no matter what ripple and average current are. Therefore, such a solution is not suitable in applications where loads require a precise average current with small shifts around its value. However, the biggest issue of the typical PCMC comes from its instability when the duty cycle grows over 50% [6]. Every time $T_{on} > T_{off}$ or, equivalently, $V_{out} > V_{in}/2$, the system would be unstable. In the application case, where $V_{in} = [4.5, 27]$ V and $V_{out} = 4.5$ V, problems would arise when $V_{in} < 9$V. Figure 4 shows what happen when crossing the limit case $D = 0.5$: if the duty cycle is increased the PCMC is no longer helping in controlling the current. The reason behind this issue is that, in typical PCMC, after a fixed period T, the power-DMOS is forced to be turned ON. Instead, in Variable-T_{off}, the power-DMOS turns ON T_{off} seconds after the current has peaked. Contrarily to the typical PCMC, if the input voltage changes, i.e. T_{on} changes, the T_{off} doesn't change (4), the ripple doesn't change and therefore the current features are not impacted. The system stays stable. Techniques as Slope compensation [4] are no longer needed.

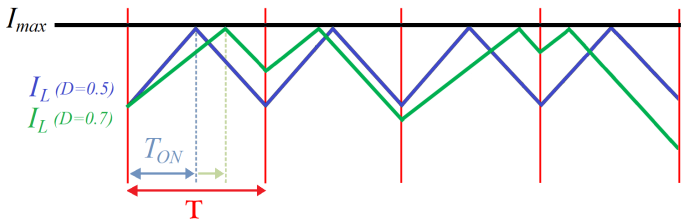


Figure 4: Output current in typical PCMC

4 Constant-T_{off} Algorithm Implementation

The circuit implementation is shown in Figure 5. The only external components are the inductor (L) and the Schottky Diode. As internal components, enclosed in blocks, there are: the Current Sensing, the Control Logic blocks, the Protection, the Internal Reference (including a bandgap voltage reference to generate I_{max} and I_{mean}), the Gate Driver to turn ON and OFF the Power-DMOS [7] and the Power Supply to generate a low voltage domain for the low-side circuit.

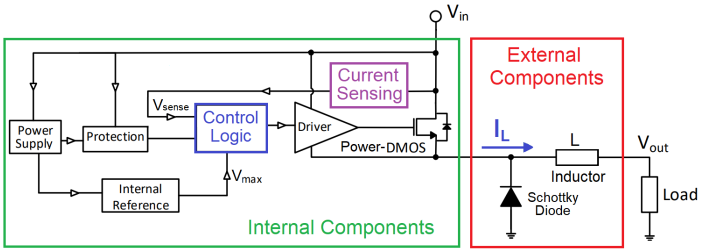


Figure 5: Concept top level

The Current Sensing block is used to read the current flowing I_L. The Power-DMOS current is here mirrored with a factor $N = 1/1000$ and flows through a sensing resistance. Based on the read voltage drop, the system is able to determine the magnitude of the current. Since the proposed technique only operates during the ON-phase (when the Power-DMOS conducts) the power consumption is lower than in the case on which the current is sensed on

the output node, always operating. Furthermore, an internal sensing block also allows to minimize the area occupation on silicon.

4.1 Control Logic block

Figure 6 shows the concept Control Logic block implementation. The result of the comparison between I_{sense} and I_{max} gives the Reset of the SR Latch. When $I_{sense} \geq I_{max}$, the Reset activates and the Power-DMOS turns OFF through the Gate Driver. The Set input is instead connected to a counter consisting of an oscillator and a chain of four D Flip-Flops. As soon as the Power-DMOS turns OFF, the counter starts counting up to T_{off} duration. Afterwards the Set is activated and the Power-DMOS turns ON. The current value at this moment, I_{min}, is stored into a register. The loop then calculates the average error on respect to the desired I_{mean} value. The error, summed to the previous $T_{off,prev}$ value returns the $T_{off,next}$ as in (6).

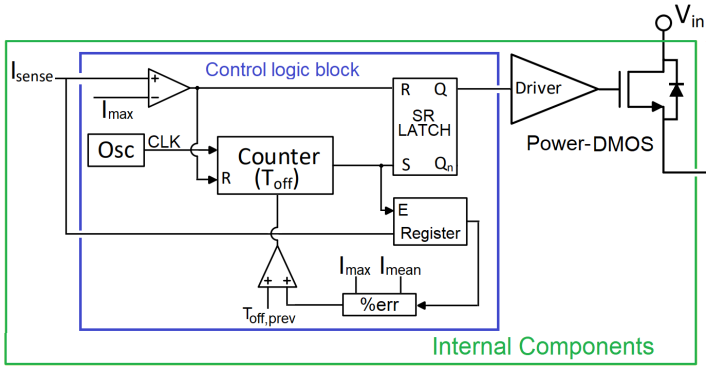


Figure 6: Variable T_{off} Current Control Loop implementation

5 Design example and Simulations results

Table 1 resumes the application specs.

Table 1: Summary of specifications

Technology	0.35 μm
V_{in}	4.5-27 V
V_{out}	4.5 V
f_{max}	500 kHz
Maximum Current Level	3.3 A
Mean Current Level	3 A
Output Current Ripple	0.60 A

By applying the formulas given in Section I, the design parameters are obtained. (9) gives the lower T_{off} limit:

$$T_{off} \simeq \frac{1}{500kHz} = 2\mu s \tag{10}$$

In fact, $1/f_{max} = T_{on,min} + T_{off}$ and (3) becomes:

$$V_{out} \cdot T_{off} = (V_{in,max} - V_{out}) \cdot (1/f_{sw,max} - T_{off}) \tag{11}$$

Therefore a more precise estimation of the minimum T_{off} value is given by:

$$T_{off} = \frac{27V - 4.5V}{27V} \cdot \frac{1}{500kHz} = 1.7\mu s \tag{12}$$

Once the ripple (Table 1) and T_{off} (12) are given, both the peak and the average current are fixed. Furthermore, by using these parameters in (4), the inductance value can be chosen as:

$$L > \frac{V_{out} * T_{off}}{\Delta I_L} = 13\mu H \rightarrow 30\mu H \tag{13}$$

(to stay large, considering also 20% of inductance value uncertainty on the nominal value and considering that actually the typical operation is centered around V_{in} = 16V).

5.1 Simulation results

The concept of Figure 5 has been firstly implemented on Simulink (MATLAB) to ideally simulate the behaviour of the high-level concept design. Simulink works with a block approach system, shown in Figure 7, which follows exactly from Figure 5. Other software are actually exploited in the design phase (such as Virtuoso-cadence). However, there, the circuit must be designed in a deeper transistor-level. Since the design may take large amounts of time, prototyping must be done previously ensuring that the chosen topology is the most suitable one for the application in question and that its worthwhile to proceed with transistor-level design and future fabrication.

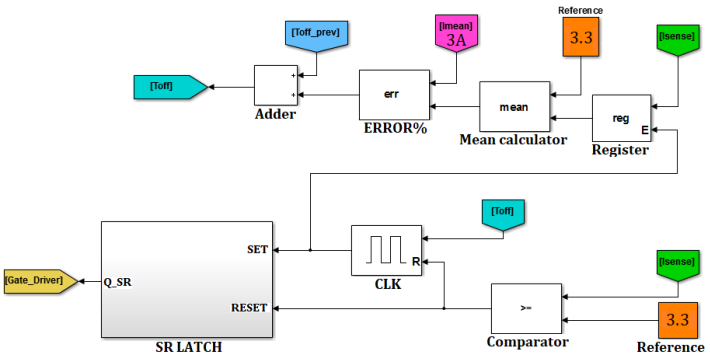


Figure 7: Simulink implementation

The simulations include also worst case scenarios to ensure the correct operation in each condition. To simulate the presence of a load which fixes the 4.5V output voltage with a 3A average current, a 1.5Ω resistor has been used.

Figure 8 (left) shows how the system works with a 7V input voltage and a peak reference reduced from 3.3A to 3.2A, i.e. with a Δ_{I_L} = 0.4A. As expected, the operation is not affected by instability issues which, in a typical PCMC, would surely arise whenever V_{in} < 9V. Later, the input voltage has been abruptly increased to 18V to also prove that a V_{in} change (as could happen during the cranking of the car) has no relevant effect on the operation. The T_{off} remains unchanged (no dependency on the V_{in}) and equal to 2.7μs as expected from (7) (with Δ_{I_L} = 0.4A).

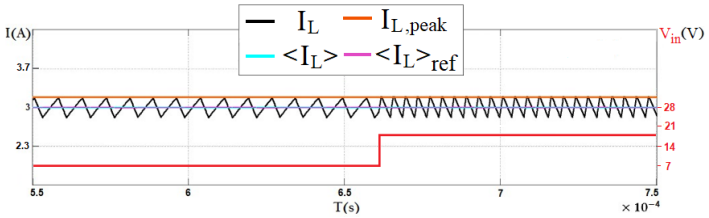


Figure 8: System response changing the input voltage

Figure 9 shows instead the transient simulation plot for V_{in} = 18V with the peak reference at 3.3A as from specification. As expected from (7), the T_{off} = 4μs. To further prove the system robustness, an average reference positive step (from 3A to 3.15A) has been added. In order to achieve the new average current, the loop gradually reduces the T_{off} to make the calculated average equal to the target. Since the peak reference stays equal, according to (8) and (9), the ripple reduces and the switching frequency increases up to 260kHz.

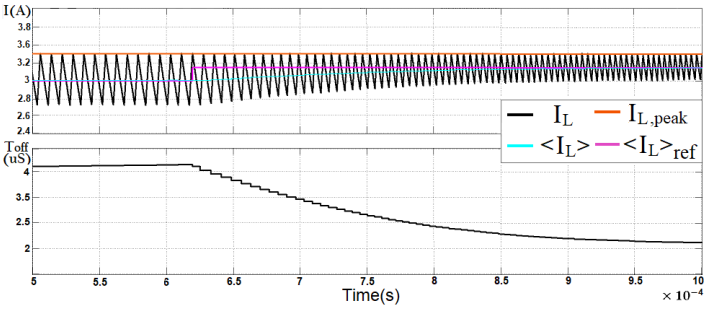


Figure 9: System response changing the average reference

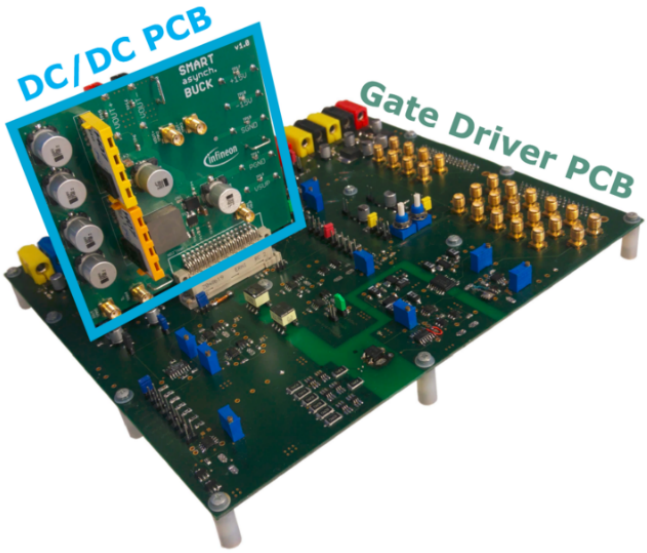


Figure 10: Buck Converter and Gate Driver PCBs

Once proven the concept works as expected, the system can be tested in laboratory by means of a rapid control prototyping called dSpace. Here, the power-DMOS and external components are no longer digital, but located on a DC-DC converter Printed Circuit Board (PCB), which is connected to a Gate driver board (Figure 10), connected in turn to dSpace.

dSpace, through its FPGA, will embody the Simulink bloch approach control logic circuit (uploaded by PC), also providing the current sensing, the internal reference and all the devices to communicate with the hardware (ADCs, DACs). dSpace allows to test whatever control circuit including all the un-idealities of the system, verifying moreover the correct operation of the power-DMOS and that one of a real load. In this way it can be eventully proven that, not only the circuit works as expected, but that is suitable for the application at stake. Through a Graphic User Interface, in Figure 11, the user can monitor the system operation using the PC. Not only to insert the protection to avoid OC and OV, but also to vary the references and other parameters and observe the system response and waveforms in each condition.

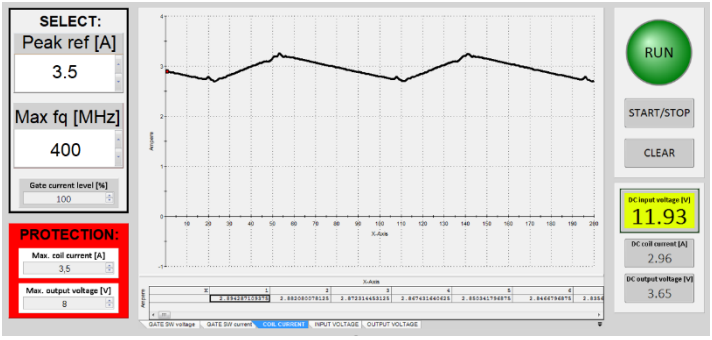


Figure 11: Graphic User Interface

Of course, for having a higher sampling precision, the same waveform can be actually taken from oscilloscope, shown in Figure 12. As expected, the circuit is able to drive the load maintain the desired current features over time and with the same theoretical behaviour of Figure 2.

Figure 12: Output current (violet) and gate driver (green) Waveforms from oscilloscope

6 Conclusions

The presented DC-DC Buck Converter Variable-T_{off} loop allows to precisely and smartly control the current features, returning a reliable car battery sourced driver. Other than the current peak, the

loop is able to control the average current. Moreover, given (2), also the ripple is indirectly controlled. With respect to other solutions, the switching period is variable, intrinsically solving the instability issues. Indeed, the T_{on} changes with the input voltage while T_{off} increases/decreases at each switching cycle accordingly to the comparison between I_{avg} and the target value, making (2) smaller/bigger (i.e. increasing/decreasing I_{avg}).

The system exploits an integrated low-side sensing resistor circuit to minimize the Printed Circuit Board usage space, avoiding extra pads to sense the current. To make the chip lighter and cheaper, the Buck Converter C_{out} has been neglected, as superfluous for the application.

The concept has been simulated with Simulink and then a more rigorous test has been conducted in laboratory with dSpace where worst condition simulations shown an excellent dynamic behaviour. The circuit appears to be robust, proving that sudden perturbations on V_{in}, I_{max} and T_{off} do not upset the operation of the circuit. Compared to [8], the presented solution is lighter and works with an input voltage range of $[4.5, 27]$V, able to drive a 3A LED with 500kHz maximum switching frequency, allowing extremely good average regulation precision and speed. A more detailed comparison of all Buck performances, including efficiency, with the state of the art is provided in [7].

Acknowledgment A part of the work has been performed in the project iDev40. The iDev40 project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 783163. The JU receives support from the European Union's Horizon 2020 research and innovation programme. It is co-funded by the consortium members, grants from Austria, Germany, Belgium, Italy, Spain and Romania.The information and results set out in this publication are those of the authors and do not necessarily reflect the opinion of the ECSEL Joint Undertaking.

References

[1] L. Dixon, "Average Current Mode Control of Switching Power Supplies," 1999.

[2] O. Gasparri, R. Di Lorenzo, P. Del Croce, A. Pidutti, A. Baschirotto, "Variable off-Time Peak Current Mode Control (VoT-PCMC) as method for average current regulation in Buck Converter Drivers," 2019.

[3] I. J. Prasuna, M. Kavya, K. Suryanarayana, B. Shrinivasa Rao, "Digital Peak Current Mode Control of Boost Converter," 2014.

[4] Z. Jiuming, L. Shulin, "Design of slope compensation circuit in peak-current controlled mode converters," 1310–1313, 2011.

[5] J. Ejury, Buck Converter Design, Infineon Technologies North America (IFNA), 2013.

[6] Fang, Chung-Chieh, Redl, "Subharmonic Instability Limits for the Peak-Current-Controlled Buck converter With Closed Voltage Feedback Loop," Power Electronics, 1085–1092, 2015.

[7] R. Di Lorenzo, O. Gasparri, A. Pidutti, P. Del Croce, A. Baschirotto, "On-Chip Power Stage and Gate Driver for Fast Switching Applications," 2019.

[8] V. Anghel, C. Bartholomeusz, A. Vasilica, G. Pristavu, G. Brezeanu, "Variable Off-Time Control Loop for Current-Mode Floating Buck Converters in LED Driving Applications," 2014.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 353-362 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Blueprint Model: An Agile-Oriented Methodology for Tackling Global Software Development Challenges

Andre Figliuolo da Cruz, Cristiano Pereira Godoy, Lanier Menezes dos Santos, Lucas Frota Marinho, Marco Santarelle Jardim, Elisangela Paiva da Silva, Cícero Augusto Pahins, Paulo Fonseca*, Felipe Taliar Giuntini

SOL department, Sidia: R & D Institute, Manaus, 69050-020,Brazil

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 23 October, 2020
Online: 20 November, 2020

Keywords:
Methodology
Agile
Scrum
Kanban
Global Software Development

ABSTRACT

In recent decades, the challenge of managing the development teams spread across different time zones has become increasingly common, raising the importance of the development of Global Software Development (GSD) techniques, in order to tackle its particular problems. This work discuss these issues in the context of Sidia, an R&D institute which implements technological solutions for global companies. The main partner of Sidia is a mobile multinational company located in overseas. The development team must cooperate with the overseas team, even though there are no overlapping working hours between both teams. Besides, the teams have a different set of skills regarding design, quality assurance, and software engineering. In order to address theses problems, we propose the Blueprint model, a Kanban and Scrum based model that supports the development of GSD systems, the allocation of tasks and teams, and the efficient communication. Finally, we discuss the aspects and lessons learned of development of project and deploy of a new model for systems development on a real-word project.

1 Introduction

Many companies adopt software development methodologies with fixed timebox periods for developing incremental deliveries, which is a strategy aligned with agile-based approaches. In these scenarios, Scrum proved to be the *de facto standard* for the most software-oriented industries, showing improved results over *waterfall* methodologies. Despite this, some groups can face difficulties when dealing with the Scrum's ceremonies in the context of Global Software Development (GSD), e.g., estimating effort during the planning sessions or even conducting daily meetings.

In this current work, we extend the definition of the Blueprint model [1], generalizing it to other scenarios and groups by furthering exploring its technical aspects and presenting add-ons to the original proposal. As described in [1] by the authors, Blueprint Model is a lightweight project management that can encourage and facilitate communication between groups in different sites when matched with the suggested group organization. In this particular work, the authors focus on reporting the main challenges when developing a project in collaboration with an external group that does not use agile methodologies, which resulted in a set of barriers to successful integration. To resolve these issues, the Blueprint Model defines its deadlines oriented by the scope of work, which leads to timeboxes of different sizes that consider the effort needed for each activity. Such an approach is different from Scrum, where *sprint* cycles have fixed size throughout the project lifecycle. In this Model, the scope of work, hereafter called *blueprint*, comprises the individual tasks required to reach the target, the necessary information needed to perform these tasks, and what is needed to integrate the work done by every group member. A blueprint defines different group layers, concerning the nature of the work performed (e.g., design group, quality assurance), and assigns the individual tasks to them. Each group layer can be further split into many other groups or even share group members. In this Model, every group must focus on a specific kind of activity since it leads to facilitated project management.

This research was developed on Sidia, an R&D institute located in Brazil, which produces and validates software for an international mobile device manufacturer. The Institute develops innovative software in a kind of areas such as machine learning, software management, and others related to mobile products. Both local and overseas groups deal with the business model integration problems and manage the Sidia's customer's standards for system and management interactions.

In order to meet its customer's demands, Sidia employs Scrum

*Corresponding author: Paulo Fonseca, Sidia: R & D Institute, Amazonas, Email: paulo.fonseca@sidia.com

in the context of a global software environment to deal with several model integration problems and manage Sidia and its customers for management interactions. In this scenario, both groups found it difficult to follow Scrum's activities and bureaucracies. In fact that the standard daily meeting system becomes arduous to follow, because, unlike Scrum recommends, groups are not multi-functional. Each group has its own set of skills, including software engineering, design, and quality assurance. Another prominent issue is to apply Scrum in large scale projects, becoming impractical with excessive management tasks and many activities to manage under a distributed scenario.

The remainder of this paper is structured as follows: Sections 2 and 3 summarizes other similar methodologies, relating them to our proposal. Section 4 shows the Blueprint approach and its implications. Section 5 shows a set of results obtained when employing our proposal on a real-world project. In Section 6 we present the conclusions and avenues for future work.

2 Background

In this section, we discuss the frameworks and methods used as references for the development of the Blueprint model, which is base on agile methodologies concepts and rules like those influenced by the values of the *agile manifest*, namely: (i) individuals and interactions over processes and tools, (ii) working software over comprehensive documentation, (ii) customer collaboration over contract negotiation, and (iv) responding to change over following a plan. As already stated, the Blueprint model shares similarities with both Scrum and Kanban.

Scrum is a structural framework created to manage work on complex products and consists of Scrum groups associated with rules, events, roles, and artifacts. Each component owned with purpose, and integration success depends on managing the relationships and interaction between them. Scrum was the primary reference when designing Blueprint.

A fundamental difference between Scrum and Blueprint is that, while the former has been used to develop software, hardware, networks of interacting function, autonomous vehicles, schools, government, and managing the operation of organizations, the latter is focused on software development. We stretch this difference in Section 4. The essence of Scrum is a small group of people while remaining highly flexible and adaptive. To achieve that, it prescribes four formal events for inspection and adaptation: (i) Sprint Planning, (ii) Daily Scrum, (iii) Sprint Review, and (iv) Sprint Retrospective. Appointed events are used in Scrum to create frequency and to reduce the meetings not defined in the regular workflow. In contrast, Blueprint is a model that reduces or eliminates the regularity of those ceremonies, once it can negatively impact some scenarios of GSD. We explore these scenarios in Section 4.4.

To task management, Blueprint employs Kanban, once it is focused on managing commitment and balancing workflow to achieve greater agility. Kanban is a method for defining, managing, and improving services that deliver knowledge work, such as professional services, creative endeavors, and the design of both physical and software products. While designing Blueprint, we make sure that it shares Kanban's same values, namely: transparency, balance, collaboration, customer focus, flow, leadership, understanding, agreement, and respect.

In summary, Blueprint deal with two typical GSD challenges:

Communication Barriers Both Scrum and Kanban define meeting and ceremonious as part of project communication, but in a GSD context, they can be challenging to coordinate, once group members and stakeholders are in different timezones.

Scope Management Although Scrum is favorable to accepting a change of scope, the typical framework implementation implies that a group is committed to the sprint scope once it begins. This scenario can cause undesirable overhead once it requires both backlog and sprint replanning.

3 Related Work

In the last two decades, agile methods increased their popularity and were established as the main software development methodologies. A large number of efforts [2–5] were focused on studying their application under different circumstances and evaluate their advantages and shortcomings. In [4] the authors conducted a systematic review of works that used Scrum guidelines in a global software development (GSD) context. The authors identified the challenges of using Scrum in such context and which are the current strategies to deal with them. The main difficulties are related to: synchronous communication, collaboration difficulties, communication bandwidth, tool support, group size and office space.

The lack of synchronized hours between the group hampers the ability of holding long meetings, such sprint planning or retrospective sessions, which can last up to four hours or more. Some groups [6, 7] handle these issues using strategies such as defining strict timeboxes for meetings' duration and executing preparation tasks in order to shorten the meeting. This approach may help teams to face GSD challenges while partially following Scrum guidelines, however, this is not possible between locations where the overlapping hours are scarce, for example, Brazil and Japan have a 12-hour offset.

Additionally, in many cases, the quality of communication is also a problem, given that distributed projects are composed by group members with diverse cultural and linguistic background, which may lead to group members not completely voicing their opinions, due to fear of being misinterpreted. A number of works [8–10] tackles these issues by proposing additional on-site meetings in multiple project phases: before starting a project, in order to reach a common understanding about the project; multiple exchange visits of group members, during the project, in order to reduce the cultural distance. These studies show that such practices are valuable to both teams as they shorten the communication gap, an important challenge in GSD. Nevertheless, these approaches have a high financial and time cost. As a way to address these issues, other approaches [11, 12] focus on enhance documentation practices in order to bridge the communication gap.

Besides these socio-cultural differences, other inherent aspects of GSD projects also cause Scrum projects to heavily rely on a diverse set of tools in order to exchange information, share group's

states and enforce Agile guidelines. In current literature, works as [13–15] used a large number of tools to achieve these goals, including enterprise wikis, shared electronic work space, global issue and bug trackers and backlog management tools, these approaches might be used in order to provide information for distributed teams in a centralized manner. However, these tools might add complexity and bureaucracy to the team operation, negatively affecting its performance.

Another common technique used in large distributed groups is to divide the group into into smaller sub-groups [6, 7, 13], according to different strategies, including: different aspects of the development group(*e.g.* design, back-end, front-end); product features and deliverable; independent modules or subsystems of the main project. This division eases group management due to the smaller gap in communication of a smaller and more focused group. Scrum may also be adapted in different ways to support the groups subdivision: a fully integrated Scrum group, where all members must participate in all meetings, with frequent meeting ensure correct and clear communication between distributed sub-groups; or separated Scrum sub-groups based on location running their own iterations while sharing key points between groups. These subdivisions tackle the management and interdependence problems in GSD because usually team members from the same aspect need to exchange information more frequently than team members from different aspects.

The majority of approaches performs only minor changes in the Scrum framework, by specifying constraints or adding features to original guidelines. Other approaches suggest more fundamental changes to Scrum guidelines or even use other Agile methodologies.

In [16], the authors investigated the impact of using a combination of Scrum and Kanban (Scrumban) on global software development and showed that challenges related to communication, cultural and strategic issues are addressed due to characteristics of Scrumban such as iterative and incremental deliveries, regular feedback ceremonies and limiting the amount of work in progress. However, there are still few challenges related to technical and security issues in GSD that are not addressed or alleviated through the use of Scrumban, which require other methodologies and tools to be mitigated.

In [17], the authors evaluated the use of Kanban on the same environment and found that some Kanban elements, such as the Kanban Board, *Inclusion Criteria*, which guarantees that all items in development deliver value to the customer, and *Reverse Items*, which enables development items to return to a previous state, are valuable to face communication challenges in GSD.

This work proposes a model called Blueprint, which is a lightweight project management that was envisioned specifically aiming global software development scenarios. This combines strengths of Scrum and Kanban, as well as the use additional tools in order to facilitate communication between groups in different sites and provide holistic project view that centralizes information for all team members. In this particular work, the authors focus on reporting the main challenges when developing a project in collaboration with an external group that did not use agile methodologies, which resulted in a set of barriers to successful integration.

4 Blueprint

4.1 Goals

The Blueprint model has been designed to deal with geographically distributed groups, handling timezone issues, which may lead to difficulties in setting up meetings, misunderstanding requirements, delay in receiving or reporting issues. Besides the timezone challenge, this model also focuses on increasing time spent on product development, parallel work, and reduce meetings overhead.

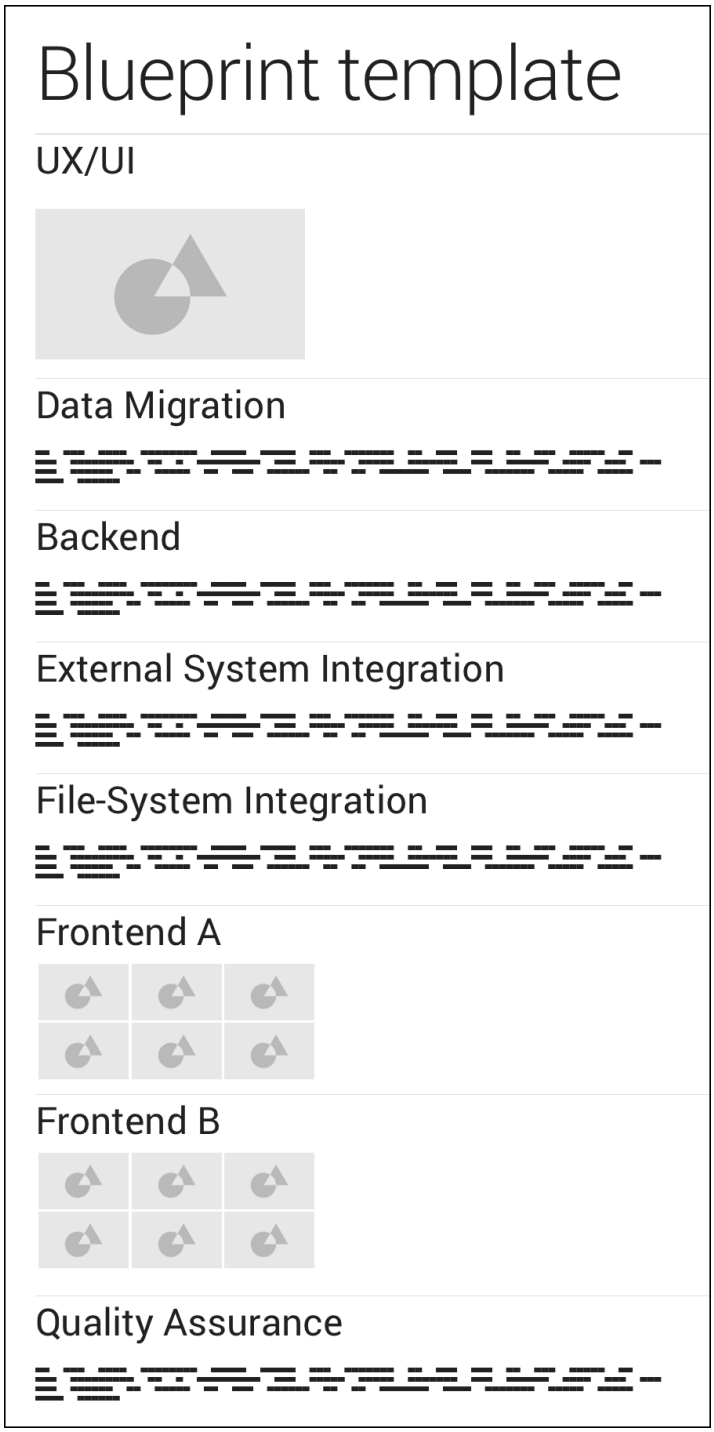


Figure 1: The *Blueprint* Layout, unifying areas *sub-teams* with information on a wiki page.

Therefore, to achieve these objectives, Blueprint seeks to simplify and improve project understanding by using visual-aid tools and helping the group prototype the main project aspects, i.e., UI and technical features. To group communication, these resources are stored in a wiki tool, available to every member so that everyone can be notified of any update.

The Blueprint model employs a hierarchical group organization of sub-groups, once it facilitates the development of unrelated requirements in parallel. Each sub-group is responsible for developing and integrating different aspects of the project into a releasable package. In contrast to other agile methodologies like Scrum, in which groups are considered self-organizing without any prior hierarchical organization [18] other than Testing, Architecture, Operations, and Business Analysis Domains, the Blueprint requires a Leader for each sub-group that is responsible for attending meetings, define requirements, and propose deadlines.

In summary, the primary purpose of Blueprint is to reduce overheads, so the model does not contain predefined which ceremonies must occur neither its regularities, which should be scheduled on demand. Blueprint suggests that each group guides itself by using a set of add-ons, e.g., a wiki tool, improving communication, and information sharing. For the coordinate task, Blueprint suggests Kanban boards, making it easy to share the information across different timezones and provide an accessible project overview.

4.2 Model

In [19], the authors describe an essential aspect of Scrum and Kanban, two popular agile methodologies used in the industry [20], is the constant expansion and improvement of its tools. A Blueprint model is a tool inspired by Scrum and Kanban that manages the project's business entities before it is developed. Blueprint introduces key adaptations to Scrum and Kanban to reduce unwanted bureaucracies and to facilitate global software development. The fundamental concepts of the Blueprint model are:

The Blueprint The concept of a blueprint is very similar to Scrum's story. While both the story and blueprint contain a brief description of the requested features, the latter contains every information needed to design, execute, evaluate, and deliver a set of features, as shown in Figure 3 1. A blueprint should be divided into meaningful sections and contain all related files and discussions to an activity, e.g., the development progress, facilitate the communication between group members and stakeholders, and lower the complexity of both maintenance and monitoring. For efficient scope management, a blueprint can be revised, merged, or divided to reflect the latest requested features and evaluate agreed scope, which should occur at regular intervals. In this document, you can find two very similar terms, the Blueprint Model and the blueprint. The first one, Blueprint Model, is directly related to the model itself. Every time the model is being addressed, it will be called as Blueprint Model, using the first letters for both words in uppercase. The other one is regarding the unity of the work of each blueprint. In this case, the first letter in lowercase refers to every blueprint created for each project, having its characteristics being explained over this document. To implement the blueprint concepts, one could use wiki tools to (i) allow discussions as recorded chats and (ii) keep a history of decisions during development. We further explore the use of wiki tools in Section 4.2.1.

Timebox and Sense of Urgency Timeboxing is a well-known time management technique employed for both project and personal management approaches, once it keeps a constant *workflow* with deliberated deadlines. The use of fixed timeboxes can be challenging, notably in projects with frequent scope changes, since it must be adopted by every group member and stakeholder, imposing difficulties in managing unexpected features requests or modifications. The Blueprint model was designed as a solution for companies and groups struggling to keep with fixed timeboxes, as suggested by agile methodologies like Scrum. In the Blueprint model, every cycle can have an independent deadline determined by the amount of work estimated by the members involved in the development, ranging from days up to weeks. Note that the total time to deliver a set of features is the sum of all activities related to a specific blueprint, similar to a Gantt chart.

Group Organization Contrary to Scrum, in which the group typically uses a flat organization, with members associated with specific roles, the Blueprint model defines a position entitle *group leader*, which is responsible for understanding the solution, define the project schedule, and evaluate deliveries. In the context of a *research and development* project, the Group Leader is also responsible for granting the paper publication. The Blueprint model organizes the group into focused sub-groups, which can be formed by specialists or multidisciplinary members. Sub-groups are advised to define a Sub-Leader to help the group leader in the project management activities, once it eases communication and offers an opportunity for every group member to express its ideas during the discussion of solutions and deadlines. We further explore the Blueprint's group organization in the following sections. Similar to other agile methodologies, the Blueprint model can be adopted by *non-coders*, e.g., testers and designers.

Contrary to the Scrum group division, Blueprint proposes a group hierarchy that can better adapt to software requirements changes during project execution. Blueprint defines its group hierarchy as *sub-groups*. The number of members per *sub-group* is not defined by Blueprint, but as an analogy to Scrum, and as discussed in [19], *sub-groups* should focus on small groups. Also, the number of *sub-groups* may change during the execution of a project.

4.2.1 Add-ons

In the following paragraphs, we present add-ons that can improve Blueprint adoption in different scenarios and groups.

Blueprint Table As explained in Section 4.2, the *blueprint table* is a management tool that should be regularly updated by the group leader to help other members realize the overall status of the project. The *tables* of Blueprint are equivalent to Kanban's boards, but instead of using *status flags* as columns, it uses the concept of *sub-group* ownership. The order of the sub-groups in the table should be defined to reflect the workflow being developed by the group, e.g., design should come first in several common situations, and quality assurance should come last in most of the scenarios.

The list of the condition is used to manage the operation of each stage described in the *blueprint table*, as shown in Table 2.

Contrary to Kanban, in which group members move tasks over the board to indicate the current status of a task, in the blueprint table, the Group Leader is responsible for moving the sub-groups status to consider the *blueprints* in the preferred order of development. The order of development must be arranged between the group, stakeholders, and Group Leader, facilitating project management. The Blueprint model defines the usage of status flags as:

- Every sub-group status should start as PENDING.
- Sub-groups without actions associated with a blueprint, assume the N/A status.
- At the point when the sub-group begins to work in a particular outline, its status should be changed to IN PROGRESS to show that not all conditions and errands are done.
- After a sub-group finishes up all tasks and conditions, the status of the sub-group ought to be set apart as DONE.
- When a sub-group concludes a blueprint, if possible, the sub-group must start working on another blueprint. If not possible, the sub-group must contact the group leader as soon as possible for resource reallocation.
- Status like *SIMILAR* and *BUG FIX* were intended to help general overview.
- SIMILAR ought to be utilized when no effort should be spent by a sub-group due to another outline. At the point when applied accurately, it can diminish modify.

Table 1: *Blueprint* suggestion of content by *subgroups*

Sub-group	Blueprint content
UX/UI Sub-Groups	▪ UX workflow annotation ▪ UI visualizations based on UX definition ▪ Business rules definition ▪ List of assets
Data Migration Sub-group	▪ Database infrastructure servers documentation ▪ Database model related to the Blueprint ▪ Migration architecture documentation ▪ Batch architecture documentation ▪ List of Scripts
Back-end Sub-group	▪ Back-end infrastructure servers documentation ▪ Back-end architecture documentation ▪ Security documentation ▪ Back-end endpoints list
External System Integration Sub-group	▪ External System infrastructure servers documentation ▪ External Integration architecture documentation ▪ External System endpoints list
File-System Integration Sub-group	▪ File-System infrastructure servers documentation ▪ File-System Integration architecture documentation ▪ File-System endpoints list
Front-end A Sub-group	▪ Front-end A infrastructure servers documentation ▪ Front-end A architecture documentation ▪ Security documentation
Front-end B Sub-group	▪ Front-end B infrastructure servers documentation ▪ Front-end B architecture documentation ▪ Security documentation
Quality Assurance Sub-group	▪ Test Cases documentation ▪ Tests Execution report ▪ Know Issue list

- BUG FIX ought to be utilized to show that a blueprint arrived at its last stage and bugs are being fixed.

Every time a new blueprint is created, abandoned or regrouped, it is advised to replan the blueprint table. The group Leader and the group must be aware of the impacts and dependencies of documentation modification during the entire project lifecycle. The same happens when remodeling sub-groups. If any sub-group is created, ceased, or regrouped, the blueprint table should be planned. In summary, a blueprint table is a tool for sub-group management that can be associated with a collection of Kanban boards for handling tasks administration.

Wiki Tools On the Blueprint model, the usage of Wiki tools is essential to organize and aggregate information on different aspects of the project, facilitating the communication between the group and stakeholders. The main points of integration are:

- The web page format makes it easy to maintain and update.
- Every blueprint should be stored as a unique wiki web page.
- The group Leader should organize all pages and ask for the group to keep every part updated.
- Every sub-group must have its own space following the same order on every blueprint page.
- The storage of technical conversations and discussions in a log format contributes to the project's maintenance.
- All related usage of blueprints in the Wiki tools used the company's infrastructure company's infrastructure and its internal Information Security policy to manage it. This means every group member was under the company's management overview and user access control rules.

Table 2: Status to control activities in the *blueprint table*.

Status	Definition
PENDING	Initial status or *sub-group* activities are on hold. It may have several causes. Descriptions should be at the Blueprint related page.
IN PROGRESS	The activities of this *Blueprint* are on development. It may contain dependencies or impediments but still in progress.
DONE	All activities are finished. Have no dependencies.
SIMILAR	When the activities of this *sub-group* are equals to the activities of the *sub-group* above. This eliminates the effort of the related activities.
BUG FIX	Fixing bugs of the *Blueprint*. The *Blueprint* is in its final status.
N/A	Not applied. It occurs when the *sub-group* have no activities for that Blueprint.

- Replanning should be easy to maintain by employing the wiki's versioning system.
- By leveraging wiki tools, groups that require necessary social distancing, home office, or are facing other adverse situations should communicate effectively by (i) integrating a Kanban board with each blueprint and (ii) using the *export* feature to share relevant information with stakeholders.

4.3 Implementing Blueprint on a Real-World Project

To show Blueprint in detail, we will use a real-world project called *The Project*, which characterized a list of *sub-groups* that can be equivalent to many webs, mobile, or embedded software projects. We first planned *The Project* with a number of *sub-groups*, which was thought to be sufficient, but at the end, the list changed to optimize group allocation, as shown in Figure 2. As showed on the Blueprint model, sub-groups can can aggregate multidisciplinary individuals or authorities with a similar profile, for example analyzers situated in Quality Assurance Sub-gathering, architects in Design Sub-gathering, engineers taking a shot at both or solely on Front end Sub-gathering and additionally Back end Sub-gathering. Note that the sub-bunch rundown can be changed by the necessities of the task to streamline bunch portion or build up another prerequisite. In *The Project*, the accompanying circumstances lead to sub-bunch rearrangement:

- To deliver the back-end features on the schedule, the sub-group was divided to help some members to work on two sets of requests.
- The client requested an alternative and entirely different UI solution from Design Sub-group that was approved to be developed, so it was decided to split the Front End Sub-group into groups *A* and *B* to allow each work on a different release.
- A new mobile solution was requested, but there is no developer with such skills in the group. Therefore, it was decided to hire new members and they will work in the entirely new Mobile Sub-group.
- After a schedule modification, the client suggested for developers to work in a new scope of migration data, and due to this, a new Migration Sub-group is created. The group Leader decided that some back-end developers and testers will work on this group part-time.

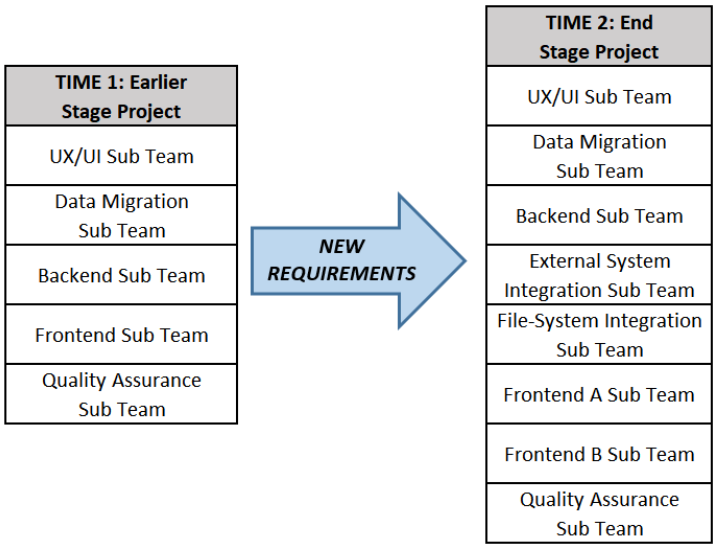


Figure 2: *Sub-groups* preparations during beginning (*time 1*) and ending (*time 2*) of *The Project* execution.

Blueprint consents to the thought that the quantity of individuals doesn't speak to quality or profitability. In the Blueprint model, each *sub-group* has a pioneer characterized as a part of mindfully, and not really work execution. Each *sub-bunch leader* is liable for the exercises and consequences of its gathering. He is additionally answerable for incorporating the errands done by the gathering into the undertaking, planning this cycle with different pioneers. All *sub-bunch leaders* react to the *group leader*, as delineated in Figure 3. This association follows the meaning of how improvement ought to be executed by the Blueprint model to permit and encourage *sub-group* communication.

After the meaning of *sub-bunch leaders*, the following stage is to execute Blueprint select and the board. This stage is a variation and conglomeration of two Scrum the board undertakings and ceremonies, in particular, *Product Backlog* and *User Stories*. All individuals are locked in to effectively take an interest in this stage during venture improvement, as it is the center of the model. As depicted in [12], the duty to characterize and keep up the Scrum's *Sprint Backlog* is in the possession of the job *Product Owner*. It is normal for the *Product Backlog* to require consistent correction and reworking. These assignments may contain business rules in the *User Stories'* depiction. A few organizations have utilized extra documentation to keep engineering definition, model plan, and framework examination refreshed by the *Product Owner*. Uniquely in contrast to Scrum, *blueprints* are business substances that are relied upon to be persistently changed, kept up, and assessed by all individuals from the task. Pioneers must have the duty to change the *blueprints*, however this assignment isn't limited to them.

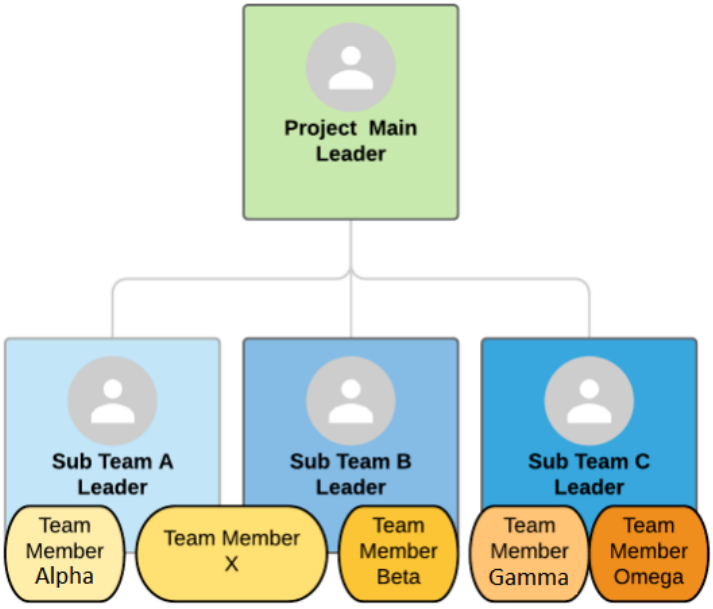


Figure 3: *Sub-group Leaders* and their relationship with the *main group Leader*.

As presented in Section 4.1, the Blueprint model was intended to help bunches in various destinations and to function admirably with distant *sub-groups*. It energizes correspondence among designers and customers by incorporating the conversation in a wiki-based website page. Each *blueprint* ought to be enlisted utilizing a wiki instrument, and its substance development relies upon each *sub-group's* work. At *The Project*, *blueprints* are wiki pages containing

all substance vital for the advancement of a module or an aspect of the undertaking. Table 2 sums up the substance of each *sub-group* obligations that was characterized in *The Project*.

Blueprints must contain all *sub-groups* related materials, as tables, diagrams or file that aggregates information to development. *Blueprints* centralizes information and communication between *sub-groups* as well as with the *group leader* and the client. Wiki tools typically allow for managing discussions between users as a recorded chat. Blueprint takes advantage of this feature to keep the history of decisions made by the group during project execution. Figure 1 illustrates a *blueprint* template that contains fields related to each *sub-group* define in *The Project*.

Blueprint maintenance is done by keeping each blueprint activity from all sub-groups up to date, which helps reflect on the project status. The usage of blueprints encourages deadlines with a focus on the agreement between sub-groups, group leaders, and stakeholders. Each sub-group deadline is a composition of all tasks planned at the blueprint, which can be replanned without impacting other blueprints. Therefore, the Blueprint model enables groups and stakeholders to control and refactor activities by proposing specific deadlines for each sub-group without re-planning activities defined.

The understanding about the hour of every movement is legitimately identified with the development of the gathering and its self-administration. After the meaning of each *blueprint* content and the principle cutoff time of *The Project*, some gathering individuals recognized the requirement for micromanagement exercises. To oblige that request, the underlying adaptation of the *blueprint table* was moved up to characterize reasonable cutoff times for all *sub-groups*. Figure 5 embodies the second form of the *The Project*'s *blueprint table*.

Blueprint encourages deadlines with a focus on *sub-group* assignments that are characterized in in agreement with the *group leader* and the client. Each *sub-group* deadline is a piece of all assignments arranged at the *blueprint*. During the development of *The Project*, bunch individuals started to demand rethinking of certain activities, and it started to impact on previously defined *blueprints*. The Blueprint empowers to gatherings and customers to control and refactor exercises by proposing specific deadlines for each *sub-group* without previously characterized.

The Blueprint model characterizes a*blueprint table* as a administration tool hat ought to be utilized to continually show the general status of an undertaking to all gathering individuals and customers. Figures 4, 5 and 6 show the *blueprint tables* during execution of *The Project*. Toward the start of *The Project*, an underlying form of the *blueprint table* was characterized by the gathering with the assistance of the customer, as observed in Figure 4. This underlying arrangement permitted all partners to concur with the interests of the task and with the substance of each *blueprint*. During the development of *The Project*, members of the *sub-groups* identified that the *SIMILAR* status helped to reduce software rework.

After different steady modifications, *The Project* characterized its last *blueprint table* as appeared in Figure 6. Note that, in examination with Figure 5, new *sub-groups* were made to oblige already impromptu prerequisites of the customer. The creation or evacuation of *sub-groups* can influence different exercises and should be all around oversaw and talked about between *sub-bunch leaders*, however never debilitated. For example, at the underlying rendition of the *blueprint table*, just one front-end gathering, called *Front-end Sub-group* was liable for all front-end exercises. At that point, when an additional necessity rose up out of the customer, another front-end *sub-group* called *Front-end B Sub-group* should have been made to build up an alternate arrangement of arrangements. The underlying front-end *sub-group* was needed to survey its exercises to help the redistribution of errands of the reestablished *blueprint table*. A portion of the gathering individuals need to chip away at bunches *A* and *B* to help extra exercises. *Front-end bunch B Sub-group* was situated at the abroad site and worked distantly with the gathering situated at Sidia. The last rundown of *blueprints* are represented in Table 3.

Table 3: List of Blueprints used in *The Project*

Blueprint Name	Example of work by blueprint
Authentication	Login page. Back-end authentication service. Front-end authentication error page.
CRUD A, B, C and D	Create, Read, Update and Delete business entity A, B, C and D. Back-end service for each. Front-end pages for each.
Extended Specification A and B	New requirement with several business workflow. Requested during the execution of the project. Replan needed.
Report A, B and C	Report generator page. Back-end service for report business model.
External Requests A and B	Unintended external integration. Requested during the execution of the project. Replan needed.
Mobile Integration and Mobile Interface	Unintended mobile version. Requested during the execution of the project. Replan needed.
Customer Service, User Overview and About	Low priority administrator pages. Gained priority due to requested changes.

The Blueprint Model as some other philosophy has its reception challenges, for instance:

- Communication hazards The Blueprint model has a great dependency on documentation process. As proposed, it is very dependent of Wiki documents in order to keep good communication between sites. This dependency combined with possible many changes requests, generates a constant necessity of Wiki documents updates. It can turn out very difficult for the group to keep the good communication under many changes requests.

- Predictability hazards If not applied as its standards by the entire group, the Blueprint model can show up as a harmful tool when it comes to deliveries and projects goals. The presented model requires management skills from each area of the group. Once the required skills are not present in the group members, the entire process can be compromised, which would impact on the completion and good quality of deliveries.

- Performance hazards Concerning software development, all sub-groups that work together have dependency of information from one another. On Blueprint model, as each sub-group is required to maintain its own Wiki page updated, the lack of update of any sub-group can generate delays, compromise

group work and even entirely block group members of working.

4.4 Use Cases

In the following paragraphs, we present some use cases in that the Blueprint model proves successful.

Global Software Development The Blueprint model was designed to support remote groups and sub-groups in different locations, since its hierarchical group organization of a Project Leader and its

Blueprint	Deadline	UI/UX	Backend	Frontend	QA
Autentication	20XX/01/10	DONE	DONE	DONE	IN PROGRESS
CRUD A	20XX/01/10	DONE	DONE	IN PROGRESS	PENDING
Report A	20XX/02/23	DONE	IN PROGRESS	PENDING	PENDING
Report B	20XX/01/16	IN PROGRESS	PENDING	PENDING	PENDING
(others)	20XX/03/15	PENDING	PENDING	PENDING	PENDING

Figure 4: Starting version of the *blueprints table*. Deadlines are characterized for *blueprints* in *The Project*.

Blueprint	UI/UX	Backend	Frontend	QA	Global Status
Autentication	20XX/01/10	20XX/01/18	20XX/01/22	20XX/01/24	20XX/01/26
	DONE	DONE	DONE	IN PROGRESS	IN PROGRESS
CRUD A	20XX/01/12	20XX/01/20	20XX/01/24		
	DONE	DONE	IN PROGRESS	PENDING	IN PROGRESS
Report A	20XX/01/14	20XX/01/22			
	DONE	IN PROGRESS	PENDING	PENDING	IN PROGRESS
Report B	20XX/01/16				
	IN PROGRESS	PENDING	PENDING	PENDING	IN PROGRESS
(others)					
	PENDING	PENDING	PENDING	PENDING	PENDING

Figure 5: Second version of the *blueprint table*. The deadlines were characterized by *sub-groups*.

ID	Blueprint	UI/UX	Backend				Frontend		QA	Global Status
			Data Migration	Backend	External migration	File integration	Frontend A	Frontend B		
1	Autentication	20XX/01/10	20XX/01/12	20XX/01/14	20XX/01/16	20XX/01/18	20XX/01/20	20XX/01/22	20XX/01/24	20XX/01/26
		DONE	DONE	DONE	DONE	DONE	DONE	DONE	DONE	DONE
2	CRUD A	20XX/01/12	20XX/01/14	20XX/01/16	20XX/01/18	20XX/01/20	20XX/01/22	20XX/01/24	20XX/01/25	20XX/01/26
		DONE	DONE	DONE	DONE	DONE	DONE	DONE	DONE	BUG FIX
3	CRUD B	20XX/01/14	20XX/01/16	20XX/01/18	20XX/01/20	20XX/01/21	20XX/01/22	20XX/01/23	20XX/01/24	20XX/01/26
		DONE	DONE	DONE	DONE	DONE	DONE	DONE	DONE	BUG FIX
4	CRUD C	20XX/01/16	20XX/01/17	20XX/01/18	20XX/01/19	20XX/01/20	20XX/01/21	20XX/01/22	20XX/01/23	20XX/01/26
		DONE	DONE	DONE	DONE	DONE	DONE	DONE	DONE	BUG FIX
5	CRUD D	20XX/01/16	20XX/01/17	20XX/01/18	20XX/01/19	20XX/01/20	20XX/01/21	20XX/01/22	20XX/01/23	20XX/01/26
		DONE	DONE	DONE	DONE	DONE	DONE	DONE	DONE	BUG FIX
6	Extended Specification A	20XX/01/16	20XX/01/17	20XX/01/18	20XX/01/19	20XX/01/20	20XX/01/21	20XX/01/22	20XX/01/23	20XX/01/26
		DONE	DONE	DONE	DONE	DONE	DONE	DONE	DONE	BUG FIX
7	Extended Specification B	20XX/01/16	20XX/01/17	20XX/01/18	20XX/01/19	20XX/01/20	20XX/01/21	20XX/01/22	20XX/01/23	20XX/01/26
		DONE	DONE	DONE	DONE	DONE	DONE	IN PROGRESS	IN PROGRESS	IN PROGRESS
8	Report A	20XX/01/16	20XX/01/17	20XX/01/18	20XX/01/19	20XX/01/20	20XX/01/21	20XX/01/22	20XX/01/23	20XX/01/26
		DONE	DONE	DONE	DONE	DONE	DONE	IN PROGRESS	IN PROGRESS	IN PROGRESS
9	Report B	20XX/01/16	20XX/01/17	20XX/01/18	20XX/01/19	20XX/01/20	20XX/01/21			20XX/01/26
		DONE	DONE	DONE	DONE	DONE	IN PROGRESS	PENDING	PENDING	IN PROGRESS
10	Report C	20XX/01/16	20XX/01/17	20XX/01/18	20XX/01/19	20XX/01/20				20XX/01/26
		DONE	DONE	DONE	DONE	DONE	PENDING	PENDING	PENDING	IN PROGRESS
11	External Requests A	20XX/01/16	20XX/01/17	20XX/01/18	20XX/01/19					20XX/01/26
		DONE	DONE	DONE	IN PROGRESS	PENDING	PENDING	PENDING	PENDING	IN PROGRESS
12	External Requests B	20XX/01/16	20XX/01/17							20XX/01/26
		DONE	IN PROGRESS	PENDING	PENDING	PENDING	PENDING	PENDING	PENDING	IN PROGRESS
13	Mobile Integration	20XX/01/16								20XX/01/26
		DONE	PENDING	PENDING	PENDING	PENDING	PENDING	PENDING	PENDING	IN PROGRESS
14	Mobile Interface	20XX/01/16								20XX/01/26
		DONE	PENDING	PENDING	PENDING	PENDING	PENDING	PENDING	PENDING	IN PROGRESS
15	Customer Service	20XX/01/16								20XX/01/26
		DONE	PENDING	PENDING	PENDING	PENDING	PENDING	PENDING	PENDING	IN PROGRESS
16	User Overview	20XX/01/16								20XX/01/26
		IN PROGRESS	PENDING	PENDING	PENDING	PENDING	PENDING	PENDING	PENDING	IN PROGRESS
16	About									
		PENDING	PENDING	PENDING	PENDING	PENDING	PENDING	PENDING	PENDING	PENDING

Figure 6: Last version of the *blueprint table*. New *sub-groups* and *blueprints* are remembered for correlation with past tables.

respective Sub-groups facilitates daily communication, in such a way that this interaction becomes more fluid and objective for adopting a particular scope for each sub-group. In our case, we deal with a customer in a country in which the timezone is at least 13 hours ahead, making real-time communication via call or chat unfeasible. In this scenario, we use the 'wiki' to keep track of what each member of the group is planning for the next few days, what they are doing, and what was done in such a way that everyone has an overview of the project's progress, being able to follow all deadlines and features that have already been developed and those to come.

Remote Work In the face of COVID-19, several development groups and companies in different industries were forced to adopt remote work to continue their work. In this scenario, Blueprint proved to be an adequate strategy to accommodate the remote work regime, regarding the difficulty in communication between groups because from the moment employees are working from home, the practicality and speed of transmitting information are both affected. Consequently, adopting a standard of documenting the information in a centralized form and bringing the division of sub-groups, makes the dissemination of internal information through the wiki page more effective by leveraging a reliable consumption source that is validated by the client.

Scope Change Control The Blueprint also helps to manage frequently *scope changes* since unlike Scrum which suggests another iteration of a set of ceremonies when facing modified requests, such as *Project Vision*, *User Stories*, *Planning*, and others, the Blueprint offers lightweight project management that avoids these ceremonies that can cause rework and further prolong delivery forecasts. To improve on that, Blueprint proposes the idea that both the developer and client will leverage the 'wiki' as a guide for (i) decision making, (ii) discussions about new releases, and (iii) requirements alignment. By documenting information in a centralized repository, the stakeholders are able to easily monitor the requested scope changes, helping to avoid noise or loss of information from typical verbal communication.

5 Implementation Findings

The design of Blueprint was mainly oriented by the GSD challenges faced during the collaboration of remote teams on developing *The Project*, which evolved from 10 members using Scrum to a successful Blueprint case more than 20 members. The increased number of members resulted from the stakeholders' higher engagement and contribution to the process and project. Once the process started running, it allowed the team to accept more work scope, which might be understood as an increase in productivity and led to a natural team increase. Through the advancement of *The Project*, different gradual modifications were made to the Blueprint Model to energize its appropriation. This cycle included both neighborhood and far off groups to facilitate the work process, encourage worldwide programming designing, and improve appropriated correspondence. Following the summarizing of *The Project*, we notice various situations to improve the Blueprint model execution to ensuing tasks.

We found that an essential aspect of a successful Blueprint implementation is to focus on the member's sense of integration and empowerment. Regarding *The Project*, most of the individuals concurred they turned out to be more critical and made more proposals to the undertaking the executives with the appropriation of the Blueprint Model. Every part had the option to share its answers and thoughts to other people, and due to Blueprint central focuses, significant correspondence issues were survived. The primary advantage

of zeroing in on the part's feeling of reconciliation and strengthening is that the group felt inspired for monitoring Blueprint's curios, prompting improved undertaking the executives.

Another significant part of a successful usage was the group's adherence to Blueprint's rehearses. Regularly, less connected with *sub-group leaders* portrayed exercises and progress ineffectively, which can be credited to the difficulties while embracing another administration model. By the by, the individuals saw a connection between (I) the time frame pages changes, (ii) nature of these pages and (iii) commitment of the sub-pioneers. We notice a connection between diminishing the nature of Blueprint pages to diminishing code quality and last on bugs appearing.

The improvement of the communication between team and stakeholders is a crucial objective of the Model, so it is crucial to make more accessible the verification and reporting of information through a *wiki tool*. The successful implementation of the Blueprint Model enables identifying disagreements and conflicts in the Wiki and not necessarily during meetings, which enabled the resolution of these problems in private. Meeting events are not often and gathers a few members, which prevents long discussions of non-related subjects from reducing the execution time and reaching its goal as quickly as possible while the rest of the team keeps working on project tasks. After adopting the BLuePrint Model, the team had performance gains related to the planned x performed activities, related to clear and accurate communication, thus ensuring a clear understanding of the changes. The fact that did not happen using the Scrum model, due to the difficulty of managing constant changes within the sprint and the difficulty of formal meetings due to the time zone. Another improvement that is possible to observe is the quality of deliveries. In the BluePrint model, team members can maintain their specialties without the need to be multidisciplinary, and brings a significant gain in side effects, especially when it came to scope changes. Still linked to the specialty of the team members, it is also possible to see the result in the motivation of the team; it means, to work with the specialty of each one, this brings motivation and engagement in the day-to-day work, and consequently, this is reflected in product quality delivered.

Differently from other agile methodologies like Scrum, the Blueprint Model exempts ceremonies and daily meetings that need the entire team to happen, making the Model more flexible. This characteristic praises the Model because it fits well with the current pandemic scenario where the new common sense of IT company members working remotely and need to keep the flow of work to meet delivery deadlines. Some Blueprint Model positive points:

- No team ceremony: The team adopted the Blueprint Model since the beginning of our development, and we kept it when everyone on the team was doing home office. We managed to keep all deadlines using only the Wiki as a reference for the division of activities.
- Sub team division: This sub-team division was of great value for the atypical home office scenario. The team needed to maintain a very similar communication level as a live-work environment, which made the team feel more comfortable maintaining contact with companions.
- Model suited to a project scope that frequently changes: Scope change is passive to happen very commonly in any software development project. The team validated the model adherence to the common practice since every project was constantly changed. Due to the very centralized information on the Wiki, there were small relevant impacts, all highly negotiable.

6 Conclusion

In this work, we presented the extended definition of the Blueprint model, a management methodology designed to accommodate the best techniques and habits of both Scrum and Kanban to facilitate (i) global software development, (ii) groups and task allocations, and (iii) effective members communication. The straight communication between groups and the high visibility of the overall status of the projects are the main highlights of the methodology, which lead to a better scope change management with low bureaucracy and an accessible understanding of the entire development. To validate the model, we discussed its adoption on a real-world project of a large mobile-related software development company in the context of GSD. The strategic use of Blueprint's add-ons enables us to coordinate, design, develop, and test a complex set of features on a highly unstable scope scenario. Blueprint usage of UI and UX for the documentation of analysis, design, and development phases in conjunction with the lightweight cycle management tool helped to facilitate the project management, once it allows sharing responsibility between the *group Leader*, *Sub-group Leaders*, and other members in documenting and reviewing the progress of the project. More than that, the Blueprint model was improved during the execution of the project through a collaborative effort of groups located in different countries in order to improve its main benefits and to lower any adoption barriers. As future work, we plan to extend the Blueprint's add-ons to integrate other easy-to-use and well-known tools of project and task management, as well as to improve the model's adherence to other types of non-code projects.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment We thank MSc. Ilsun Hong, Taehee Kim, Rafael S. Zerbini, Marcos R. G. B. Junior and Alison L. Garantizado for their valuable help during the development of this framework. This work was partially supported by Samsung Eletrônica da Amazônia Ltda., under the auspice of the informatics law no 8.387/91.

References

[1] C. P. Godoy, L. M. Santos, A. F. Cruz, R. S. Zerbini, E. P. Silva, C. A. L. Pahins, "Blueprint Model: A New Approach to Scrum Agile Methodology," in Proceedings of the 14th International Conference on Global Software Engineering, ICGSE '19, 85–89, IEEE Press, 2019, doi:10.1109/ICGSE.2019.00014.

[2] I. Zada, S. Shahzad, "Issues and implications of scrum on global software development," Bahria University Journal of Information & Communication Technologies (BUJICT), **8**(1), 2015.

[3] L. Silva, C. Santana, F. Rocha, M. Paschoalino, G. Falconieri, L. Ribeiro, R. Medeiros, S. Soares, C. Gusmão, "Applying XP to an Agile–Inexperienced Software Development Team," in International Conference on Agile Processes and Extreme Programming in Software Engineering, 114–126, Springer, 2008, doi:10.1007/978-3-540-68255-4_12.

[4] E. Hossain, M. A. Babar, H.-y. Paik, "Using scrum in global software development: a systematic literature review," in 2009 Fourth IEEE International Conference on Global Software Engineering, 175–184, IEEE, 2009, doi:10.1109/ICGSE.2009.25.

[5] J. López-Martínez, R. Juárez-Ramírez, C. Huertas, S. Jiménez, C. Guerra-García, "Problems in the adoption of agile-scrum methodologies: A systematic literature review," in 2016 4th international conference in software engineering research and innovation (conisoft), 141–148, IEEE, 2016, doi: 10.1109/CONISOFT.2016.30.

[6] M. Paasivaara, S. Durasiewicz, C. Lassenius, "Distributed agile development: Using scrum in a large project," in 2008 IEEE International Conference on Global Software Engineering, 87–95, IEEE, 2008, doi:10.1109/ICGSE.2008.38.

[7] J. Sutherland, A. Viktorov, J. Blount, N. Puntikov, "Distributed scrum: Agile project management with outsourced development teams," in 2007 40th annual Hawaii international conference on system sciences (HICSS'07), 274a–274a, IEEE, 2007, doi:10.1109/HICSS.2007.180.

[8] M. Summers, "Insights into an Agile adventure with offshore partners," in Agile 2008 Conference, 333–338, IEEE, 2008, doi:10.1109/Agile.2008.37.

[9] E. Therrien, "Overcoming the challenges of building a distributed agile organization," in Agile 2008 Conference, 368–372, IEEE, 2008, doi:10.1109/Agile.2008.9.

[10] M. Cottmeyer, "The good and bad of Agile offshore development," in Agile 2008 Conference, 362–367, IEEE, 2008, doi:10.1109/Agile.2008.18.

[11] M. Vax, S. Michaud, "Distributed Agile: Growing a practice together," in Agile 2008 Conference, 310–314, IEEE, 2008, doi:10.1109/Agile.2008.44.

[12] M. Hron, N. Obwegeser, "Scrum in practice: an overview of Scrum adaptations," in Proceedings of the 51st Hawaii International Conference on System Sciences, 2018, doi:10.24251/HICSS.2018.679.

[13] W. Williams, M. Stout, "Colossal, scattered, and chaotic (planning with a large distributed team)," in Agile 2008 Conference, 356–361, IEEE, 2008, doi:10.1109/Agile.2008.25.

[14] J. Cho, "Distributed Scrum for large-scale and mission-critical projects," AMCIS 2007 Proceedings, 235, 2007.

[15] J. Sutherland, G. Schoonheim, M. Rijk, "Fully distributed scrum: Replicating local productivity and quality with offshore teams," in 2009 42nd Hawaii International Conference on System Sciences, 1–8, IEEE, 2009, doi: 10.1109/HICSS.2009.225.

[16] A. Banijamali, M. O. Ahmad, J. Similä, M. Oivo, K. Liukkunen, et al., "Empirical investigation of scrumban in global software development," in International Conference on Model-Driven Engineering and Software Development, 229–248, Springer, 2016, doi:10.1007/978-3-319-66302-9_12.

[17] M. Tanner, M. Dauane, "The Use Of Kanban To Alleviate Collaboration And Communication Challenges Of Global Software Development," Issues in Informing Science & Information Technology, **14**, 2017, doi:10.28945/3716.

[18] S. V. Spiegler, C. Heinecke, S. Wagner, "Leadership gap in agile teams: how teams and scrum masters mature," in International Conference on Agile Software Development, 37–52, Springer, Cham, 2019, doi:10.1007/978-3-030-19034-7.

[19] H. Lei, F. Ganjeizadeh, P. K. Jayachandran, P. Ozcan, "A statistical analysis of the effects of Scrum and Kanban on software development projects," Robotics and Computer-Integrated Manufacturing, **43**, 59–67, 2017, doi: 10.1016/j.rcim.2015.12.001.

[20] M. Düchting, D. Zimmermann, K. Nebe, "Incorporating user centered requirement engineering into agile software development," in International Conference on Human-Computer Interaction, 58–67, Springer, 2007, doi: 10.1007/978-3-540-73105-4_7.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 363-369 (2020)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Vietnamese Text Classification with TextRank and Jaccard Similarity Coefficient

Hao Tuan Huynh[1], Nghia Duong-Trung[2], Dinh Quoc Truong[1], Hiep Xuan Huynh[*1]

[1]*College of Information & Communication Technology, Can Tho University, Can Tho city, 94000, Vietnam*

[2]*Software Engineering Department, FPT University, Can Tho city, 94000, Vietnam*

ARTICLE INFO

Article history:
Received: 03 September, 2020
Accepted: 28 October, 2020
Online: 20 November, 2020

Keywords:
TextRank
Vietnamese
Jaccard Similarity
Representative Keywords
Text Classification

ABSTRACT

Text classification is considered one of the most fundamental and essential problems that deal with automatically classifying textual resources into pre-defined categories. Numerous algorithms, datasets, and evaluation measurements have been proposed to address the task. Within the era of information redundancy, it is challenging and time-consuming to engineering a sizable amount of data in multi-languages manually. However, it is time-consuming to consider all words in a text, but rather several key tokens. In this work, the authors proposed an effective method to classify Vietnamese texts leveraging the TextRank algorithm and Jaccard similarity coefficient. TextRank ranks words and sentences according to their contribution value and extracts the most representative keywords. First, we collected textual sources from a wide range of Vietnamese news websites. We then applied data preprocessing, extracted keywords by TextRank algorithm, measured similarity score by Jaccard distance and predicted categories. The authors have conducted numerous experiments, and the proposed method has achieved an accuracy of 90.07% on real-world datasets. We have proved that it is entirely applicable in practice.

1 Introduction

This paper is an extension of work initially presented in IEEE RIVF 2020 [1] as an invited paper.

Within the era of information redundancy, it is challenging and time-consuming to manually engineer a sizable amount of multilingual data. For example, an electronic library might quickly identify documents for archiving and management. Besides, it is often less accurate and overwhelming if humans are involved in the process. Therefore, apply machine approaches to automate the procedure of text classification is mandatory. It makes the classified results more reliable and less subjective. It helps alleviate human involvement and information overload and enhances knowledge retrieval efficiency. There are many text classification studies such as, e.g. Bayes [2], decision trees [3], K-nearest neighbor [4], and neural network [5]. The literature shows that automated text classification is one mainstream research in natural language processing [6, 7]. Many research papers have been conducted to solve such problems as email-messages filtering [8], topic modeling [9], geo-localization [10], and document categorization [11]. The flowchart of standard text classification is presented in Figure 1.

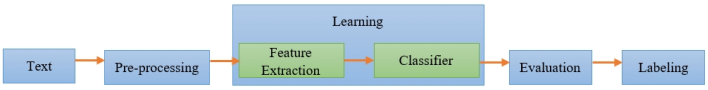


Figure 1: The flowchart of regular text classification where pre-processing and featuring perform an essential role.

The curse of dimensionality affects accuracy and computing time that arise when analyzing and organizing textual data in high-dimensional spaces [9]. However, machine learning models do not need to learn all tokens in the texts to categorize them. Instead, the text's label can be identified through its necessary tokens, which contribute the most to the text's meaning. Consequently, if we can extract the text source's main keywords, we might accurately classify it into an appropriate topic. The TextRank algorithm [12, 13] allows extracting the list of representative keywords of textual contents. On that consideration, the authors propose the method of automatic Vietnamese text classification based on the representative keyword analysis of the text. Textual datasets have been downloaded from several news websites with 15 main topics. Then, keyword sets that represent each topic are built. The system will extract that text's specific keywords when it is necessary to identify the topic

[*]Corresponding Author: Hiep Xuan Huynh, College of Information & Communication Technology, Can Tho University, Can Tho city, 94000, Vietnam, Email: hxhiep@ctu.edu.vn

of a text. The class of the text will be determined by computing the coverage of the representative keywords. It would be the right way for text classification applications in areas such as electronic library management. Although many Vietnamese documents are electronically available, no previous research has applied keywords-based text classification to them.

This work portrays our endeavor to construct a practical framework for addressing Vietnamese texts' classification task on the news. To the best of our knowledge, in this extended article, the authors have made several contributions. First, we discuss the paper [1] in a lot more detail by extending the related works and technical background. Second, we extend the experiments by investigating more scenarios.

2 Related Work

An early effort to address the task of Vietnamese text classification was conducted more than a decade ago [14]. In that paper, the authors solved automatically categorizing the problem, given textual sources into predefined categories. A comparison between statistical N-Gram language modeling and bag of words approaches has been investigated on their collected dataset. Several researchers have applied the idea of spam filtering into Vietnamese text sources [15]. Short messages such as conversational texts have also been exploited by addressing the task of suggestion intents [16]. The authors proposed a user suggestion intent definition in general from conversational texts at a functional segment unit. The task of automatic text categorization has been studied by comparing the performance of several term weighting schemes rather than analyzing the actual classification task [17]. Regarding Vietnamese sources, full-text representation has been exploited by many other research papers [18, 19, 20]. Thoughtfully learning the literature, we could claim that this is the first attempt to featuring Vietnamese texts using the idea of representative keywords.

3 Proposed System

3.1 Design Concept

The Vietnamese text classification system proposed by the authors includes two main components. One is the keyword extraction module, and the other is the comparison with the training set to identify the topic of a new document. We present the overall design of the system in Figure (**??**). The process begins with the text sources featured in the representative keyword vectors. For the training set, the vectors will be marked with the topic label. For the test data, the keyword extraction module is applied to convert the original text data into its keyword-based representation. Next, the system compares it to training data by calculating similarity scores. Finally, the prediction is assigned to the test data.

3.2 Text Pre-processing

Vietnamese is the only language in which every syllable is pronounced separately and is represented by a written word. This feature is evident in all aspects of phonetics, vocabulary and grammar. Data pre-processing is the first important step of any data mining process. It makes data in its original form easier to observe and explore. For the problem of text classification, due to specific characteristics, each language has its own challenges. The preprocessing process will help improve sorting efficiency and reduce the complexity of the training algorithm. Depending on the purpose of the classifier, we will have different preprocessing methods, such as

- Convert text to lowercase and correct spelling errors.
- Remove punctuation marks (if no sentence separation is performed).
- Remove special characters ([], [.], [,], [:], [”], [”], [;], [/], [[]], [$\tilde{]}$, [$\acute{]}$, [!], [@], [#], [$], [%], [$\hat{]}$, [&], [*], [(], [)]).
- Separate tokens by compound words (Vietnamese).
- Remove the stopwords, e.g. the words that appear most in the text that are not meaningful when participating in text classification. We utilize a list of 1942 Vietnamese stopwords[1] in our data processing.
- For the tokenization step, we utilize vnTokenizer [21] in our research. The comparison of tokenization accuracy achievable with different software is beyond the scope of this research paper.

3.3 Vietnamese Text Tokenization

Phonetic characteristics In Vietnamese, there is a special type of unit called ”tieng” or a sound of the thing. Phonetically, each ”tieng” is a syllable. For example, the word ”student” is translated into two syllables ”sinh vien” which are two separate words. As a result, these two words should come together to form a meaning token.

Vocabulary characteristics Each ”tieng”, in general, is a meaningful element. Continuing the previous example, the words ”sinh” and ”vien” have their own meaning when coming alone. But when they come together to form a single word ”sinh vien”, it has the meaning of student as in English. The vocabulary of Vietnamese is based on single words (one syllable) and the countless combination of them. Creating new words is very easy and flexible. If we pronounce a sound, we could write it down as a word.

Grammatical characteristics The Vietnamese words do not change morphology. For example, verbs in Vietnamese do not have -ed, -s, -ing forms. This feature will dominate other grammatical characteristics. When words combine with words into sentences, it is important to know the word order, the word phrase, and keywords for tenses recognition. Sorting words in a certain order is one of the main processes to express syntax relations.

[1] https://github.com/stopwords/vietnamese-stopwords/blob/master/vietnamese-stopwords.txt

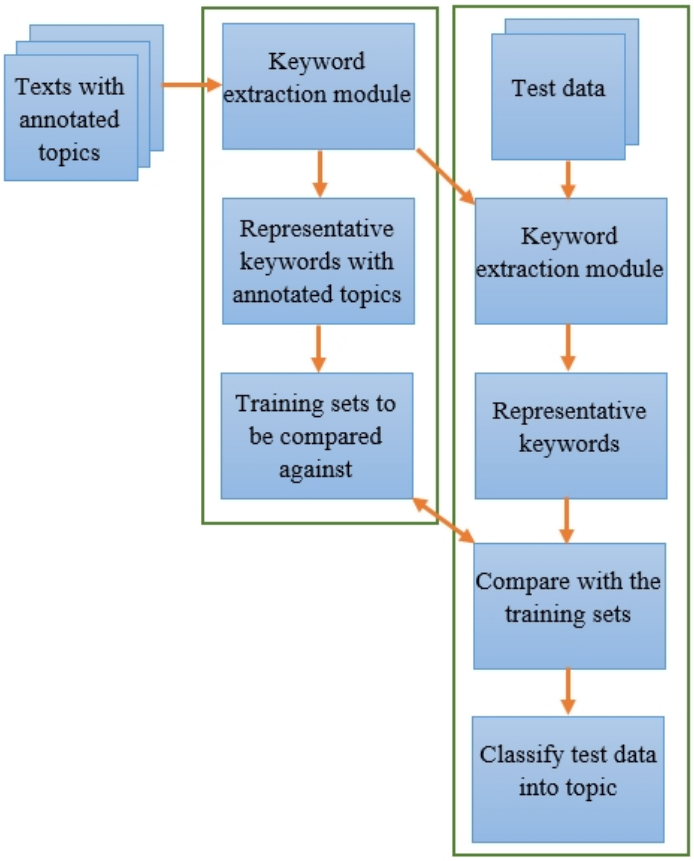


Figure 2: Our proposed Vietnamese text classification procedure.

3.4 Keyword Extraction by TextRank

Algorithm TextRank [22] was developed based on the main idea of the PageRank algorithm that Google search engine uses to rank website cite zhou2019chip, langville2008google, berkhout2016google. The bottom line of the TextRank algorithm is to use graphs to represent text and to score important information about the structure in which the text is represented by keywords. In other words, the TextRank algorithm processes a group of keywords representing the entire text. TextRank ranks words by their importance, arranges them in descending order of computed value, and extracts the most important words. The number of important words is a hyperparameter which is determined by the user prior to the TextRank algorithm execution. This algorithm is successfully applied for keyword extraction based on key value from a single text and this is also the advantage of TextRank.

The TextRank algorithm represents the textual source as a graph $G = (V, E)$ where V is the set of vertices and E is the set of edges in the graph. E is derived from a subset of V x V. Each vertex of the graph G corresponds to one word extracted from the text. An edge between any two vertices is created when their words appear in the text at any position between 2 and N. The value for the importance of the vertex V_i is calculated using the following formula:

$$S(V_i) = (1-d) + d * \sum_{j \in \text{In}(V_i)} \frac{1}{|\text{Out}(V_j)|} S(V_j)\,, \tag{1}$$

where $d = [0, 1]$. In our experiments, d is set to 0.85 by default [22]. $\text{In}(V_i)$ is the collection of vertices that point to it, and $\text{Out}(V_i)$ is the collection of vertices that vertex V_i points to it. The TextRank algorithm can be presented in the Algorithm (1).

Algorithm 1 Implementation of TextRank algorithm.

Input: Textual data
Output: Extracted keywords K

1: Build the graph $G = (V, E)$
2: Compute the edges' importance score by Equation (1)
3: Sort the edges by their scores
4: Select top K edges

3.5 Keyword Extraction for Topics

At this stage, we build a sample set of keywords for each topic from a set of keywords extracted from the subject-labeled texts in the training set. The model of building a set of keyword samples for topics can be summarized as Figure (3). The keywords of the topic are calculated by the statistical method of the number of occurrences of each word in the list of keyword sets of training text. Words that are keywords of one topic cannot be keywords of another topic.

3.6 Similarity Measurement by Jaccard Distance

Mathematically, there are many ways to calculate the similarity between any two keyword lists R_i and any R_j, provided they are of the same length. However, in the context of the similarity between two documents, we do not need to include all words in the text but only the representative keywords T. The number of keywords represented will be a lot less than the entire word in the text. Then, a weighted version of the Jaccard [23] distance is determined as follows:

$$\text{Jaccard}(R_i, R_j) = \frac{1}{T} \sum_{d=1}^{T} \frac{\left|R_{i,d} \cap R_{j,d}\right|}{\left|R_{i,d} \cup R_{j,d}\right|}, \tag{2}$$

4 Experiments

4.1 Datasets

Data are collected from highly reputable Vietnamese websites. We used Teleport pro [2] software for automated data collection. The downloaded data is converted to plain text file and saved to the corresponding folders with the folder name as predefined theme name. Specifically, data is downloaded from the website [3] [4] [5] [6] with 15 main topics. These topics are summarized in Table (1). The

[2] http://tenmax.com
[3] http://vnexpress.net
[4] http://dantri.vn
[5] http://tuoitre.vn
[6] http://yahoo.com.vn

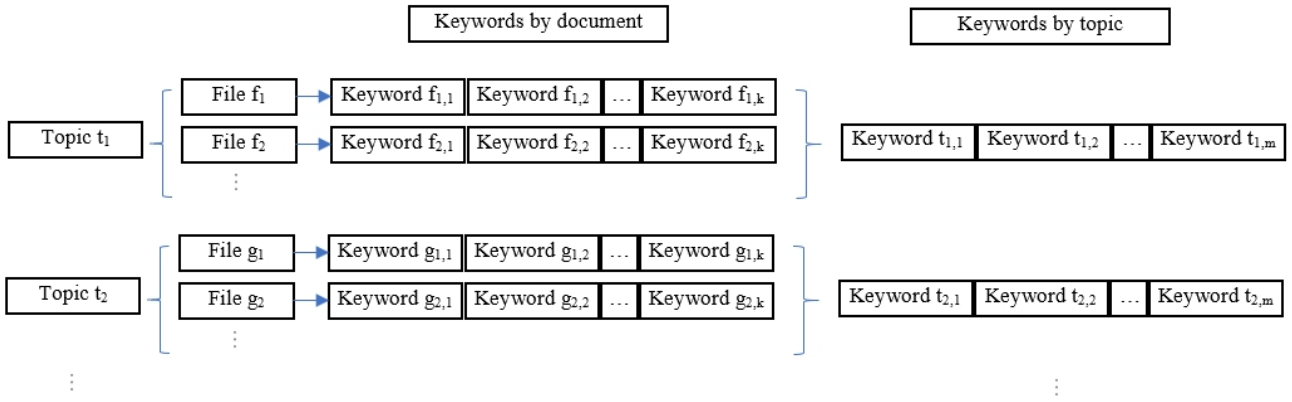


Figure 3: The keyword extraction for topics.

collected data was distributed into 80% training data group and 20% test data group.

Table 1: Summary of 15 topics

#	Topic	Description
1	Music	Singers, songs, musicians, performers, audiences.
2	Cuisine	Food, restaurants, cooks, menus.
3	Movies	Cinemas, genres, actors/actresses.
4	Tourism	tourist destinations, resorts, attractions, hotels.
5	Family	Marriage, family, grandpa, father, mother.
6	Education	Education and training, enrollment, students.
7	Science	Research, inventions, scientific discoveries.
8	Business	market, buying and selling, gold prices, currencies.
9	Beauty	beauty salon, care, therapies.
10	Motorcycle	motorbikes, cars, prices, comparison, repair.
11	Law	Legal, criminal, police events.
12	Health	health information, hospitals.
13	Sports	soccer matches, games, scores, comments.
14	Fashion	Field of fashion, costumes, designers, stylists.
15	Computers	Digital technology, computers, operating systems.

4.2 Evaluation Metric

We define the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). We also define m_+ the total of condition positives, m_- the total of condition negatives, $\hat{m}_+$ the total predicted condition positives, $\hat{m}_-$ the total predicted condition negatives, and m the total population. Then, we compute the sensitivity or recall by using:

$$\text{recall} = \frac{\text{TP}}{m_+} . \tag{3}$$

We compute precision as follows:

$$\text{precision} = \frac{\text{TP}}{\hat{m}_+} . \tag{4}$$

Then, we compute F1-score as follows:

$$\text{F1-score} = 2\frac{\text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}} . \tag{5}$$

Table 2: Statistics on all experimental scenarios.

# Keywords	Accuracy (%)	Training time (s)	Test time (s)
10	75.27	2155	1111
15	82.47	2156	1116
20	85.87	2162	1114
25	87.40	2168	1115
30	88.13	2178	1116
35	88.00	2177	1118
40	89.40	2184	1119
45	89.90	2191	1119
50	90.07	2199	1120
55	90.07	2205	1112
60	90.03	2213	1123
65	90.06	2219	1127
70	90.07	2307	1125
all	**95.40**	9279	1362

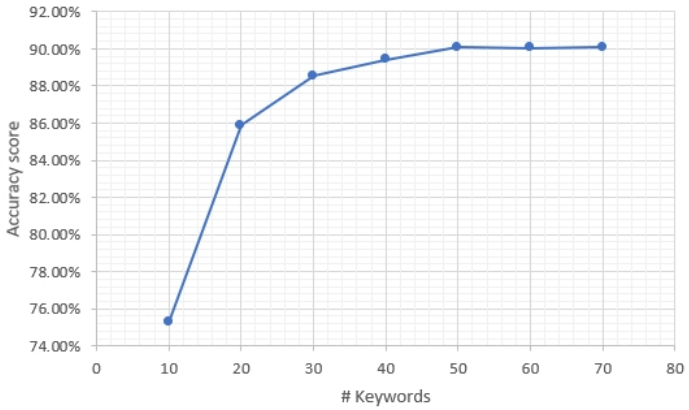


Figure 4: The correlation between the number of keywords and accuracy score.

4.3 Experimental Results

In the experiments, the authors have designed several scenarios where the number of keywords is varied to test the model's performance. The number of keywords $T = \{10, 20, 30, 40, 50, 60, 70\}$ and

in an extreme scenario where all the keywords selected by the TextRank algorithm are used. The experimental results are presented in Table (3), Table (4), Table (5), Table (6), Table (7), Table (8), Table (9), and Table (10) respectively.

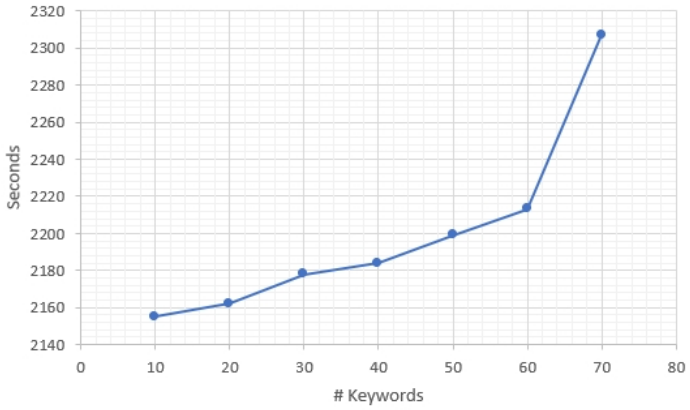


Figure 5: The correlation between the number of keywords and training time.

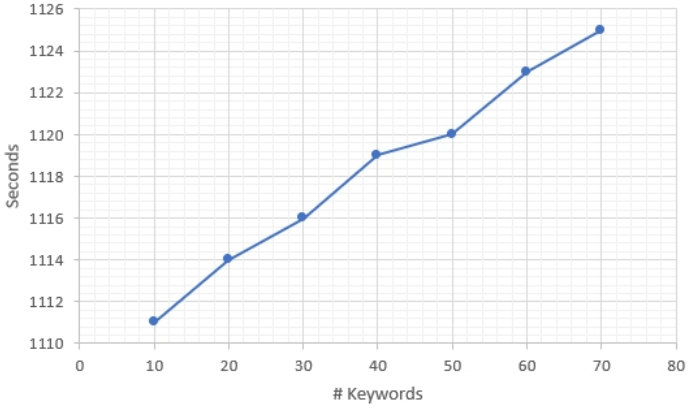


Figure 6: The correlation between the number of keywords and test time.

Table 3: Test set's confusion matrix with 10 keywords selected by the TextRank algorithm.

Topic ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Accuracy (%)
1	**87**	0	4	0	7	0	0	0	0	0	1	0	0	1	0	87
2	0	**73**	0	2	6	0	11	2	4	0	0	0	0	2	0	73
3	6	0	**93**	0	1	0	0	0	0	0	0	0	0	0	0	93
4	5	2	3	**60**	14	0	0	6	2	0	2	1	2	3	0	60
5	4	2	2	0	**78**	2	0	2	0	0	6	2	0	0	2	78
6	4	2	0	0	2	**80**	0	0	0	0	10	0	2	0	0	80
7	4	2	3	5	9	5	**49**	0	7	2	0	1	3	1	9	49
8	4	1	0	2	7	0	2	**77**	0	2	0	1	2	0	2	77
9	0	1	1	0	2	0	4	1	**83**	0	0	6	2	0	0	83
10	0	0	0	1	0	0	0	4	1	**91**	1	0	1	1	0	91
11	0	1	1	1	7	0	1	6	0	0	**82**	1	0	0	0	82
12	4	0	0	3	9	1	5	0	6	1	5	**66**	0	0	0	66
13	2	1	0	1	18	1	0	0	1	0	0	0	**74**	1	1	74
14	6	0	10	2	2	0	2	2	3	4	1	0	1	**67**	0	67
15	2	0	2	0	1	0	1	10	3	10	0	0	0	2	**69**	69
Average accuracy																**75**

The authors also make statistics on all experimental scenarios in Table (2). Note that *all* in the Table means all the tokens in a particular text. The texts' length is varied, and the average number of tokens is 500. We plotted the correlation between the number of keywords and accuracy scores in Figure (4). The accuracy of 90% is stable at 50 keywords. In Figure (5), we observe a considerable increase from 60 to 70 keywords. While in Figure (6), the test time grows eventually.

Table 4: Test set's confusion matrix with 20 keywords selected by the TextRank algorithm.

Topic ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Accuracy (%)
1	**96**	0	1	0	3	0	0	0	0	0	0	0	0	0	0	96
2	1	**85**	0	1	5	0	3	2	0	1	1	0	0	1	0	85
3	0	0	**95**	1	3	0	0	0	1	0	0	0	0	0	0	95
4	0	0	1	**83**	5	0	0	10	0	0	0	0	0	1	0	83
5	0	0	0	2	**92**	0	0	0	0	0	6	0	0	0	0	92
6	0	0	2	0	7	**77**	0	2	0	0	10	0	0	0	2	77
7	0	0	0	3	5	2	**78**	3	3	2	0	2	1	0	1	78
8	0	0	2	1	2	0	2	**89**	0	1	2	0	0	0	1	89
9	0	0	0	0	4	0	2	1	**81**	3	0	7	0	2	0	81
10	1	0	0	0	0	0	1	1	0	**95**	0	0	1	1	0	95
11	1	0	0	0	8	0	0	5	0	0	**85**	1	0	0	0	85
12	0	1	0	2	14	2	5	2	6	1	3	**63**	0	0	1	63
13	0	0	0	0	4	0	0	0	0	0	0	0	**96**	0	0	96
14	5	0	3	1	3	0	0	0	1	1	0	0	0	**86**	0	86
15	1	0	2	0	0	0	2	3	0	5	0	0	0	0	**87**	87
Average accuracy																**86**

Table 5: Test set's confusion matrix with 30 keywords selected by the TextRank algorithm.

Topic ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Accuracy (%)
1	**95**	0	2	0	3	0	0	0	0	0	0	0	0	0	0	95
2	0	**91**	0	2	7	0	0	0	0	0	0	0	0	0	0	91
3	0	0	**95**	0	3	0	0	0	0	0	0	0	0	2	0	95
4	0	0	2	**84**	5	0	0	8	0	0	1	0	0	0	0	84
5	0	0	0	0	**94**	0	0	0	0	0	4	2	0	0	0	94
6	0	0	0	0	12	**74**	2	6	0	2	4	0	0	0	0	74
7	0	0	0	0	2	0	**88**	5	3	1	0	1	0	0	0	88
8	0	0	2	1	1	1	3	**91**	0	0	1	0	0	0	0	91
9	0	0	0	0	4	0	0	0	**88**	2	0	4	0	2	0	88
10	0	0	0	0	0	0	0	3	0	**96**	0	0	1	0	0	96
11	0	0	0	0	5	0	0	7	0	0	**88**	0	0	0	0	88
12	0	3	0	2	15	1	5	5	5	0	2	**62**	0	0	0	62
13	0	0	0	0	4	0	0	0	0	0	0	1	**95**	0	0	95
14	4	0	2	1	0	0	0	2	1	0	0	0	0	**90**	0	90
15	0	0	0	0	1	0	1	5	0	2	0	0	0	0	**91**	91
Average accuracy																**88**

Table 6: Test set's confusion matrix with 40 keywords selected by the TextRank algorithm.

Topic ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Accuracy (%)
1	**91**	0	1	0	7	0	0	0	0	0	0	0	0	1	0	91
2	0	**95**	0	2	3	0	0	0	0	0	0	0	0	0	0	95
3	0	0	**94**	1	3	0	0	0	0	2	0	0	0	0	0	94
4	0	0	1	**87**	1	0	0	11	0	0	0	0	0	0	0	87
5	0	0	0	0	**92**	2	0	0	0	0	4	2	0	0	0	92
6	0	0	0	0	12	**78**	2	2	0	0	6	0	0	0	0	78
7	0	0	1	1	0	0	**90**	4	2	1	0	1	0	0	0	90
8	0	0	1	0	2	0	0	**94**	0	1	2	0	0	0	0	94
9	0	0	0	1	3	0	0	0	**90**	1	0	2	0	3	0	90
10	0	0	0	0	0	0	0	4	1	**95**	0	0	0	0	0	95
11	0	0	1	1	5	0	0	3	0	0	**89**	0	0	0	1	89
12	0	3	0	3	14	0	3	11	5	0	0	**61**	0	0	0	61
13	0	0	0	0	4	0	0	0	0	0	0	0	**96**	0	0	96
14	2	0	1	0	0	0	0	2	0	0	0	0	1	**94**	0	94
15	0	0	0	0	1	1	0	2	0	1	0	0	0	0	**95**	95
Average accuracy																**89**

Table 7: Test set's confusion matrix with 50 keywords selected by the TextRank algorithm.

Topic ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Accuracy (%)
1	**94**	0	1	0	4	0	0	0	0	0	0	0	0	1	0	94
2	1	**96**	0	1	1	0	0	1	0	0	0	0	0	0	0	96
3	6	0	**87**	0	2	0	0	1	0	2	0	0	0	2	0	87
4	0	0	0	**89**	3	0	0	8	0	0	0	0	0	0	0	89
5	0	0	0	0	**96**	0	0	0	0	0	2	2	0	0	0	96
6	0	0	0	0	14	**82**	2	0	0	0	2	0	0	0	0	82
7	0	0	1	3	0	0	**87**	3	1	4	0	1	0	0	0	87
8	0	0	0	2	1	0	0	**95**	0	0	2	0	0	0	0	95
9	0	0	0	0	1	0	0	0	**90**	1	0	4	1	3	0	90
10	0	0	0	1	0	0	1	4	1	**93**	0	0	0	0	0	93
11	0	0	1	0	6	0	0	2	0	0	**91**	0	0	0	0	91
12	0	1	0	1	13	0	3	6	3	0	1	**72**	0	0	0	72
13	0	0	0	0	4	0	0	0	0	0	0	0	**96**	0	0	96
14	3	0	2	0	1	0	0	0	0	0	0	0	0	**94**	0	94
15	2	0	0	0	1	1	0	5	1	1	0	0	0	0	**89**	89
Average accuracy																**90**

Table 8: Test set's confusion matrix with 60 keywords selected by the TextRank algorithm.

Topic ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Accuracy (%)
1	**93**	0	1	0	4	0	0	1	0	0	0	0	0	1	0	93
2	0	**94**	0	2	4	0	0	0	0	0	0	0	0	0	0	94
3	2	0	**89**	2	4	0	0	1	0	0	0	0	0	2	0	89
4	0	0	0	**87**	3	0	0	10	0	0	0	0	0	0	0	87
5	0	0	0	0	**94**	2	0	0	0	0	4	0	0	0	0	94
6	0	0	1	1	12	**74**	2	2	0	0	8	0	0	0	0	74
7	0	0	1	0	0	0	**87**	5	0	6	0	1	0	0	0	87
8	0	0	2	2	1	0	0	**93**	0	0	2	0	0	0	0	93
9	0	1	0	1	0	0	0	0	**93**	1	0	2	0	2	0	93
10	0	0	0	1	0	0	1	3	0	**95**	0	0	0	0	0	95
11	0	0	0	1	1	0	0	3	0	0	**95**	0	0	0	0	95
12	0	1	0	2	7	0	5	6	4	0	3	**72**	0	0	0	72
13	0	0	0	0	4	0	0	0	0	0	0	1	**95**	0	0	95
14	3	0	1	0	0	0	0	2	0	0	0	0	0	**94**	0	94
15	0	0	0	0	2	0	0	2	0	1	0	0	0	0	**95**	95
Average accuracy																**90**

Table 9: Test set's confusion matrix with 70 keywords selected by the TextRank algorithm.

Topic ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Accuracy (%)
1	**94**	0	1	0	2	0	1	1	0	0	0	0	0	1	0	94
2	0	**93**	0	1	3	0		0	0	0	0	2	0	1	0	93
3	0	0	**87**	2	6	0	0	1	0	0	0	0	0	4	0	87
4	0	0	0	**94**	0	0	0	6	0	0	0	0	0	0	0	94
5	0	0	0	0	**94**	0	0	0	0	0	4	2	0	0	0	94
6	0	0	0	3	10	**75**	2	4	0	0	6	0	0	0	0	75
7	0	0	0	0	0	0	**91**	5	0	4	0	0	0	0	0	91
8	0	0	1	1	1	0	0	**95**	0	0	2	0	0	0	0	95
9	0	0	0	1	3	0	0	0	**90**	1	0	2	0	3	0	90
10	0	0	0	1	0	0	1	4	0	**94**	0	0	0	0	0	94
11	0	1	0	0	1	0	0	2	0	0	**95**	1	0	0	0	95
12	0	1	0	0	10	1	3	6	8	0	4	**67**	0	0	0	67
13	0	0	0	0	4	0	0	0	0	0	0	1	**95**	0	0	95
14	2	0	2	0	0	0	1	1	0	0	0	0	0	**94**	0	94
15	0	0	0	0	1	0	0	4	1	0	0	0	0	1	**93**	92
Average accuracy																**90**

5 Conclusion

In this article, the authors have described a proposed approach that allows text classification based on the solution of extracting specific representative keywords of the text. We discussed the proposed system in detail from the abstract design, text pre-processing, and Vietnamese characteristics. Then we described the TextRank algorithm based on graphs to score important information about the text's structure. Intensive experiments have been conducted to prove the stability and robustness of the proposed system. High accuracy of 90.07% has been achieved. Although many Vietnamese documents are electronically available, this is the first to conduct text classification based on keywords. This research portrays our endeavor to construct a practical framework for addressing Vietnamese texts' classification tasks on news websites.

Table 10: Test set's confusion matrix with all keywords selected by the TextRank algorithm.

Topic ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Accuracy (%)
1	**92**	0	0	0	8	0	0	0	0	0	0	0	0	0	0	92
2	0	**92**	0	2	6	0	0	0	0	0	0	0	0	0	0	92
3	0	0	**86**	2	12	0	0	0	0	0	0	0	0	0	0	86
4	0	0	0	**96**	2	0	0	2	0	0	0	0	0	0	0	96
5	0	0	0	0	**99**	0	0	0	0	0	1	0	0	0	0	99
6	0	0	0	0	16	**84**	0	0	0	0	0	0	0	0	0	84
7	0	0	0	0	0	0	**99**	1	0	0	0	0	0	0	0	99
8	0	0	1	0	0	0	0	**99**	0	0	0	0	0	0	0	99
9	0	0	0	0	1	0	0	0	**99**	0	0	0	0	0	0	99
10	0	0	0	1	1	0	1	0	0	**97**	0	0	0	0	0	97
11	0	0	0	0	1	0	0	0	0	0	**99**	0	0	0	0	99
12	0	0	0	0	2	0	0	0	1	0	0	**97**	0	0	0	97
13	0	0	0	0	4	0	0	0	0	0	0	0	**96**	0	0	96
14	0	0	1	0	0	0	0	0	0	0	0	0	0	**99**	0	99
15	0	0	0	0	1	0	1	0	0	0	0	0	0	0	**98**	98
Average accuracy																**95**

References

[1] H. T. Huynh, N. Duong-Trung, X. S. Ha, N. Q. T. Tang, H. X. Huynh, D. Q. Truong, "Automatic Keywords-based Classification of Vietnamese Texts," in 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), 1–3, IEEE, 2020, doi:10.1109/RIVF48685.2020.9140761.

[2] L. Zhang, L. Jiang, C. Li, G. Kong, "Two feature weighting approaches for naive Bayes text classifiers," Knowledge-Based Systems, **100**, 137–144, 2016, doi:10.1016/j.knosys.2016.02.017.

[3] C. C. Aggarwal, Data classification: algorithms and applications, CRC press, 2014.

[4] S. Jiang, G. Pang, M. Wu, L. Kuang, "An improved K-nearest-neighbor algorithm for text categorization," Expert Systems with Applications, **39**(1), 1503–1509, 2012, doi:10.1016/j.eswa.2011.08.040.

[5] J. Nam, J. Kim, E. L. Mencía, I. Gurevych, J. Fürnkranz, "Large-scale multi-label text classification—revisiting neural networks," in Joint european conference on machine learning and knowledge discovery in databases, 437–452, Springer, 2014, doi:10.1007/978-3-662-44851-9_28.

[6] V. B. Kobayashi, S. T. Mol, H. A. Berkers, G. Kismihók, D. N. Den Hartog, "Text classification for organizational researchers: A tutorial," Organizational research methods, **21**(3), 766–799, 2018, doi:10.1177/1094428117719322.

[7] N. Duong-Trung, Social Media Learning: Novel Text Analytics for Geolocation and Topic Modeling, Cuvillier Verlag, 2017.

[8] G. Xu, J. Qi, D. Huang, M. Daneshmand, "Detecting spammers on social networks based on a hybrid model," in 2016 IEEE International Conference on Big Data (Big Data), 3062–3068, IEEE, 2016, doi:10.1016/j.neucom.2015.02.047.

[9] N. Duong-Trung, L. Schmidt-Thieme, "On Discovering the Number of Document Topics via Conceptual Latent Space," in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM '17, 2051–2054, Association for Computing Machinery, New York, NY, USA, 2017, doi:10.1145/3132847.3133086.

[10] N. Duong-Trung, N. Schilling, L. Schmidt-Thieme, "Near real-time geolocation prediction in twitter streams via matrix factorization based regression," in Proceedings of the 25th ACM international on conference on information and knowledge management, 1973–1976, ACM, 2016, doi:10.1145/2983323.2983887.

[11] D. Kim, D. Seo, S. Cho, P. Kang, "Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec," Information Sciences, **477**, 15–29, 2019, doi:10.1016/j.ins.2018.10.006.

[12] R. Niu, B. Shen, "Microblog User Interest Mining Based on Improved TextRank Model," Journal of Computers, **30**(1), 42–51, 2019, doi:10.3966/199115992019023001005.

[13] N. Akhtar, M. S. Beg, H. Javed, "TextRank enhanced Topic Model for Query focussed Text Summarization," in 2019 Twelfth International Conference on Contemporary Computing (IC3), 1–6, IEEE, 2019, doi:10.1109/IC3.2019.8844939.

[14] V. C. D. Hoang, D. Dinh, N. Le Nguyen, H. Q. Ngo, "A comparative study on vietnamese text classification methods," in 2007 IEEE International Conference on Research, Innovation and Vision for the Future, 267–273, IEEE, 2007, doi:10.1109/RIVF.2007.369167.

[15] T.-H. Pham, P. Le-Hong, "Content-based approach for Vietnamese spam SMS filtering," in 2016 International Conference on Asian Language Processing (IALP), 41–44, IEEE, 2016, doi:10.1109/IALP.2016.7875930.

[16] T.-L. Ngo, K. L. Pham, H. Takeda, S. B. Pham, X. H. Phan, "On the identification of suggestion intents from vietnamese conversational texts," in Proceedings of the Eighth International Symposium on Information and Communication Technology, 417–424, 2017, doi:10.1145/3155133.3155201.

[17] V. T. Nguyen, N. T. Hai, N. H. Nghia, T. D. Le, "A Term Weighting Scheme Approach for Vietnamese Text Classification," in International Conference on Future Data and Security Engineering, 46–53, Springer, 2015, doi:10.1007/978-3-319-26135-5_4.

[18] N. H. D. Tri, N. T. Quan, N. V. Tien, N. T. Hung, "Xay dung mo hinh phan tan cho phan lop khoi luong lon van ban theo chu de (in English: building distributed model for classification massive text data by topic)," PROCEEDING of Publishing House for Science and Technology, 2017, doi:10.15625/vap.2016.000104.

[19] B. K. Linh, N. T. T. Ha, N. T. N. Tu, D. T. Tinh, "Phan loai van ban tieng Viet dua tren mo hinh chu de (in English: vietnamese text classification based on topic modeling)," PROCEEDING of Publishing House for Science and Technology, 2017, doi:10.15625/vap.2016.00065.

[20] T. Ngoc Phuc, P. Tran Vu, P. Cong Xuyen, N. Vu Duy Quang, "Phan loai noi dung tai lieu Web tieng viet (in English: classification of vietnamese texts on the web)," Vietnam Journal of Science and Technology, **51**(6), 669–680, doi:10.15625/2525-2518/51/6/11629.

[21] N. T. M. Le Hong Phuong, A. R. Huyen, A. Ho Tuong Vinh, "Hybrid Approach to Word Segmentation of Vietnamese Texts," 2008, doi:10.1007/978-3-540-88282-4_23.

[22] R. Mihalcea, P. Tarau, "Textrank: Bringing order into text," in Proceedings of the 2004 conference on empirical methods in natural language processing, 404–411, 2004.

[23] G. Rebala, A. Ravi, S. Churiwala, An introduction to machine learning, Springer, 2019.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 370-390 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Intrusion Detection and Classification using Decision Tree Based Key Feature Selection Classifiers

Manas Kumar Nanda[1,*], Manas Ranjan Patra[2]

[1]*Department of Computer Application, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, 751030, India*

[2]*Department of Computer Science, Berhampur University, Berhampur, 760007, India*

ARTICLE INFO

Article history:
Received: 26 August, 2020
Accepted: 31 October, 2020
Online: 20 November, 2020

Keywords:
Attribute
Confidentiality
Features
Intrusion
Rank
Spam

ABSTRACT

Feature selection method applied on an intrusion dataset is used to classify the intrusion data as normal or intrusive. We have made an attempt to detect and classify the intrusion data using rank-based feature selection classifiers. A set of redundant features having null rank value are eliminated then the performance evaluation using various feature selection algorithms are done to determine the behavior of attributes. We can distinguish the key features which plays an important role for detecting intrusions. There are 41 features in the dataset, out of which some features play significant role in detecting the intrusions and others do not contribute in the detection process. We have applied different feature selection techniques to select the predominant features that are actually effective in detecting intrusions.

1. Introduction

Intrusion Detection System (IDS) works as an application or device which identifies some hostile activities or as policy violations by the intruder in network. IDS is used to analyze the network traffic and detect some possible intrusive activities in the computer network. Mainly misuse detection system and anomaly detection system are the two types of intrusion detection systems. It is capable of detecting probable attacks from the known patterns or signatures, and identify some intrusive activities which deviates from normal behavior in a monitored system, and can detect some unknown attacks [1]. The most popular IDSs are SourceFire, McAfee, and Symantec, which plays an important role for network surveillance and monitoring, and functions like a network security guard. IDS can be categorized as Network based Intrusion Detection System (NIDS) and Host based Intrusion Detection System (HIDS) [2]. In NIDS, the Intrusion Detection System (IDS) is installed before and after the firewall to capture network traffic for the entire network segment, but in HIDS, the Intrusion Detection System (IDS) is applied on a specific host to analyse packets, logs and system calls. As compared to NIDS, HIDS is more suitable for identifying the internal attacks.

We have applied a number of techniques to analyze the intrusion data and build a system that has higher detection rate.

*Corresponding Author: Manas Kumar Nanda, SOA University, 9437296663 & manasnanda@soa.ac.in

2. Data Mining-based Approach

Data mining is a method of discovering a way of systematic relationship of data and an approach of determining the fundamental information of data. It is broadly divided into two categories such as supervised and unsupervised approach. Classifications and clustering are the best examples of supervised and unsupervised algorithms respectively. In a clustering approach, the group of unique objects are based on the characteristics of such data points [3]. Where these data points in a cluster is similar to other data points in the cluster and is dissimilar to the data points in different cluster. By grouping such similar data points into one cluster which shows the abnormality identification. Hence this approach may be responsible for potentially increase of the false alarm rate. The performance of IDS is highly dependent on the low false alarm rate, which may degrade the performance when it generate high false alarm [4]. Classification is one of the best supervised approach used for classifying the benign or anomalous data, for reducing the false alarm rate. It has the ability to differentiate unusual data pattern, which may be suitable for identifying new attack patterns [5]. Classification is widely used for its strong ability in identifying the normal structure very accurately, which contribute towards its reducing false detection [6]. These ensemble techniques are used to combine several classifiers which obtain better prediction for its accuracy in performance [7].

2.1. Classification

A classification technique (also known as classifier) is a systematic approach to build the classification models from an input dataset. Some of the techniques like Decision Tree based classifiers, Rule based classifiers, Neural Networks, Naïve Bayes classifiers and Support Vector Machines etc., each of the technique employs a learning-based algorithm to identify a model that best suits the relationship between the set of attributes and class label of input data. The generated model by learning algorithm should fit both input data well and also correctly predict class labels of records. The primary objective of the learning based algorithm is to build a model with good generalization and capability; i.e., the models that accurately predict the class labels of previously unknown records. This classification is done using a training set which consists of records whose class labels are known and must be provided. To build a classification model, using this is subsequently applied to the test set, which consists of records with unknown class labels.

The performance of a classification model is evaluated based on the classification of test records correctly and incorrectly predicted by the model. The counts of the predicted values are tabulated in a table known to be confusion matrix. For the learning and classification, we have used various machine learning techniques. There are two major categories of machine learning techniques, namely, Supervised and unsupervised technique, supervised technique requires an initial training phase where the algorithm is trained using existing dataset with appropriate classification. The algorithm then uses this knowledge to perform the real-time classification of test data. Conversely, the unsupervised technique does not require any existing classification method and basically use multiple runs to fine tune the classification patterns.

2.2. Decision Tree

Classifying the test record is a straightforward approach once a decision tree is being constructed. Basically starting from a root node, we go on applying the test condition to the records and follow to the appropriate branches based upon the outcome of the test result. This will lead to us either to the internal node, to which the new test condition is being applied or to the leaf node. This class label which is associated with the leaf node is then assigned to the record.

There are many decision trees, which can be constructed from a given set of attributes, where some of the trees are more accurate than others, and finding a optimal tree is computationally infeasible because of exponential size of the size of the search space. A number of algorithms have been developed to induce, with a reasonably accurate and albeit suboptimal decision tree constructed in a reasonable amount of time. Such algorithms employ a greedy strategy that design a decision tree by taking a series of locally optimum decisions about which, attributes are to be used for partitioning the data. One of such algorithm is Hunt's algorithm, which is the basis of most of the existing decision tree induction algorithms, including, C4.5, Classification and Regression Trees (CART), and Iterative Dichotomiser 3 (ID3). An efficient algorithm to build a decision tree is C4.5, used for classification (also known as statistical classifier), can be described as "a landmark of decision tree program that is probably the machine learning workhorse and most widely used in practice to date".

2.2.1. Random Forest

The most versatile tree-based machine learning algorithm, which is used to build several trees (or decision trees), and then combining each of the output to improve the generalization ability of the building model. This method of combining the trees (i.e combining weak learners (or individual trees) to produce a strong learner (a Forest)) is also known as an ensemble method. Random Forest algorithm can be used to solve the regression problems (where the dependent variables are continuous) and classification problems (where the dependent variables are categorical).

In a given data frame, the tree stratifies or partitions the data, based on rules (such as if-else), then these rules divide the dataset into a number of distinct and non-overlapping regions. Such rules are determined by the use of variable's contribution to the homogeneity or pureness of the respective resultant child nodes. In the regression trees (where the output is predicted by the mean of observations at the terminal nodes), the splitting decision is based on minimizing the RSS. The variable, which leads to the greatest possible reduction in RSS, is chosen in the root node. The tree splitting takes a top-down greedy approach (which is also known as recursive binary splitting) , because the algorithm cares to make the best split, at the current step rather than saving a split for better result on future nodes.

2.3. Feature Selection

Feature selection is a process of selecting a subset of M features from a set of N features, so that the feature space is optimally reduced based on a certain evaluation criteria. The objective of feature subset selection method is to find a optimum set of features such that, the resulting probability distribution of the data classes is as close as possible to the original distribution obtained using all the attributes [8]. Feature selection process used to improve the classification performance by searching for a subset of features, which best classifies the training data. In a high dimensional feature space some of the features may be redundant or irrelevant, which may deteriorate the performance of classifiers [9]. It is very much important to remove these redundant or irrelevant features. It is necessary to find the subset of features involves in the selection process to improve the prediction of accuracy or decrease the size of the features in the dataset without significantly decreasing the prediction accuracy of a classifier which is built using only the selected features.

Feature selection is the foundation of machine learning [10], a process of discovering the most useful and prominent features for the learning-based algorithm. It is very much important to extract the set of redundant or irrelevant features need to prevent the classifiers from being biased and required to minimize the feature selecting error so as to improve the abnormal behavior and detection rate. This is because of the application of appropriate algorithm and its effectiveness highly dependent on the feature selection process. Filters and wrappers [11] are the two generalized methods used in feature selection. Filter-based subset evaluation (FBSE) method is used to remove the redundant features inside the filter ranking [12]. This process examines the complete subset in a multivariate way and select the relevant features and explores the degree of relationship between the features.

Henceforth FBSE is a heuristics approach which involves the probability and statistical measures for searching and evaluation of the usefulness of identified features. The wrapper based subset evaluation (WBSE) method use the classifier for estimating the worth of each feature subset. Basically WBSE methods have greater predictive accuracy as compared to FBSE method, because the selection approach is optimized, when evaluating each of the feature subset using classification algorithm. So more or less the WBSE method use a classification based algorithm for evaluation of each set of features. But WBSE method becomes uncontrollable, at the time of dealing with large databases with many features [13]. Hence, WBSE methods are highly associated with the classifier algorithm which makes it more difficult at the time of shifting from one classifier to another classifier because of the re-initiation of selection process. But the FBSE method uses distance measures and correlation functions for selection criteria of features [14]. Where FBSE method do not need re-execution of different learning based classifiers and hence its execution process is more faster than WBSE. So FBSE is favorable for large databases environment which contains many features.

The statistical based detection approach introduced by [15] based on the collection of data for creating normal behavior profile. Here traffic data over a period of time is collected for utilization of intrusion detection. In Packet Header Anomaly Detection (PHAD), the abnormal patterns are recognized using the packet characteristics and behaviors. The normal profile is constructed using statistical measurement of activity history[16]. A set of traffic can be defined as intrusive, that are deviated from normal profile and behaved abnormally. PHAD uses all of the 33 attributes of a packet header that represents information of data link, network and transport layers of 7 layer OSI model without using IP address and port number, The probability of each packet being benign or tending towards abnormal behavior is measured by the information contain in each attribute. For any such dissimilarity detected at the time of matching against the training data, an anomaly score is given. The anomaly score for each packet is summed-up, and if the score surpasses the preset threshold then it is flagged as anomalous.

Network-based and host-based are two different environments in Protocol based Packet Header Anomaly Detection (PhPHAD) of conventional PHAD system [17]. TCP, UDP and ICMP are the three main protocols used to construct normal profile. The Light weight Network Intrusion Detection System (LNID) has been proposed to identify the malicious packets in Telnet traffic. In LNID, the behavior is extracted from the training data to construct the normal profile which is further used for computing the anomaly score. This anomaly score is used to match between training and testing data. Then surpassed preset threshold score packets are treated as malicious packets. To reduce computational cost the insignificant features from the training data are removed during preprocessing phase.

Rank based feature selection: Feature ranking method calculate the score of each attribute and arrange them in descending order according to their score. The performance of six ranking methods used for feature selection is divided into entropy based attribute evaluator and statistical attribute evaluator technique. Entropy based attribute evaluator technique is used in information theory to characterize the purity of an arbitrary collection of samples. Information Gain (IG), Gain Ratio (GR) and Symmetrical Uncertainty (SU) are the entropy based attribute evaluators used to measure system's unpredictability. Whereas One Rule (One R), Chi Squares and Relief-F Attribute evaluators are statistical attribute evaluator techniques.

3. Rank-based Classification

We have used most of the efficient data mining classification algorithms used for IDS. The Best-First Decision Tree based (BFT) classifier, basically used for binary splitting of both normal and numeric valued attributes, the decision tree learner based on imprecise probabilities and uncertainty measures (CDT), the class implementing decision Forest Algorithm (FPA) using bootstrap samples and penalized attributes [18] , the building of Functional Trees (FT) for classification, more specifically functional trees uses logistic regression based functions at inner nodes and leaves. This algorithm can also deal with the binary as well as multiclass target variables, along with nominal attributes and numeric and with missing values [19]. A Hoeffding Tree (VFDT) is incremental based, anytime decision-tree induction algorithm, which is capable of learning from a massive data stream. Hoeffding trees exploit the fact that, an optimal splitting attribute can often be chosen from a small sample [20], the class for generating a pruned or unpruned C4 (J48) [21] and the class for generating a pruned or unpruned C4.5 Consolidated Tree Construction (CTC) algorithm (J48Consolidated) in which a set of subsamples are used to build a single tree, where the Resampling Method (RM) [22] is built with a few new options added to the J48 class, whereas the class for generating a grafted (pruned or unpruned) C4 (J48graft) [23] and the class for generating a multi-class alternating decision tree using LogitBoost strategy (LADTree) [24]. The classifier for building a Logistic Model Trees (LMT), in which the classification trees are the logistic regression functions of the leaves. This algorithm can also deal with the binary as well as multiclass target variables, along with nominal attributes and numeric and with missing values [25]. The class for generating decision tree with Naïve Bayes classifiers at the leaves (NBTree) [26]. A class for constructing a forest with Random Trees (RF) [27] and the class considers K number of randomly chosen attributes at each level of node (RT) for constructing a tree. The. Fast decision tree learner (REPT) which builds a decision tree or regression tree using the information gain or variance and prunes it using the reduced error pruning method with backfitting, and the Implementation of the decision forest algorithm SysFor (SF) [28].

4. Dataset used

The KDDCUP99 dataset is derived from the DARPA98 network traffic data in 1999, which assembled individual TCP packets into TCP connections. Each of the TCP connection having 41 features along with a label that specifies a specific type of attack or normal as a status of a connection. Dataset consists of 38 numeric features and three symbolic features, which are again classified into following four different categories [29]. First nine (f1-f9) features are used to describe each TCP connection. In this category all the attributes are being extracted from a TCP/IP connection, and these features lead to an implicit-delay in detection. The second thirteen (f10-f22) are domain knowledge related content features used to indicate that suspicious behavior in the network traffic having no sequential patterns. But unlike

Table 1: The Various types of Attacks and their Classifications

Attack Category/ Attack Name		Attack Description
Denial of Service (DoS)		In such attack an attacker tries to make the system's computing/ memory resources too busy or full to handle the legitimate requests, or denies such legitimate users to access a system. The most possible ways to launch the DoS attacks are by abusing the computers for legitimate features, and by targeting the bugs, or by exploiting the system's misconfiguration.
	Back	A DoS attack against the apache web server, in which a client requests URL containing many backslashes, that slows down server response
	Land	A DoS attack where remote host sends a UDP packet with same source and destination, freezes the machine
	Neptune	Syn-flood DoS attack on one port or on more ports
	Ping of Death	DoS ping-of-death
	Smurf	A DoS attack in which a large number of ICMP (Internet Control Message Protocol) packets with its intended victim's spoofed source IP may broadcast to computer network using IP broadcast address. By which the victim's computer may slow down for devices on a network intending to send a reply to the source IP, in a flooded traffic of large no of packets.
	Teardrop	A program, sends IP fragments to a machine which is connected to a network or internet. It is a DoS attack that exploit an overlapping IP fragment bug which is present in Windows 95, Windows NT and Windows 3.1 Operating System Machines. This bug causes TCP/ IP fragmentation reassembly code for improperly handle overlapping IP fragments, which needs a reboot for preferred remedy.
Remote to Local (R2L)		In such attack an attacker without having a registered account in a remote machine, that send packets to machine on a network and exploits the vulnerability for illegally gain local access as a user on that machine.
	Ftp_write	The remote FTP user creates .rhost file in the world writeable anonymous FTP directory that obtains local login which gains user access.
	Guess_passwd	An attacker tries to gain access to the user account, by repeatedly guessing possible passwords.
	Imap	A remote buffer-overflow using Imap port that leads to root shell which gains root access.
	Multihop	Multi-day scenario by which a user first breaks into a machine.
	Phf	The exploitable CGI script that allows a client to execute the arbitrary commands on a machine using misconfigured web server.
	Spy	That sends packets to a machine over a network through it which doesn't have an account in the target machine.
	Warezclient	The user used to download illegal software that was previously posted using anonymous FTP by the warezmaster
	Warezmaster	Exploits a system bug associated with FTP Server.
User to Root (U2R)		Using such attack, an attacker used to access to a normal user account attempts to exploit system vulnerabilities to gain root access to the system. A class of such attacks are the regular buffer overflows, that are caused by the regular programming mistakes and the environmental assumptions.
	Buffer_overflow	Such type of attacks are designed to trigger the arbitrary code execution using a program and by sending it to more than that it supposed to receive.
	Load Module	Non-stealthy load module attack, that resets IFS for the normal user and that creates a root shell.
	Perl	The perl attack sets a user-ID to root in a perl script and it creates a root shell.
	Root kit	A Multi-day scenario where the user installs one or more components of a rootkit.
Probes		In such attack an attacker who scans network of computers to get information or to find known vulnerabilities. An attacker with the information of map of machines and services over a network that can be used to exploit. There are different types of probes, few of them abuse computer's legitimate features or social engineering techniques. These class of attacks are the most commonly heard, which requires little technical knowledge or expertise.
	Satan	A publicly available tool which probes the network for security vulnerabilities and for misconfigurations.
	Ipsweep	The surveillance sweep is performing either a port sweep or ping on multiple host which identifies the active machine.
	Nmap	Network mapping using nmap tool, identifies active ports on a machine.
	Portsweep	The Surveillance sweep is performing either a port sweep or ping on multiple host address.

DOS and Probing attacks, the R2L, and U2R attacks don't contain any intrusion frequent sequential patterns, because DoS and Probing attacks involve many connections to the host(s) in a short period of time, so the R2L and U2R attacks are embedded with the data portions of packets and normally involve a single connection. Hence, we need some features to detect such attacks, which may look in the data portion for the suspicious behavior, (as an example a number of failed login attempts) are called content features. The third nine (f23-f31) time-based traffic features are designed to capture the properties that mature over a two second temporal window, (as an example the number of connections to the same host over a two second interval). Final fourth ten (f32-f41) host-based traffic features basically utilize a historical window, which are estimated over a number of connections instead of time, such features are designed to access attacks, with span intervals are longer than 2 seconds [30]. There are 41 feature attributes for each connection record plus one class label, out of which 38 are numeric and 3 are symbolic. Symbolic attributes are protocol type, service, and flag. Discrete data can be numeric but it can also be categorical, continuous data are always numeric and are not restricted to define separate values and can take any value within a range [31]. The various types of attack with their categories are discussed in the table [32]. The severity of the attack and the classification is discussed.

There are 125973 number of records of the dataset is; out of which 53.48 percent are normal and 46.52 percent of records are of intrusive types. There are 24 different types of attack which can be mainly classified into four categories, such as Danial of Services (DoS), Remote to Local (R2L), User to Root (U2R), and Probing.

Table 2: Different Types of Attack and their Class Occurrences

Class Type		Number of Instances	% of Attack Type	% of Attack Class	% of Total Class Occurences
DoS		45927	-	78.33	36.45
1	neptune	41214	89.74	70.3	32.72
2	teardrop	892	1.94	1.52	0.71
3	smurf	2646	5.76	4.51	2.1
4	pod	201	0.44	0.34	0.16
5	back	956	2.08	1.63	0.76
6	land	18	0.04	0.03	0.01
Probes		11656	-	19.88	9.25
1	ipsweep	3599	30.88	6.14	2.86
2	portsweep	2931	25.15	5	2.33
3	nmap	1493	12.81	2.55	1.19
4	satan	3633	31.17	6.2	2.88
R2L		995	-	1.7	0.78
1	warezclient	890	89.45	1.52	0.71
2	guess_passwd	53	5.33	0.09	0.04
3	ftp_write	8	0.8	0.01	0.01
4	multihop	7	0.7	0.01	0.01
5	imap	11	1.11	0.02	0.01
6	warezmaster	20	2.01	0.03	0.02
7	Phf	4	0.4	0.01	0
8	Spy	2	0.2	0	0
U2R		52		0.09	0.04
1	rootkit	10	19.23	0.02	0.01
2	buffer_overflow	30	57.69	0.05	0.02
3	loadmodule	9	17.31	0.02	0.01
4	Perl	3	5.77	0.01	0
Total Attacks		58630	-	-	46.52
Normal		67343	-	-	53.48
Total Instances		125973			100

In the intrusive data set around 36.45 percent data are of DoS type, about 9.25 percent data are of Probes type, 0.78 percent of R2L type and 0.04 percent of U2R type.

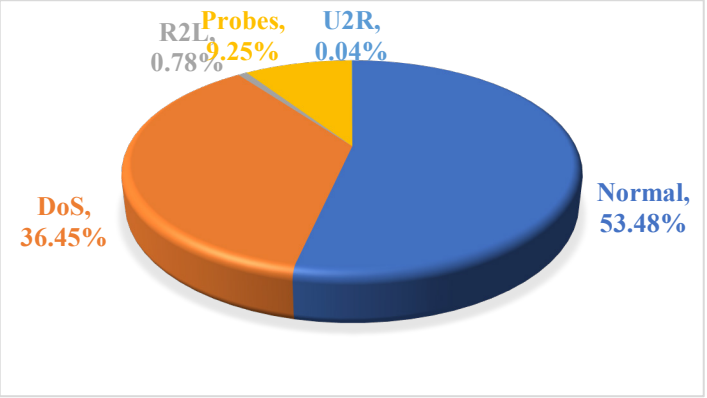


Figure 1: Percentage of Instances of Various Attack Classes

The figure describes various attack categories, around 78.33 percent of data are of DoS type out of which 70.3 percent are of Neptune type, 1.52 percent teardrop, 4.51 percent smurf, 0.34 percent pod, 1.63 percent pod and 0.03 percent of land type attacks. There are 19.88 percent Probes type, where 6.14 percent are ipsweep, 5 percent are portsweep, 2.55 percent are nmap and 6.2 percent are satan type of attack. Nearly 1.7 percent are R2L type of attack, in which 1.52 percent are warezclient, 0.09 percent are guess_passwd, 0.01 percent are ftp_write, 0.01 percent are multihop, 0.02 percent are imap, 0.03 percent are warezmaster, 0.01 percent are phf, and 0.0 percent are spy. In the 0.09 percent of U2R attack, there are 0.02 percent rootkit, 0.05 percent are buffer_overflow 0.05 percent are loadmodule, and there are 0.01 percent perl type of attack.We have applied the rank based feature selection method to compute the rank of the features and the order of the features based on their rank is as follows. The order of the features varies with the different feature selection approach. The rank based feature selection approach applied on the various set of data set of NSL KDD Cup data set is as follows. The rank of the features varies with the

change of rank based feature selection approach applied on various classes of data. The order of the features varies based on their rank values and feature selection mechanism. The rank for different features varies based on their contribution in the selection process is mentioned below. The rank value signify the contribution of the feature to the different class of data.

With the application of rank based feature selection approach, we found the rank value of the features in the above mentioned table is '0' (Null), So these features do not contribute in the process of intrusion classification. The rest of the features mentioned in the below table are responsible in the classification of the Intrusion data.

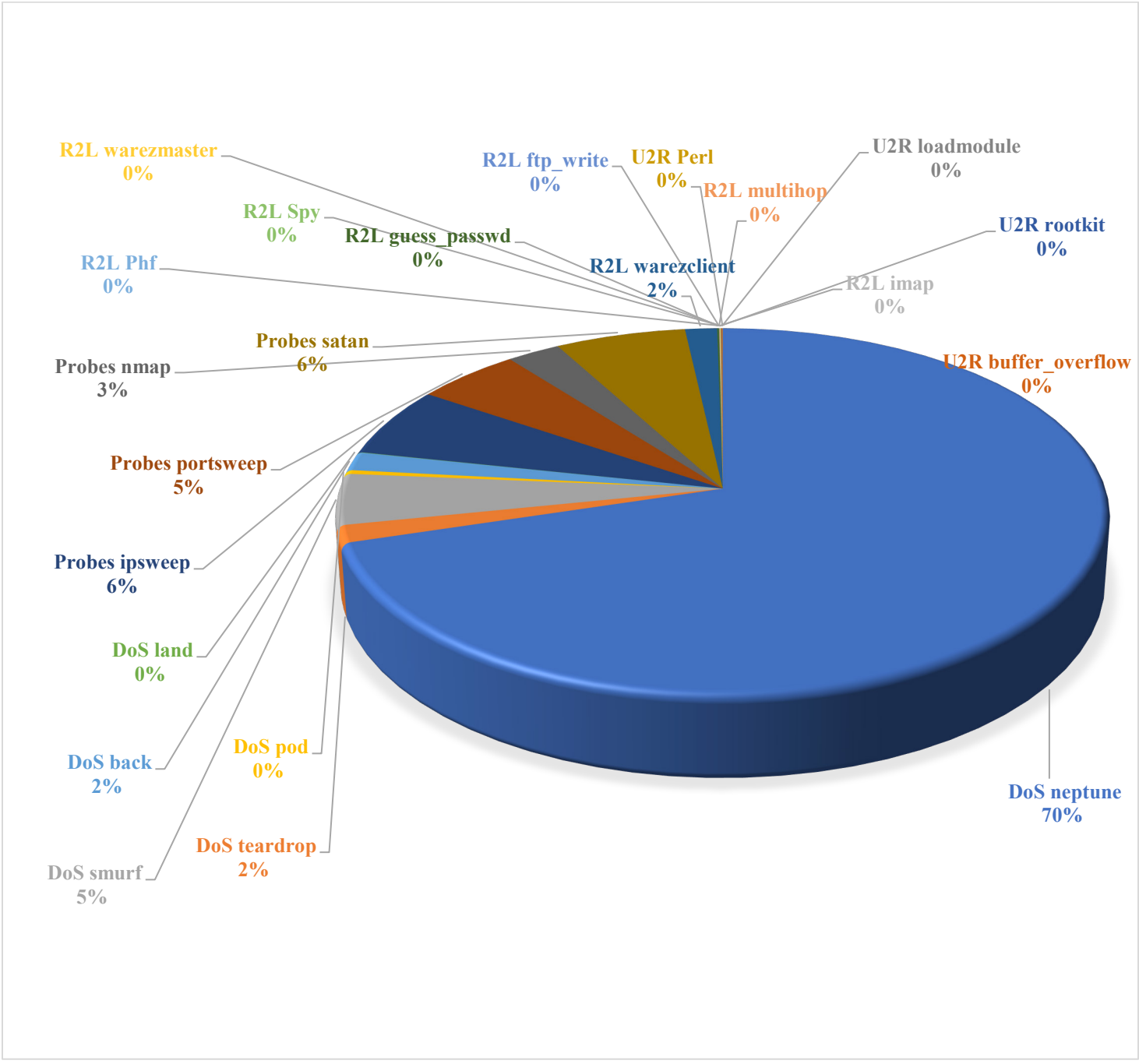


Figure 2: Percentage of Instances of Various Attack Types

Table 3: Number of features do not Contribute in Selection Process

Feature Selection Approach	Number of Features	Feature Selection	Name of Feature	Rank Value
NSL KDD'99 (All)	2	f20, f21	Nu_ob, Is_ho_lg	0
NSL KDD'99 (DOS)	11	f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	N_f_login, Nu_ob, Is_gu_lg, N_shell, Nu_ac_fl, Num_f_cr, R_shell, Urgent, Su_attem, Num_roo, Is_ho_lg	0
NSL KDD'99 (PROBES)	14	f7, f20, f9, f22, f8, f19, f18, f11, f17, f16, f15, f14, f13, f21	Land, Nu_ob, Urgent, Is_gu_lg, W_frag, Nu_ac_fl, N_shell, N_f_login, Num_f_cr, Num_roo, Su_attem, R_shell, Num_com, Is_ho_lg	0
NSL KDD'99 (R2L)	8	f29, f2, f20, f30, f15, f7, f8, f21	Sa_srv_rt, Pro_type, Nu_ob, Di_srv_rt, Su_attem, Land, W_frag, Is_ho_lg,	0
NSL KDD'99 (U2R)	27	f15, f14, f13, f41, f11, f9, f8, f7, f21, f19, f33, f30, f31, f37, f20, f38, f39, f29, f28, f27, f26, f40, f22, f23, f24, f25, f1	Su_attem, R_shell, Num_com, D_hsrr, N_f_login, Urgent, W_frag, Land, Is_ho_lg, Nu_ac_fl, Ds_ho_sr, Di_srv_rt, Sr_di_h0, Ds_d_h_rt, Nu_ob, D_h_sr, Ds__h_r, Sa_srv_rt, Sr_rr_rt, Rer_rt, Se_se_rt, Ds_hrr, Is_gu_lg, Count, Sev_coun, Ser_rate, Duration	0

Table 4 : Number of Features do not Contribute in DOS Attack

	Features do not Contribute			Features Contribute			
	Number	Features	Value	Number	Features	Name	Key Feature
NSL KDD'99 (All)	2	f20, f21	0	39	f5, f4, f2, f3, f29, f36, f30, f24, f35, f34, f23, f33, f38, f39, f25, f26, f40, f8, f41, f12, f6, f10, f13, f32, f28, f27, f31, f37, f7, f1, f11, f22, f18, f19, f17, f14, f9, f15, f16	-	0
NSL KDD'99 (DOS)	11	f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	0	30	f5, f4, f2, f3, f29, f36, f30, f24, f35, f34, f23, f33, f38, f39, f25, f26, f40, f8, f41, f12, f6, f10, f13, f32, f28, f27, f31, f37, f7, f1	-	0
NSL KDD'99 (DOS-neptune)	15	f8, f10, f13, f7, f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	0	26	f5, f4, f2, f3, f29, f36, f30, f24, f35, f34, f23, f33, f38, f39, f25, f26, f40, f41, f12, f6, , f32, f28, f27, f31, f37	-	0
NSL KDD'99 (DOS-teardrop)	21	f39, f26, f41, f12, f10, f13, f28, f31, f37, f7, f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	0	20	f5, f4, f2, f3, f29, f36, f30, f24, f35, f34, f23, f33, f38, f25, , f40, f8, , f6, f32, f27, , f1	Src_bytes	f5
NSL KDD'99 (DOS-smurf)	26	f30, f39, f25, f26, f8, f41, f12, f6, f10, f13, f28, f27, f31, f37, f7, f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	0	15	f5, f4, f2, f3, f29, f36, f30, f24, f35, f34, f23, f33, f38, f39, f25, f26, f40, f8, f41, f12, f6, f10, f13, f32, f28, f27, f31, f37, f7, f1, f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	Same_srv_rate	f29
NSL KDD'99 (DOS-pod)	23	f30, f39, f25, f26, f41, f12, f6, f10, f13, f28, f27, f7, f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	0	18	f5, f4, f2, f3, f29, f36, f24, f35, f34, f23, f33, f38, , f40, f8, f32, f31, f37, f1,	Same_srv_rate	f29
NSL KDD'99 (DOS-back)	15	f35, f8, f37, f7, f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	0	26	f5, f4, f2, f3, f29, f36, f30, f24, f34, f23, f33, f38, f39, f25, f26, f40, f41, f12, f6, f10, f13, f32, f28, f27, f31, f1	Logged_in	f12

NSL KDD'99 (DOS-land)	19	f5, f8, f41, f12, f6, f10, f13, f28, f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	0	22	f4, f2, f3, f29, f36, f30, f24, f35, f34, f23, f33, f38, f39, f25, f26, f40, f32, f27, f31, f37, f7, f1	Land	f7

Table 5: Number of Features do not Contribute in U2R Attack

Feature Selection Approach	Features do not Contribute			Features Contribute			
	Number	Features	Value	Number	Features	Key Feature	Name
NSL KDD'99 (All)	2	f21, f20	0	39	f35, f34, f32, f36, f18, f17, f6, f5, f10, f16, f12, f3, f2, f4, f15, f14, f13, f41, f11, f9, f8, f7, f19, f33, f30, f31, f37, f38, f39, f29, f28, f27, f26, f40, f22, f23, f24, f25, f1	-	Nu_ob, Is_ho_lg
NSL KDD'99 (U2R)	27	f15, f14, f13, f41, f11, f9, f8, f7, f21, f19, f33, f30, f31, f37, f20, f38, f39, f29, f28, f27, f26, f40, f22, f23, f24, f25, f1	0	14	f35, f34, f32, f36, f18, f17, f6, f5, f10, f16, f12, f3, f2, f4	f12, f29	Logged_in, Same_srv_rate
NSL KDD'99 (U2R-buffer_overflow)	15	f35, f18, f15, f11, f9, f8, f7, f21, f19, f31, f20, f38, f39, f26, f22,	0	26	f34, f32, f36, f17, f6, f5, f10, f16, f12, f3, f2, f4, f14, f13, f41, f33, f30, f37, f29, f28, f27, f40, f23, f24, f25, f1	f12	Logged_in
NSL KDD'99 (U2R_loadmodule)	16	f15, f11, f9, f8, f7, f21, f31, f20, f38, f39, f28, f27, f26, f40, f22, f25	0	25	f35, f34, f32, f36, f18, f17, f6, f5, f10, f16, f12, f3, f2, f4, f14, f13, f41, f19, f33, f30, f37, f29, f23, f24, f1	f12	Logged_in
NSL KDD'99 (U2R_perl)	21	f10, f15, f13, f41, f11, f9, f8, f7, f21, f19, f30, f31, f37, f20, f38, f39, f28, f27, f26, f22, f25	0	20	f35, f34, f32, f36, f18, f17, f6, f5, f16, f12, f3, f2, f4, f14f33, f29, f40, f23, f24, f1	f12	Logged_in
NSL KDD'99 (U2R_rootkit)	15	f18, f15, f8, f7, f21, f19, f30, f31, f20, f38, f28, f27, f26, f22, f25	0	26	f35, f34, f32, f36, f17, f6, f5, f10, f16, f12, f3, f2, f4, f14, f13, f41, f11, f9, f33, f37, f39, f29, f40, f23, f24, f1	f29	Same_srv_rate

Table 6: Number of Features do not Contribute in R2L Attack

Feature Selection Approach	Features do not Contribute			Features Contribute			
	Number	Features	Value	Number	Features	Key Feature	Name
NSL KDD'99 (All)	2	f20, f21	0	39	f6, f5, f3, f12, f39, f28, f4, f10, f11, f38, f40, f41, f36, f27, f1, f32, f33, f35, f37, f34, f19, f31, f26, f25, f24, f17, f23, f14, f22, f16, f13, f18, f9, f29, f2, f30, f15, f7, f8	-	-
NSL KDD'99 (R2L)	8	f29, f2, f20, f30, f15, f7, f8, f21	0	33	f6, f5, f3, f12, f39, f28, f4, f10, f11, f38, f40, f41, f36, f27, f1, f32, f33, f35, f37, f34, f19, f31, f26, f25, f24, f17, f23, f14, f22, f16, f13, f18, f9	-	-
NSL KDD'99 (R2L-ftp_write)	17	f39, f28, f11, f38, f40, f41, f27, f26, f25, f14, , f18, f20, f30, f15, f7, f8, f21	0	24	f6, f5, f3, f12f4, f10, f36, f1, f32, f33, f35, f37, f34, f19, f31, f24, f17, f23, f22, f16, f13, f9, f29, f2	f29	Same_srv_rate

NSL KDD'99 (R2L_guess_passwd)	15	f35, f19, f31, f17, f14, f16, f13, f18, f9, f20, f30, f15, f7, f8, f21	0	26	f6, f5, f3, f12, f39, f28, f4, f10, f11, f38, f40, f41, f36, f27, f1, f32, f33, f37, f34, f26, f25, f24, f23, f22, f29, f2, f20	f29	Same_srv_rate
NSL KDD'99 (R2L_imap)	16	f11, f41, f27, f37, f19, f17, f14, f22, f18, f9, f20, f30, f15, f7, f8, f21	0	25	f6, f5, f3, f12, f39, f28, f4, f10, f38, f40, f36, f1, f32, f33, f35, f34, f31, f26, f25, f24, f23, f16, f13, f29, f2	f29	Same_srv_rate
NSL KDD'99 (R2L_multihop)	16	f39, f28, f11, f38, f41, f27, f37, f26, f25, f9, f20, f30, f15, f7, f8, f21	0	25	f6, f5, f3, f12, f4, f10, f40, f36, f1, f32, f33, f35, f34, f19, f31, f24, f17, f23, f14, f22, f16, f13, f18, f29, f2	f29	Same_srv_rate
NSL KDD'99 (R2L_phf)	22	f39, f11, f38, f40, f41, f36, f27, f37, f26, f25, f17, f22, f16, f13, f18, f9, f20, f30, f15, f7, f8, f21	0	19	f6, f5, f3, f12, f28, f4, f10, f1, f32, f33, f35, f34, f19, f31, f24, f23, f14, f29, f2	f29	Same_srv_rate
NSL KDD'99 (R2L_spy)	21	f28, f10, f11, f40, f41, f36, f27, f37, f31, f26, f25, f14, f22, f16, f13, f9, f20, f30, f7, f8, f21	0	20	f6, f5, f3, f12, f39, f4, , f38, f1, f32, f33, f35, f34, f19, f24, f17, f23, f18, f29, f2, f15	f29	Same_srv_rate
NSL KDD'99 (R2L_warezclient)	13	f11, f19, f17, f14, f16, f13, f18, f9, f20, , f15, f7, f8, f21	0	28	f6, f5, f3, f12, f39, f28, f4, f10, f38, f40, f41, f36, f27, f1, f32, f33, f35, f37, f34, f31, f26, f25, f24, f23, f22, , f29, f2, f30	f12	Logged_in
NSL KDD'99 (R2L_warezmaster)	21	f39, f28, f11, f41, f27, f37, f19, f31, f26, f25, f14, f16, f13, f18, f9, f20, f30, f15, f7, f8, f21	0	20	f6, f5, f3, f12, f4, f10, , f38, f40, f36, f1, f32, f33, f35, f34, f24, f17, f23, f22, f29, f2	f29	Same_srv_rate

Table 7: Number of features do not Contribute in PROBES Attack

Feature Selection Approach	Features do not Contribute			Features Contribute			
	Number	Features	Value	Number	Features	Key Feature	Value
NSL KDD'99 (All)	2	f20, f21	0	39	f32, f35, f34, f37, f23, f27, f4, f40, f5, f2, f3, f30, f29, f36, f33, f31, f41, f28, f24, f38, f25, f1, f26, f39, f6, f12, f10, f7, f9, f22, f8, f19, f18, f11, f17, f16, f15, f14, f13,	-	-
NSL KDD'99 (PROBES)	14	f7, f20, f9, f22, f8, f19, f18, f11, f17, f16, f15, f14, f13, f21	0	27	f32, f35, f34, f37, f23, f27, f4, f40, f5, f2, f3, f30, f29, f36, f33, f31, f41, f28, f24, f38, f25, f1, f26, f39, f6, f12, f10	-	-
NSL KDD'99 (PROBES_nmap)	20	f27, f40, f41, f28, f12, f10, f7, f20, f9, f22, f8, f19, f18, f11, f17, f16, f15, f14, f13, f21	0	21	f32, f35, f34, f37, f23, f4, f5, f2, f3, f30, f29, f36, f33, f31, f24, f38, f25, f1, f26, f39, f6	-	-
NSL KDD'99 (PROBES_ipsweep)	13	f26, f10, f7, f20, f9, f22, f8, f19, f18, f11, f15, f14, f21	0	28	f32, f35, f34, f37, f23, f27, f4, f40, f5, f2, f3, f30, f29, f36, f33, f31, f41, f28, f24, f38, f25, f1, f39, f6, f12f17, f16, f13	-	-
NSL KDD'99 (PROBES_portsweep)	14	f7, f20, f9, f22, f8, f19, f18, f11, f17, f16, f15, f14, f13, f21	0	27	f32, f35, f34, f37, f23, f27, f4, f40, f5, f2, f3, f30, f29, f36, f33, f31, f41, f28, f24, f38, f25, f1, f26, f39, f6, f12, f10	-	-
NSL KDD'99 (PROBES_satan)	9	f7, f20, f9, f8, f19, f18, , f15, f14, f21	0	32	f32, f35, f34, f37, f23, f27, f4, f40, f5, f2, f3, f30, f29, f36, f33, f31, f41, f28, f24, f38, f25, f1, f26, f39, f6, f12, f10, f22, f11, f17, f16f13	-	-

5. Experimental Setup

We have tested the various machine learning methods on KDDCUP'99 dataset. In this experiment we have used a computing environment of core i7 processor, 2.6 GHz, 8 GB RAM, 1TB hard disk and windows 10 (64 bit) operating system. The various tree based classifiers have been used to classify different types of DOS, Probes, U2R and R2L attacks.

Table 8: Classification of Various DOS Attacks Using Different Classifiers

Sl. No.	Classifier	Classified Percentage	Unclassified Percentage
1	BFT	99.9935	0.0065
2	CDT	99.9956	0.0044
3	FPA	99.9913	0.0087
4	FT	99.9978	0.0022
5	HT	99.9673	0.0327
6	J48	99.9826	0.0174
7	J48C	99.9935	0.0065
8	J48G	99.9804	0.0196
9	LADT	99.9826	0.0174
10	LMT	99.9978	0.0022
11	NBT	99.9956	0.0044
12	RF	99.9913	0.0087
13	RT	99.9760	0.0240
14	REPT	99.9956	0.0044
15	SF	99.9826	0.0174
16	Min	99.9673	0.0022
17	Max	99.9978	0.0327
18	Avg	99.9882	0.0118

The result shows that FT and LMT has classified about 99.9978 percent. The performance of the tree based classifier has been observed on the KDD CUP'99 DOS attack type dataset. The classified result has been plotted in the below mentioned graph.

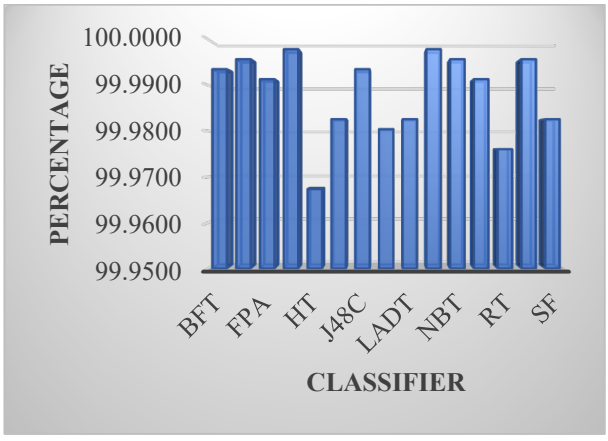


Figure 3: The Classification of Various Attack Type of DOS Attack

We have observed the result of the various classifiers applied on the rank-based feature selection methods. Information Gain (IG), Gain Ratio (GR) and Symmetrical Uncertainty (SU) feature selection methods have been applied to measure the performance and classification of various attack type. The accuracy and classification rate of few tree based classification algorithm is very high. The performance of the various algorithms and the classification percentage is reflected in the plotted graphs.

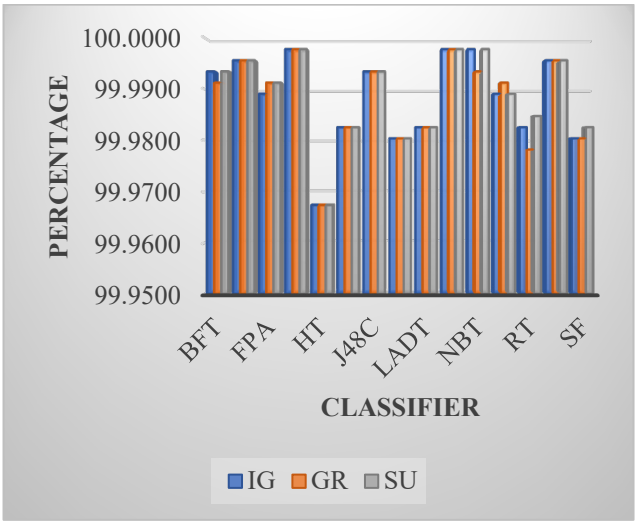


Figure 4: The Classification of Various Attack Type of DOS Attack Using Feature Selection

Further in the rank-based feature selection classification process we have applied a range of features to monitor the change in classification process. We have selected a number of features from the set of 41 features in the dataset and found there are 30 number of features participating or contributing in the classification process. We have applied various tree based classification algorithms to classify the NSL-KDDCUP'99 (DoS) Dataset and found the rank of the features are contributing in the classification process. There are 11 number of features whose rank value is null, which are not participating in the classification process. The performance of the J48 algorithms and the classification percentage is reflected in the plotted graphs.

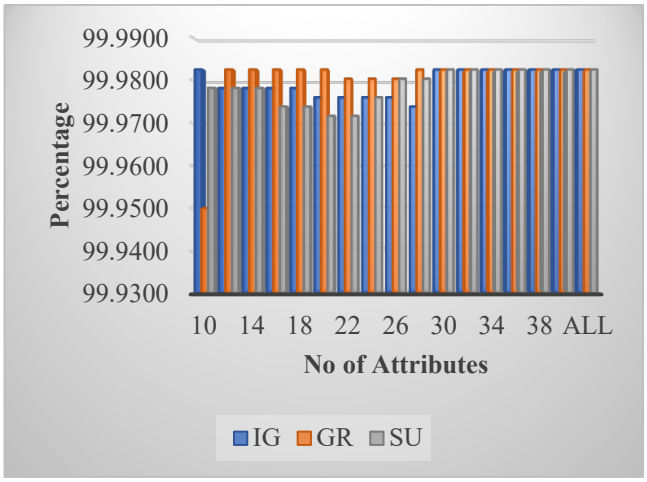


Figure 5: The Classification of Various Attack Type of DOS Attack Using Feature Selection and Attribute Selection

We have applied tree based feature selection classification technique on the NSL KDD'99 data set, The null valued redundant features are eliminated and the classification technique is applied to compute the percentage of accuracy of different attack types.

We have applied the J48, Random Forest and Functional Trees which are one the best decision tree based classification algorithm classifies the optimum accuracy of the different types of attack.

The below mentioned features are eliminated and the other non-null features are used to classify the different classes of intrusive data. The percentage of classification of different classes of attack are mentioned below.

The graph reflects the percentage of classification of different types of DOS attacks. The redundant null valued features are eliminated and the most suited tree-based algorithms are applied to find the optimum classified result.

The graph reflects the percentage of total classification of various DOS attacks. The redundant null valued features are eliminated and the most suited tree based algorithms are applied to find the optimum classified result.

In the classification process there are a number of key features which plays an important role to classify and determine the various attack types in the class of intrusive dataset. We have applied the same algorithms to determine the percentage of classification of various attack types using the key features.

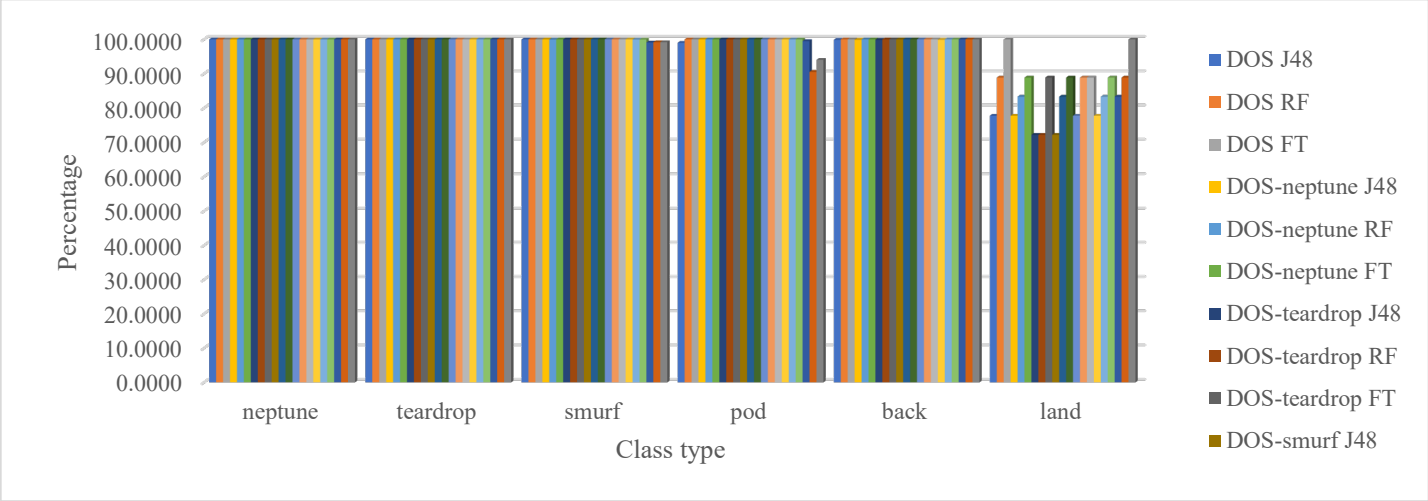


Figure 6: The Classification of Various Attack Class Types of DOS Attack by Removing Redundant Features

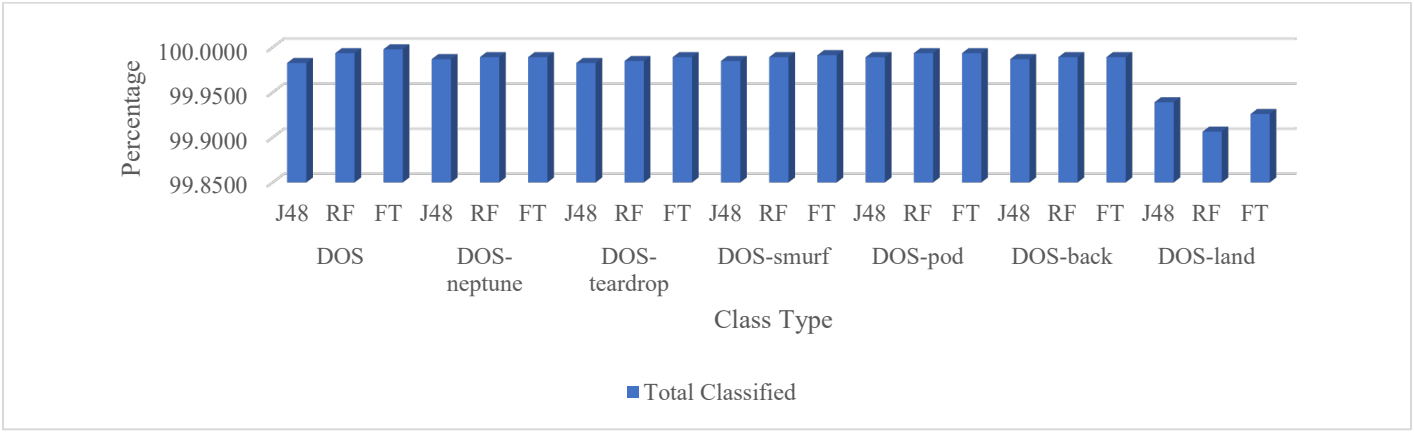


Figure 7: The Total Classification of Various Attack Class Ty pes of DOS Attack by Removing Redundant Features

Table 9: Classification of Various DOS Attacks Using Feature Reduction of Rank Based Feature Selection Classifiers

Sl. No	Class Type	Removed Features	Algorithm Used	Total No of Instances	Total Classified	Total Unclassified
1	NSL KDD'99 (DOS)	f11,f20, f22, f18, f19, f17 , f14, f9, f15, f16, f21	J48	45927	99.9826	0.0174
			RF	45927	99.9935	0.0065
			FT	45927	99.9978	0.0022
2	NSL KDD'99 (DOS-neptune)	f8, f10, f13, f7, f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	J48	41214	99.9869	0.0131
			RF	41214	99.9891	0.0109
			FT	41214	99.9891	0.0109
3	NSL KDD'99 (DOS-teardrop)	f39, f26, f41, f12, f10,	J48	892	99.9826	0.0174
			RF	892	99.9848	0.0152

		f13, f28, f31, f37, f7, f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	FT	892	99.9891	0.0109
4	NSL KDD'99 (DOS-smurf)	f30, f39, f25, f26, f8, f41, f12, f6, f10, f13, f28, f27, f31, f37, f7, f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	J48	2646	99.9848	0.0152
			RF	2646	99.9891	0.0109
			FT	2646	99.9913	0.0087
5	NSL KDD'99 (DOS-pod)	f30, f39, f25, f26, f41, f12, f6, f10, f13, f28, f27, f7, f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	J48	201	99.9891	0.0109
			RF	201	99.9935	0.0065
			FT	201	99.9935	0.0065
6	NSL KDD'99 (DOS-back)	f35, f8, f37, f7, f11, f20, f22, f18, f19,f17, f14, f9, f15, f16, f21	J48	956	99.9869	0.0131
			RF	956	99.9891	0.0109
			FT	956	99.9891	0.0109
7	NSL KDD'99 (DOS-land)	f5, f8, f41, f12, f6, f10, f13, f28, f11, f20, f22, f18, f19, f17, f14, f9, f15, f16, f21	J48	18	99.9390	0.0610
			RF	18	99.9064	0.0936
			FT	18	99.9260	0.0740

Table 10: Classification of Various DOS Attacks Based on Key Feature Selection of Rank Based Feature Selection Classifiers

Sl. No	Class Type	Selected Features	Key Feature	Algorithm Used	Total No of Instances	Total Classified	Total Unclassified
1	NSL KDD'99 (DOS)	f5, f4, f2, f3, f29, f36, f30, f24, f35, f34, f23, f33, f38, f39, f25, f26, f40, f8, f41, f12, f6, f10, f13, f32, f28, f27, f31, f37, f7, f1	_	J48	45927	99.9826	0.0174
				RF	45927	99.9935	0.0065
				FT	45927	99.9978	0.0022
2	NSL KDD'99 (DOS-neptune)	f5, f4, f2, f3, f29, f36, f30, f24, f35, f34, f23, f33, f38, f39, f25, f26, f40, f8, f41, f12, f6, f10, f13, f32, f28, f27, f31, f37, f7, f1	_	J48	41214	99.9869	0.0131
				RF	41214	99.9891	0.0109
				FT	41214	99.9891	0.0109
3	NSL KDD'99 (DOS-teardrop)	f5, f4, f2, f3, f1	f5	J48	892	99.9586	0.0414
				RF	892	99.9586	0.0414
				FT	892	99.9586	0.0414
4	NSL KDD'99 (DOS-smurf)	f4, f2, f3, f1, f29	f29	J48	2646	99.5275	0.4725
				RF	2646	99.5275	0.4725
				FT	2646	99.5275	0.4725
5	NSL KDD'99 (DOS-pod)	f4, f2, f3, f1, f29	f29	J48	201	99.5275	0.4725
				RF	201	99.5275	0.4725
				FT	201	99.5275	0.4725
6	NSL KDD'99 (DOS-back)	f4, f2, f3, f1, f12	f12	J48	956	99.5275	0.4725
				RF	956	99.5275	0.4725
				FT	956	99.5275	0.4725
7	NSL KDD'99 (DOS-land)	f4, f2, f3, f1, f7	f7	J48	18	99.5275	0.4725
				RF	18	99.5645	0.4355
				FT	18	99.5667	0.4333

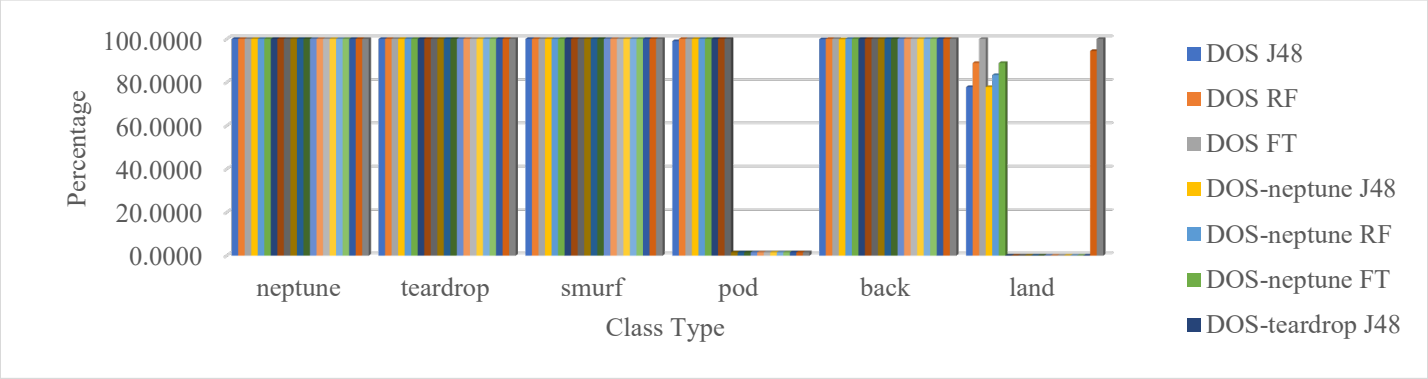


Figure 8: The Classification of Various Attack Class Types of DOS Attack by Selecting Key Features

The graph reflects the percentage of classification of various DOS attacks. The appropriate key features for DOS attack are applied and the redundant null valued features are eliminated then the most suited tree based algorithms are applied to find the optimum classified result.

The graph reflects the percentage of total classification of various DOS attacks. The appropriate key features for DOS attack are applied and the redundant null valued features are eliminated then the most suited tree based algorithms are applied to find the optimum classified result.

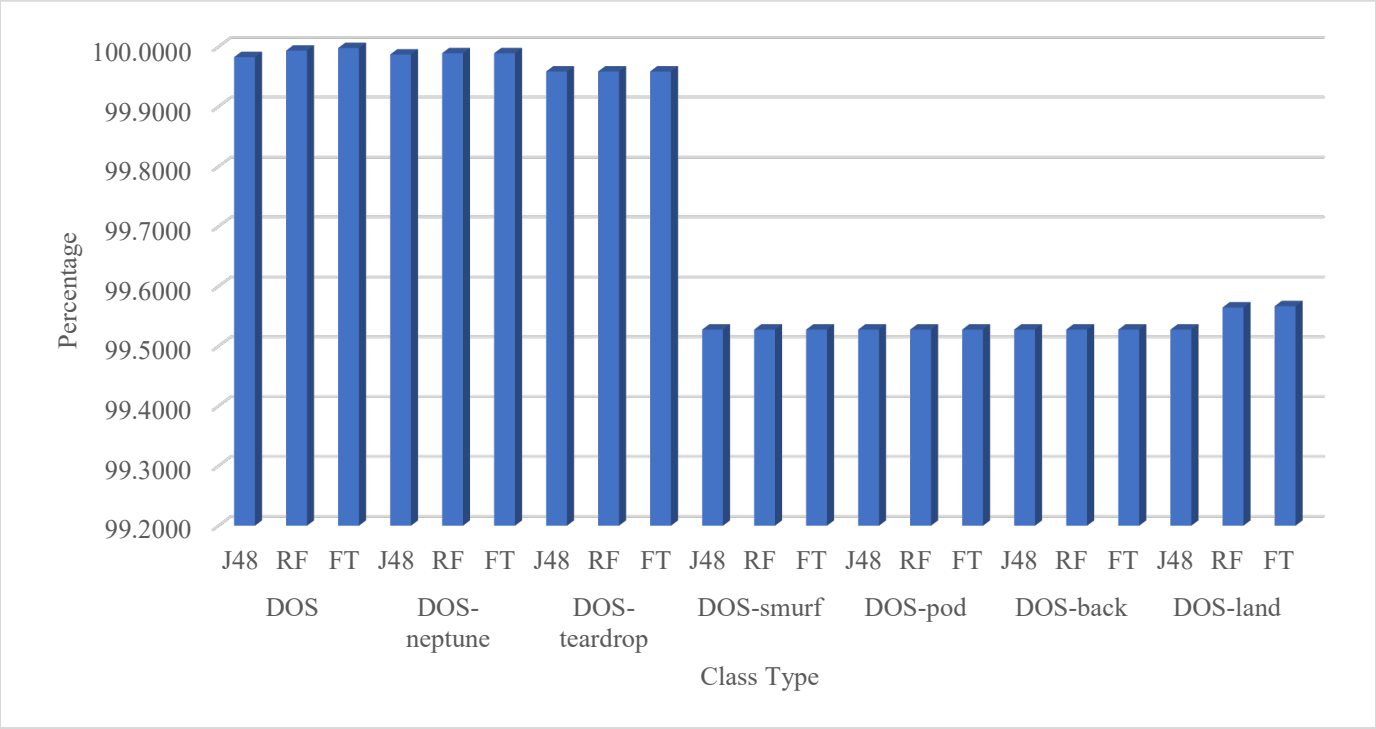


Figure 9: The Total Classification of Various Attack Class Types of DOS Attack by Selecting Key Features

Table 11: Classification of Various PROBES Attacks Using Feature Reduction of Rank Based Feature Selection Classifiers

Sl. No	Class Type	Removed Features	Algorithm Used	Total No of Instances	Total Classified	Total Unclassified
1	NSL KDD'99 (PROBES)	f7, f20, f9, f22, f8, f19, f18, f11, f17, f16, f15, f14, f13, f21	J48	11656	99.5281	0.4719
			RF	11656	99.6826	0.3174
			FT	11656	99.4852	0.5148
2	NSL KDD'99 (PROBES_ nmap)	f27, f40, f41, f28, f12,f10, f7, f20, f9, f22, f8, f19,f18, f11, f17, f16, f15, f14, f13, f21	J48	1493	99.5281	0.4719
			RF	1493	99.6482	0.3518
			FT	1493	99.5367	0.4633
3	NSL KDD'99 (PROBES_ ipsweep)	f26, f10, f7, f20, f9, f22, f8, f19, f18, f11, f15, f14, , f21	J48	3599	99.5281	0.4719
			RF	3599	99.6654	0.3346
			FT	3599	99.4595	0.5405
4	NSL KDD'99 (PROBES_ portsweep)	f7, f20, f9, f22, f8, f19, f18, f11, f17, f16, f15, f14, f13, f21	J48	2931	99.5281	0.4719
			RF	2931	99.6740	0.3260
			FT	2931	99.4852	0.5148
5	NSL KDD'99 (PROBES_ satan)	f7, f20, f9, f8, f19, f18, f15, f14, f21	J48	3633	99.5281	0.4719
			RF	3633	99.6740	0.3260
			FT	3633	99.4852	0.5148

The graph reflects the percentage of classification of different types of PROBES attacks. The redundant null valued features are eliminated and the most suited tree based algorithms are applied to find the optimum classified result.

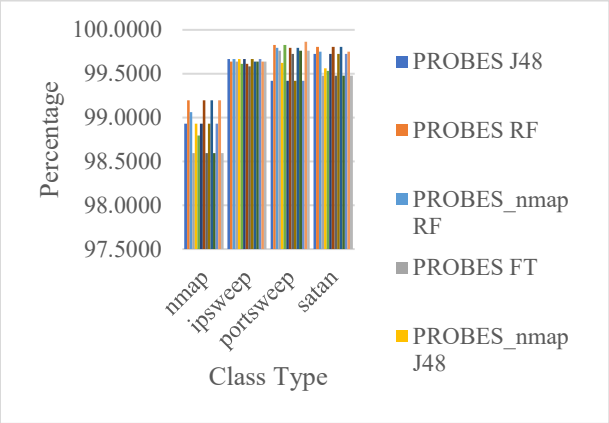


Figure 10: The Classification of Various Attack Class Types of PROBES Attack by Removing Redundant Features

The graph reflects the percentage of total classification of various PROBES attacks. The redundant null valued features are eliminated and the most suited tree based algorithms are applied to find the optimum classified result.

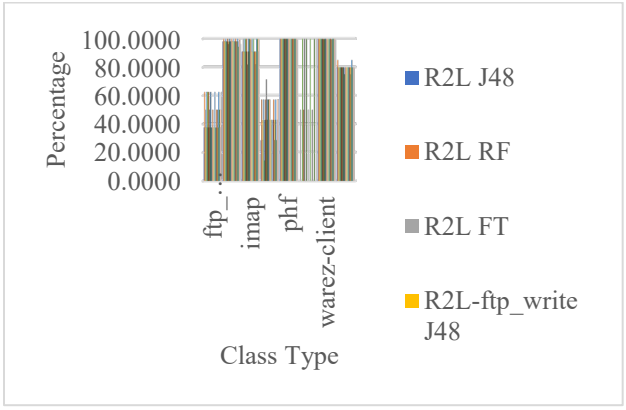


Figure 11: The Total Classification of Various Attack Class Types of PROBES Attack by Removing Redundant Features

Table 12: Classification of Various R2L Attacks Using Feature Reduction of Rank Based Feature Selection Classifiers

Sl. No	Class Type	Removed Features	Algorithm Used	Total No of Instances	Total Classified	Total Unclassified
1	NSL KDD'99 (R2L)	f29, f2, f20, f30, f15, f7, f8, f21	J48	995	98.0905	1.9095
			RF	995	98.7940	1.2060
			FT	995	98.4925	1.5075
2	NSL KDD'99 (R2L-ftp_ write)	f39, f28, f11, f38, f40, f41, f27, f26, f25, f14, , f18, f20, f30, f15, f7, f8, f21	J48	8	98.2915	1.7085
			RF	8	98.7940	1.2060
			FT	8	98.5930	1.4070
3	NSL KDD'99 (R2L_guess_ passwd)	f35, f19, f31, f17, f14, f16, f13, f18, f9, f20,f30, f15, f7, f8, f21	J48	53	98.1910	1.8090
			RF	53	98.8945	1.1055
			FT	53	98.4925	1.5075
4	NSL KDD'99 (R2L_ imap)	f11, f41, f27, f37, f19, f17, f14, f22, f18, f9, f20, f30, f15, f7, f8, f21	J48	11	98.0905	1.9095
			RF	11	98.7940	1.2060
			FT	11	98.5930	1.4070
5	NSL KDD'99 (R2L_ multihop)	f39, f28, f11, f38, f41, f27, f37, f26, f25, f9, f20, f30, f15, f7, f8, f21	J48	7	98.1910	1.8090
			RF	7	98.6935	1.3065
			FT	7	97.9899	2.0101
6	NSL KDD'99 (R2L_phf)	f39, f11, f38, f40, f41, f36, f27, f37, f26, f25, f17, f22, f16, f13, f18, f9, f20, f30, f15, f7, f8, f21	J48	4	98.1910	1.8090
			RF	4	98.6935	1.3065
			FT	4	98.1910	1.8090
7	NSL KDD'99 (R2L_spy)	f28, f10, f11, f40, f41, f36, f27, f37, f31, f26, f25, f14, f22, f16, f13, , f9, f20, f30, f7, f8, f21	J48	2	98.2915	1.7085
			RF	2	98.5930	1.4070
			FT	2	97.9899	2.0101
8	NSL KDD'99 (R2L_ warezclient)	f11, f19, f17, f14, f16, f13, f18, f9, f20, , f15, f7, f8, f21	J48	890	98.0905	1.9095
			RF	890	98.8945	1.1055
			FT	890	98.3920	1.6080
9	NSL KDD'99 (R2L_w arezmaster)	f39, f28, f11, f41, f27, f37, f19, f31, f26, f25, , f14, f16, f13, f18, f9, f20, f30, f15, f7, f8, f21	J48	20	98.1910	1.8090
			RF	20	98.6935	1.3065
			FT	20	98.5930	1.4070

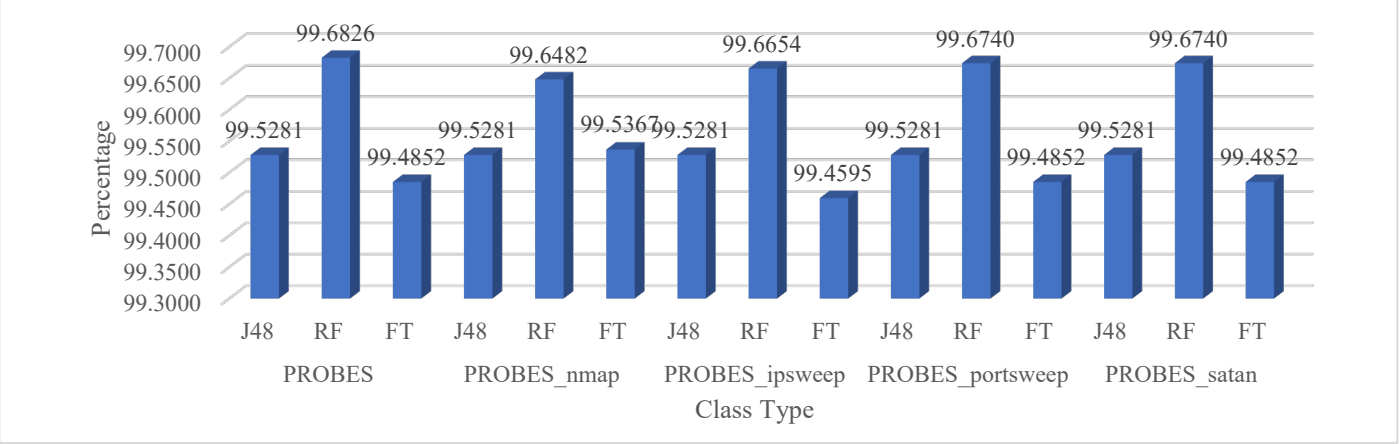


Figure 12: The Classification of Various Attack Class Types of R2L Attack by Removing Redundant Features

The graph reflects the percentage of classification of different types of R2L attacks. The redundant null valued features are eliminated and the most suited tree based algorithms are applied to find the optimum classified result.

The graph reflects the percentage of total classification of various PROBES attacks. The redundant null valued features are eliminated and the most suited tree based algorithms are applied to find the optimum classified result.

Table 13: Classification of Various R2L Attacks Based on Key Feature Selection of Rank Based Feature Selection Classifiers

Sl. No	Class Type	Selected Features	Key Feature	Algorithm Used	Total No of Instances	Total Classified	Total Unclassified
1	NSL KDD'99 (R2L)	f6, f5, f3, f12, f39, f28, f4, f10, f11, f38, f40, f41, f36, f27, f1, f32, f33, f35, f37, f34, f19, f31, f26, f25, f24, f17, f23, f14, f22, f16, f13, f18, f9		J48	995	98.0905	1.9095
				RF	995	98.7940	1.2060
				FT	995	98.4925	1.5075
2	NSL KDD'99 (R2L-ftp_write)	f3, f4, f1, f29, f2	f29	J48	8	98.0905	1.9095
				RF	8	98.4925	1.5075
				FT	8	96.8844	3.1156
3	NSL KDD'99 (R2L_guess_ passwd)	f3, f4, f1, f29, f2	f29	J48	53	98.0905	1.9095
				RF	53	98.4925	1.5075
				FT	53	96.8844	3.1156
4	NSL KDD'99 (R2L_imap)	f3, f4, f1, f29, f2	f29	J48	11	98.0905	1.9095
				RF	11	98.4925	1.5075
				FT	11	96.8844	3.1156
5	NSL KDD'99 (R2L_ multihop)	f3, f4, f1, f29, f2	f29	J48	7	98.0905	1.9095
				RF	7	98.4925	1.5075
				FT	7	96.8844	3.1156
6	NSL KDD'99 (R2L_phf)	f3, f4, f1, f29, f2	f29	J48	4	98.0905	1.9095
				RF	4	98.4925	1.5075
				FT	4	96.8844	3.1156
7	NSL KDD'99 (R2L_spy)	f3, f4, f1, f29, f2	f29	J48	2	98.0905	1.9095
				RF	2	98.4925	1.5075
				FT	2	96.8844	3.1156
8	NSL KDD'99 (R2L_ warezclient)	f3, f12, f4, f1, f2	f12	J48	890	98.0905	1.9095
				RF	890	98.2915	1.7085
				FT	890	98.0905	1.9095
9	NSL KDD'99 (R2L_ warezmaster)	f3, f4, f1, f29, f2	f29	J48	20	98.0905	1.9095
				RF	20	98.4925	1.5075
				FT	20	96.8844	3.1156

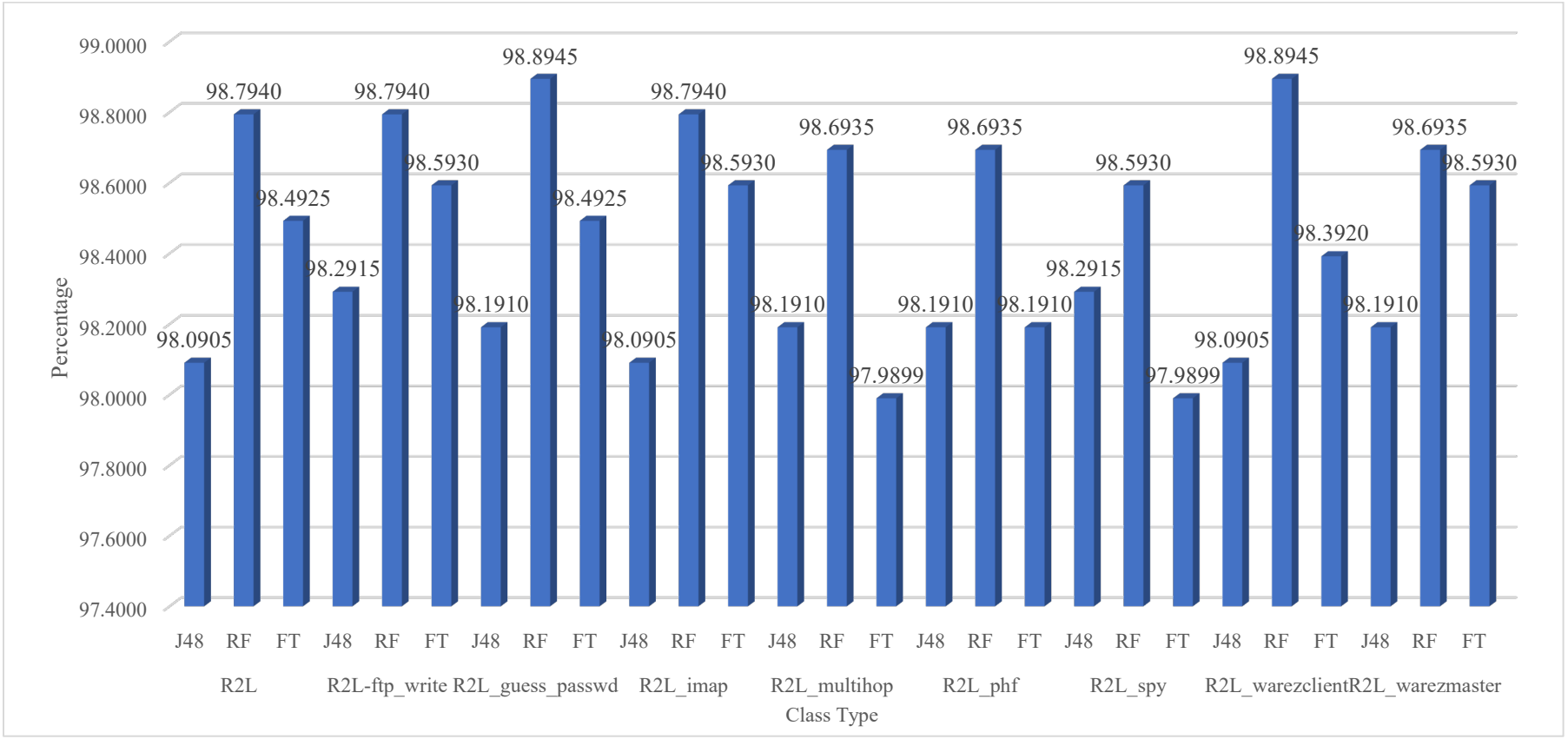


Figure 13: The Total Classification of Various Attack Class Types of R2L Attack by Removing Redundant Features

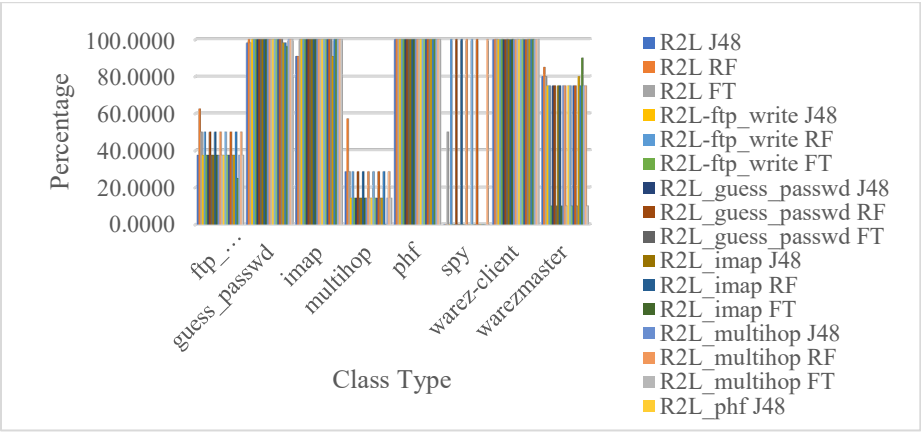


Figure 14: The Classification of Various Attack Class Types of R2L Attack by Selecting Key Features

The graph reflects the percentage of classification of various R2L attacks. The appropriate key features for R2L attack are applied and the redundant null valued features are eliminated then the most suited tree based algorithms are applied to find the optimum classified result.

The graph reflects the percentage of total classification of various R2L attacks. The appropriate key features for R2L attack are applied and the redundant null valued features are eliminated then the most suited tree based algorithms are applied to find the optimum classified result.

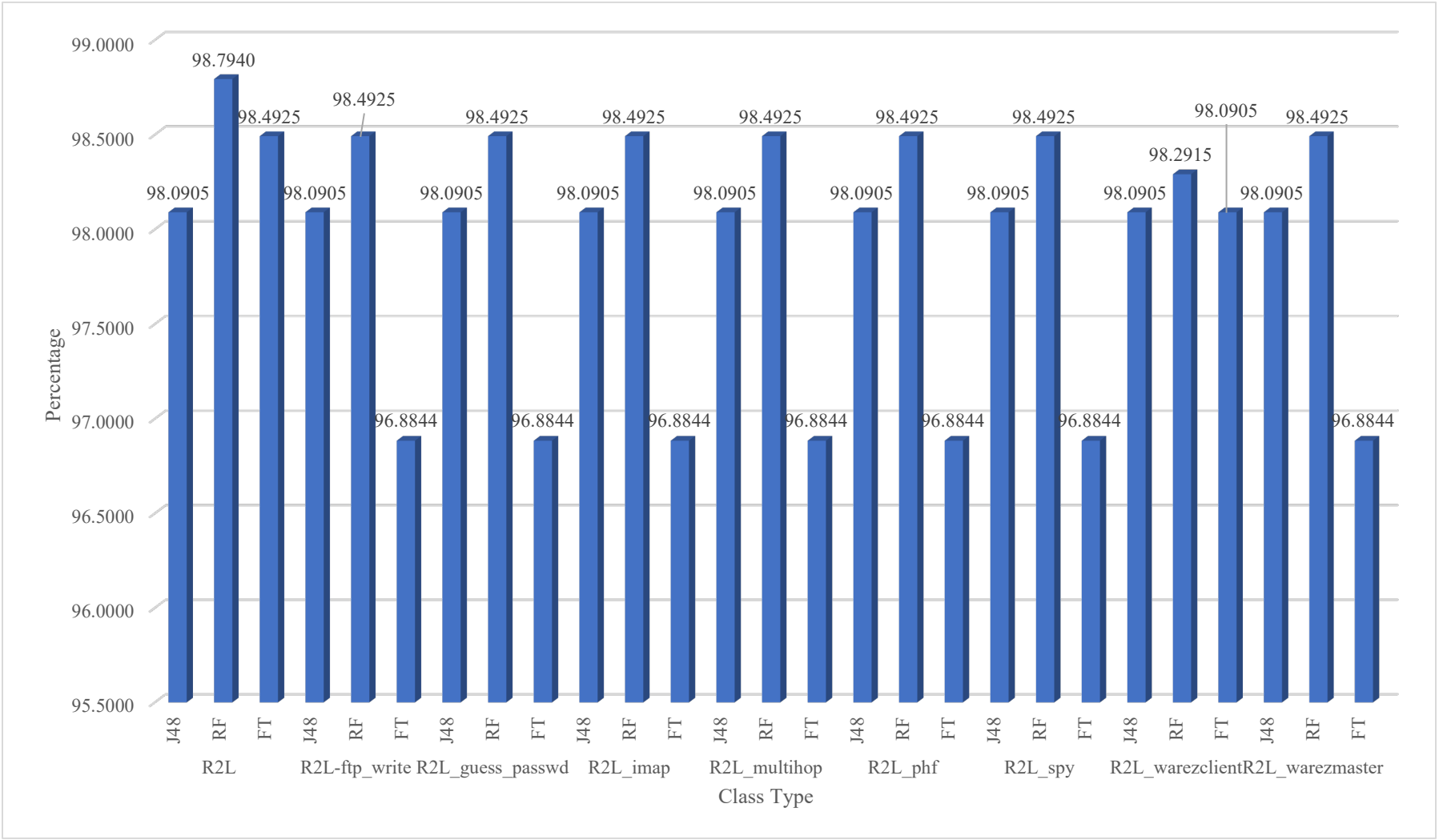


Figure 15 : The Total Classification of Various Attack Class Types of R2L Attack by Selecting Key Features

Table 14: Classification of Various U2R Attacks Using Feature Reduction of Rank Based Feature Selection Classifiers

Sl. No	Class Type	Removed Features	Algorithm Used	Total No of Instances	Total Classified	Total Unclassified
1	NSL KDD'99 (U2R)	f15, f14, f13, f41, f11, f9, f8, f7, f21, f19, f33, f30, f31, f37, f20, f38, f39, f29, f28, f27, f26, f40, f22, f23, f24, f25, f1	J48	52	80.7692	19.2308
			RF	52	84.6154	15.3846
			FT	52	84.6154	15.3846
2	NSL KDD'99 (U2R-buffer_ overflow)	f35, f18, f15, f11, f9, f8, f7, f21, f19, f31, f20, f38, f39, f26, f22,	J48	30	76.9231	23.0769
			RF	30	80.7692	19.2308
			FT	30	78.8462	21.1538
3	NSL KDD'99 (U2R_loadmodule)	f15, f11, f9, f8, f7, f21, f31, f20, f38, f39, f28, f27, f26, f40, f22, f25	J48	9	82.6923	17.3077
			RF	9	84.6154	15.3846
			FT	9	88.4615	11.5385
4	NSL KDD'99 (U2R_ perl)	f10, f15, f13, f41, f11, f9, f8, f7, f21, f19, f30, f31, f37, f20, f38,f39, f28, f27, f26, f22, f25	J48	3	78.8462	21.1538
			RF	3	78.8462	21.1538
			FT	3	86.5385	13.4615
5	NSL KDD'99 (U2R_ rootkit)	f18, f15, f8, f7, f21, f19, f30, f31, f20, f38, f28, f27, f26, f22, f25	J48	10	80.7692	19.2308
			RF	10	80.7692	19.2308
			FT	10	90.3846	9.6154

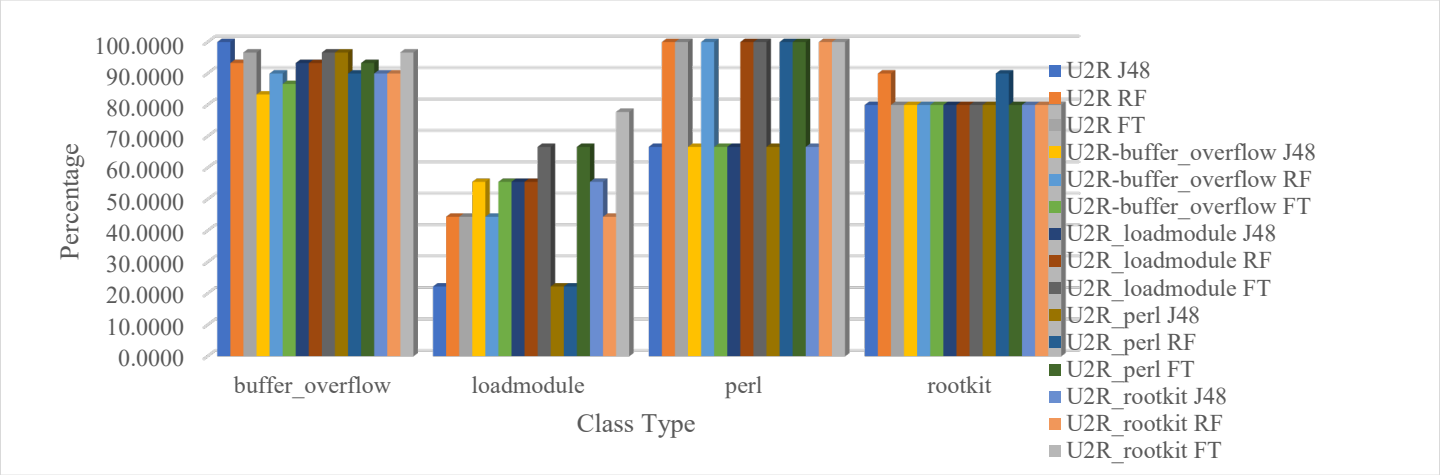


Figure 16: The Classification of Various Attack Class Types of U2R Attack by Removing Redundant Features

The graph reflects the percentage of classification of different types of U2R attacks. The redundant null valued features are eliminated and the most suited tree-based algorithms are applied to find the optimum classified result.

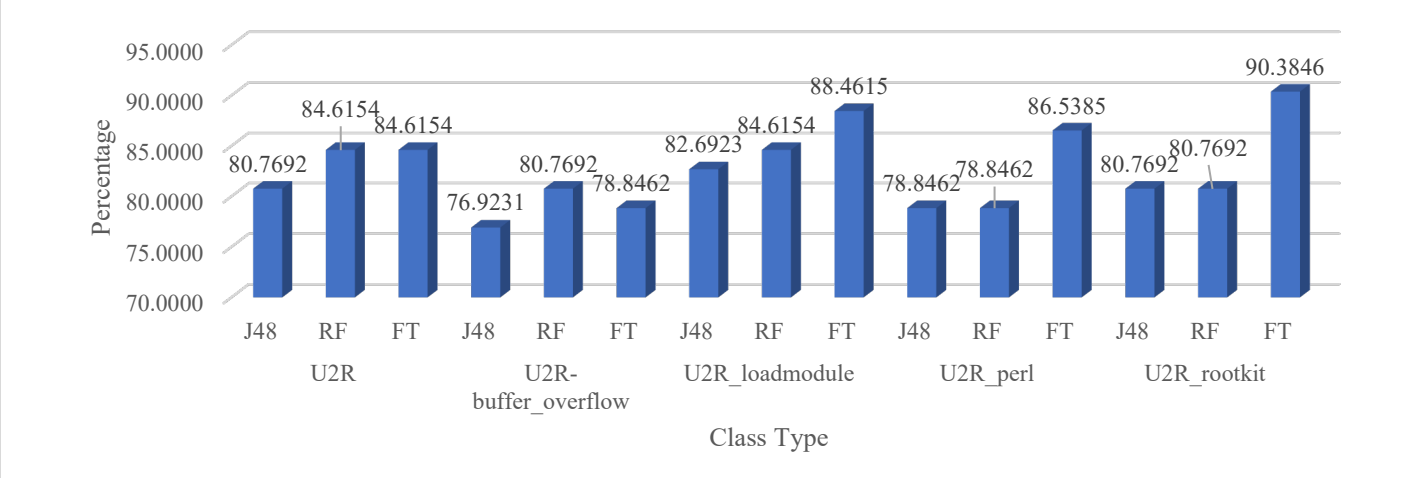


Figure 17: The Total Classification of Various Attack Class Types of U2R Attack by Removing Redundant Features

The graph reflects the percentage of total classification of various U2R attacks. The redundant null valued features are eliminated and the most suited tree based algorithms are applied to find the optimum classified result.

The graph reflects the percentage of classification of various U2R attacks. The appropriate key features for U2R attack are applied and the redundant null valued features are eliminated then the most suited tree based algorithms are applied to find the optimum classified result

Table 15: Classification of Various U2R Attacks Based on Key Feature Selection of Rank Based Feature Selection Classifiers

Sl. No	Class Type	Selected Features	Key Feature	Algorithm Used	Total No of Instances	Total Classified	Total Unclassified
1	NSL KDD'99 (U2R)	f12, f3, f2, f4, f29, f1	f12, f29	J48	52	63.4615	36.5385
				RF	52	65.3846	34.6154
				FT	52	59.6154	40.3846
2	NSL KDD'99 (U2R-buffer_overflow)	f12, f3, f2, f4, f1	f12	J48	30	67.3077	32.6923
				RF	30	57.6923	42.3077
				FT	30	61.5385	38.4615
3	NSL KDD'99 (U2R_loadmodule)	f12, f3, f2, f4, f1	f12	J48	9	67.3077	32.6923
				RF	9	57.6923	42.3077
				FT	9	61.5385	38.4615
4	NSL KDD'99 (U2R_perl)	f12, f3, f2, f4, f1	f12	J48	3	67.3077	32.6923
				RF	3	57.6923	42.3077
				FT	3	61.5385	38.4615
5	NSL KDD'99 (U2R_rootkit)	f3, f2, f4, f29, f1	f29	J48	10	53.8462	46.1538
				RF	10	65.3846	34.6154
				FT	10	57.6923	42.3077

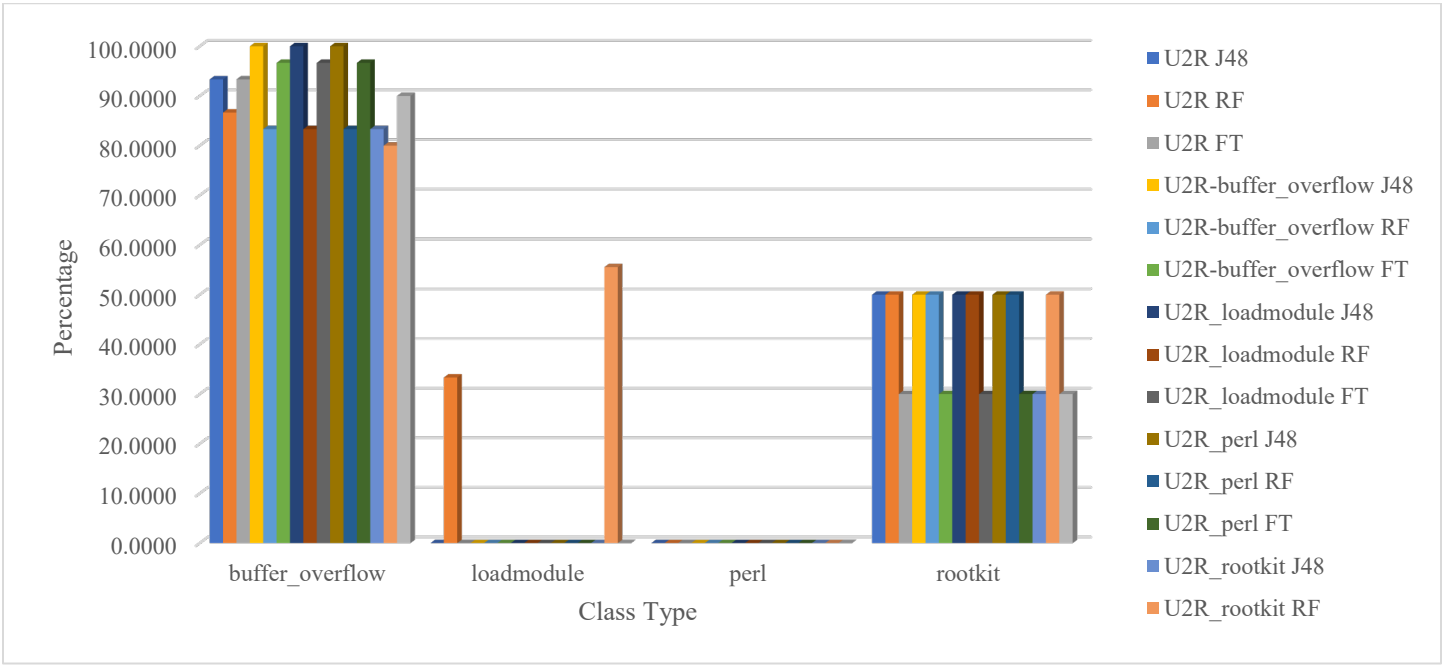


Figure 18: The Classification of Various Attack Class Types of U2R Attack by Selecting Key Features

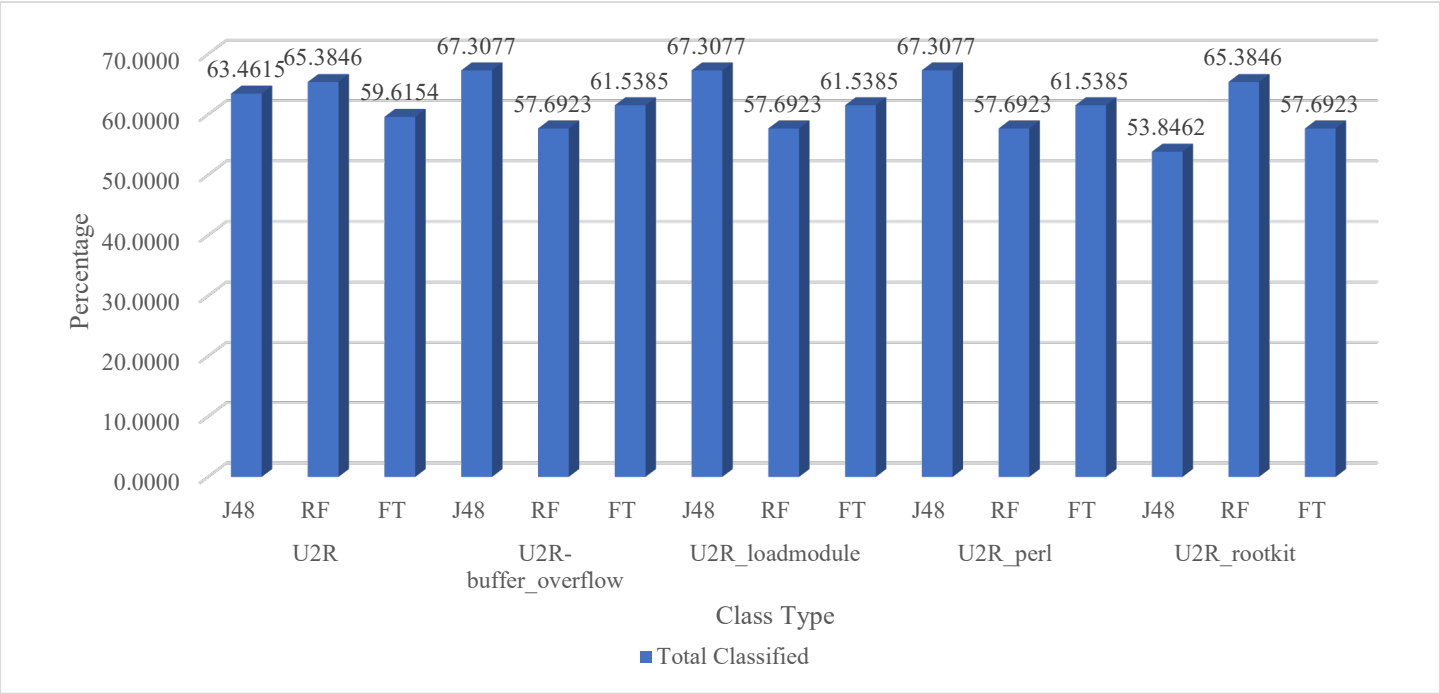


Figure 19: The Total Classification of Various Attack Class Types of U2R Attack by Selecting Key Features

The graph reflects the percentage of total classification of various U2R attacks. The appropriate key features for U2R attack are applied and the redundant null valued features are eliminated then the most suited tree based algorithms are applied to find the optimum classified result.

6. Conclusion

The classification of various attack types is based on the percentage of total classified. The null rank valued features are eliminated and the selected features are applied to find the optimum classification result. The various attack classes are selected and the attack types are classified on the basis of classification result. Various tree-based classification algorithms are used and the results are optimized based on their classification percentage. The optimized result is compared with the result computed by selected key feature. The computed result is merely dependent and influential by the selected key feature. The null valued features are eliminated and the contributing features are selected. Finally, the key feature is selected from the list of contributing features and the result is optimized.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, M. Rajarajan, "A survey of intrusion detection techniques in cloud," Network Computer Application, **36**(1), 42–57, 2013. Doi: https://doi.org/10.1016/j.jnca.2012.05.003

[2] Z. Muda, W. Yassin, M. N. Sulaiman, N. I. Udzir, "Intrusion detection based on K-means clustering and OneR classification," International Conference Information Assurance Security IAS 2011, 192–197, 2011. DOI: 10.1109/ISIAS.2011.6122818

[3] W. Koff, P. Gustafson, "CSC leading edge forum data revolution," CSC leading edge forum, 68, 2011.

[4] S. V. Thakare, D. V. Gore, "Comparative study of CIA," International Conference Communication System Network Technology, 713–718, 2014.

[5] U. Fiore, F. Palmieri, A. Castiglione, A. De Santis, "Network anomaly detection with the restricted Boltzmann machine," Neurocomputing, 122, 13–23, 2013.

[6] Z. Muda, W. Yassin, M. N. Sulaiman, N. I. Udzir, "Intrusion detection based on K-means clustering and Naïve Bayes classification," 7th Int. Conf. IT Asia Intrusion, 1–6, 2011.

[7] G. Folino, P. Sabatino, "Ensemble based collaborative and distributed intrusion detection systems: a survey," Netw. Comput. Appl., 66, 1–16, 2016. Doi: https://doi.org/10.1016/j.jnca.2016.03.011

[8] S. Fakhraei, H. Soltanian-Zadeh, F. Fotouhi, "Bias and stability of single variable classifiers for feature ranking and selection," Expert Syst. Appl., **41**(15), 6945–6958, 2014. Doi: https://doi.org/10.1016/j.eswa.2014.05.007

[9] Y. Freund, R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Comput. Learn. theory, 55, 119–139, 1995.

[10] S. Alelyani, J. Tang, H. Liu, "Feature selection for clustering : a review," Data Clust. Algorithms Appl., 1–37, 2013.

[11] F. Amiri, M. Rezaei Yousefi, C. Lucas, A. Shakery, N. Yazdani, "Mutual information-based feature selection for intrusion detection systems," Netw. Comput. Appl., **34**(4), 1184–1199, 2011. Doi: https://doi.org/10.1016/j.jnca.2011.01.002

[12] S. Solorio-Fernandez, J. A. Carrasco-Ochoa, J. F. Martinez-Trinidad, "A new hybrid filter wrapper feature selection method for clustering based on ranking," Neurocomputing, 214, 866–880, 2016.

[13] M. A. Hall, "Correlation-based feature subset selection for machine learning," Hamilton, New Zeland, 1999.

[14] N. Cleetus, "Genetic algorithm with different feature selection method for intrusion detection," First Int. Conf. Comput. Syst. Commun., 220–225, 2014.

[15] D. E. Denning, "Intrusion-detection model," IEEE Trans. Softw. Eng., 2, 222–232, 1987.

[16] M. V. Mahoney, P. K. Chan, "PHAD: packet header anomaly detection for identifying hostile network traffic," Florida Technol. Tech. Rep. CS-2001, 2001.

[17] S. B. Shamsuddin, M. E. Woodward, "Modeling protocol based packet header anomaly detector for network and host intrusion detection systems," Proc. 6th Int. Conf. Cryptol. Netw. Secur., 209–227, 2007.

[18] Md. Nasim Adnan, Md. Zahidul Islam, "Forest PA: constructing a decision forest by penalized attributes used in previous trees," Expert Systems With Applications, 89, 389-403 2017.

[19] J. Gama, "Functional trees," Machine Learning, **55**(3), 219-250.

[20] G. Hulten, Laurie Spencer, Pedro Domingos, “Mining time-changing data streams,” in ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 97-106, 2001.
[21] R. Quinlan, C4.5: programs for machine learning, Morgan Kaufmann Publishers, San Mateo, CA.
[22] J. M. Perez, Javier Muguerza, Olatz Arbelaitz, Ibai Gurrutxaga, Jose I. Martin, “Combining multiple class distribution modifies subsamples in a single tree,” Pattern recognition Letters, **28**(4), 414-422, 2007. Doi: https://doi.org/10.1016/j.patrec.2006.08.013
[23] G. Webb, “Decision tree grafting from the all-tests-but-one partition,” in San Francisco, CA, 1999.
[24] G. Holmes, Bernhard Pfahringer, Richard Kirkby, Eibe Frank, Mark Hall, “Multiclass alternating decision trees,” in ECML, 161-172, 2001.
[25] N. Landwehr, Mark Hall, Eibe Frank, “Logistic model trees,” Machine Learning, **95**(1-2), 161-205, 2005.
[26] R. Kohavi, “Scaling up the accuracy of Naïve-Bayes classifiers,” A Decision-Tree Hybrid, in Second International Conference on Knowledge Discovery and Data Mining, 202-207, 1996.
[27] L. Breman, “Random forests,” Machine Learning, **45**(1), 5-32.
[28] I. Giggins, “Knowledge discovery through sysfor – systematically developed forest of multiple decision trees,” in Australasian Data Mining Conference (AusDM 11), Ballarat, Australia, 195-204, 2011.
[29] M. Nanda, M. Patra, "Network intrusion detection and monitoring in cloud based systems," in 2019 International Conference on Applied Machine Learning (ICAML), Bhubaneswar, India, 197-204, 2019. doi: 10.1109/ICAML48257.2019.00045,
[30] M. K. Nanda, M. R. Patra, Chapter 17, Intrusion detection and classification using decision tree-based feature selection classifiers, Springer Science and Business Media LLC, 2021.
[31] H.S. Hota, Dinesh K. Sharma, A.K. Shrivas, "Development of an efficient classifier using proposed sensitivity-based feature selection technique for intrusion detection system", International Journal of Information and Computer Security, 2018.
[32] M.H. Kamarudin, Carsten Maple, Tim Watson, Nader Sohrabi Safa, "A new unified intrusion anomaly detection in identifying unseen web attacks", Security and Communication Networks, 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 391-397 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Comparison of Gaze Points Among Viewing Conditions (Resolution, Display Size, Viewer Position) During Video Viewing

Miho Shinohara*, Yusuke Nosaka, Reiko Koyama, Riko Nakagawa, Takuya Sarugaku, Mitsuho Yamada

Graduate School of Information and Telecommunication Engineering, Tokai University, Tokyo, 108-0074, Japan

ARTICLE INFO

Article history:
Received: 16 July, 2020
Accepted: 01 November, 2020
Online: 20 November, 2020

Keywords:
Ultra-High Resolution
Average Gaze Position
Viewing Position

ABSTRACT

Ultra-high resolution broadcasting such as 4K and 8K are becoming more popular for at-home use. The position of the viewer's mean gaze point when they are looking at a larger display depends on the video clip, viewing position, and viewing distance. The International Telecommunication Union provided the document BT-2022; as the standard viewing conditions for the subjective evaluations of flat-panel TV displays. However, it is rare that a television is viewed at home under standard viewing conditions. In most homes, a TV of any size is watched from the viewer's desired position according to the layout size of the room hosting the TV. No study has been conducted to determine how the movement of a viewer's line of sight changes when a display with different resolution is viewed from a position different from the standard viewing condition. If the gaze position when viewing TV differs due to the difference in resolution, the size of the display, and the difference in observation position, it affects the method of TV production. This would also affect consumers considering the resolution, installation location, and viewing location of a TV. Here, we clarified that the mean gaze position and its standard deviation are almost the same viewing conditions (i.e., resolution, display size, and viewing position) are changed, even when the same contents are viewed. Our findings demonstrate that consumers could buy a display of their preferred resolution and size according to their room layout.

1. Introduction

In Japan, high-definition broadcasting with a resolution of 2K has become widespread, and the public use of ultra-high-definition 4K and 8K broadcasting began on December 1, 2018. In the International Telecommunication Union (ITU-R) document 'BT-2022' ('General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays'), the standard viewing distance for a television's display is defined as the maximum distance at which a person with a visual acuity of 1.0 can see the pixel structure [1]. Since the current HDTV broadcasting has approx. 2 million pixels at 1920×1080, the standard viewing distance when watching HDTV images is 3H (H; Display height). In contrast, 4K broadcasting has ~ 8 million pixels at 3840×2160, and that of 8K broadcasting has ~ 33 million pixels at 7680×4320.

It was reported that when the resolution becomes higher, the pixel structure cannot be seen even when one is approaching it, and thus it is possible to view from a closer distance and enjoy a wide-angle image [2]. The ITU-R stipulates that a viewer should watch from the center of the front of the screen even when approaching the screen, but at each home, the viewing position differ depending on the room layout, the number of viewers, the size of the TV purchased, and more. The display is frequently viewed from sites that are not at the center of the screen. According to a survey of TV viewing positions, the positions spread horizontally ±60° and vertically ±30°, and 37% of viewers look at the display from the right side of the TV, only 3% from the center, and 60% from the left side [3]. When watching a movie on a large screen, the impression of the movie varies greatly depending on the position of the viewer's seating, such as the center of the screen, the left and right edges of the front seats, and the rearmost seats. However, no research has examined how the position of the content where a viewer views on a TV display changes depending on the position of the viewer's seating. If the content is viewed differently depending on the seating position, the producers of televised content must consider the effects of the viewing position when producing the content. Consumers seeking a TV to purchase also need to consider where to put the TV and furniture.

*Corresponding Author: Miho Shinohara, 0cjnm010@mail.u-tokai.ac.jp

A study of the differences in the viewing of contents depending on the TV's resolution compared NTSC; analog color system with (525 scanning lines) and high-definition (1,125 scanning lines) [4]. In the low-resolution NTSC, the viewer's gaze points are distributed in the center of the screen because multiple cameras are switched in order to provide the camera shot that the content's director wants to show in the center. In high-definition, the entire screen is displayed, and the viewers' gaze points are spread over the entire screen. Several research groups have verified the differences in viewing distance by using parameters such as screen size, display brightness, resolution, and video content [5-8]. In those studies, the contents were viewed from the center position of the display, and the viewing position was not changed in the horizontal or vertical direction. In addition, an NTSC or phase alternating line (PAL) system was used (not higher-definition 4K), and the results cannot always be applied to a 4K or 8K display. In the present study, using 4K high-definition images that have never been used, we examined the differences in viewers' lines of sight when they were watching television from a distance that is not the standard viewing distance and from various positions (right, center or left). Viewing 4K or 8K ultra-high-definition TV with large-screen images from a wide field of view at a short distance has become established in for home use. As a result, viewers are more likely to see at various viewing positions, not limited to the center of the screen, as seen in movie theaters, but there is no study that examined the effect of viewing position on viewing behavior from eye movements. Therefore, our present findings will provide meaningful data for both viewers and video producers to know whether or not the observation position affects the gaze position.

The purpose of our research is to objectively evaluate the psychological effects of changes in display resolution, size, and viewing position on viewing behavior from eye movements. For that purpose, two kinds of experiments were conducted.

In Experiment 1, we analyzed the influence of the display resolution on the viewing positions of individuals watching three video clips on the display. We showed the subjects a video clip whose resolution was 4K (3840×2160), 2K (1920×1080) or 1K (960×540) at the viewing distance of 0.75H. In Experiment 2, we analyzed the influence of the size of the screen on the viewing position. We showed the subjects a video clip whose screen size was 1/1, 1/2 or 1/4 at 0.75 H as the viewing distance [9].

2. Experimental Method

2.1. Experimental equipment

When viewing a 4K image at the viewing distance of 1.5H, the field of view is 60°. The range that a human being can see by using only saccades, which is a high-speed eye movement, is up to 15° [10]. Head movements are used together with eye movements to exceed the 15° range. It was reported that head movements contributed twice as much as the eye movements when moving the line of sight to fixate at one point [11]. It is also known that in the ratio of head movement to the line of sight, the head share becomes 60%-80% when the line of sight moves outside a 30° range [12, 13]. We thus surmised that in this experiment, it was necessary to detect changes in the line of sight that included the addition of the subjects' head movements to their eye movements. We used the line-of-sight measuring device developed in our laboratory [14]. The accuracy of the horizontal direction was approx. 1.5° to -1.7° at ±30° corresponding to the visual angle of the display edge at 1.5H. The accuracy of the perpendicular direction was approx. 2.6° to -1.4° at ±18°, corresponding to the visual angle of the display edge at 1.5H [4]. We measured eye movements (EMR-8B; Nac Image Technology, Tokyo) and head movements (3DM-GX4; Microstrain, Williston, VT, USA). Data were recorded at 60 Hz. In the experiment, the subjects looked at the display in a free posture without being fixed on the head-chin rest of the experimental set-up.

2.2. Viewing conditions

The viewing distance of the display was fixed at 1.5H, and the subject was allowed to view the display from three positions: the center (display center), the right (1/4 to the right from the display center), and the left (1/4 to the left from the display center) as shown in Figure 1. The standard viewing distance for image-quality evaluations is not 1.5H when the image resolution is 2K or 1K, or when the screen size is 1/2 or 1/4: however, in a most households, the room layout is prioritized, and the viewing distance often does not change even if the screen resolution or size changes. We therefore fixed the viewing distance. Since the subjects' head movements were measured simultaneously with their eye movements, the subject's head movements were not restricted and the subject was instructed to watch the TV in a natural posture. The subjects watch videos with three types of resolution, three screen sizes, and three viewing positions, but each subject was prevented from seeing the same image more than once. This experiment was conducted according to BT-2022 and followed the Ethical Provisions of Tokai University's guidelines for "Research using humans."

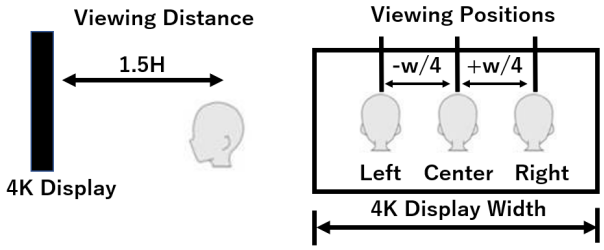


Figure 1: Viewing conditions.

2.3. Laboratory environment

The video clips were presented to a 4K projector (VPL-VW 500ES; Sony, Tokyo) using a personal computer (PC) as the signal source. The height of the screen was 147 cm, the width of the screen was 262 cm, and the screen brightness was 0.04 cd/m^2 at the minimum and 104 cd/m^2 at the maximum. The illuminance in the room was 69.7 Lux. The layout of the laboratory, the procedure of the experiment, and the selection of subjects were based on BT-2022.

2.4. Changes in gaze position due to the difference in resolution (Experiment 1)

The details of the video clips used in Experiment 1 are shown in Table 1 and in Figures 2-4 (scenes used for video clips 1-3). The 4K video clips consisted mainly of natural scenes (mountains, flowers, seas) and were downloaded from a free video clips material website [15]. The 4K video clips were recorded in the

60p/XAVCS format, which is also used for broadcasting, and have low compression. Each video clip length was 40 sec. The series of 4K video clips was edited to a length of approx. 4 min with the use of nonlinear editing software (VegasPro13; Sony). The 2K and 1K video clips were down-converted from 4K video clips using the same nonlinear editing software. A still video showing the fixation point in the center was inserted before and after each video clip presented to the subject, so that it could be recognized even if the calibration of the eye movement was misaligned. The subjects were 25 college students (10 males and 15 females) in their twenties, seven of whom had the visual acuity score of >1.0: the other 18 had the visual acuity score of 0.7 (including corrected vision). We analyzed the data of 21 of the 25 subjects (seven males, and 14 females) because four subjects had a large number of measurement errors during the experiment. No subject viewed the same image more than once. Each subject was randomly determined which image to view at which resolution and viewing position.

Table 1: Video clips used in Experiment 1

Video clip	Time	Resolution
Video clip 1: River, ship or bridge	4 min 14 sec	4K, 2K and 1K
Video clip 2: Cherry blossoms or Mt. Fuji	4 min 2 sec	4K, 2K and 1K
Video clip 3: Town or animal	4 min 9 sec	4K, 2K and 1K

Figure 2: Scenes used for video clip 1.

Figure 3: Scenes used for video clip 2.

Figure 4: Scenes used for video clip 3.

2.5. *Changes in gaze position due to the difference in screen size (Experiment2)*

The details of the video clips used in Experiment 2 are shown in Table 2 and Figures 5-7 (scenes used for video clips 1-3). The video clips were edited as in Experiment 1, and 1/4-size and 1/2-size video clips were generated by reducing the number of pixels from 4K video clips using the nonlinear editing software. The subjects were 24 college students (nine males and 15 females) in their twenties, eight of whom had a visual acuity score of >1.0: the other 16 had a visual acuity score of 0.7 (including corrected vision).

Nineteen of the 24 subjects also participated in Experiment 1. There were no measurement errors in the data of all 24 subjects, and all were used for analysis. No subject viewed the same image more than once. Each subject was randomly determined which image to view at which screen size and viewing position.

Table 2: Table 2. Video clips used in Experiment 2

Video clip	Time	Screen size
Video clip 1: Flowers	4 min 7 sec	1/4, 1/2, 1/1 size of 4K
Video clip 2: Okinawa	4 min 2 sec	1/4, 1/2, 1/1 size of 4K
Video clip 3: Yokohama	3 min 14 sec	1/4, 1/2, 1/1 size of 4K

Figure 5: Scenes used for video clip 1.

Figure 6: Scenes used for video clip 2.

Figure 7: Scenes used for video clip 3.

3. Results

The distribution of the gaze points of each subject was analyzed while the subject viewed the video clips. We used 10 deg/s as the threshold speed for extracting the gaze point from the gaze movement [16]. The horizontal and vertical components of gaze were calculated by the following formulas.

Horizontal components of Gaze = Horizontal components of Eye movement + Horizontal components of Head rotation

Vertical components of Gaze = Vertical components of Eye movement + Vertical components of Head rotation

The mean gaze position and its standard deviation (std. dev.) during viewing have often been used to compare the gaze behaviors of individual watching video clips [4]. We thus calculated the mean gaze position and its std. dev. from the remaining 4 minutes of data excluding the 1 minute immediately after the presentation of the video clips, at which point the data were not stable.

The mean gaze point position and the std. dev. of the mean gaze point distribution showed the same tendencies in all three video clips, and we therefore analyzed the data of the three video clips together. We performed a two-way analysis of variance (ANOVA) to analyze the results. The level of significance was set at 10%.

3.1. Results of Experiment 1: Change in gaze position due to the difference in display resolution (Experiment 1)

During the measurement of a person's gaze, there may be many blinks, or the eye movement sensor may shift, causing an error in eye movement. Of the 21 subjects test in Experiment 1, the seven subjects who had almost no measurement error during the measurements for all the 1K, 2K, and 4K video clips were used for the analysis in order to perform statistical analysis with all the accurate data, because, it is extremely difficult to perform all experiments completely without measurement errors due to the nature of eye movement measurement. Figure 8 shows the averages and standard deviations of the mean gaze point positions of these seven subjects. In the figure, the center of the screen is 0°, and the angles displayed on the horizontal and vertical axes represent the viewing angles used by the subject. Figure 9 shows the mean gaze positions obtained by summarizing the results of the three viewing positions for each resolution.

The mean gaze position at the left and right viewing positions was around the center of the screen, and the average gaze position at the central viewing position was slightly left of the center of the screen regardless of the resolution, as shown in Figure 8. In the summary of the results of the three viewing positions, the mean gaze position is located at the center of the screen regardless of the differences in resolution, as shown in Figure 9. These results suggest that there was a slight difference in the mean gaze position depending on the viewing position, but it was not due to the difference in display's resolution.

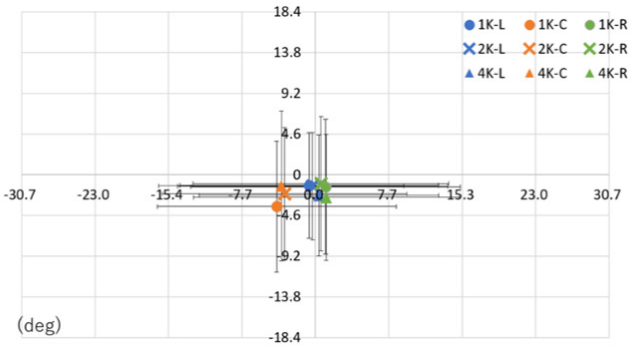


Figure 8: Averages of the mean gaze position and std. dev. (resolution and viewing position)

The results of the two-way ANOVA for mean gaze position with resolution and viewing position as variables are illustrated in Figure 10. No significant differences were observed for resolution, but a multiple comparison analysis revealed a significant difference in viewing position in the right-central and left-central positions. No significant differences were detected in the std. dev. of mean gaze position as shown in Figure 11; however, there was a slight but significant difference in viewing positions at the 10% level of significance.

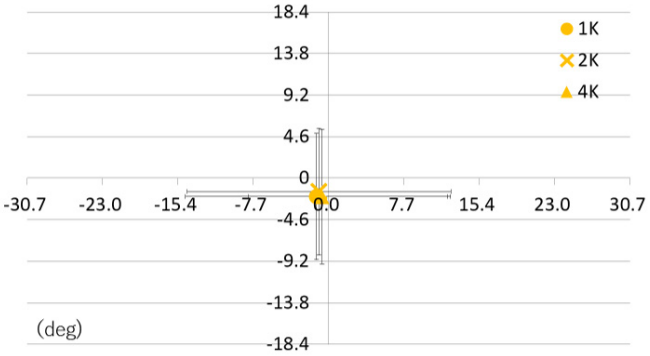


Figure 9: Averages of the mean gaze position and std. dev. (resolution).

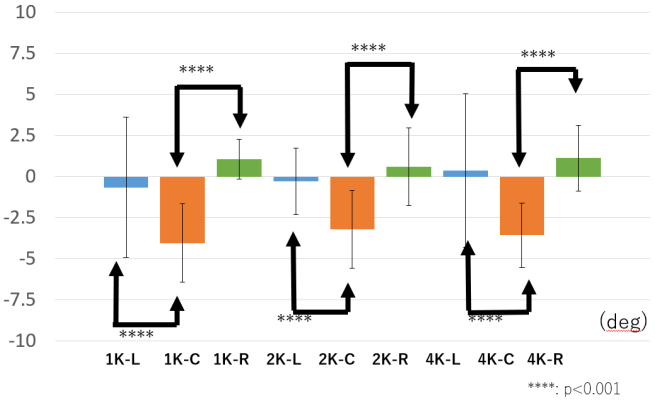


Figure 10: Results of the two-way ANOVA for mean gaze position (resolution).

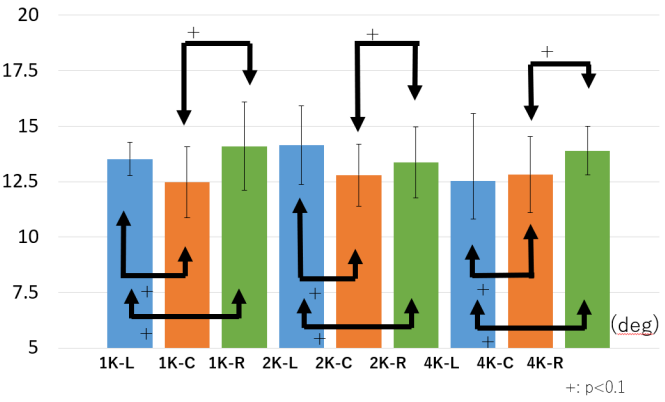


Figure 11: Results of the two-way ANOVA for the std. dev. of mean gaze position (resolution).

3.2. Results of Experiment 2: Change in gaze position due to the difference in screen size (Experiment 2)

In Experiment 2, of the 24 subjects, eight who had almost no measurement error during the measurements for all of the 1K, 2K,

and 4K video clips were used for the analysis to perform the same accurate statistical analysis as in Experiment 1. The average mean gaze position of these eight subjects who viewed 1/4-, 1/2- and 1/1-size video clips at each viewing position is shown in Figure 12. Figure 13 provides the mean gaze positions obtained by summarizing the results of the three viewing positions for each resolution. In this experiment, since only the screen size was changed without changing the viewing distance, the angle of viewing the screen expected from the subjects would be 1/4 and 1/2 for the 1/4 and 1/2 sizes, respectively. Therefore, to examine whether each subject gazed at the same place in the display and had the same gaze range, we normalized these graphs to the 4K 1/1 size.

The results for a 1/4-size image (Fig. 12) indicated that the subjects viewed the area to the right of the center screen from the left and central viewing positions, and they viewed the screen center from the right viewing position. The results for a 1/2-size image (Fig. 12) revealed that the subjects viewed the area to the right of the center screen from the right and central viewing positions, and they viewed the lower right of the screen from the left viewing position. The results for a 1/1-size image (Fig. 12) demonstrated that the subjects viewed the area to the right of the center screen from the right and central viewing positions, and they viewed the screen center from the left viewing position. The results shown in Figure 12 suggest that the mean gaze position depending on the viewing position tended to change slightly when the size of the screen becomes small. According to Figure 13, the average of the results for all subjects and all viewing position regardless of the screen size shows that the mean gaze position falls in the area to the right of the center screen.

The results of the two-way ANOVA for mean gaze position are illustrated in Figure 14, and those of the two-way ANOVA for the standard deviation of the mean gaze position are shown in Figure 15. No significant differences in the size of the screen or in viewing position were revealed by the two-way ANOVA. Although the mean gaze position tended to change slightly depending on the viewing position when the screen size and viewing position changed, the mean gaze position and the standard deviation of the mean gaze position did not show significant differences.

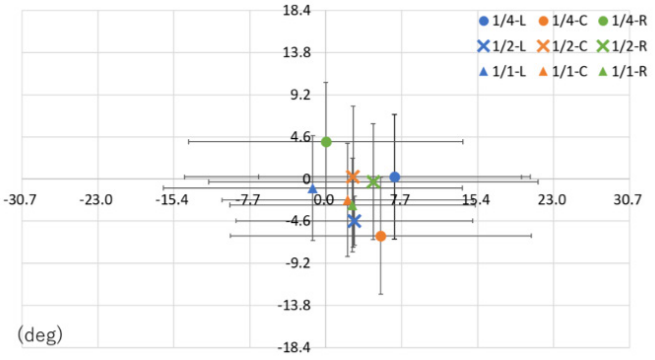


Figure 12: Averages of the mean gaze position and std. dev. (screen size and viewing position).

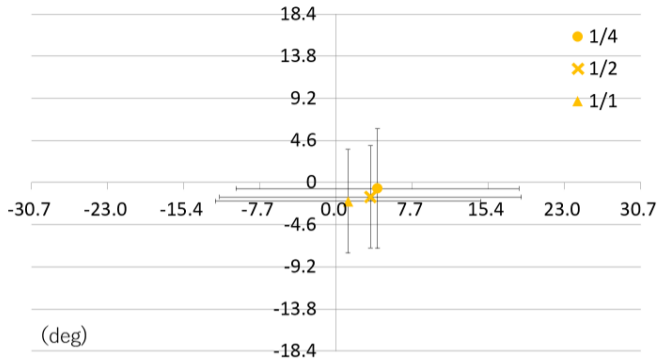


Figure 13: Averages of the mean gaze position and std. dev. (screen size).

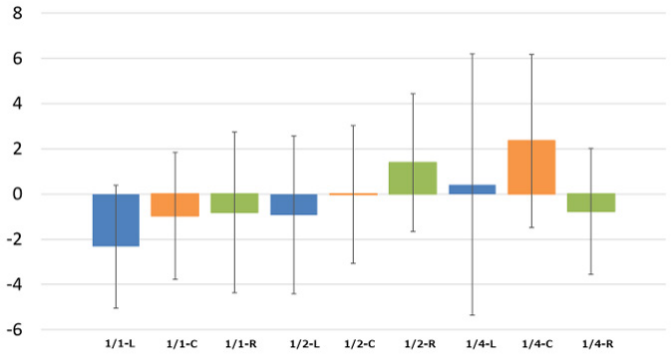


Figure 14: Results of the two-way ANOVA for mean gaze position (screen size).

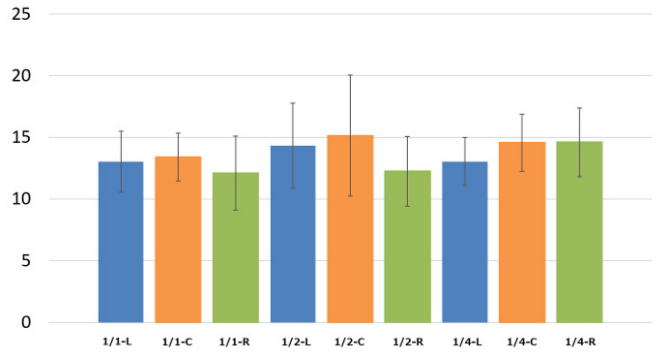


Figure 15: Results of the two-way ANOVA for the std. dev. of the mean gaze position (screen size).

4. Discussion

There were no significant differences in the mean gaze position or the standard deviation of the average gaze position due to differences in resolution, even when the viewing position was changed. Similarly, even when the viewing position was changed, no significant differences were found in the average gaze position or the standard deviation of the mean gaze position due to differences in screen size.

The tendency for the initial gaze position to fall at the center of the image when looking at a still image is called center bias. Photographer bias and viewing strategies have been suggested to be major factors in center bias [7]. Photographer bias is the practice of placing an object or the actor in whom the photographer is most interested near the center of the screen. With respect to viewing

strategies, people usually turn their eyes to the center of the image, having learned through frequent experience with photographer bias that objects that attract attention are generally arranged around the center of the screen. We believe that our present results in which the subjects watched the screen center regardless of differences in resolution and the size of the screen demonstrate a tendency similar to center bias; that is, while center bias is defined as a tendency to look to the center of a still image, our subjects looked to the center of video clips, rather than still images.

5. Conclusion

Broadcasting in 4K and 8K ultra-high-definition video has started, and viewers can enjoy high-definition, wide-angle video near the display. The conditions for subjectively evaluating a display are strictly defined by the ITU-R. Subjective evaluations of displays by broadcasters and manufacturers are performed under these conditions. However, in the home where the display is actually located, the display is viewed at a free position according to the size of the TV, the layout of the room, and structure of the household. For example, when watching a movie in a large-screen movie theater, the viewers experience a different sense of presence and emotion after viewing from the seats at the left and right ends of the front row, the seats at the center of the movie theater, and the seats at the rear. Television broadcasting is also expected to be viewed from various positions in the home by viewing a high-definition, wide-field display. If the impression produced by the image differs depending on the viewing position, it is necessary for the consumers as well as the image producers and the display manufacturers to take this into consideration.

However, the evaluations of displays are performed according to the ITU-R and are rarely performed under the different conditions. Our present experiments show that the mean gaze position during viewing depends on the video clips, the viewing position, and the viewing distance, with the size and resolution of the display fixed and the viewing position as a parameter [17]. This research was conducted with a constant display size and resolution, but in reality, displays with various resolutions and sizes are found in various homes, and it has been necessary to investigate the effects of resolution and size. Therefore, in Experiment 1, we used the resolution and the viewing position as parameters, and in Experiment 2, the screen size and the viewing position were used as parameters, and the mean gaze position and its standard deviation were analyzed. In analyses of the gaze behavior of still images such as photographs, the average gaze position and the variance of gaze points are examined by heat maps and other methods, and there is a photographer bias that places the part that the photographer wants to show to the viewer in the center [18]. It is thus known that there is a center bias in which viewers empirically tend to see the center of the photograph [19].

In addition, regarding evaluations of the viewing of TV images, the mean gaze position and its standard deviation are often used to compare gaze behaviors while subjects watch video clips [4, 17], and we thus applied the mean gaze position and its standard deviation to the analysis of viewing behavior when the resolution, screen size, and viewing position were changed. When 4K-resolution content is viewed from 1.5H, the field of view seen from the eye is 60°. It was reported that the range of visual information that can be captured only by eye movements in daily activities is at most 15° [10]. We therefore, measured the line of sight (head movement + eye movement) with the subjects in a free posture without fixing the head in the device that we developed that can measure head movement and eye movement at the same time [14].

In Experiment 1, the resolution was changed to 1K, 2K, and 4K, and the mean gaze position and its standard deviation were obtained by changing the viewing position using the same-size display. The difference in the mean gaze position was seen by the difference in the viewing position. Notably, the mean gaze position was different between the central viewing position and the left and right viewing positions. There was a slight change in the standard deviation of the mean gaze position due to the difference in the viewing position. However, the changes in the mean gaze position and its standard deviation were slight.

In Experiment 2, the display size was changed to 1/4, 1/2, and 1/1, and the mean gaze position and its standard deviation were obtained by changing the viewing position using the same resolution. The mean gaze position and its standard deviation did not change even when the screen size was changed.

With the introduction of 4K and 8K broadcasting, broadcasting media and video contents have diversified. However, in each home, displays with various resolutions and sizes are being viewed at any viewing position. Our present results demonstrated that when the same content was viewed at different resolutions, screen sizes, and viewing positions, there was almost no change in the mean gaze position, which is one of the viewing behaviors. This means that video producers can engage in program production without needing to being aware of the resolution of the display to be viewed, the screen size, or the viewing position. In addition, consumers can purchase a display according to the size and layout of the room and their preferences, and watch their television from any position.

References

[1] Recommendation BT.2022 https://www.itu.int/rec/R-REC-BT.2022/en
[2] Y. Shishikui, “Super Hi-Vision R\&D,” NHK STRL R\&D, **137**, 4-9, January, 2013.
[3] S. Kubota, A. Shimada, S. Okada, Y. Nakamura, E. Kido, “Television Viewing Conditions at Home,” The Journal of The Institute of Image Information and Television Engineers, **60**(4), 597-603, 2006, doi: 10.3169/mta.7.112.
[4] M. Yamada, T. Fukuda, “Quantitative Evaluation of Eye Movements as Judged by Sight-Line Displacements,” SMPTE Journal, **95**(12), 1230-1241 1986, doi: 10.5594/J04053.
[5] N. Narita, M. Kanazawa, F. Okano, “Optimum Screen Size and Viewing Distance for Viewing Ultra High-Definition and Wide-Screen Images,” The Journal of The Institute of Image Information and Television Engineers, **55**(5), 773-780, May, 2001, doi: 10.3169/itej.55.773.
[6] M. Arbito, “Studies of the Infkuence of Display Size and Picture Brightness on the Preferred Viewing Distance for HDTV Program,” SMPTE J., **103**(8), 517-522, August, 1994, doi: 10.5594/J06461.
[7] A.M. Lund, “The Influence of Video Image Size and Resolution on Viewing-Distance Preferences,” SMPTE J., **120**(5), 406-415, May, 1993, doi: 10.5594/J15915.
[8] S. Kubota, K. Kishimoto, S. Goshi, S. Imai, Y. Igarashi, T. Matsumoto, H. Haga, T. Nakatsue, Y. Mano, Y. Kobayashi, “Preferred Viewing Distance for High Definition Television LCDs,” The Journal of The Institute of Image Information and Television Engineers, **65**(8), 1215-1220, August, 2011, doi: 10.3169/itej.65.1215.
[9] R. Koyama, R. Nakagawa, M. Suganuma, S. Mochiduki, M. Yamada, “The

Effect of Viewing Position by Changing Display Resolution and Display Size," IMQA 2018, 103-109, September, 2018.
[10] A.T. Bahill, D. Adler, L. Stark, "Most naturally occurring human saccades have magnitudes of 15 deg or less," Invest. Ophthalmol., **14**(6), 468-469, 1975.
[11] B. Biguer, C. Prablance, M. Jeannerod, "The contribution of coordinated eye and head movements in hand pointing accuracy," Experimental Brain Research, **55**(3), 462-469, 1984, doi: 10.1007/BF00235277.
[12] M.A. Gresty, "Coordination of head and eye movements to fixate continuous and intermittent targets," Vision Research, **14**(6), 395-403, 1974, doi: 10.1016/0042-6989(74)90238-7.
[13] G.R. Barns, "Vestibulo-ocular function during co-ordinated head and eye movements to acquire visual targets," The Journal of Physiology, **287**, 121-147, 1979, doi: 10.1113/jphysiol.1979.sp012650.
[14] H. Takahira, K. Kikuchi, M. Yamada, "A System for Measuring Gaze Movement and Hand Movement Simultaneously for Hand-Held Devices," The Institute of Electronics, Information and Communication Engineers Trans. Commun., **E98-B**(1), 51-61, January, 2015, doi: 10.1587/transcom.E98.B.51.
[15] "Royalty Free 4K HD Footage (XAVC S format).", http://footage3.opensp c2.org/HDTV/ footage/4K/60p/XAVCS/., May 10, 2018.
[16] R. Fukuda, M. Sakuma, E. Nakamura, T. Fukuda, "An Experimental Consideration on the Definition of a Fixation Point," The Japanese Journal of Ergonomics, **32**(4), 197-204, 1996, doi: 10.5100/jje.32.197.
[17] H. Takahira, S. Mochiduki, M. Yamada, "Analysis of Gaze Movement When Changing Viewing Distance and Viewing Position," The IEICE Transactions on Information and Systems (Japanese Edition), **J99-D**(3), 293-306, March, 2016, doi: 10.14923/transinfj.2015PDP0005.
[18] P. H. Tseng, R. Carmi, I. G. M. Cameron, D. P. Munoz, L. Itti, "Quantifying Center Bias of Observers in Free Viewing of Dynamic Natural Xcenes," Journal of Vision, **9**(7-(4)), 1-16, 2009, doi: 10.1167/9.7.4.
[19] D. J. Parkhurst, E. Nieber, "Scene content selected by active vision," Spatial Vision, **16**(2), 125-154, 2003, doi: 10.1163/15685680360511645.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 398-402 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Method for Improving the Quality of the Product Obtained by Abrasive Treatment with Impregnated Tools

Viktor Butenko[1], Liana Gusakova[1], Dmitry Durov[2], Boris Safoklov[*,2], Oleg Dolgov[2]

[1]*Department Design of Special Aviation Complexes, Moscow Aviation Institute (National Research University), Moscow, 125993, Russia*

[2]*Department Technological Design and Quality Management, Moscow Aviation Institute (National Research University), Moscow, 125993, Russia*

ARTICLE INFO

Article history:
Received: 28 July, 2020
Accepted: 31 October, 2020
Online: 20 November, 2020

Keywords:
Abrasive Tools
Impregnation
Chromium Diiodide
Surfactants

ABSTRACT

The article describes a way to increase the efficiency of an abrasive tool by its impregnation in an aqueous solution of chromium diiodide. The device scheme for impregnating the instrument is proposed. Research results of durability of tools impregnated with chromium diiodide and structural condition of material of a surface layer of details after processing are resulted. Within of the researches presented in the article the grinding processing a shafts made from various iron-carbon steels with prepared abrasive tools was conducted. As a result of the machining the scattering fields of the batch sizes of the parts within the grinding wheel resistance were determined. This data for processing with standard grinding wheels and impregnated by chromium diiodide were done. Using data from controlled grinding of workpieces, approximate values of the K_{Δ} coefficient were determined depending on the diameter of the workpiece surface, the required size accuracy of part and the method application of the surfactants. It is shown that regardless of the material being processed, the use of a grinding wheel impregnated with chromium diiodide leads to a reduction in the specific value of the accumulated deformation energy E_{spec}. Taking into account these results it is possible to predict up to twofold reduction of the cost of machining with the abrasive tool of grinding parts, which will reduce the cost of their manufacture by 20-25%. The developed process of impregnating grinding wheels with chromium diiodide can be used without large economic costs in many machine-building company also during repair works.

1. Introduction

The problem of increase of efficiency of abrasive processing of surfaces of machine parts remains actual in machine-building manufacture and one of the perspective directions of its decision is impregnation (saturation) of abrasive tools with surfactants.

There are known methods [1] of abrasive instruments [2] impregnation [3,4] with paraffin, ozokerite, phenolformaldehyde resin, sulfur, rosin, cerezin, stearin, potassium ethylxantogenate, sodium silicate solution, aromatic chlorine-containing compounds and other substances [5], compositions of several elements including fillers [6] (graphite, molybdenum disulfide, fluoroplastics, micropowders, etc.). Shortcomings of these compounds are reflected in the limited applicability and the need to select an impregnating compound depending on the material and tool, in some cases, the toxicity to humans and the use of impregnating compounds solely on a water-based, corrosive activity and insufficient retention of chemicals in the pores of the grinding wheel, reducing its hardness, etc, possible increase in quality control points and increase in the prime cost of the final product.

When grinding with impregnated wheels, there are fast processes of destruction and restoration of solid lubricating films in the contact zone between the grinding tool and the workpiece due to high speeds of molecules movement at high temperatures and pressures. The substances located in the pores of the wheel can pass into different aggregate state, and the impregnation efficiency

[*]Corresponding Author: Boris Safoklov, Email: safoklovbb@mai.ru

depends on the products of its decomposition, their penetration ability and the activity of interaction with the processed material [7].

High reactivity at interaction with metals, especially with high adhesion to abrasive material, halogens [8] and among a large number of types of surfactants, as well as methods of their application in the blasting of surfaces of machine parts the most effective is the method of impregnation of grinding wheels in an aqueous solution of chromium diiodide [9, 10], molecular formula CrI_2 [11], we have considered all the various ways to improve the efficiency of an abrasive tool, and developed a new effective method for impregnating grinding wheels.

2. Materials and Methods

To increase the efficiency of the abrasive tool there was developed an effective method of impregnating grinding wheels in the aqueous solution of chromium diiodide [11], the main essence of which is that the abrasive tool on a ceramic bond is placed in a container with aqueous solution of chromium diiodide (20-25 g per liter of water) and within one minute it is informed about the vertical oscillating motion with a certain frequency ν, which provides its complete immersion in the aqueous solution and complete removal from it [12,13]. After impregnation, the impregnated instrument is subjected to convective drying when rotating at a speed of (0.3-0.5) s^{-1} for 1.5-2 hours at a temperature of (40-50) ºC.

Vertical vibrating movements of the impregnated grinding wheel with a given frequency v and amplitude A were carried out using a special installation, shown in Fig. 1 and containing an electric motor 1, gearbox 2, cam 3, pusher 4, rocker arm 5, mobile support 6, rod 7, to which the impregnated abrasive tool 8 is attached, and a container with an aqueous solution of chromium diiodide 9.

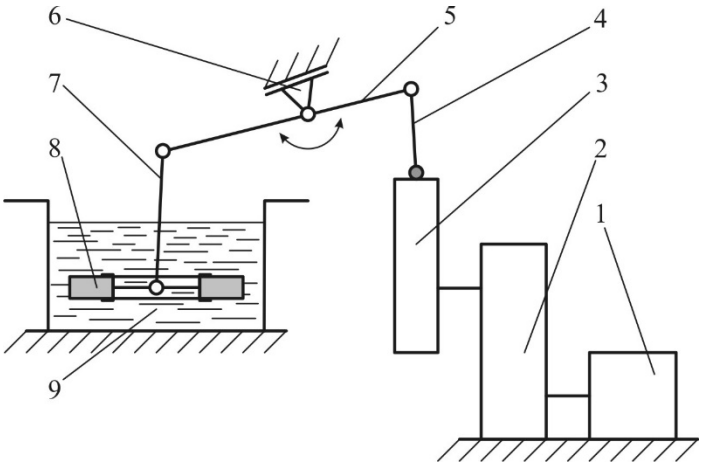


Figure 1: Plant diagram for impregnating grinding wheels

Research of efficiency of application of grinding wheels impregnated with chromium diiodide was carried out by processing of samples from steel 12CrNi3A (Table.1) with diameter 30 mm and length 200 mm after finishing turning (surface roughness before grinding Ra = 8.0-12.5 μm). Grinding of samples was carried out on a circular grinding machine mod. 3U12VF11 PP 400×50×203 14AF60K7V GOST R 5281-2007 with cutting speed V_{gw} = 35 m/s, sample rotation speed V_m = 0.314 m/s, longitudinal feed S_{lon} = 0.02 m/s, cutting feed S_{cut} = 0.01 mm / double stroke, the number of double strokes m = 3. The 5% aqueous solution of emulsion Ukrinol-14 was used as a cooling lubricant [14,15]. The change of firmness of grinding wheels T after their dressing was determined.

Table 1: Steel Type 12CrNi3A, chrome-nickel structural alloy steel products GOST 4543-2016.

C, %	Si, %	Mn, %	P, %	S, %	Cr, %	Ni, %
0.09-0.16	0.17-0.37	0.3-0.6	≤0.025	≤0.025	0.6-0.9	2.75-3.15

With the use of scanning electron microscope "Quanta-200" were carried out spectral studies of the state of the surface layer of samples made of steel 12CrNi3A, subjected to grinding by impregnated chromium diiodide abrasive tool.

In [12,16,17,18] we have shown that the most important characteristic of the operational properties of the surface layer material of processed parts is its energy state, which can be estimated by the specific value of the accumulated energy of deformation E_{spec}. For experimental determination of the value of E_{spec} a special unit was developed, the description of which is given in [17, 19]. One of the most important indicators of the effectiveness of the use of grinding wheels is to ensure the necessary dimensional accuracy of the workpieces during the entire period of tool life. In this regard, a controlled treatment was carried out by grinding shafts of various iron-carbon alloys.

Dispersion fields for batch sizes of parts within the grinding wheel resistance were determined. A large number of shafts with diameters from 20 mm to 40 mm and tolerances of IT6-8 were subjected to controlled machining over a long period of time. Based on the obtained results of controlled machining, the scattering fields of diameters of the parts processed without the use of surfactants ω_1 and with the use of various methods of feeding the impregnative agent – chromium diiodide into the machining zone (impregnation, additive to lubricant-and-cooling technological agent, introduction into the abrasive mass) ω_2 were determined. The scattering fields were determined after processing each batch of 10 pieces. The scattering fields were compared when the first batch ω_1^b, ω_2^b was processed and the last batch ω_1^e, ω_2^e was processed.

3. Results

3.1. The effectiveness of the use of chromium diiodide

The results of research into the effectiveness of the use of chromium diiodide impregnated grinding wheels are presented in the form of histograms in Fig. 2, from the analysis of which it follows that the resistance of chromium diiodide impregnated grinding wheels in the process of dressing decreases: first significantly, then on average by 3-5% after each dressing, and then starts to grow again. This is apparently due to the continuous washing out by the lubricating and cooling fluid of the impregnator chromium diiodide from the intergranular space of the grinding wheel.

3.2. The analysis of spectrograms

The analysis of spectrograms of electronic microscope "Quanta-200" of surfaces of details from a steel 12CrNi3A in an

initial condition and after grinding by impregnated chrome diiodide [17] speaks that in the course of grinding active iodine interacts with juvenile surfaces of details, forming on them the iron iodides having layered structure and possessing low coefficient of a friction [9, 20].

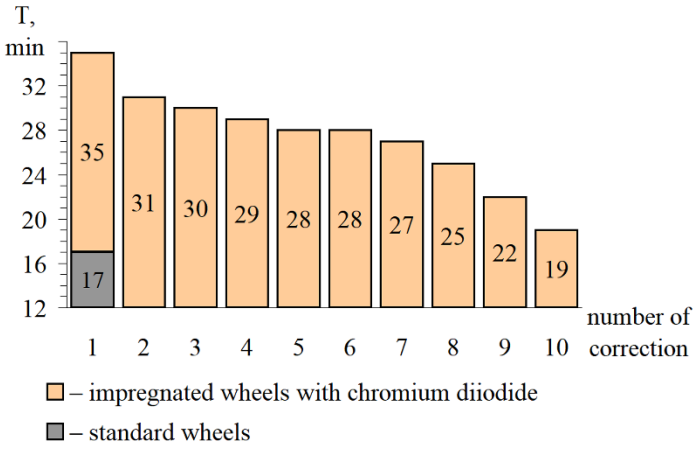


Figure 2: Histograms of changes in resistance of impregnated grinding wheels as they are corrected

3.3. Results of change of scattering fields.

Results of changing scattering fields of sizesits details ω_1 and ω_2 for the period of tool firmness at an admissible range of change of sizes of processed details in the generalized form are presented in fig. 3, which allows to trace the pattern of their change.

It was found that in all cases $\omega^b > \omega^e$, which seems to be associated with a decrease in the number of cutting grains and an increase in smoothing and pressing [21]. This is demonstrated by decreasing in the roughness parameter R_a of the treated surfaces of parts (Fig. 4) and an increasing in the value of technological residual tensile stresses in the material of the surface layer σ_{rem} (Fig. 5) with simultaneous reduction in metal removal.

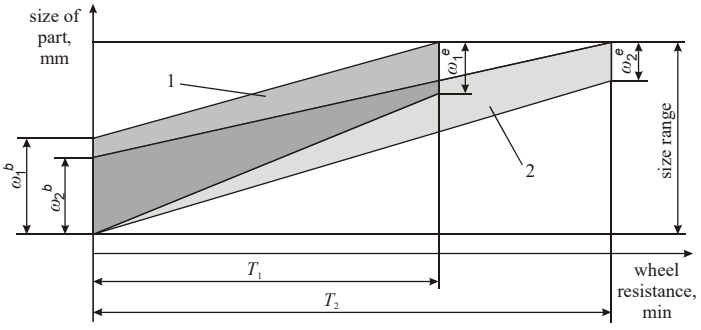


Figure 3: Diagrams of scattering fields for batch sizes of parts within the wheel resistance, processed with standard (1) and impregnated (2) tools.

Analysis of the data shown in Fig. 3, indicates that $\omega_1^b > \omega_2^b$ and $\omega_1^e > \omega_2^e$. If we take $\Delta\omega_1 = \omega_1^b - \omega_2^b$ and $\Delta\omega_2 = \omega_1^e - \omega_2^e$, then $\Delta\omega_1 \geq \Delta\omega_2$.

Using the coefficient $K_\Delta = \Delta\omega_1 / \Delta\omega_2 \geq 1$, then by its value it will be possible to decide on the degree of influence of the adopted type and method of surfactant application on the accuracy of the obtained dimensions of the processed part. Using the data of controlled machining by grinding of workpieces, approximate values of coefficient K_Δ depending on diameter of processed surface, required qualification of size accuracy and application method of impregnator (in this case, chromium diiodide) were determined, which are given in Table 2. These K_Δ values can be used during the selection a surfactant application method.

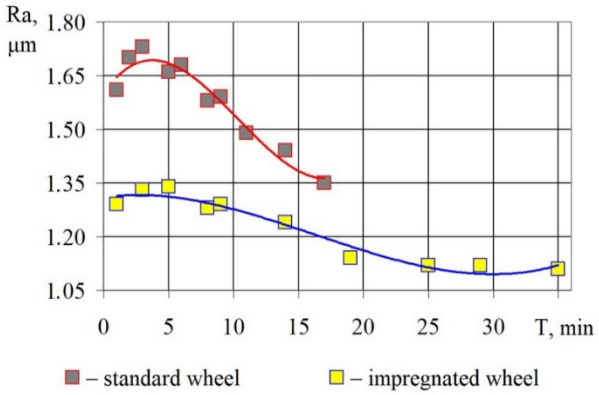


Figure 4: Changing the roughness parameter of the machined surface of the part for the period of the grinding wheel resistance T when using standard and impregnated tools (steel 12CrNi3A)

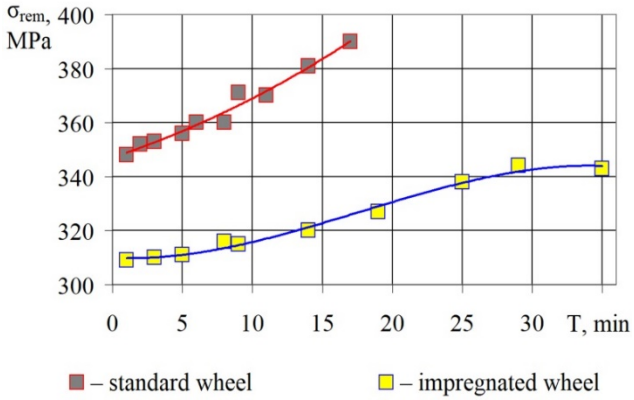


Figure 5: Variation of the value of technological residual stresses in the material of the surface layer of the machines parts for the period of grinding wheel resistance T during the use of standard and impregnated tools (steel 12CrNi3A)

Table 2: Values of coefficient K_Δ by processing parts from iron-carbon alloys with yield point 250-400 MPa

Workpiece diameter, mm	Accuracy qualification	How to use the surfactant	Coefficient K_Δ
20-30	6	Impregnation	1.4
		Additive to cooling lubricant	1.6
		Introduction to abrasive mass	1.5
	7	Impregnation	1.6
		Additive to cooling lubricant	1.7
		An injection into the abrasive mass	1.8
	8	Impregnation	1.9
		Additive to cooling lubricant	1.7
		An injection into the abrasive mass	2.0
31-40	6	Impregnation	1.5
		Additive to cooling lubricant	1.7
		An injection into the	1.6

		abrasive mass	
	7	Impregnation	1.8
		Additive to cooling lubricant	1.7
		An injection into the abrasive mass	1.8
	8	Impregnation	2.0
		Additive to cooling lubricant	1.7
		An injection into the abrasive mass	2.1

To determine the degree of influence of grinding wheel characteristics, its condition and composition of the used cooling lubricant on the specific accumulated energy of deformation E_{spec}, a significant amount of experimental studies on samples from various iron-carbon alloys, subjected to grinding on the regimes recommended in the reference [22], the results of which are reflected in Table 3. From their analysis it follows that the energy state of the material of the surface layer of the workpiece being grinded depends to a great extent on the state of the abrasive tool used and the composition of the cooling lubricant. Regardless of the material to be treated, using of a grinding wheel impregnated with chromium diiodide leads to a reduction in the specific value of the accumulated deformation energy of E_{spec}. At the same time, the injection of 20 grams per liter of chromium diiodide into the composition of the cooling lubricant also leads to a decrease in the value of E_{spec}.

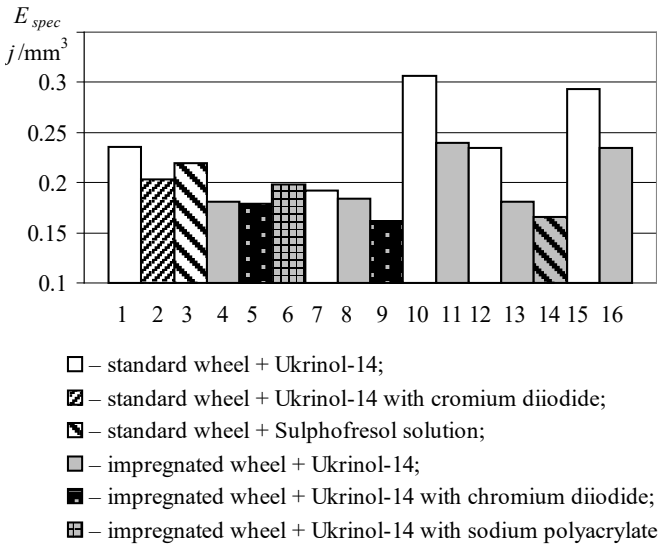


Figure 6: E_{spec} histogram of changes from the state of the grinding wheel and the composition of the used lubricating and cooling process agent

For clarity on the figure 6 is shown a histogram of E_{spec} value changes from the state of the wheel and the composition of the used the cooling lubricant when grinding samples from steel 50 (1-6), steel 12CrNi3A (7-9), alloy 45Cr25Ni20Si2 (10, 11) [23] with a wheel of 14AF60K7V and steel 50 (12-14) and alloy 45Cr25Ni20Si2 (15, 16) when grinding with a wheel of 54AF60 M7V. The histogram shows that the minimum values of E_{spec} can be achieved with the simultaneous use of chromium diiodide impregnated grinding wheel and the cooling lubricant, which consists of 5% aqueous solution of emulsion "Ukrinol-14" with the addition of 20 g/l of chromium diiodide. Apparently in this case, chromium diiodide, which is a part of lubricating and cooling technology, not only reduces the coefficient of friction in the area of contact of the abrasive grain with the processed material, but also contributes to the permanent restoration of chromium diiodide in the impregnated wheel [24-31].

Table 3: Depending of E_{spec} values on the material to be treated, grinding wheel characteristics and the composition of the lubricant and coolant used.

# Ex.	Processed material	Grinding wheel characteristic	Grinding wheel condition	Composition of the lubricant and coolant used	$E_{spec.}$ j / mm^3
1	Steel 50	14AF60K7V	standard	5-% emulsion aqueous solution Ukrinol-14	0.236
2				5-% emulsion aqueous solution Ukrinol-14 + 20 g/l chromium diiodide	0.203
3				5-% sulphofresol solution	0.219
4			impregnated	5-% emulsion aqueous solution Ukrinol-14	0.181
5				5-% emulsion aqueous solution Ukrinol-14 + 20 g/l chromium diiodide	0.179
6				5-% emulsion aqueous solution Ukrinol-14 + 20 g/l sodium polyacrylate	0.198
7	Steel 12CrNi3A		standard	5-% emulsion aqueous solution Ukrinol-14	0.192
8			impregnated	5-% emulsion aqueous solution Ukrinol-14	0.184
9				5-% emulsion aqueous solution Ukrinol-14 + 20 g/l chromium diiodide	0.162
10	Steel		standard	5-% emulsion aqueous solution Ukrinol-14	0.306
11	45Cr25Ni20Si2		impregnated		0.240
12	Steel 50	54AF60M7V	standard		0.235
13			impregnated		0.181
14				5-% sulphofresol solution	0.166
15	Steel		standard	5-% emulsion aqueous solution Ukrinol-14	0.293
16	45Cr25Ni20Si2		impregnated		0.235
17	Steel 50	64AF70P10V	standard		0.230
18			impregnated		0.177
19				5-% sulphofresol solution	0.172
20	Steel		standard	5-% emulsion aqueous solution Ukrinol-14	0.298
21	45Cr25Ni20Si2		impregnated		0.234

4. Conclusions

- The energy state of the material of the surface layer of the grinded part depends largely on the condition of the abrasive tool used and the composition of the lubricant coolant. Regardless of the material to be treated, the use of a grinding wheel impregnated with chromium diiodide leads to a reduction in the specific value of the accumulated deformation energy of E_{spec}. The injection of 20 grams per liter of chromium diiodide into the lubricating and cooling process also leads to a decrease in the value of E_{spec}.
- Thus, the processing of parts with impregnated chromium diiodide wheels allows to increase up to two times the resistance of used abrasive tools, it is quite natural to increase the size accuracy of the processed parts, to improve the

quality of the surface layer and to create on the processed surfaces of parts – rotational bodies, such as rollers, axles, fingers, etc., structures with a low coefficient of friction within 0.08-0.12 depending on the operating conditions, which is very important for the parts of tribosystems.

- Taking into account these results it is possible to predict up to twofold natural reduction of the cost of machining with the abrasive tool of grinding parts, which will reduce the cost of their manufacture to 20-25%. The developed process of impregnating grinding wheels with chromium diiodide can be used in any machine-building production, including repair, without large economic costs.

Nomenclature

E_{spec} – specific latent energy of deformation j/mm^3;
K_Δ – coefficient $K_\Delta = \Delta\omega_1 / \Delta\omega_2$;
R_a – arithmetic mean deviation of the profile (μM);
V_{gw} – cutting speed, m/s;
V_m – rotation speed, m/s;
S_{lon} – longitudinal feed, m/s;
S_{cut} – cutting feed, mm / double stroke;
Ukrinol-14 – Cooling and lubricating fluids (oil);
ω_1 – the scattering fields of diameters of the parts processed without the use of surfactants;
ω_2 – the scattering fields of diameters of the parts processed with the use of the impregnation.

14AF60K7V GOST R 5281-2007 – Grinding wheel Russian standard is adopted from the European, code example*:*

14A*F60K7V* (type of abrasive, ***14A*** – normal electro corundum);
*14A****F60****K7V* (***F*** **–** standard FEP (Federation of European Producers of abrasives) and ISO 8486, **F60** average grain size is 0.26 mm);
*14AF60****K****7V* (Hardness, ***K*** – medium soft);
*14AF60K****7****V* (***7*** – structure number indicates how much of the grinding wheel volume is occupied by abrasive grains, **7** – 48...50%);
*14AF60K7****V*** (type of bundle, ***V*** – a ceramic bundle).

References

[1] Y. Rudometov,. "Application of the abrasive instruments impregnated with impregnator suspensions," (in Russian) STIN, **11,** 34-37, 2012.

[2] W. Kapłonek, K. Nadolny, M. Sutowska, M.Mia, D.Y. Pimenov, M.K. Gupta, "Experimental Studies on MoS2-Treated Grinding Wheel Active Surface Condition after High-Efficiency Internal Cylindrical Grinding Process of INCONEL® Alloy 718," Micromachines **10**(4), 255, 2019, doi.org/10.3390/mi10040255

[3] G.V. Chirkov, "Impact of grinding wheel impregnation on processing quality," Engineering Technology, **2**, 22-23, 2007.

[4] A.V. Nikitin, "Grinding of Hard Materials by Impregnated Wheels as a Way to Increase Their Cutting Properties," Tools and Technologies, **28**, 52-58, 2010.

[5] W. Kapłonek, K. Nadolny, K. Rokosz, J. Marciano, M. Mia, D.Y. Pimenov, O. Kulik, M.K. Gupta, "Internal Cylindrical Grinding Process of INCONEL® Alloy 600 Using Grinding Wheels with Sol–Gel Alumina and a Synthetic Organosilicon Polymer-Based Impregnate," Micromachines **11**(2), 115, 2020, doi.org/10.3390/mi11020115

[6] M.-Y. Tsai S.-X. Jian "Development of a micro-graphite impregnated grinding wheel," International Journal of Machine Tools and Manufacture, **56**, 94-101, 2012, doi.org10.1016/j.ijmachtools.2012.01.007

[7] V.A. Nosenko, A.P. Mitrofanov, A.A. Krutikova. "Increase of Grinding Efficiency with the Use of Halogen-forming Impregnator," Izvestia vysokhraneniya. Mechanical Engineering. **8** (665), 65-72, 2015, doi.org/10.18698/0536-1044-2013-1-50-54

[8] Morrison R., Boyd R. Organic Chemistry, Prentice Hall India, 2002.

[9] V.I. Butenko, Finishing of the parts surfaces: methods, devices, tools (in Russian), Rostov-on-Don: DSTU Publishing Centre, 2016.

[10] V.I. Butenko, L.V. Gusakova. Increase of the machine parts surface grinding efficiency (in Russian) Taganrog: SFU Publishing House, 2012

[11] "Royal Society of Chemistry APIs" [Chromium(II) iodide]. Available Online at: http://www.chemspider.com/Chemical-Structure.13318420.html

[12] V.I. Butenko, L.V. Gusakova, A.D. Kulinskiy "Perspective directions of increase of efficiency of grinding of surfaces of machine parts," Vestnik of Bryansk State Technical University, **2**(50), 112-120, 2016, doi.org/10.12737/20254

[13] H.T. Nguyen, Q.C. Hsu, "Surface Roughness Analysis in the Hard Milling of JIS SKD61 Alloy Steel," Applied Sciences, **6,** 172, 2016, doi.org/10.3390/app6060172

[14] V.I. Butenko, D.S. Durov, R.G. Shapovalov. "Science-intensive technologies for creation of the high resource machine parts," Taganrog SFU Publishing House, 2014.

[15] E. Benedicto, D. Carou, E.M. Rubio, "Technical, Economic and Environmental Review of the Lubrication/Cooling Systems Used in Machining Processes," Procedia Engineering, **184**, 99-116, 2017, doi.org/10.1016 / j. proeng.2017.04.075

[16] V.I. Butenko, Structure and potential of the polycrystalline materials tribocontact (in Russian), Rostov-on-Don: DSTU Publishing Centre, 2018.

[17] V.I. Butenko, Technological compatibility of the functional layers and coatings (in Russian), Rostov-on-Don, 2020.

[18] V.A. Lebedev, Technology of the dynamic methods of the surface plastic deformation (in Russian), Rostov-on-Don: DSTU Publishing Center, 2006.

[19] W. Li, Q Tian, B. Zhang. "Impregnant for impregnated grinding wheels and application thereof in preparing impregnated grinding wheels," Patent no. 101791786 CN, B 24 D 3/34. 2010.

[20] V.I. Butenko, K.V. Davidenko "Method of impregnation of abrasive tools," Patent RF № 2703063. MPK B24 D 3/34/– Bulletin 29, 2019.

[21] A.A. Ryzhkin, Physical bases of the metal processing by cutting: a training manual (in Russian) Rostov-on-Don: Publishing centre. DSTU, 1996.

[22] A.M. Dalskiy, A.G. Suslov, A.G. Kosilov "Directory of the machine-building technologist" Moscow: Mashinostroenie-1, 2003.

[23] F. Ju "Special Steel Co.," Available Online at: https://www.jfs-steel.com/ru/steelDetail/-1Cr25Ni20-Si2/GB-1Cr25Ni20-Si2.html

[24] B.K. Sato, J.C. Lopes, A.E. Diniz, A.R. Rodrigues, H.J. de Mello, L.E.A. Sanchez, P.R. Aguiar, E.C. Bianchi, "Toward sustainable grinding using minimum quantity lubrication technique with diluted oil and simultaneous wheel cleaning," Tribology International, **147**:106276, 2020, doi.org/10.1016/j.triboint.2020.106276.

[25] S. Agarwal, "On the mechanism and mechanics of wheel loading in grinding," Journal of Manufacturing Processes, **41**, 36-47, 2019, doi.org/10.1016/j.jmapro.2019.03.009

[26] H. González, A. Calleja, O. Pereira, N. Ortega, L.N. López de Lacalle, M. Barton, "Super Abrasive Machining of Integral Rotary Components Using Grinding Flank Tools," Metals, **8**, 24, 2018, oi.org/10.3390/met8010024

[27] K. Nadolny, M. Wojtewicz, W. Sienicki, D. Herman, "An analysis of centrifugal MQL supply system potential in the internal cylindrical grinding process," Archives of Civil and Mechanical Engineering 15, 639–649, 2015, doi.org/10.1016/j.acme.2014.08.009

[28] M. Wojtewicz, W. Sienicki, K. Nadolny, "Vitrified grinding wheels impregnation methods with use solid lubricants," Mechanik. **September** 1108-1109, 2016, doi.org/10.17814/mechanik.2016.8-9.271.

[29] B.P. Ravuri, B.K. Goriparthi, R.S. Revuru, V.G. Anne, "Performance evaluation of grinding wheels impregnated with graphene nanoplatelets," The International Journal of Advanced Manufacturing Technology, **85**(9), 2235–2245, 2016, doi.org/10.1007/s00170-015-7459-6.

[30] R.B. Pavan, A.V. Gopal, M. Amrita, B.K. Goriparthi "Experimental investigation of graphene nanoplatelets–based minimum quantity lubrication in grinding Inconel 718", Proceedings of the Institution of Mechanical Engineers, Part B, Journal of Engineering Manufacture. **233**(2), 400-410, 2019, doi.org/10.1177/0954405417728311

[31] W. Kapłonek, K. Nadolny, M. Wojtewicz, W. Sienicki, "Characterisation of abrasive tools active surface after the impregnation process by modified ARS method based on imaging and analysis of the scattered light". International Journal of Machining and Machinability of Materials, **17**(5), 397 – 417, 2015, doi.org/10.1504/IJMMM.2015.073149

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 403-407 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Level of Resilience and Family Functionality in Adolescents of two Educational Institutions of a Vulnerable Area in Lima Province

Rosa Perez-Siguas[1], Hernan Matta-Solis[1], Eduardo Matta-Solis[1], Melissa Yauri-Machaca[2], Anika Remuzgo-Artezano[3], Lourdes Matta-Zamudio[4]

[1]*Research and Intellectual Creativity Direction, Universidad María Auxiliadora, Lima, 15408, Perú*

[2]*Research and Technology Direction, BM Tech, Lima, 15408, Perú*

[3]*Faculty of Health Sciences, Universidad María Auxiliadora, Lima, 15408, Perú*

[4]*Health Sciences, Instituto Peruano de Salud Familiar, Lima, 15408, Perú*

ARTICLE INFO

Article history:
Received: 04 August, 2020
Accepted: 28 October, 2020
Online: 20 November, 2020

Keywords:
Resilience
Family
Adolescent
Family functionality

ABSTRACT

Family functionality is when the family meets the needs of its members, especially adolescents. Resilience is the adolescent's ability to face situations in which there are difficulties, life problems and thus be able to overcome them. The purpose of the study is to determine the Level of Resilience and Family Functionality in Adolescents of two Educational Institutions of a Vulnerable Area in Lima Province. It is a non-experimental study, descriptive of a quantitative approach, it is a cross-sectional correlational investigation, with a population of 204 adolescents, also demographic data and the resilience scale (CD-RISC) and family APGAR. In the results regarding to the level of resilience and family functionality, adolescents present a mild family dysfunction with a medium level of resilience (58.0%), in comparison of the Educational Institutions, in both cases, the majority of students are located in the medium level of resilience, in the case of students from Honorio Manrique Nicho school 48.5% and in the case of students from the Andrés Avelino Cáceres school 54.2%. It is recommended in these institutions to carry out activities that promote the social formation of the adolescent, since this will allow them to communicate with society and the environment.

1. Introduction

Family plays a fundamental role in the adolescent, since it is an influential factor for the relationship with society [1]. One of the stages where one seeks to find changes in society is adolescence, adolescents try to know each other better and also seek difference from themselves with other family members, this process that the adolescent does is a change abruptly, but it is the way in which they will become autonomous and independent people, this process has passed all the members of the family [2].

Family functionality is understood when the family meets the needs of its members, especially adolescents, since it is a stage where there are differentiated changes with other family members and also adapt to situations of change [3]. So, a family is considered functional when it is able to look for solutions to any problem, so that it does not affect the satisfaction that family members have, including the adolescent [4].

Resilience is understood as the adolescent's ability to face situations in which there are difficulties, life problems and thus be able to overcome them [5]. Resilience is very important for adolescents, because it will allow them to adapt and recover from adversities of the problem so that it continues projecting itself in its future [6].

In [7], the author from Portugal presented that it was evidenced that adolescent males are those who have higher levels of resilience compared to adolescent females; in the self-

*Corresponding Author: Rosa Perez-Siguas, Mrs., +51 1 996826968 & rosa.perez@uma.edu.pe

confidence component, males obtained a score of 35.08% and in the relationship with peers, 26.28%.

In [8], carried out in Taiwan, the author mentioned that in children and adolescents, it was found that there is a risk of family dysfunction where approximately 40% of families have accumulated family risk due to family stress that occurs within their home and also 30% of the mothers have low levels of education or health problems. Lastly, around 7% of the participants, children or adolescents, come from large families that have 4 or more children living in a home.

In [9], carried out in South Korea, the author described that it was found that family dysfunction related to domestic violence and parental addiction was related to smartphone addiction, it was found that the observed association between family dysfunction and Smartphone addiction was not significant in adolescents with self-control or high friendship quality, therefore, adolescents exposed to domestic violence or with addicted parents had a higher probability of smartphone addiction, but had high self-control or friendship quality that could protect them.

The objective of the study is to determine the level of family resilience and functionality in adolescents from two educational institutions of a vulnerable area in Lima Province, in which it will allow to observe what is the level of adolescent resilience and family functionality. This study is important since it will give relevant and real data about family resilience and functionality as adolescents enter a stage of sudden changes where they seek solutions, where they can adapt and manage the situation in the most appropriate way.

In the study, the data collection instrument Connor-Davidson Resilience Scale (CD-RISC) and Gabriel Smilkstein's Family APGAR were applied, which have been useful to assess the level of resilience and family functionality of adolescents from the secondary level of educational institutions. The data collection was processed through the survey of adolescents from educational institutions, the data to be entered was performed in a data matrix that will be designed in the statistical program SPSS (Statistical Package for the Social Sciences) in its version 24.0, in which it will allow a better data processing to make statistical tables so that they can be described and interpreted in results and discussions, respectively.

The research work is structured as follows: In section II, the development of the data collection process of each nursing student will be presented, as well as the guidelines to consider so that they are in the research work. In section III, the results will show the level of resilience and family functionality of adolescents according to the specified dimensions of the instruments in the measurement of the variables. In section IV, we present the discussions and conclusions of the research work.

2. Methods

In this section, the type and design of the research will be developed, as well as the population that will be carried out in the research work, the inclusion and exclusion criteria will be explained in detail, and finally the technique and the instrument for collecting data.

2.1. Research type and design

The present research work is non-experimental, descriptive of a quantitative approach, it is a correlational research [10].

2.2. Population

The population is made up of 204 adolescents from the two educational institutions.

Inclusion criteria:

- Adolescents who normally attend the educational institution.
- Adolescents who have informed consent and assent.
- Adolescents who have the authorization signed by their parents and agree to voluntarily join the study.

Exclusion criteria:

- Adolescents who do not regularly attend to the educational institution.
- Adolescents who do not have the authorization of their parents.

2.3. Technique and Instrument

The technique used is the survey, using the questionnaire or data collection instrument (CD-RISC) and Family APGAR which aims to measure the level of family resilience and functionality in adolescents from two educational institutions of a vulnerable area in Lima Province.

The Connor - Davidson Resilience Scale (CD-RISC) was built by Connor and Davidson in the United States in 2003. Its application is for the adolescent and adult population from 11 to 80 years old, the scale evaluates the following dimensions: Persistence-tenacity-self-efficiency (8 items); Control under pressure (7 items); Adaptability and Ability to Recover (5 items); Control and purpose (3 items) and Spirituality (2 items). It consists of 25 structured items, on a summative Likert scale where 0 "absolutely", 1 "rarely", 2 "sometimes", 3 "often" and 4 "almost always". The scale range goes from 0 to 100, the higher score means more resilience [11].

The family APGAR was designed by Gabriel Smilkstein in 1978 at the University of Washington, it evaluates the following dimensions: adaptation, participation, resources, affectivity and capacity, it consists of 5 items structured on a summative Likert scale where 0 "never", 1 "almost never", 2 "sometimes", 3 "almost always" and 4 "always". The range is from 0 to 20, the higher the score mean greater the family functionality [1].

2.4. Place and Application of the Instrument

The questionnaire was carried out to measure the level of family resilience and functionality in adolescents, it was carried out at the educational institutions Andres Avelino Caceres and Honorio Manrique Nicho.

First, the necessary permissions for research are carried out, coordinated with the directors of both educational institutions to carry out the surveys, then, the permission to each classroom teacher explaining about the survey and why the research work is being carried out to they have knowledge about what is going to be done.

Table 1: Resilience level in adolescent students of both Educational Institutions of Lima Province (N=204)

		Resilience Score				
		Count	%	Mean	Minimum	Maximum
Resilience Dimension (grouped)	Low	46	22,5%	52,28	32,00	60,00
	Medium	105	51,5%	70,28	61,00	79,00
	High	53	26,0%	85,75	80,00	95,00
	Total	204	100,0%	70,24	32,00	95,00

Table 2: Family functionality of adolescent students of both Educational Institutions of a Vulnerable Area in Lima Province (N = 204)

		Family Functionality Score				
		Count	%	Mean	Minimum	Maximum
Family Functionality Dimension (grouped)	Family dysfunction	10	4,9%	7,10	2,00	9,00
	Moderate family dysfunction	29	14,2%	11,90	10,00	13,00
	Mild family dysfunction	100	49,0%	15,59	14,00	17,00
	Good family function	65	31,9%	19,15	18,00	20,00
	Total	204	100,0%	15,78	2,00	20,00

The questionnaire was carried out in consecutive shifts during the morning, at the same time as the adolescents from these institutions, the questionnaire was taken to the adolescents with an approximate time of 25 minutes in each section of secondary level (according to the inclusion and exclusion criteria) in the research work, concluding with good satisfaction when collecting the questionnaires as the adolescents supported our research.

It is important to emphasize the presence of the teacher from each surveyed classroom and the medical staff at the time of filling questionnaire of the adolescent, since most of them showed signs of having a health problem, because both educational institutions are in an area where they are exposed to natural or anthropogenic events where the probability of an event that puts adolescents in educational institutions at risk can happen and also the damage can be considerable.

3. Results and Discussions

The following tables from the surveys carried out following the guidelines corresponding to the research work will be shown in table 1.

In Table 1, the level of resilience in adolescents of both Educational Institutions of a Vulnerable Area in Lima Province, it can be seen that of the 204 adolescents surveyed, 105 (51.5%) of adolescents present a medium level of resilience, followed by 53 (26.0%) of the adolescents present a high level of resilience and lastly 46 (22.5%) of the adolescents present a low level of resilience.

In Table 2, the family functionality of the adolescents of both Educational Institutions of a Vulnerable Area in Lima Province, it can be seen that of the 204 adolescents, 100 (49%) of the adolescents present a mild family dysfunction, followed by 65 (31.9 %) adolescents present a good family function, 29 (14.2%) of the adolescents present a moderate family dysfunction and 10 (4.9%) of the adolescents present a family dysfunction.

Table 3 shows that adolescents with good family function have a high level of resilience (49.22%), adolescents with mild family dysfunction have a medium level of resilience (58.0%), adolescents with a moderate family dysfunction has a medium level of resilience (55.2%) and adolescents with family dysfunction have a low level of resilience (60.0%).

To determine the level of resilience and family functionality, the non-parametric Person's chi-square test was applied, which reported an $X^2 = 44.842$ and a p-value = 0.000 per l indicating that there is a significant relationship between these variables.

In Table 4, both cases, most students are located at the medium level of resilience, in the case of students from Honorio Manrique Nicho, 48.5%, and in the case of students from the Andrés Avelino Cáceres school, 54.2%. So, students with a high level of resilience are located at the Honorio Manrique Nicho school with 30.9% compared to students from Andrés Avelino Cáceres who are only 21.5%.

Both studies with the results of the research work can interpret that the levels of resilience and family functionality in adolescents

Table 3: Relationship between Resilience and Family Functionality in adolescent students of both Educational Institutions of a Vulnerable Area in Lima Province (N = 204)

		Resilience Dimension (grouped)							
		Low		Mild		High		Total	
		Count	%	Count	%	Count	%	Count	%
Family Functionality Dimension (grouped)	Family dysfunction	6	60,0%	4	40,0%	0	0,0%	10	100,0%
	Moderate family dysfunction	13	44,8%	16	55,2%	0	0,0%	29	100,0%
	Mild family dysfunction	21	21,0%	58	58,0%	21	21,0%	100	100,0%
	Good family function	6	9,2%	27	41,5%	32	49,2%	65	100,0%
	Total	46	22,5%	105	51,5%	53	26,0%	204	100,0%

Pearson's chi-square tests		
Family Functionality Dimension (grouped)	Chi-square	44,842
	df	6
	Sig.	,000*

Results are based on nonempty rows and columns in each innermost sub table.
*The chi-square statistic is significant at the .05 level.

Table 4: Comparison of Resilience Level of adolescent students between both Educational Institutions of a Vulnerable Area in Lima Province (N = 204)

		Evaluated School					
		Honorio Manrique Nicho		Andrés Avelino Cáceres		Total	
		Count	%	Count	%	Count	%
Resilience Dimension (grouped)	Low	20	20,6%	26	24,3%	46	22,5%
	Medium	47	48,5%	58	54,2%	105	51,5%
	High	30	30,9%	23	21,5%	53	26,0%
	Total	97	100,0%	107	100,0%	204	100,0%

are linked to the promotion of mental and family health of adolescents, taking into account that resilience is very relevant because it allows adolescents to have the ability to adapt and also in making decisions about the negative situations that arise. It also focuses on family functionality because in the adolescent, it will allow it to maintain an affective interaction with the members of its family, and also to has the ability to make changes in order to overcome the difficulties that arise in the family. It is vitally important to have a high level of resilience and, have good family functionality since the development of adolescents must be observed, since there is little knowledge about the topics which allows us to obtain a review of the scientific information related to the topic.

In the present study, the issue of level of family resilience and functionality is raised from the point of view of promoting adolescent health, which seeks to contribute to educational institutions as part of adolescent training, which it seeks to include programs that allow adolescents to find ways to enhance their adaptation capacities and make their own decisions, both in society and family, so that they can improve their lives in the future.

In the level of resilience in adolescents from educational institutions, the medium level predominated, followed by the high and low. These results can be based on how adolescents live with their parents sharing their own experiences, in which it is a factor that favors the adolescent since it allows them to easily excel and overcome any anticipated problem. In the study by Sulimani Y. [12], on the results of resilience, he referred to the adolescent's distinction between formal and informal social ties and, also, the ability to seek help and guidance from adult figures, as well as strong relationships and supporting parents and family members is a debatable factor as a major component of resilience for adolescents.

In relation to family functionality in adolescents from Cajatambo educational institutions, mild family dysfunction predominated, this is due to the lack of empathy, understanding and sensitivity towards the adolescent by the family is scarce, where scenes such as intra-family conflicts and parental misbehavior are reflected in the adolescent adapting to such actions, allowing the adolescent to think that this normally happens in all other families. Suarez P. and collaborators [2], argue that the development of values and actions of adolescents will be reflected in the role played by the family in their environment, factors such as care, affectivity and fluid communication are criteria that support the adolescent development.

In relation to the level of resilience and family functionality in adolescents from Cajatambo's educational institutions, mild family dysfunction predominated with a medium level of resilience, it is important to emphasize the intervention of the family in the development of resilience in adolescents, since it is mainly the parents who will help build the adolescent's resilience by providing the necessary support that will allow them to overcome certain obstacles. The family will influence positively or negatively in the construction of the adolescent's own resilience, but some adolescents mostly present a high resilience. Tian L. et al. [13], mention that parental support was more strongly related to adolescent resilience, and that interventions should be made to promote adolescent resilience with greater emphasis on parental support in the future of their children.

In comparison the level of resilience between the two Educational Institutions in Lima Province, it can be seen that in the Honorio Manrique Nicho Educational Institution and in the Andres Avelino Caceres Educational Institute, the middle level predominates more, this is because after the family as Resilience builder is also the school, at school the adolescents experiences certain conditions that make them promote resilience, because the school also has the capacity to build it in the adolescent. The way in which resilience can be built is based on the trust that the adolescent has in front of the parents, teachers or friends that generate security to solve any obstacle without doubts.

It is concluded that the particular way in which adolescent education is carried out, whether in the family or at school, should be taken into account; it could be investigated about resilience and family function through other studies to determine this study in greater depth.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

[1] L. Higuita-Gutiérrez, J. Cardona-Arias, "Percepción de funcionalidad familiar en adolescentes escolarizados en instituciones educativas públicas de Medellín (Colombia), 2014," Revista CES Psicología, **9**(2), 167–178, 2016, doi:10.21615/cesp.9.2.11.

[2] P. Suárez, M. Vélez, "El papel de la familia en el desarrollo social del niño: una mirada desde la afectividad, la comunicación familiar y estilos de educación parental," Revista Psicoespacios, **12**(20), 173–198, 2018.

[3] D. Paternina, M. Pereira, "Funcionalidad familiar en escolares con comportamientos de riesgo psicosocial en una institución educativa de Sincelejo (Colombia)," Salud Uninorte, **33**(3), 429–437, 2017, doi:10.14482/sun.33.3.10932.

[4] A. Adebayo, E. Ajuonu, B. Betiku, "Family functionality and parental characteristics as determinants of sexual decision-making of in-school youths in a semi-urban area of Southwest Nigeria," International Journal of Adolescent Medicine and Health, **28**(4), 413–418, 2016, doi:10.1515/ijamh-2015-0029.

[5] H. Bravo, J. López, "Resiliencia familiar: una revisión sobre artículos publicados en español," Revista de Investigación En Psicología, **18**(2), 172, 2015, doi:10.15381/rinvp.v18i2.12089.

[6] M. Simón-Saiz, R. Fuentes-Chacón, M. Garrido-Abejar, M. Serrano-Parra, E. Larrañaga-Rubio, S. Yubero-Jiménez, "Influence of resilience on health-related quality of life in adolescents," Enfermeria Clinica, **28**(5), 283–291, 2018, doi:10.1016/j.enfcli.2018.06.003.

[7] C. Albuquerque, J. Almeida, M. Cunha, A. Madureira, A. Andrade, "Protective Resilience Factors in Institutionalised Portuguese Adolescents," Procedia - Social and Behavioral Sciences, **171**(1), 276–283, 2015, doi:10.1016/j.sbspro.2015.01.122.

[8] L. Yi-Ching, S. Dong-Chul, "Cumulative family risks across income levels predict deterioration of children's general health during childhood and adolescence," PLoS ONE, **12**(5), 1–13, 2017, doi:10.1371/journal.pone.0177531.

[9] K. Hye-Jin, M. Jin-Young, M. Kyoung-Bok, L. Tae-Jin, Y. Seunghyun, "Relationship among family environment, self-control, friendship quality, and adolescents' smartphone addiction in South Korea: Findings from nationwide data," PLOS ONE, **13**(2), 1–13, 2018, doi:10.1371/journal.pone.0190896.

[10] C. Fernández, P. Baptista, Metodología de la Investigación, 6ta edición, Mc Graw Hill Education/Interamericana, México, 2014.

[11] G. Guihard, L. Deumier, B. Alliot-Licht, L. Bouton-Kelly, C. Michaut, F. Quilliot, "Psychometric validation of the French version of the Connor-Davidson Resilience Scale," Encephale, **44**(1), 40–45, 2018, doi:10.1016/j.encep.2017.06.002.

[12] Y. Sulimani-Aidan, "Promoting resilience among adolescents in care from their social workers' perspectives," Children and Youth Services Review, **94**(1), 43–48, 2018, doi:10.1016/j.childyouth.2018.09.005.

[13] L. Tian, L. Liu, N. Shan, "Parent–Child Relationships and Resilience Among Chinese Adolescents: The Mediating Role of Self-Esteem," Frontiers in Psychology, **9**(1), 1–11, 2018, doi:10.3389/fpsyg.2018.01030.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 408-415 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

To the Question of Multi-Criteria Optimization of Aircraft Components in Order to Optimize its Life Cycle

Sergey Alekseevich Serebryansky*, Alexander Vladimirovich Barabanov

Department Aircraft Design and Certification, Moscow Aviation Institute (National Research University), Moscow, 125993, Russia

ARTICLE INFO

Article history:
Received: 20 August, 2020
Accepted: 21 October, 2020
Online: 20 November, 2020

Keywords:
Complex Methodology
Optimization of The Design Cycle
Aerodynamic Drag
Rational Parameters
Nose of The Aircraft

ABSTRACT

The paper examines the impact on the product life cycle of reducing the duration of the development stages of aviation technology through the use of multi-criteria optimization. The subject of the study is the nose of a supersonic aircraft and the process of linking it. This example describes in detail the application of a comprehensive design methodology for aircraft components using top-level criteria as a way to significantly reduce the development time and the risks of obtaining negative test results when entering a full-scale sample.

1. Introduction

The great Russian aircraft designer Pavel Osipovich Sukhoi said: "We must not stop at what we have achieved, but must constantly move forward. After all, what was new and advanced yesterday will only be satisfactory tomorrow." This statement is still relevant until now because to take leadership positions in aviation, engineering, Economics, and business, you must constantly work ahead of time and your competitors. The current trend in the field of mechanical engineering is a significant (from 10-15 to 5-8 years) reduction in the development time of new aircraft systems due to the use of mathematical modeling, high unification and the use of scientific and technical reserve. The use of complex design techniques is one of the elements based on scientific and technical groundwork and allows you to reduce the design time.

2. Aircraft product life cycle

Within the lifecycle of the aircraft as a product, the aircraft goes through the following stages (gates):

Figure 1 shows the typical life cycle of an aircraft product, which is divided into internal and external design stages. The focus of the current work is optimization of internal design stages, so we will analyze them in more detail.

It is worth noting that key technical decisions are made at the stages of the preliminary project and preliminary technical design. The work is carried out by a relatively small number of OKB employees, the minimum amount of financial resources from the project budget is used, and the maximum number of technical decisions is made as shown in Figure 2. According to the results of these stages, the appearance of the designed aircraft is fixed.

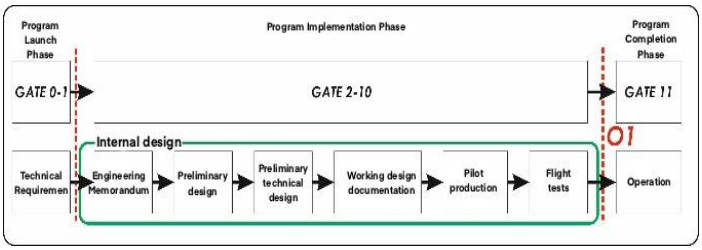


Figure 1: The typical lifecycle of the aviation system

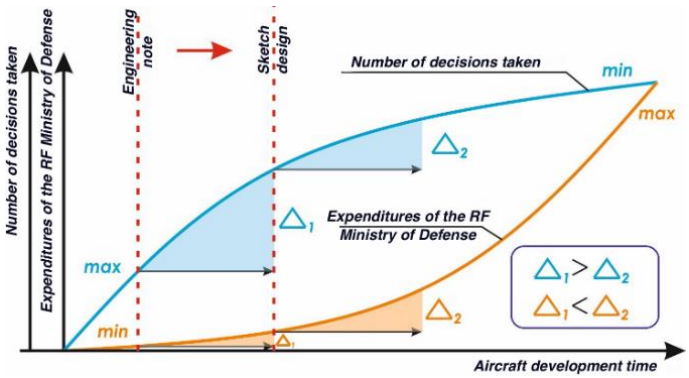


Figure 2: Dependence of the number of decisions made and the customer's costs on the project implementation time

*Corresponding Author: Sergey Alekseevich Serebryansky, s-s-alex@mail.ru

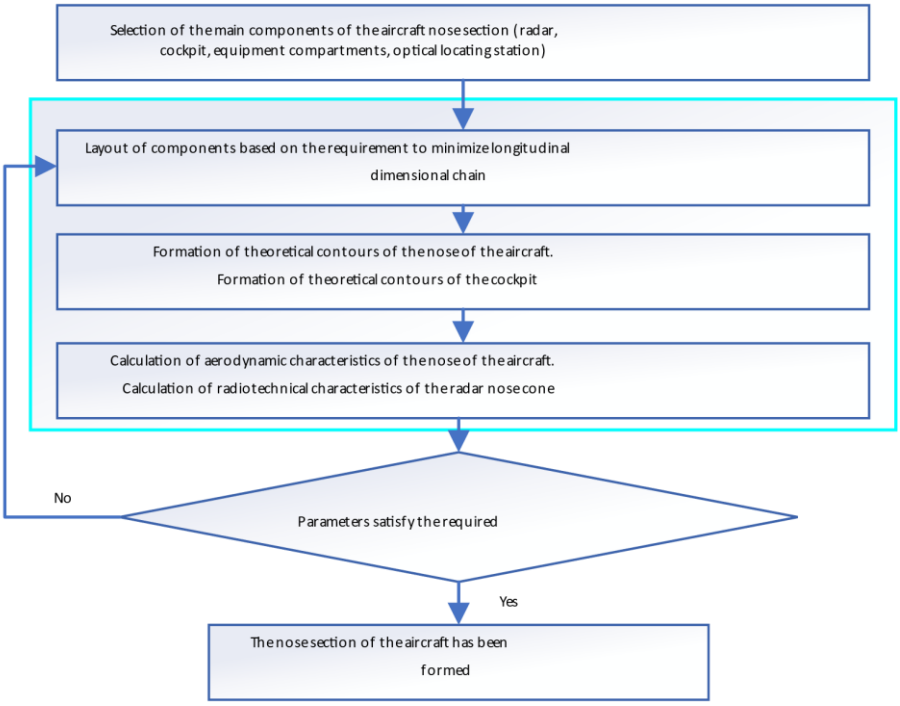


Figure 3: Typical algorithm for developing the nose of the aircraft

Thus, the stages of project initiation are the most effective in terms of cost and number of technical decisions:

$$\left.\begin{matrix}\Delta_i^{expenses}\\ \Delta_i^{time}\end{matrix}\right\} \rightarrow \min(opt)$$

However, these stages have the greatest technical risks, and errors in the technical decisions made may become fatal for the further fate of the project.

As can be seen from Figure 2, the initial stages of project implementation are the most effective in terms of the ratio of the customer's costs to the number of decisions made. Simultaneously, the technical decisions taken at these stages determine the further technical appearance of the project and can significantly affect its results.

Figure 3 shows an algorithm for linking the nose of the aircraft, consisting of sequential operations and, in case of errors, having an iterative character.

To complete the formation of the theoretical contour of the head part of the aircraft it must validate the obtained electronic and aerodynamic characteristics of the selected form of the radio waves fairing requirements to ensure the characteristics of the radar. Meanwhile, the speed and range characteristics of the aircraft will depend on the aerodynamic characteristics in the future, and the potential capabilities of the locator will depend on the radio technology. Thus, such a seemingly insignificant thing as the shape of the nose of an aircraft can determine its key characteristics of combat use. Moreover, the shape of the nose part of the minimum resistance and the shape of the maximum radar-transparency of a thing are incompatible because the best radio-technical characteristics are usually of small elongation nosecone (the ratio of the square of the length of the nosecone to the area of the initial cross-section), such a shape allows you to ensure the uniformity of the distribution of the angles of incidence of the electromagnetic wave front on the surface of the nosecone wall, the best aerodynamic characteristics on the contrary have the body with maximum elongation, it allows reducing significantly the resistance at supersonic flight speeds, see Figure 4:

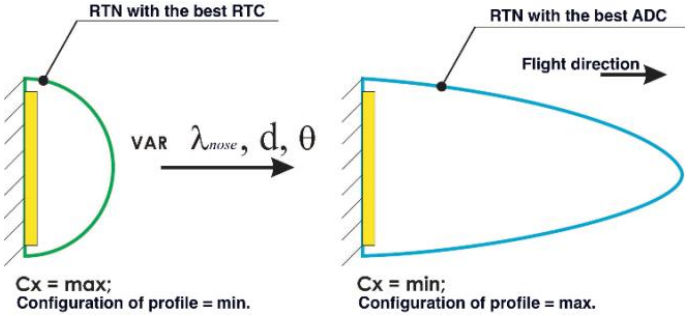


Figure 4: Comparison of nosecone shapes of the aircraft

From Figure 4 we can conclude that the final shape of the nosecone will be a compromise between the two shown in the figure. In this case, the wrong form at this stage may lead to the need to repeat the whole cycle of flight tests in case of its change after the construction of prototypes (stages of pilot production and flight tests).

3. Optimization of the aircraft life cycle

The scientific novelty of the work consists in carrying out multi-criteria optimization of the geometric parameters of the studied RPO using the Russian-made LOGOS software product and in terms of functions, which is an analog of the FLUENT product. The results of mathematical modeling and evaluation of aerodynamic characteristics have been validated with existing empirical methods, and the software product itself is widely used in Russian industrial projects.

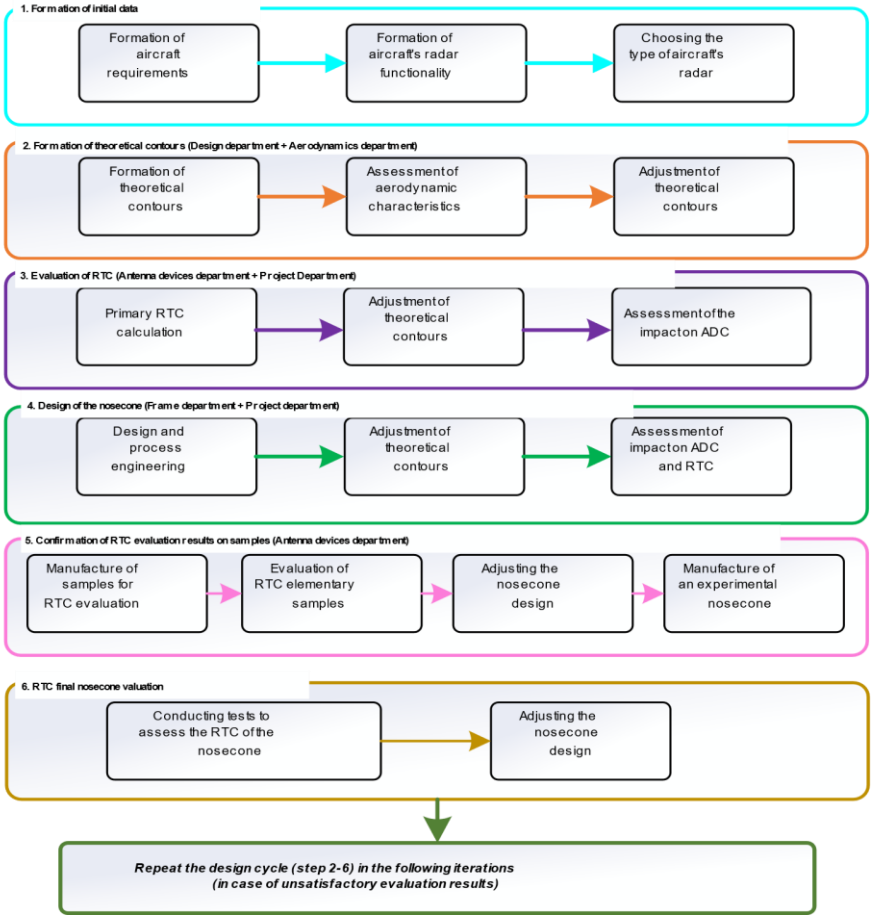


Figure 5: The existing algorithm for the development of the fairing

Reducing the timing of development work can be achieved by reducing the development and testing of the aircraft complex. Concurrently, correct and timely technical decisions made at the development stage will also reduce the time required for ground and flight tests. Let's look at an example of optimizing the development stages of a nose radio-transparent fairing (RPO). The current state of the RPO design algorithm is shown in Figure 5.

After analyzing Figure 4, we can conclude that the design of the radio-transparent fairing according to the existing methodology does not meet the requirements for the development of new aircraft systems. To meet the customer's requirements for the development of new AK in 5-8 years, it is necessary to significantly optimize this algorithm. To do this, you must: To optimize this algorithm, it is necessary:

1) In accordance with [1], [2] select a comprehensive criterion for evaluating the characteristics of RPO radar, both in terms of radio engineering (RTC) and aerodynamic (ADC) characteristics;

2) Form a research algorithm, in which to establish the relationship between the parameters of the shape of the fairing and its ADC and RTC.

3) Based on the results of the study, formulate technical recommendations for choosing rational parameters of the fairing shape, depending on the initial data and technical requirements.

As a complex criterion of the upper level (1), having analyzed the scientific and technical background in the field of research of radio technical characteristics (RTC) [3] and aerodynamic characteristics (ADC) [4], [5], [6], it is possible to take the ratio of the criterion of the first level of aerodynamic characteristics to the criterion of the first level of radio engineering characteristics:

$$X^{*} = ArgMax\left(\frac{ArgMax(RTC)}{ArgMin(ADC)}\right) \quad (1)$$

In this case, we will choose as the criterion of the first level for RTC nosecone its transmission coefficient (TC)—a dimensionless value equal to the ratio of the radiation flow through the medium (nosecone) to the radiation flow to its surface. The first level criterion for ADC is its drag coefficient (Cx), a dimensionless value reflecting how the body (nosecone) reacts with the environment (air).

According to this criterion, it is necessary to conduct research on possible options for the geometric appearance of the fairing and choose rational parameters that meet all the requirements. The algorithm for conducting these studies is shown in Figure 6.

Based on the results of the above studies, the method for selecting the geometric parameters of the fairing will allow optimizing the stages of RPO development. Adjusted according to the results of the optimization algorithm for the design of the fairing is presented in Figure 7.

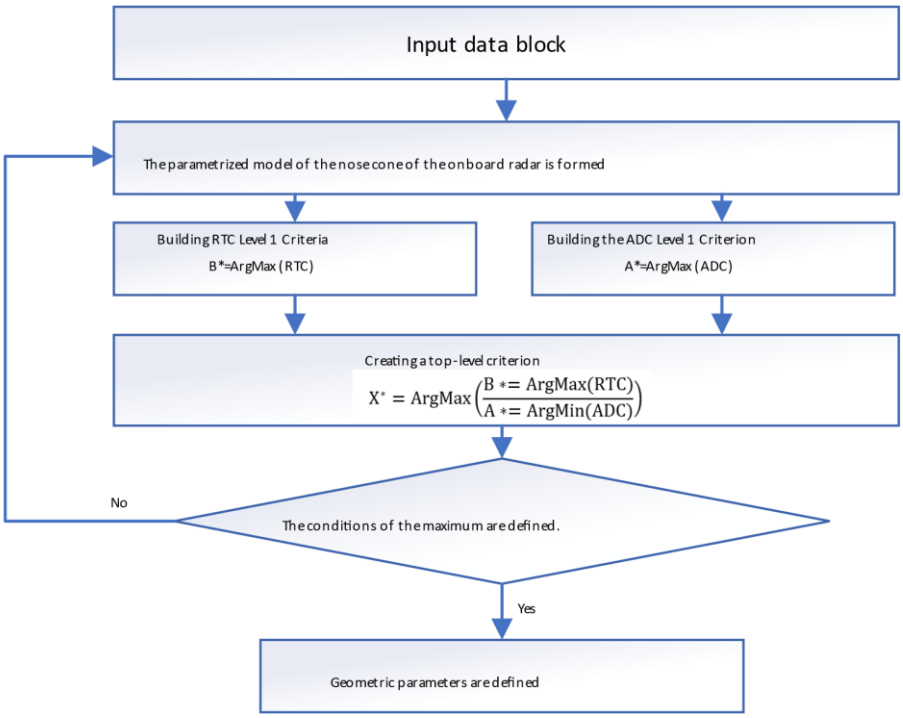


Figure 6: Algorithm for studying geometric parameters of RPO

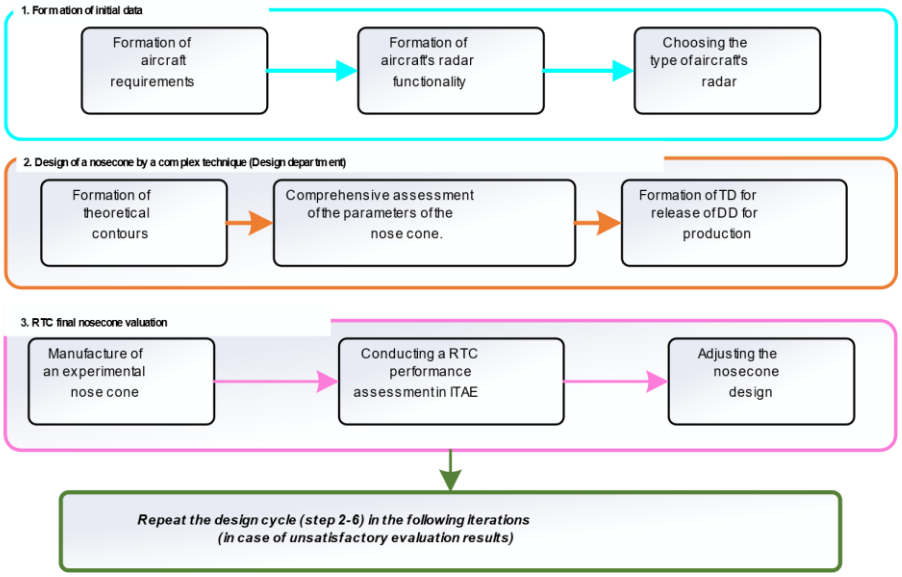


Figure 7: Adjusted the algorithm design of the fairing

Thus, the development of such methods for designing aircraft components can significantly reduce the number of design stages, as well as the involvement of company personnel, which in turn reduces the time and cost of product development.

4. Forming an optimal fairing with minimal drag

A typical example of multi-criteria optimization of an aircraft component to significantly reduce the product life cycle is the optimization of the geometric parameters of the radio-transparent fairing in order to obtain a body of minimal resistance. For this optimization, you must:

Firstly, to choose a complex criterion for evaluation of aerodynamic characteristics;

Secondly, to make up an algorithm for solving this problem and determine a method for evaluating aerodynamic characteristics that will have the required level of sensitivity to changes in key geometric parameters and the required accuracy;

Third, using the selected algorithm, to conduct research and, using a complex criterion, select rational geometrical parameters of the nose section of the aircraft for a given flight mode.

4.1. Choosing a comprehensive criterion

As noted earlier in this work, the issue of selecting rational geometric parameters for the nose section of the aircraft for the supersonic cruise mode is under consideration $M \sim 1.5$. To select the main criterion of aerodynamic characteristics of the nose

section, it is necessary to analyze the existing aerodynamic characteristics of the aircraft components as a whole.

According to [7], there are three main criteria for the aerodynamic characteristics of an object: C_x – aerodynamic resistance coefficient, C_y – lifting factor, Δmz_0 – component contribution to the longitudinal static torque of the aircraft as a whole:

1) Coefficient C_{xnose} as shown in Part 1. The introduction is one of the essential parts of the aerodynamic drag coefficient of the pressure of the aircraft as a whole $C_{xfpressure}$ in particular for $M > 1$.

2) In doing so C_{ynose} does not make a significant contribution to is not more than 0.1% of the total C_y as a whole.

3) It should be noted that the nose part of the aircraft can be shaped as a body rotation, and as a body of complex shape in the cross section. So for the body of rotation the contribution to Δmz_0 the aircraft as a whole tends to 0, and for a complex body it requires deep analysis and has a significant impact due to the large removal of the aerodynamic focus of the fore part of the aircraft from the center of the masses, which leads to losses on balancing and reduction of fuel efficiency of the aircraft in cruising mode of flight. Because for the current work is considered a cruise flight of the aircraft in the mode of $M \sim 1.5$ let's look at the fore part in the shape of a rotational body.

Thus, it is possible to draw an unequivocal conclusion that the criterion of a choice of rational geometrical parameters of a nose part of an airplane is the minimum value of the aerodynamic resistance coefficient, depending on basically geometrical options. This criterion can be written mathematically as follows (see (2)):

$$A^* = ArgMin\,(C_{x\,nose}(\lambda_{nose}, d, \theta)). \quad (2)$$

4.2. Selection of performance evaluation method

At the present time a significant scientific and technical reserve has been accumulated in the field of research of aerodynamic characteristics of bodies of various shapes and geometry, the main branch institute and center of competence of aerodynamics in the Russian Federation is the Central Aerodynamic Institute named after Professor N.E. Zhukovsky (TsAGI). Based on TsAGI's extensive work to determine the aerodynamic performance of various components of the aircraft, empirical dependencies of the AP on a few geometric parameters were obtained. These methods and dependencies are presented in [7] and are used to determine the AP (C_x and C_y) of the main components of the aircraft (wing, fuselage, plumage) and build integral AP of the aircraft. The empirical approach is widely used by the country's leading aviation and missile companies.

However, because these dependencies have been obtained from numerous wind tunnel experiments on wind tunnel models, there are certain assumptions in the wind tunnel estimates due to model manufacturing tolerances and measurement errors.

The trend formed in the XXI century of constant toughening of requirements to newly developed aircraft complexes (ACs) by the State Customer, related to ensuring competitive edge over foreign ACs, requires minimization of design reserves in the development of AKs and a more accurate assessment of the AP of both components and the aircraft as a whole.

In recent years, the main trend and element of product life cycle management in the aircraft industry is the digitalization of the life cycle stages in a single information space, including various types of tests [8] and the creation of electronic twins of the aircraft, including aerodynamic, through the application of mathematical simulation of geometry flow by finite element method [9]. The models received by results of similar digital tests are verified and possess necessary sensitivity to change of geometrical characteristics, are shown as algorithm in Figure 2, also allow to receive result with high accuracy and can be parametrized that essentially reduces terms of research.

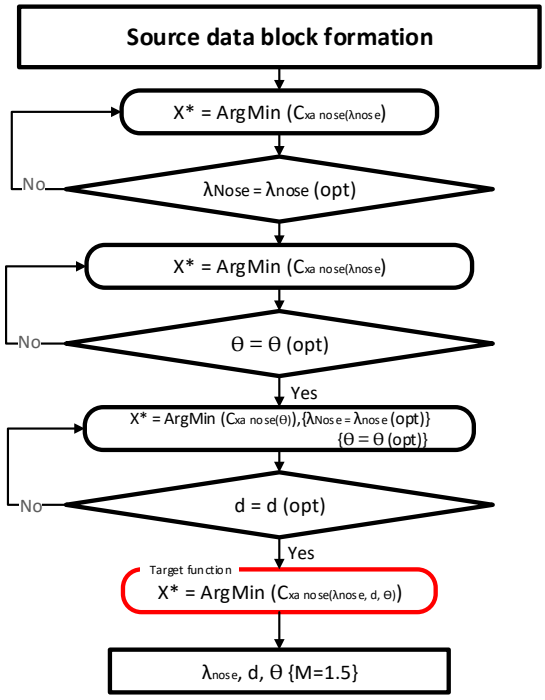


Figure 8: Algorithm for solving the problem of determining the rational parameters of the aircraft fuselage nose section

4.3. Problem solving algorithm

For effective solution of the problem within the framework of this work it is necessary to apply methods of mathematical modeling with the use of parametrized mathematical models of the surface, where the main geometrical parameters of the fore-body indicated earlier are used as variables of parametrization.

Main calculation parameters are accepted as a block of input data: altitude and speed parameters of flight, values of angles of attack, roll and slide, as well as atmospheric parameters at a selected altitude and speed.

The problem is solved iteratively, and at the first stage is the optimal value of the parameter of elongation of the nose part of the aircraft (λ_{nose}) as a most comprehensive parameter. Next, for an optimum elongation value, the optimum angle of tangent to the forming curve is selected (θ) and the final step is to determine the optimal value of the discriminant in the formation curve. (d).

It should be noted that for each stage a mathematical model of the investigated surface is built, and the range of parameter values can change based on the conditions of ensuring the smoothness of the fore-body surface (absence of double curvature) [10], [11].

4.4. Selecting the number of settlement cases

Calculated flight conditions are shown in Table 1. During the calculation the atmospheric parameters are accepted in accordance with the international standard atmosphere.

Table 1: Calculated conditions for solving the problem

	Parameter	Value
1	Estimated number of M flight	1.5 M
2	Estimated flight altitude	11 000 m
3	Calculated Flight Angle of Attack (α)	2^0
4	Calculated sliding angle (β)	0^0
5	Calculated roll angle (γ)	0^0
6	Atmospheric parameters	In accordance with the ISA

The calculation is performed in the selected calculation conditions for each of the values of the parametrization variables and in accordance with the algorithm presented in Part 4. The range of values of the parameterization variables and the step are shown in Table 2.

Table 2: Value range of variables

Parameterization parameter values								
λ_{nose}			$\theta_{,0}$			d		
min	max	step	min	max	step	min	max	step
2	5	1 (0.5 from 2 to 3)	20	45	5	0.5	0.8	0.1

The total number of settlement cases for the calculation is determined by formula (3):

$$N = n_\lambda + n_\theta + n_d, \quad (3)$$

Where n_λ – fuselage fore-nose extension parameter number, n_θ – number of angle values of the tangent curvature parameter, n_d – discriminant parameter number of the constitutive curve.

Thus, to solve the problem it is necessary to make 15 calculation experiments.

In accordance with the research algorithm (see Figure 8), a cycle of computational experiments to determine the dependence of values has been performed in this work $C_{x\,nose}$ the nose of the aircraft from its extension (λ_{nose}), the results of calculations are recorded in Table 3, the dependence of parameters is shown in Figure 9.

Table 3: Results of the dependency calculation C_{xnose} on λ_{nose} .

	θ	d	λ_{nose}	$C_{x\,nose}$
1	25^0	0.5	2	0.21
2			2.5	0.15
3			3	0.11
4			4	0.06
5			5	0.04

As can be seen from Table 3 and Figure 6, there is an inverse relationship between λ_{nose} and C_{xnose} for the nose section of the aircraft, so that the body with elongation has minimal aerodynamic resistance $\lambda_{nose} = 5$.

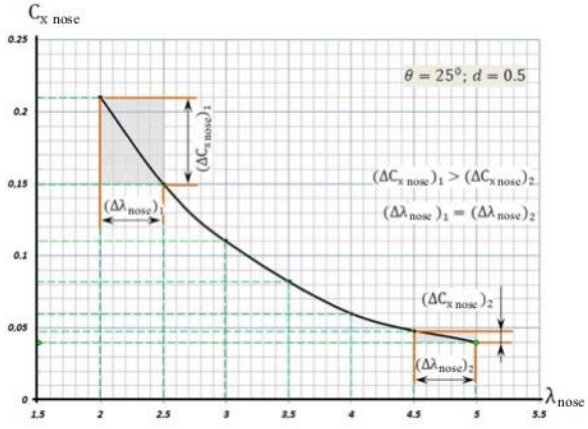


Figure 9: Dependency C_{xnose} on λ_{nose}

Taking into account these results in accordance with the algorithm (see Figure 8), we have estimated the change in the value of C_{xnose} change angle θ . In this case, to increase the sensitivity of the model to changes in the angle of tangent in the geometric model of the object surface, a spherical blunting of the spout with the $r = 15\,mm$ in this case the influence of the coefficient appears for the model $C_{xattenuation}$ in case of blunting, which is for M = 1.5 $C_{xattenuation} = 0.68$, and for $r = 15\,mm$ $C_{xattenuation} \sim 0.22$. The results of calculations are given in Table 4, the dependence is shown in Figure 10.

Table 4: Results of the dependency calculation $C_{x\,nose}$ on θ .

	λ_{nose}	d	$\theta_{,0}$	$C_{x\,nose}$
1	5	0.5	20	0.062547
2			25	0.062555
3			30	0.062565
4			35	0.062575
5			40	0.062587
6			45	0.062598

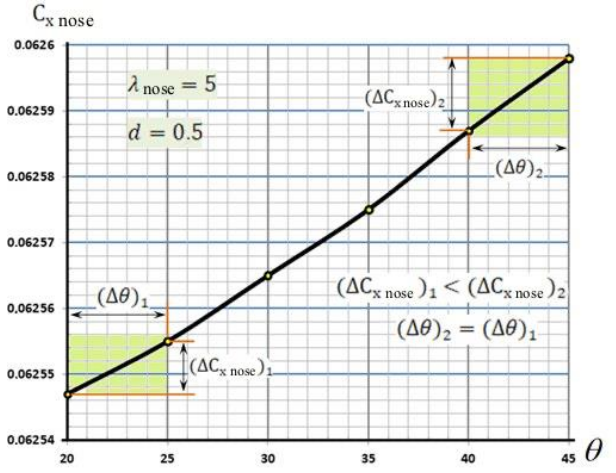


Figure 10: Dependency $C_{x\,nose}$ at θ

As can be seen from Table 4 and Figure 10, the aerodynamic resistance of the nose of the aircraft increases when the angle of tangent to the forming nose part of the aircraft. However, the difference between $C_{x\,nose}$ for $\theta = 20^{O}$ and $C_{x\,nose}$ for $\theta = 45^{O}$ at $\lambda_{nose} = 5$ not exceeding 1%. Based on these results, it can be concluded that the angle θ does not significantly affect the aerodynamic resistance of the nose section of the aircraft at $M = 1.5$. For the next phase of the study, we will accept $\theta = 20^{O}$ as a parameter with minimal aerodynamic resistance.

The next step in the study of the dependence of aerodynamic characteristics of the nose section of the aircraft on its geometrical parameters is to determine the dependence of $C_{x\,nose}$ from the curvature of the constituent – d. To determine the effect of curvature on the $C_{x\,nose}$ aircraft, a comparative analysis of the conical nose AP and the parabolic nose (at d=0.5) in determining the dependence of the $C_{x\,nose}$ on λ_{nose}. The results of this analysis are listed in Table 5, the dependencies are shown in Figure 11.

Table 5: Comparison $C_{x\,nose}$ at d.

	λ_{nose}	$C_{x\,parabola}$	$C_{x\,cone}$
1	2	0.21	0.23
2	2.5	0.15	0.17
3	3	0.11	0.13
4	4	0.06	0.08
5	5	0.04	0.059

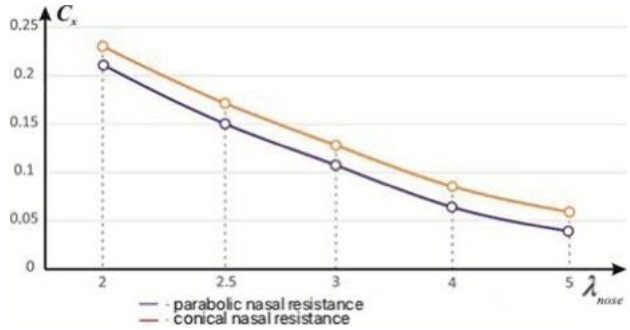


Figure 11: Comparison of nasal AP with conical and parabolic forming parts

As can be seen from Table 4 and Figure 8, both for conical and parabolic forming dependencies obtained in Figure 6 are fair. However, for the parabolic formative level $C_{x\,nose}$ lower than conical for all values λ_{nose}. Based on these results, it can be concluded that the nose section of the aircraft with parabolic formation has the lowest aerodynamic resistance.

To determine the rational value of the curvature forming within the framework of this work the dependence analysis has been performed $C_{x\,nose}$ at d for the nose section with the previously selected rational parameters λ_{nose} and θ, the results of this assessment are shown in Table 5, the dependence is shown in Figure 12.

Table 6: Dependency $C_{x\,nose}$ at d parabolic forming.

	λ_{nose}	θ	d	$C_{x\,nose}$
1	5	20^0	0.5	0.062547
2			0.6	0.062551
3			0.7	0.062558
4			0.8	0.062564

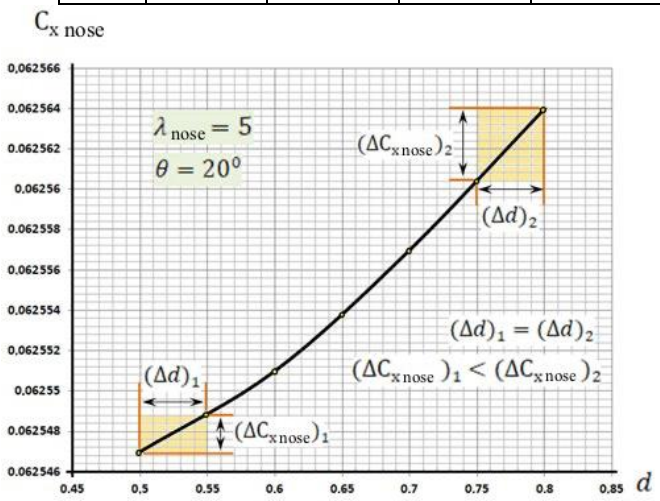


Figure 12: dependency $C_{x\,nose}$ at d

As can be seen from Table 5 and Figure 12, when the curvature of the aircraft nose increases, the aerodynamic resistance of the nose part increases. However, the difference between $C_{x\,nose}$ at $d = 0.5$ and $C_{x\,nose}$ for $d = 0.8$ at $\lambda_{nose} = 5$ and $\theta = 20^{O}$ not more than 1%. Based on these results, it can be concluded that parameter d does not significantly affect the aerodynamic resistance of the nose section of the aircraft at $M = 1.5$. As a rational value d at $M = 1.5$ take d =0.5 as a parameter with minimal aerodynamic resistance.

4.5. Optimal parameters

Within the framework of the work performed by the methods of mathematical modeling the dependences of aerodynamic resistance of the nose part of the aircraft on the main geometrical parameters varying in the selected range were obtained for supersonic cruise of the aircraft at $M = 1.5$. The following conclusions can be drawn from these studies:

1. Parameter of elongation of the nose of the aircraft λ_{nose} has a significant impact on the aerodynamic characteristics of the nose section of the aircraft. Dependence $C_{x\,nose}$ from λ_{nose} is reversed: when the parameter value is increased λ_{nose} level $C_{x\,nose}$ is decreasing. Reasonable value λ_{nose} to ensure minimum aerodynamic resistance in the selected flight mode is $\lambda_{nose} = 5$.
2. The parameter of the angle of tangent curvature to the forming nose part of the aircraft θ does not have a significant impact on the aerodynamic characteristics of the nose section of the aircraft when $M = 1.5$. Dependency $C_{x\,nose}$ for θ is direct: when the parameter value is increased θ level is increasing $C_{x\,nose}$. Rational importance θ to ensure minimum aerodynamic resistance in the selected flight mode is $\theta = 20^{o}$.
3. The nose section of the aircraft with a parabolic formation has a lower level of aerodynamic resistance than the nose section with a conical formation 3.

4. The parameter of curvature of the forming nose part of the aircraft d does not have a significant impact on the aerodynamic characteristics of the nose part of the aircraft at $M = 1.5$. Dependency $C_{x\,nose}$ from d has a direct character: when the value of parameter d increases, the level $C_{x\,nose}$ increases. The rational value d for minimum aerodynamic resistance in the selected flight mode is $d = 0.5$.

Thus, based on the conclusions (1–4) can be concluded that the main geometric parameter of the nose of the aircraft, which determines the level of its aerodynamic characteristics, for $M = 1.5$ is an extension of the nose of the aircraft λ_{nose}.

The nose section of the aircraft with rational geometrical parameters has minimal aerodynamic resistance: $\lambda_{nose} = 5$, $\theta = 20^{\circ}$, $d = 0.5$.

5. Conclusion

The main conclusion from the materials presented in this paper is that optimization of the development cycles of aircraft components leads to optimization of the product life cycle as a whole. The considered example of forming an optimization basis for choosing rational geometric parameters of RPO clearly showed a significant reduction in the RPO design cycle due to the use of complex techniques.

In this work, a total of about 90 computational experiments were performed, with an average labor intensity (if there is free computing power) of ~5 h/h. Thus, using the results of this work in the design of a new supersonic aircraft will save ~450 h/h of research work, which is ~2.5 months of work for one specialist of the organization.

The introduction of such an integrated approach in all industries and for all types of products will make it possible to make a new leap in the field of mechanical engineering in the shortest possible time. [12].

References

[1] E.S. Wentzel, Operation research, Moscow: Science, 1988.

[2] A.V. Barabanov, S. A. Serebryansky, "Approach to determining the rational shape of a supersonic aircraft using mathematical modeling methods", Vestnik SURGU, 2020

[3] V.A. Kaplun, Radar domes, Moscow: Sov. Radio, 1974.

[4] L.G. Artamonova, A.V. Kuznetsov, N.N. Pesetskaya, Verification of aerodynamic characteristics of the aircraft. Moscow, MAI, 2010

[5] G.S. Byushgens, Aerodynamics and flight dynamics of trunk aircraft, TSAGI, 1995.

[6] G.S. Byushgens, Aerodynamics, stability and control of supersonic aircraft, Phismatlit, 1998.

[7] G.A. Kolesnikov, V.K. Markov, A.A. Michailyuk, "Aerodynamics of aircraft". Moscow, Machinostroenie, 1993

[8] A.V. Barabanov, S.A. Serebryansky, "Algorithm for the formation of rational geometric shape of the radio-transparent nose radar fairing of the aircraft," Engineering J.: Science and innovation, **1**, 2020, doi: 10.18698/2308-6033-2020-1-1948

[9] S.A. Serebryanskii, A.V. Barabanov, "To the Question of Optimizing Product Life Cycle STAGES," Proc. of 12th International Conference Management of Large-Scale System Development, MLSD, doi: 10.1109/MLSD.2019.8911045, 2019

[10] D.D. Kulikov, V.S. Babanin, "Creating a parametric part model in the CAD system environment", Sci. and Tech. J. of Information Technologies, Mechanics and Optics, **4** (74), 2011.

[11] P.A. Koporushkin, A.S. Partin, D.V. Kurennov, "Algorithm for calculating parametrized objects in the system of geometric modeling of machine-building products", Automation and modern technologies, 2008.

[12] S. A. Serebryansky, A.V. Barabanov, "An approach to determining the rational shape of the nose of a supersonic aircraft using mathematical modeling methods", Bulletin of SUSU, mechanical engineering, **20(**3), 2020.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 416-422 (2020)

www.astesj.com

Special Issue on Innovative Research in Applied Science, Engineering and Technology

ASTESJ

ISSN: 2415-6698

The Probe Mark Discoloration on Bond Pad and Wafer Storage

Wen-Fei Hsieh*, Henry Lin, Vincent Chen, Irene Ou, Yung-Song Lou

Ardentec Cooperation Quality Assurance, No. 3, Gongye 3 rd Rd., Hsin-Chu Industrial Park, Hukou Township, Hsinchu County 30351, Taiwan, R.O.C.

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 15 October, 2020
Online: 20 November, 2020

Keywords:
Probe Mark Discoloration
Fluorine Contamination
Wafer Storage

ABSTRACT

In this report, a probe mark discoloration with donut/circle shape was found. The corrosion formed on Aluminum bond pads. The SEM/EDS, TEM/EELS & AES surface techniques were used to identify this corrosion product. All these material analysis results pointed out the role of fluorine to this discoloration. The HAST testing, all the chip samples obtained from wafer center/middle area, were simulated to have been exposed in air for 1 year. HAST test result showed no fluorine source came from the center/middle zones of impact wafers. Another full wafer-level FOSB storage test showed the source of corrosive fluorine came from wafer itself. EDS mapping results, from extreme wafer edge zone, indicated the fluorine element distribution. The mechanism of this novel discoloration was proposed. However, more evidences were necessary for understanding the formation of donut discoloration corrosion.

1. Introduction

In order to obtain semiconductor device electrical performance, the circuit probing (CP) tests were performed by pin probing on bond pad before device package process. The typical bond pad materials are Aluminum (Al) with low content of silicon and copper to improve the resistance of electron migration. The interconnection from aluminum pad surface to external interface frame play a critical role in semiconductor package process. It is well known the Aluminum would form corrosion protective oxide film in oxygen condition. However, in the past three decades, the formation of fluorine (F) corrosive film on aluminum bond pad surface [1,2] and non-stick on pad (NSOP) failure issue that induced broad quality issues concern were reported by some researchers before. [3-7] These corrosive film was found to suffer semiconductor reliability. The early researches indicated two main root causes for these fluorine contained contamination. One is the fluorinated silicone glass (FSG) film could release fluorine into closed spacing, and then contaminate the bond pad surface. [1] The FSG layer is still an important low K dielectric layer used in device process, so the F leakage from FSG layer is still a unavoidable issue in semiconductor dielectric layer process. The second fluorine source is the fluorine embedded on Al pad surface, and the fluorine residue from previous pad etching and cleaning gas. [9] So the fluorine source for contamination divided into two groups, direct F embedded on Al top surface, and the indirect F from FSG layer. Therefore, the wafer storage become an important factor for fluorine contamination. The early F-induced contamination studies focused on the morphology and mechanism to induce corrosion product. No probe mark and electric testing effects were reported before. P. Chang [8] had stated the early probe mark discoloration study of yellow discoloration Al bond pad corrosion. However, the probe mark discoloration had not been fully report yet. [10, 11] In this paper, the donut/circle discoloration corrosion and a compressive study would be present.

In the beginning of this case, the probe mark discoloration was found at the stage of blue tape. All the impacted wafers finished all the CP tests, and stored in the Front Opening Shipping Box (FOSB) for more than two months. There are two types of probe mark discoloration forming on the bonding pad surface. Figure 1 (a) and (c) show these typical circle/donut discoloration optical microscope (OM) photos. The special feature of discoloration morphology is that probe mark is the center of cyclic/donut center. The OM inspection for discoloration corrosion indicated most of the corrosion was cyclic, and only few was donut shaped. Figure 1 (b) show that two touchdown would not induce two discoloration. The size of cyclic discoloration is not constant, however, the largest diameter of a suffered wafer, up to 35 um, is nearly the same. With the help of full wafer dice Auto Vision Inspection (AVI), most of the corrosive dice locate near the wafer edge zone. The worst fail rate would reach up to 15-18% of all the available dice.

*Corresponding Author: Wen-Fei Hsieh, Email: michael.hsieh@ardentec.com

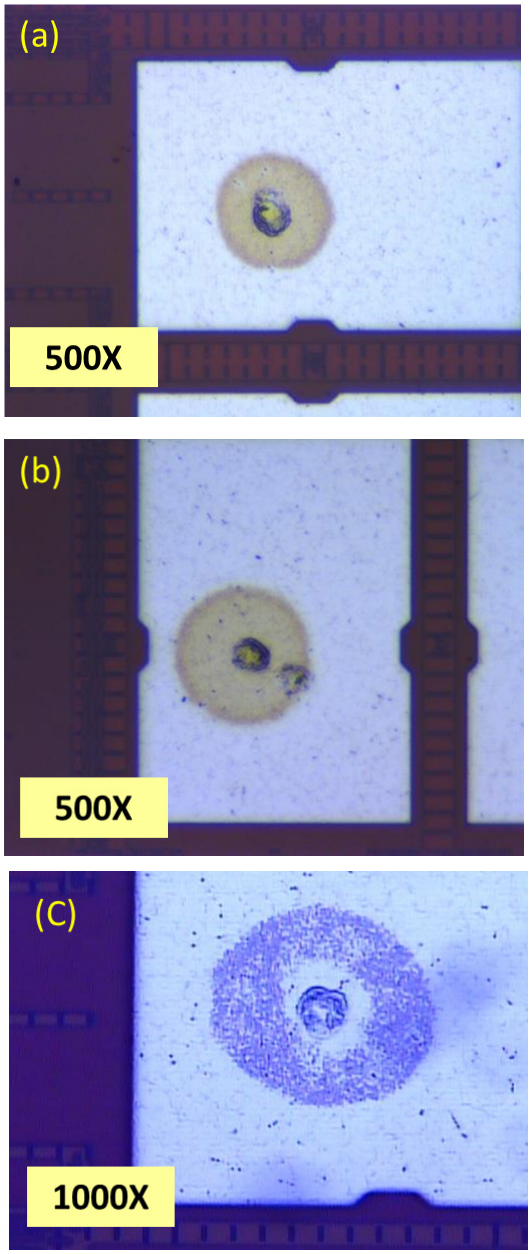


Figure 1: (a) show the typical probe mark discoloration OM image. (b) show the formation of discoloration on Al pad under dual pin touch down. No second discoloration is found. (c) show donut-shaped probe mark discoloration. To clearly show donut shape discoloration, (c) was magnified to 1000X, that is larger than Figure 1 (a) & (b).

2. Experiments and Results

In this study, the samples were checked in plane-view & cross-sectional approaches. The Scanning Electron Microscopy (SEM) images was obtained by 5 KeV electron energy. And the Energy Dispersive Spectrum (EDS) result was obtained by 10 KeV electron energy. Transmission Electron Microscopy (TEM) operated at 200 KeV with Electron Energy Loss Spectrum (EELS). SEM and TEM could get the information of morphology and corresponding EELS analysis for the normal and abnormal locations. Auger Electron Spectroscopy (AES) is a traditional surface analysis technology for bond pad corrosion case. So AES depth profile results would be included.

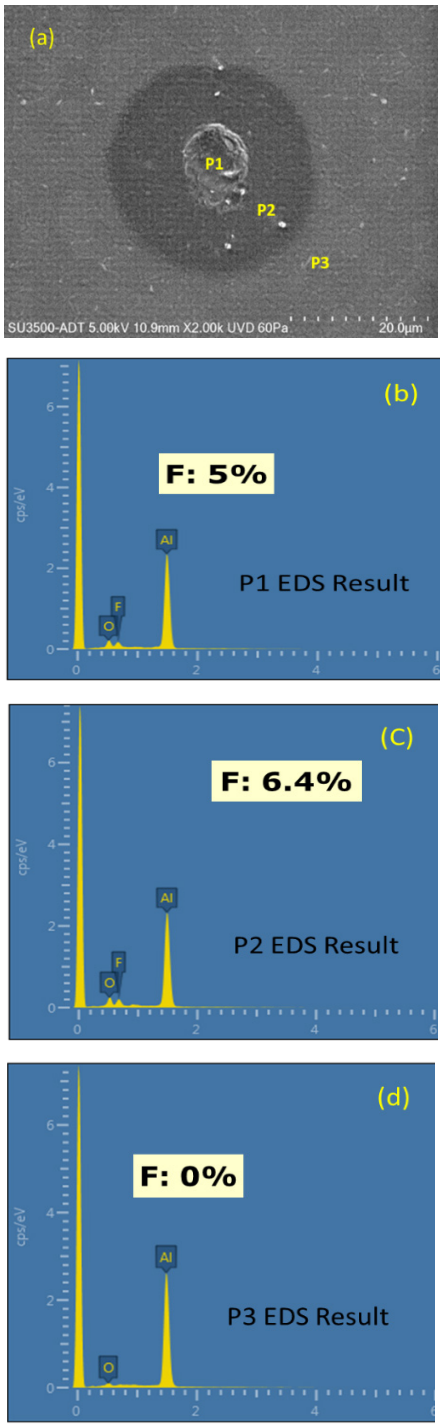


Figure 2: (a) is the SEM micrograph of probe mark discoloration corrosion. P1 is the zone of probe mark. P2 is the discoloration zone. P3 is the normal Al pad surface. (b),(c) & (d) the corresponding EDS analysis result of P1 to P3 various zones on Al pad surface. The oxygen element was found for each locations. The Fluorine was only found at the discoloration zones.

Based on the literature review, the various storage conditions would impact corrosion behavior. Highly Accelerated Stress Test (HAST) was applied with test conditions of temperature at 130ºC and RH 85% for 21 hours. This test is to simulate all the bond pad exposed in air without Nitrogen Cabinet storage or Moisture Barrier Bag (MBB) storage for 1 year. This test would be helpful to evaluate the impacted dies corrosion behavior. The secondary storage test, full wafer deposited in the regular FOSB, is to identify

the standard FOSB storage effect, and find out the source of F element.

2.1. SEM/EDS Analysis Results

Figure 2 show the top-view SEM photo and corresponding EDS analysis results of various feature locations on Al pad surface. The SEM photo clearly show the boundary of discoloration boundary on Al pad surface. Based on the SEM inspection, the rough size of this discoloration was more than 25 micro meter. The probe mark and corresponding discoloration zone were easily identified. The SEM photo also show various morphology of probe mark surface deformation. According to the EDS results of discoloration zone, the P1 & P2 of Fig. 2(a), where fluorine and oxygen contents are found on the center and middle of circle region. No Fluorine element was detected for normal P3 Al pad surface. The content of fluorine element, 5-6 %, is noted. That is, the probe mark discoloration is a kind of aluminum-fluorine-oxygen corrosion. As for the normal zone, aluminum oxide form on Al pad surface. Since this novel corrosion product has not been fully reported before, the EDS analysis is helpful to improve the key role of fluorine element.

Figure 3 show the OM photo and its corresponding EDS result. P1 to P4 of Figure 3(a) are the various locations of donut-shaped discoloration on Al pad surface. P1 is the probe mark zone. P2 is the cyclic annular zone without discoloration. P3 is the cyclic annular zone with discoloration corrosion. And P4 is the normal Al pad surface. The EDS result clearly indicate fluorine element is only found at the discoloration corrosive zone. No fluorine is found at P1- probe mark, P2- cyclic annular without discoloration, and P4 - normal Al pad surface zones. Based on these donut-shaped discoloration EDS results, the root cause of the discoloration event is due to fluorine contamination. However, due to the basic characteristic of high energy electron stopping power, the X-ray signal of EDS analysis is collected not only from Al pad top surface, but also the larger undersurface bulk volume. The distribution of fluorine element could not be identified by this EDS analysis. TEM and AES are necessary for further identification.

2.2. TEM/EELS Analysis Results

The cross-sectional TEM inspection is helpful to identify the thickness variation of corrosion film. Figure 4 show the full cross sectional TEM inspection photo of discoloration area. The thickness of discoloration corrosion film is just a few nm. Therefore, it could not be clearly identified by Figure 4. The magnified TEM photo indicate the thickness of discoloration film is uniform except for probing mark zone. Figure 5 show the High Angle Annular Dark Field (HAADF) TEM image of discoloration film structure. The HAADF image indicate the thickness of circle shape discoloration is within 5 nm. The corresponding Electron Energy Loss Spectrum (EELS) line scan results show the F & O distribution of this Al-O-F film. The fluorine is found on the top surface of corrosion film. No F & O could deeply penetrate into the Al. Figure 4 show the corrosion film of probe mark is complex. Figure 6 show the high magnification cross-sectional TEM image and its corresponding EELS data of probe mark oxidation. The variation of corrosion film thickness might be up to over 10X. The thicker film formed on probe mark zone is also the AL-O-F compound as other discoloration zone. The EELS fine structure analysis indicate discolored Al_2O_3 spectrum is a combination of

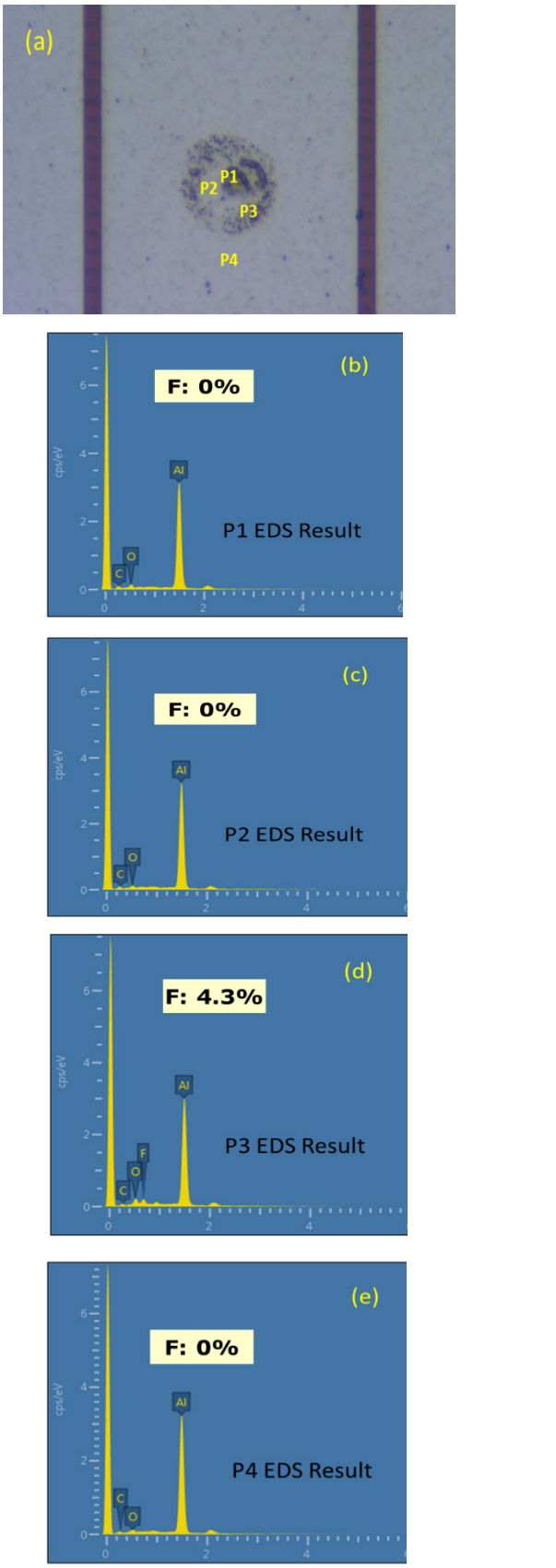


Figure 3: (a) is the OM image of donut-shaped probe mark discoloration. P1 is the zone of probe mark. P2 is the annular zone without discoloration. P3 is the annular zone with discoloration. P4 is the normal Al pad surface. (b), (c), (d) & (e) the corresponding EDS analysis results of various zones on bond pad surface. Fluorine is only found at the cyclic annular discoloration zone.

pure Al_2O_3 with OF_3. That is, the fluorine diffuse into the Aluminum oxide film. F atom form a new bonding to replace O bonding. These new formed Al-O-F must be less protective than original Al_2O_3 film.

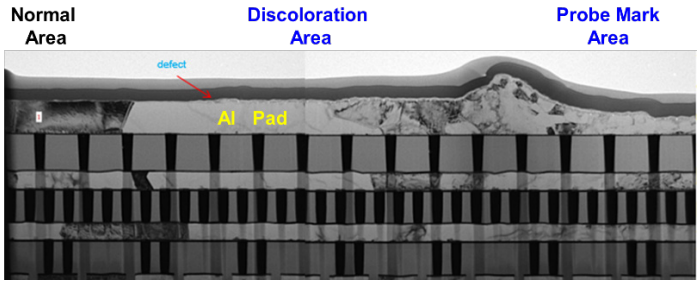


Figure 4: The low magnification full cross-sectional TEM image of discoloration defect corrosion.

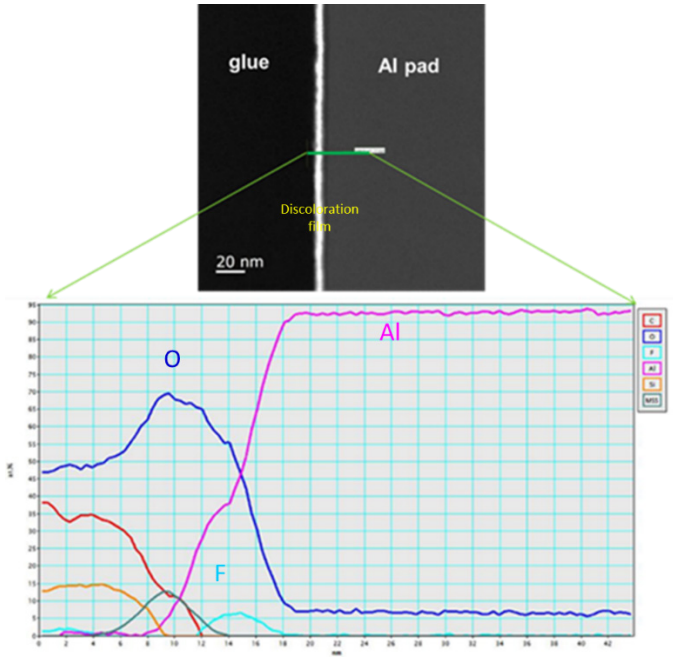


Figure 5: The cross sectional HAADF TEM image of discoloration zone and its corresponding EELS line scan result. The line scan result clearly show the distribution of fluorine and Oxygen just on the Al pad top surface.

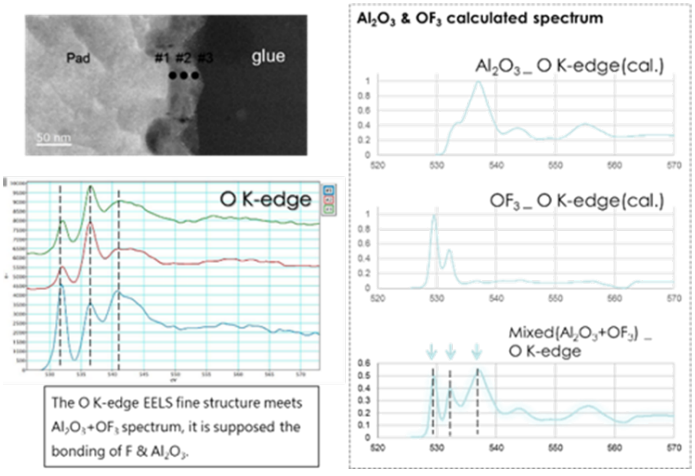


Figure 6: The TEM image and EELS data of probe mark discoloration defect film & its corresponding simulation EELS spectrum.

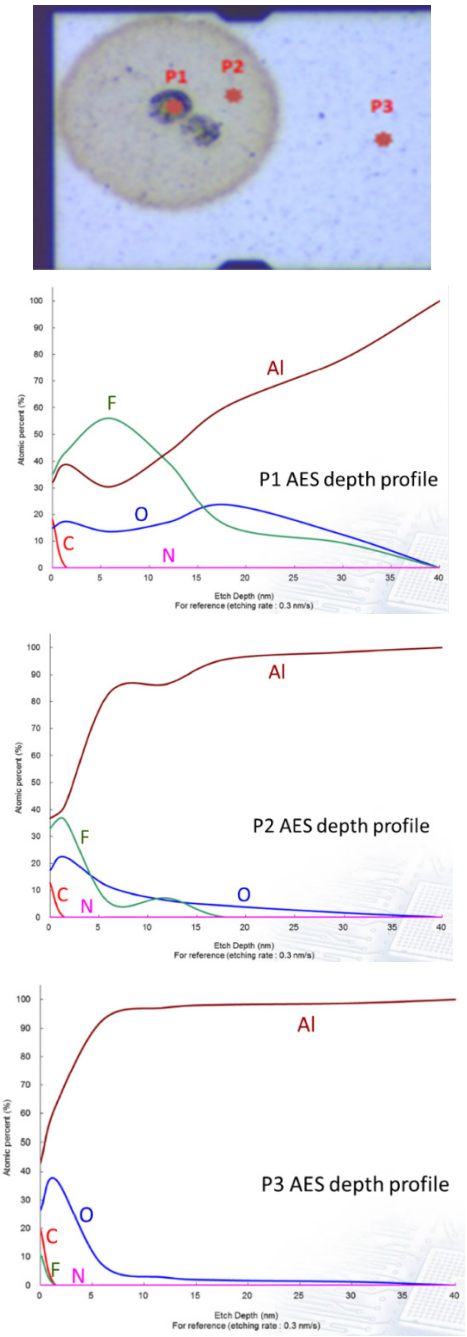


Figure 7: The OM photo and the corresponding AES depth profile of P1, P2 & P3, the P1, P2, & P3 are marked on discolored Al pad surface. P1 is the probe mark zone. P2 is the discoloration zone. P3 is the normal Al pad surface. The F element is found to co-exist with Al_2O_3 at discoloration zones.

2.3. AES Analysis Results

For a long time, the AES analysis is treated as the standard technology for Al pad surface contamination. Figure 7 (b) to (d) show the specific Auger Electron Spectrum (AES) depth profile result of normal/discoloration area of bond pad surface. The AES result indicate the main elements of probe mark and cyclic discoloration area are Al, F, and O. That is the same as previous EDS & EELS results. The AES depth profile results, P1, P2 & P3 of Figure 7, show the discoloration corrosion film is the F element co-existed with AlxOy. BTW, X-ray Photon Spectrum (XPS) is also an available surface analysis technology for oxidation state of

Al compound. However, due to the X-ray focusing capability limitation, XPS is not used in this bond pad case. The Secondary Ion Mass Spectrum (SIMS) analysis meet the similar limitation.

Two key factors decide the success of this contamination Failure Analysis. The first one is to identify the component of the corrosion product. The second one, sometimes it is more critical in probe testing FAB daily operation, is to find out the corrosion/contamination source, and then to stop this event.

Based on previous material analysis finding, the probe mark discoloration contamination is a special type of Al-F-O corrosion. Since the nature protective Al-O film would automatically form on bond pad top, therefore, the fluorine element play an important role in this corrosion event. Where do the fluorine come from? How do this discoloration develop?

The fluorine induced corrosion on Al bond pad is not an new issue[1,2]. The source of fluorine might come from wafer itself or external environment. The fluorine come from wafer itself is supposed to be the etching residue on Al pad surface.[9] The F-contained etching gas, like CF4, is broadly used in Al pad process loop. The fluorine element is supposed to easily residue on Al pad top surface after etching. Generally speaking, the 5 % fluorine residue on Al pad surface is regarded as a safe margin. However, the previous AES depth profile data indicate no fluorine element is found below 1.5nm of top surface on no discoloration area of Al pad. The source of fluorine need to clarify. Therefore, two various storage tests are necessary to find out the fluorine source. The first storage tests, HAST, the dice-level test is a high temperature corrosion accelerated test. The second test, FOSB storage test, is a full wafer-level test under daily standard storage operation. The experiment results show as below:

2.4. HAST Storage Test Result

In order to clarify the effects of long time storage and probing touchdown number on this corrosion. Four groups of various die samples were applied for the HAST. The nine discolored and nine normal dice were the first and secondary groups for HAST. The nine dice, with five and twenty probing touchdowns, set as the third and the forth groups. Therefore, in this HAST, total four groups dice obtained from impacted wafers were tested. The HAST testing preformed at the condition of temperature 130ºC and RH 85% for 21 hours. All the dice samples simulated to have exposed in air without N2 cabinet storage or MBB bag storage for 1 year. Table 1 consolidate the result of HAST storage test. The result of group one and two indicate that there is no visual change, by 100% OM inspection, after HAST storage simulation testing. No probe mark discoloration corrosion size increase after HAST. The secondary group data indicate no probe mark discoloration would form. The result of third and four group imply that the heavy deformation on Al pad surface would not automatic induce discoloration corrosion. According to the previous analysis, the fluorine is necessary for the formation of discoloration. The HAST result show no extra fluorine source exist in HAST tested dice.

Table 1: The HAST testing Result. Four various samples were tested. The size of all samples are die level. The discoloration dice show no deterioration after the test. No other tested dice show clear change. In this table, Y means discoloration; and N means no discoloration.

Group	1	2	3	4
Samples	Discolored 2 touchdown	Normal 2 touchdown	Normal 5 touch down	Normal 20 touch down
Dice Quantity	9	9	9	9
Discolored Before	Y	N	N	N
Discolor After	Y	N	N	N
Result	Discolor no deterioation	No Discolor	No Discolor	No Discolor

2.5. FOSB Storage Test Result

Based on the previous HAST result, all these dice were obtained from wafer center or middle area, and no extra fluorine was found. The FOSB storage test preformed for the full wafers deposited in normal FOSB at room temperature. The Ardentec AVI inspection was applied for the tested wafer before FOSB storage testing. Then these wafers, after some storage time, would be viewed by Ardentec standard AVI checking to identify the discoloration behavior on the same Al pad. After 4 weeks' storage time, obvious variation was found. Figure 8 clearly indicate the variation of two OM photos. The probe mark discoloration would deteriorate in FOSB storage test. That is to say, the fluorine-induced corrosion in the FOSB volume would be continuous. It is a manifest evidence the source of F element enhancing discoloration corrosion is due to the full wafers FOSB storage. The HAST result indicated no extra F was found from wafer center and middle area. Therefore, the source of F element of FOSB test is highly suspected to be the wafer edge or wafer extreme edge area. The EDS mapping was an excellent technology to show the various element distribution of semiconductor top surface. It would be helpful to identify the distribution of fluorine element of abnormal wafer extreme edge area.

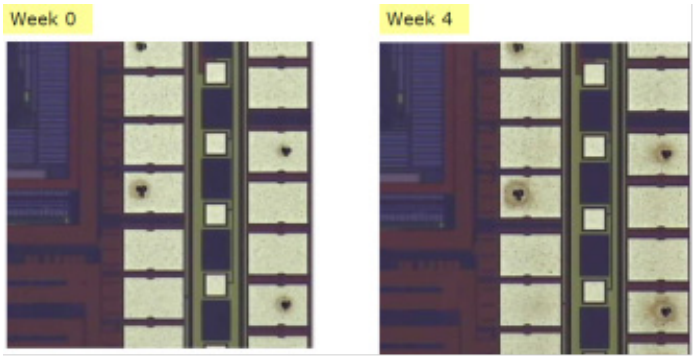


Figure 8: show the OM inspection result after FOSB storage test. The original probe mark discoloration is the OM micrograph on the left. After 4 week's FOSB storage, the size of probe mark discoloration became larger. The right hand side OM photo clearly show this deterioration.

Figure 9 show the EDS mapping result of extreme edge area of discoloration wafer. The EDS data clearly show the distribution of F, Al and other main elements on semiconductor wafer top surface. And the original location of this EDS mapping, the right hand side OM photo of Figure 9, was obtained at the extreme edge of full

wafers. There is a clear boundary between extreme edge, the upper part of OM image, and the internal die area. The fluorine-induced corrosion source is clearly identified from extreme edge of full wafers. It is supposed that the FSG layers exposed to FOSB space would be the F source of this corrosion.

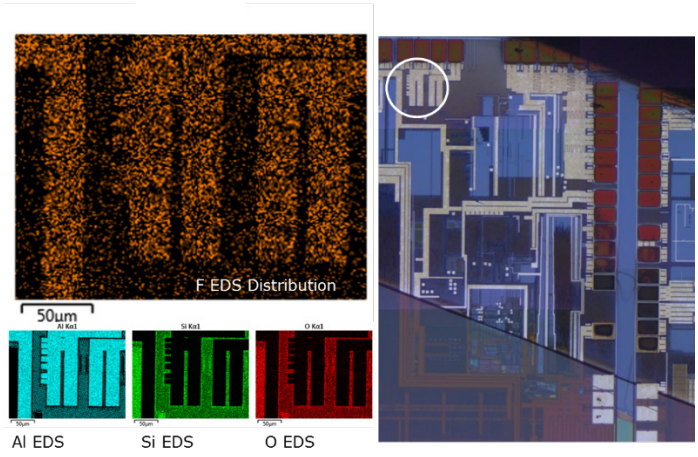


Figure 9: The EDS mapping result of Al pad on the discoloration wafer extreme edge area. The left hand side EDS mapping result show the distribution of F, Al, Si and O. The white circle area of right-hand side OM image shows the location of this EDS area of the wafer.

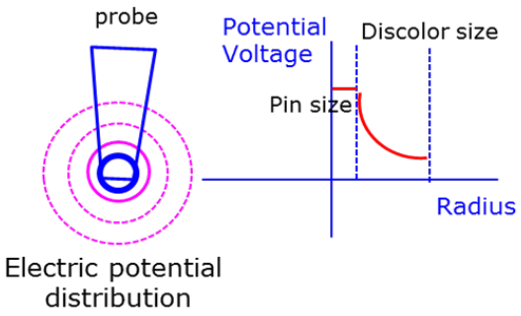


Figure 10: The critical CP testing and electrical testing potential field effect on Al bond pad surface.

The mechanism of this probe mark discoloration was proposed as follow: The CP testing would activate the embed fluorine. The formation of fluorine-contained oxide film induce this special discoloration.

The special point of this novel discoloration is the corrosion morphology with probe mark as its center. Some factors inducing the fluorine activation would spontaneously lead the probe mark as the discoloration cyclic center. In the beginning, the CP probing action would induce a permanent mechanical damage on the Al bond pad surface. This damage produce great deformation and heavy strain on the Al pad top surface, and this deformation and strain region, chemical activated, would be full of dislocation-like defects. The second mechanical effect was that probing test itself would break down the original protective Al-O film, and then create large amount of fresh aluminum surface. Both the mechanical effects would induce extremely active sites on Al pad surface. These would be easily for fluorine to react with Aluminum. The other impact of the probing action is the corresponding electric field potential, built by Circuit Probing test, would have an effect on the surrounding area of probe mark under the voltage/current variations. Under certain temperature, like the temperature of 125ºC setting at CP testing, and humidity, the fluorine contamination would be excited. Here the water is supposed to play an important role, leading to forming of hydrogen and oxygen, and after serial chemical reaction with fluorine, the AlxOyFz compound form. This kind of AlxOyFz compound is the discoloration by OM inspection. The threshold activation energy for all the chemical reactions was limited by electric potential field built by testing program, therefore, electric potential field introduced the circle shape discoloration region centered by probe mark. Figure 10 show the electric potential field distribution build by probing test.

The formation of probe mark discoloration on Al pad surface was not just due to single factor. There are three complex factors to induce this corrosion. The first factor is the sufficient fluorine supply in wafer storage environment. The fluorine might come from bond pad top surface or wafer inter metal dielectric layer. The second one is the environmental condition. The suitable temperature and enough humidity are critical to accelerate the fluorine corrosion reaction. The third factor induce probe mark discoloration formation was the probing touchdown and its sequential current and electric field. According the early study [3], the Al-O-F compound on Al pad surface show different morphology as this corrosion event. No previous report had pointed this Al-O-F compound discoloration morphology with probe mark as its circle center. Therefore, the characteristic of this contamination closely related to electric probing test condition. Then this novel probe marked discoloration need to combine all the above corresponding factors. These factors are materials property, wafer testing/storage environments and probing testing. Figure 11 show this idea.

According to the above mechanism, the formation of probe mark discoloration could stop in some ways. The material itself decide the source of fluorine. Therefore, the content of fluorine embedded on bond pad top surface or the distribution of fluorine on wafer edge is an important index to find the fluorine source. Since in CP and FT daily operation, no fluorine contained chemical materials are used. The CP & FT Fab is better to take them into regular monitor index for future discoloration event.

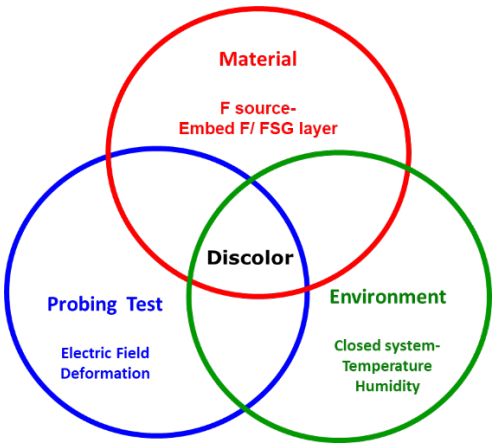


Figure 11: The factors to induce bond pad probe mark discoloration. These critical factors are material property, probing test and environment.

In the previous discoloration mechanism, the formation of donut shaped discoloration corrosion would be a weak point. Since the material analysis of donut discoloration indicate the fluorine

contained discoloration would only form on the ring zone, and no other discoloration would form on the other Aluminum pad surface. Therefore, the idea, discoloration originally form at probe mark area, and then fluorine diffusion to deteriorate discoloration, would not fully explain the formation of donut-shaped discoloration corrosion film. The EDS analysis result of donut-shaped discoloration imply that no fluorine exist at probe mark area. This impressive result also indicate the probe mark area with heavily deformation and high dislocation density would not spontaneously form corrosive fluorine contained Al-O-F film. The limitation of fluorine supply might be the key factor for this donut-shaped discoloration. There is no enough fluorine to form Al-O-F film on probe mark area, but enough fluorine to form Al-O-F film on annular area of Al top surface. In this system, there are two fluorine source, the embed fluorine and leakage form FSG. It is supposed that the embed fluorine would spontaneously form Al-O-F during CP testing under suitable electric potential activation. The fluorine supply from FSG leakage would be slowly deteriorate the corrosion due to diffusion into spacing is a time consuming process. That is to say, discoloration might be separate into two stages. In the first stage, embed fluorine play the key role in Al-O-F corrosion film formation. As for the secondary stage, the fluorine accumulation, leakage from FSG, and diffusion to Al-O-F interface play the key role. The secondary stage is a time consuming stage, therefore, storage environment is the controlled factor at this stage. The formation of cyclic discoloration corrosion would be that there is enough embed fluorine on the Al bond pad surface, and following with enough fluorine from FSG layer. As for the donut shaped discoloration, it would be that no enough embed fluorine on bond pad surface, so there is no complete cyclic discoloration formation. The key factor for donut-shaped discoloration formation is the supply of fluorine. No fluorine to react with Al metal. This explain that no Al-O-F film form at probe mark area and the formation of annular corrosion film. The possible root cause for donut discoloration might due to fluorine source complete consumption. However, no experiment would prove this proposal now.

3. Conclusions

- A new donut/circle shaped discoloration on Al bond pad is reported.
- Based on SEM/EDS, TEM/EELS and AES analysis results, the novel discoloration is due to form an Aluminum fluorine-oxide compound. The fluorine plays a key role in the formation of probe mark discoloration.
- With the help of HAST, FOSB storage test, and EDS mapping, the source of fluorine of this case is found.
- Since probe mark is the center of circle for discoloration, the formation of this corrosion is closely relative with the CP electric field.

Conflict of Interest

The authors declare no conflict of interest for this work. Formal "Conflict of Interest" document file is attached.

Acknowledgment

The authors would like to thank Polly Liu and Alex Si for their long time effort on discoloration impacted wafer management, and Berton Hsieh and Wei-Cheng Chen for their long time supporting on FA analysis.

References

[1] K. Ernst, D. Grman, R. Hauert, E. Holländer "Fluorine-induced corrosion of Aluminum microchip bond pads", Surface and Interface Analysis, **21**(10), 691 – 696, 2004, doi: 10.1002/sia.740211003

[2] W. A. P. Claassen and P. Krarner, "Aluminum Corrosion in the IC-MOS Production Line", Internal Report. Philips, Nijmegen, 1988.

[3] H. Younan, Shailesh Redkar, Lau Chi-Kwan and Mo Zhiqiang, " A Study on Non-Stick Aluminum Bond Pads due to Fluorine Contamination using SEM, EDS, TEM, IC, Auger, XPS & TOP-SIMS Techniques". in The Proceeding from the 28th International Symposium for Testing & Failure Analysis (ISTFA 2002), Phoenix, USA, 495-504, 2002.

[4] H. Younan, Shailesh Redkar, Ron Dickinson, and Peh Shirley, "A study on Fluorine-Induced Corrosion on Microchip Aluminum Bond Pads". in The Proceeding from the 29th International Symposium for Testing & Failure Analysis (ISTFA 2003), Phoenix USA, 249-255, 2003, doi: 10.1002/ sia.740211003

[5] K. Ernst, R. Hauert, E. Holländer, "The Fluorine-induced corrosion of Aluminum microchip bond pads", Surface and Interface analysis, vol.21, 691-696, 1994, doi: 10.1002/sia.740211003

[6] R.S. Sethu, "Reducing non-stick on pad for wire bond: A review" Australian Journal of Mechanical Engineering, **9**(2), 147-160, 2012, doi: 10.7158/M11-771.2012.9.2

[7] H. Younan, Nistala Ramesh Rao and Teong Jennifer, "Studies of Fluorine-induced Corrosion and Defects on Microchip Al Bond pads and Elimination Solutions" in The Proceedings from the 34th International Symposium for Testing and Failure Analysis, Portland USA, 285-290, 2008, doi: 10.1361/cp2008istfa285

[8] P.Chang, H.L Hsiao, C.H Cheng, S.I.Chu, H.W.Tai, Y.Y Shieh, C.Y Sun, "Investigation of Fluorine Induced Probe Marker Discoloration," in *2013* IEEE 15th Electronics Packaging Technology Conference (EPTC 2013), Singapore, 139-142, 2013, doi: 10.1109/EPTC.2013.6745700

[9] J.H. Linn, W. E. Swartz Jr., "An XPS study of the plasma etching of aluminum with CCl4", Spectroscopy Letters, **18**(5), 335-352, 1985, dio: 10.1080/00387018508062236

[10] W.F. Hsieh, Henry Lin, Vincent Chen, Irene Ou and Y.S. Lou, "The Investigation of Fluorine Induced Novel Probe Marker Discoloration", in The Proceeding from the 44th International Symposium for Testing & Failure Analysis (ISTFA 2018), Phoenix USA, 413-417, 2018.

[11] W. F. Hsieh, H. Lin, V. Chen, I. Ou and Y. S. Lou, "The probe marker discoloration on Al pad and wafer storage," 2019 IEEE 26th International Symposium on Physical and Failure Analysis of Integrated Circuits (IPFA), Hangzhou, China, 1-4, 2019, doi: 10.1109/IPFA47161.2019.8984909.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 423-438 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

A Simulation Based Proactive Approach for Smart Capacity Estimation in the Context of Dynamic Positions and Events

Naeem Ahmed Haq Nawaz[*,1], Hamid Raza Malik[1], Ahmed Jaber Alshaor[1], Kamran Abid[2]

[1]*Department of Computer Science, Umm Al-Qura University, Makkah Al-Mukarmah, 24381, Saudi Arabia*

[2]*Department of Electrical Engineering, University of the Punjab, Lahore, 54590, Pakistan*

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 31 October, 2020
Online: 20 November, 2020

Keywords:
Smart
Capacity
Estimation
WSN
IoT
Sensor topologies
Position
Proactive approach
Cloud

ABSTRACT

Technologies are growing with the passage of time and providing solutions for the existing problems. One of such problems is to manage the crowd according to available capacity. Especially when the crowd density is dynamic because of dynamic position of the persons (Pilgrims). Not only dynamic position of the pilgrims makes crowd capacity dynamic but also special events increase and decrease in number of pilgrims that affect the capacity in a specific place. Therefore, it is not an easy job to estimate the available capacity according to the dynamic position and event. To overcome such problems of dynamic position and event-based crowd management different techniques and approaches are adopted. To solve the above-mentioned problem, this paper proposes a proactive approach to estimate the space occupy by the pilgrims in different positions in a zone or level. Maximum capacity in each zone and level is define on the basis of each position. On the basis of maximum and occupy capacities, available (remaining) capacity has been estimated in a zone and level according to the events such as Pilgrimage, Ramadan, Jummah etc. The occupied and available capacities in a zone, level or in the whole building can be estimated with the help of technologies such as Wireless Sensor Network (WSN), cloud computing, Internet of Things (IoT) and sensor topologies according to position and event. Simulation shows different scenarios according to the zones and different levels of building to prove the proactive approach. According to the results, it is concluded that zones, levels and point of entrances should be allocated to the pilgrims to avoid the congestion problems. Further for massive crowd and large area multi sink solution is better than single sink solution to estimate available capacity.

1. Introduction

Capacity estimation plays a vibrant role to maintain equilibrium among the maximum, occupied, and remaining capacities so that number of the persons can be allowed to enter in a zone, level or in a building [1]. Number of the persons permitted to enter an area directly dependent on the remaining capacity [2]. Although image-based method does not depend on the crowd to carry any type of device and counting can be performed without handling of device by the crowd. But the crowd counting done on the basis of images, does not provide result in case of dark or fade environment [3]. To provide the population awareness globally, crowd analysis can be done on the basis of video imagery to calculate and estimate the density of the crowd. It is considered an important tool of analysis for crowd density estimation at public places. Other methods also involve counting by performing the detection method or by using the clustering method for crowd estimation. Due to the more crowded environment one of the recent ways is regression. It has the capability to handle large crowded area [4]. The proposed proactive approach for capacity estimation is to calculate the number of the persons entering and exiting from the certain area. It is estimated according to the maximum capacity on the occasion of different positions and events. The sensor devices have the capability to sense and collect data at a given zone, level, building or an area [5]. In a specific scenario, data may include person ID (pilgrim ID), device ID, zone and level number, date and time for entry and exit etc. The sensed data then can be forwarded to the cluster head, then to a sink or

[*] Corresponding Author: Naeem A Nawaz, Ummal Qura University, nanawaz@uqu.edu.sa

server at each level and finally to central cloud storage location in order to calculate the remaining capacity of the areas. For different statistics, cloud computing can play a vibrant role to use the stored data. With the help of IoT data can be forwarded to different stakeholder for timely information to manage the crowd according to the position and events.

To overwhelm the problem of estimation of the capacity in different position and event, proposed proactive approach is one of the efficient solutions to provide advantages including:

- Counting number of pilgrims without human intervention
- Set the maximum capacity for different positions (Prayer, Walking, Sitting, Standing)
- Capacity estimation for different zone, Mataf area and on each level
- Capacity estimation according to Extreme crowded events such as, Jummah prayer, Ramadan, Eidain, Pilgrimage
- To restrict the pilgrims if they already performed Umrah so that remaining pilgrims can get opportunity
- Lost pilgrims can be traced within the coverage area

One of the central pillars of Islam is also Pilgrimage. Approximately 3.5 million pilgrims assemble to perform pilgrimage in the holy city of Makkah [6]. In the month of Ramadan millions of people gather to perform Umrah and other prayers [7]. Pilgrims perform different supplications at different places and time to achieve the Pilgrimage or Umrah requirements. To perform the Umrah during Ramadan people persist around the Masjid Al-Haram. The estimation of available capacity inside the Masjid Al-Haram, may be calculated by number of pilgrims in a specific zone, or level minus maximum threshold of the capacity. It has been observed that it is problematic to estimate capacity for a massive dynamic crowd, the number of pilgrims entered in Masjid Al-Haram and what is the remaining capacity inside? When it exceeds to maximum capacity? In this paper, we try to solve the above-mentioned problem by exploiting WSN. We proposed a proactive approach to process the data for different zones and levels to estimate available capacity. As I myself perform the Jummah Prayer many times, The Holy Masjid Al-Haram was so much overcrowded that it became difficult to perform Sajood and Tashahud. It is because of wrong capacity estimation of zone or level.

This paper is an extension of our previous work accepted in The International Arab Conference on Information Technology ACIT'2019 [1]. The extension includes capacity estimation of dynamic position, event, study of sensor topologies and proactive approaches.

1.1 Paper Layout

In section 2, we briefly discussed the technologies used in our proposed proactive approach that includes the technologies such as cloud computing, IoT, clustering and WSN. In Section 3, there are briefly given the approaches or strategies at present used for capacity estimation. Section 4 depicts the proposed proactive approaches according to zone, level and for the whole building. Section 5 elaborates the use case, section 6 offers comparison between existing and proposed capacity estimation, section 7 defines the algorithms, section 8 simulation and results, section 9 gives discussion and section 10 conclusion and future directions.

2. Technologies for Proposed Framework and Proactive Approach

2.1. Cloud Computing

Cloud computing is one of the most important technologies for proposed proactive approach. In cloud computing, cloud resources are located at different physical places. But via internet cloud computing makes efficient cloud services available globally to a large number of consumers located on many geographical locations [8], [9]. In a cloud computing environment, wearable devices are suggested for supportive communication between wearables and cloud server. Wearable device can be accessed remotely via the Internet [10]. Moreover, its everything-as-service-model (XaaS) impacts its clients and creators with many benefits [11]-[13]. Cloud computing is cost effective for the customers and uses the strategy "pay as you go" [14][15]. To process and access information in a more accurate way, cloud based software architecture has been proposed [16], [17]. Sensors, actuators and many other embedded devices are considered as a data generation source that provide help to enable IoT edge computing to process data [18]. Cloud can store the data according to the position, event and zones for future planning to control the massive crowd.

2.2. Internet of Things (IoT)

IoT lets people/anybody and things/devices connect anyone, anything, anytime, anywhere, for any service. Although it is flexible, complex and dynamic network infrastructure as given in [19]. In many fields such as healthcare, smart homes, smart cities, retail, traffic, security, and agriculture, IoT offers multiple solutions [20]. With the help of an IoT we can monitor the elements such as wind, crops, soil, atmosphere, lights and water over a big area [21]**.** For the growth of the IoT, there is a need of plentiful resources for the development of the intelligent devices and the WSNs. The core building blocks are the same for network communication in between devices to build an Internet of the Things setup [22]. With up to date development and advancement in the field of information and communication technologies, it also brought positive changes in the advancement and development of IoT. Through low-cost electronic devices and communication IoT paradigm improves human collaboration in the physical world [23], [24]. In the same way, it is also increasing the importance of WSNs to integrate and make connections with the Internet to develop different commercial and industrial applications [25]. In order to achieve different objectives, updated information is required efficiently during pilgrimage and Umrah. To achieve the above goal, IoT can play a vital role to get updated information as needed at a particular zone or level. Information received from IoT may be used by the management such as security management, and resource management (water, chairs, Quranic scripts etc…). So that necessary actions can be performed according to updated information on time in proactive way.

2.3. Clustering and WSN

There are some limitations in WSN such as energy limitation, coverage area, network lifetime, and bandwidth. These limitations

can be overcome with the help of some clustering schemes to get efficient solution [26]. Amongst the other factors, clustering is one of the most important factors. Clustering can also play an important role to control the performance of a network [27]. Wearable devices provide help to sense and disseminate information, such as physical activities, blood pressure, blood sugar level, heart rate) [28]. To monitor diverse environmental situations such as humidity, temperature, sunlight, and pressure, a smart sensing platform based on NodeMCU-ESP8266 microcontroller board has developed and given in [29]. With the implantation of WSN data can be sensed, stored, disseminated and processed to estimate the capacity [30], [31]. As an industrial IoT is an important infrastructure, WSNs play a crucial role in it. In IoT applications, efficiency balanced verification scheme for WSNs becomes a big contest, due to the resource limitation feature of sensor [32].

A proactive approach and framework can play an important role to enable reliable and an efficient management for WSN. One of the examples is Web Services based on clouds and lightweight cross-layer design [25].

2.4. Sensor Topologies

The topologies can differ in complication from a single node connected to a collector node to fully meshed network covering a large area. Flat or hierarchical architecture are used to design sensor topologies. In flat (peer-to-peer) architecture, each node either can be a sink or sensor node that has the same communication and computational abilities in the network. In hierarchical architecture, the nodes link with their particular cluster head in close proximity. Normally, cluster heads own more storage capacity and processing power as compared to any ordinary sensor node. Commonly line, star or bus topologies are used by the sensor networks that are physically wired together [33].

In Figure 1, point-to-point, Bus and linear topologies are given. A hardwired connection between two points is called a permanent point-to-point topology. A switched connection is a type of point-to-point connection in which control can be transferred between end nodes.

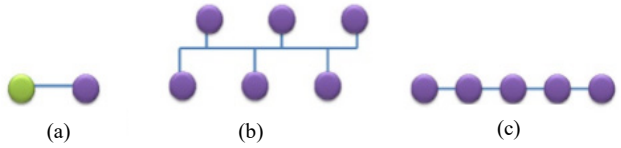


Figure 1: (a) Point-to-point, (b) Bus, and (c) Linear Network Topologies

In Bus topology, nodes are associated with a communication bus for communication. The advantages of the Bus topology are that it is simple and easy to install. But disadvantage is that the failure of central bus is the failure of the whole network, which is called a single point of failure. In Linear Topology, there is a two-way connection between the nodes. Only two terminating nodes has a single connection to their neighboring nodes at the end of the network, but all other nodes have link or connection to left and right nodes. In case of failure of a node, there is a disconnection from the network of any node connected to that node.

In Figure 2, ring, star and tree topologies are given. Network set up in a circular fashion is called ring topology. Each node in this configuration is connected to precisely two other nodes. Direction of flow of data is from the source to each node in clockwise or anticlockwise until it reaches to the intended recipient. In order to overcome this issue, a second communication ring is added by many networks that can transfer data in the opposite direction. In star topology, each node or device is connected to a single central device called hub or switch. Star topology has advantages such as easy to plan, implement, and add or remove the nodes. An intelligent node is needed as all data traffic flows through the central node. Entire network fails as a result of the failure of this node. In Tree topology, single “root node” is at the top of the hierarchy and other nodes are connected to the lower level in hierarchy in a tree manner structure. Many levels of nodes can be contained by a tree topology. This topology is extendable, and identification of the fault is easy and can be fixed easily because of simple structure. With the growth of the tree networks, it becomes more difficult to manage it.

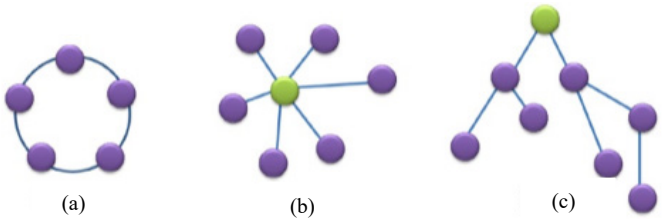


Figure 2: (a) Ring, (b) Star, and (c) Tree Network Topologies

In Figure 3, partially and fully mesh topologies are given. There are two arrangements of mesh topology: partially mesh and fully mesh. In partially mesh, some nodes are connected to more than one other node is called a partially connected mesh, while in fully mesh, every node is connected to every other node.

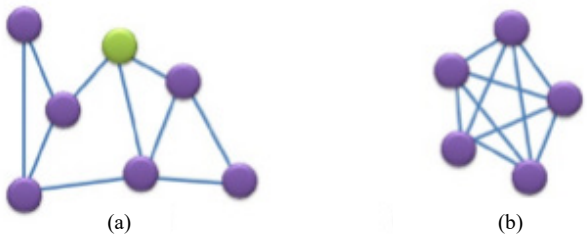


Figure 3: (a) Partially Connected and (b) Fully Connected Mesh Network Topologies

Cluster Tree topology consists of a single cluster or a multi-cluster network. Only one cluster-head (CH) is contained by a single cluster network. When all the nodes are connected to the cluster-head with single hop, the network topology becomes a star topology. More than one cluster-heads are contained by a multi-cluster network. An upper level sub-network is formed by all the cluster-heads. Nodes in different clusters communicate among themselves via their cluster heads but cannot directly talk to each other. Figure 4 illustrates the cluster tree topology, which has hierarchy architecture with the cluster-head network at the upper level and the clusters nodes at the bottom level [34]. The sink behaves as an entrance to the sensor nodes, gathering and updating the information. The sink plays a role as an interface among WSN, the external network and Internet, where the data are disseminated and processed. Sink can be single or multi sink depending upon the

requirement of the application. The goal of the single sink formation is to minimize the forwarding time and route information in the direction of a unique sink. Multi sinks are utilized for applications where tasks are distributed to different sinks. Multi sinks are used where there are high density crowd, large coverage area, redundancy, high latency, and possible high energy consumption [35]. In our scenario cluster tree or hierarchical cluster structure is suitable to collect the data from zones and levels during different positions and events.

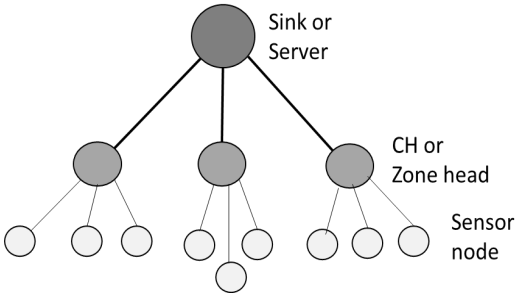


Figure 4: sensor cluster tree

During the events of Pilgrimage, Ramadan and Umrah the Holy city of Makkah (KSA) is crowded. Especially, to perform Umrah during the Ramadan all activities are around the Masjid Al-Haram, therefore it is need of the hour to organize, control and manage the crowded area smartly. The above objective can be achieved by proactive approach in smart way with the support of WSN, IoT, cloud computing and the management phases.

3. Present Capacity-Estimation Approaches

At present, capacity estimation is being done by management by taking view from the cameras or by the observation inside the Masjid Al-Haram. By cameras they discover the empty zone or level and update to the security officers to transfer the crowd to the empty or less crowded zone or level. They use the wireless system to give instructions to the security officers. By observation, the security officers check the capacity is available or not in an area where they are assigned. But they do not have auto capacity estimation approach so that they can measure how many pilgrims can be allowed in an empty space. They have just rough estimation of the available capacity. If the Zone or a level is reached to maximum capacity, the remaining crowd is diverted to the less crowded zone or level. If the Masjid Al-Haram is full then gates are covered with barrier and security officers do not permit pilgrims to get in, but they can make exit. No entry sign is turned on outside the gates. After that zones outside the Masjid Al-Haram are occupied. Sometimes zones and levels are overcrowded, and pilgrims face problem to perform the prayers or Tawaf. But currently, they do not have the smart estimation of the available capacity in a zone, level or in the whole Masjid Al-Haram. If there is smart capacity estimation with proactive approach, then pilgrims can be allowed according to the capacity available. Otherwise security stops the pilgrims at the gate and move to specified waiting area so that they do not overcrowd Masjid Al-Haram and do not cause any casualty.

4. Proposed Smart Capacity Estimation Framework and Proactive Approach

The Figure 5 represents the proposed smart capacity estimation proactive approach and framework. It provides help to estimate the capacity for zone and level to accurate the management. In order to ensure the smart capacity estimation in specific zone or level, the architecture uses WSN so that available capacity can be estimated for management automatically. The architecture provide help to understand the flow of data that is required to estimate the available capacity. It provides help to pilgrims to find out a specific zone or level where capacity is available. Figure 5 shows the overall picture of proposed framework and proactive approach for smart capacity estimation with the help of WSN, cloud storage and interaction among users by IoT for dynamic crowd management. It includes different phases: First phase involves smart capacity estimation according to time and preferable activity by pilgrims in a zone. Second phase involves smart capacity estimation according to time and their preferable activity by pilgrims at a level. Third phase involves smart capacity estimation according to time and their preferable activity by pilgrims at different levels, in order to calculate overall capacity occupied and available capacity in all zones or levels.

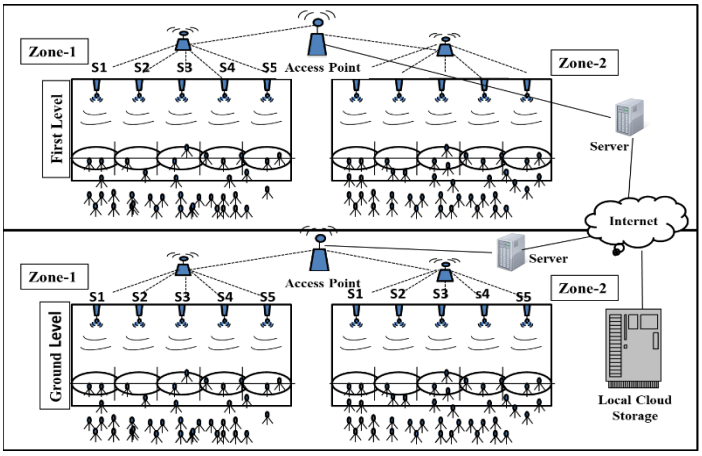


Figure 5: Proposed smart capacity estimation framework and proactive approach [1]

In the first phase, the calculation for the maximum capacity, capacity occupied by pilgrims and available capacity in the zone is done. In the second phase, head nodes or access point nodes fixed on the different zones at different levels will collect the data and forwarded to the server of different levels as update. In the third phase, data from each level from servers will be forwarded to the local cloud storage for further processing and management tasks. Detail of each proactive approach is given in subsection 4.1.

4.1. Proactive Approaches at Zone and Different Levels

Proactive approaches use to estimate the maximum, occupied and available (remaining) capacity with respect to zone, level and in a building.

The proactive approach for smart capacity estimation of zone is given as:

1- The Figure 6 depicts the steps to estimate the capacity based on a zone.

- Sensor nodes are fixed on each entry and exit point for each zone that can collect data about number of the pilgrims getting in and out of the zone.

- This information is updated to the zone access point or head node and from zone access point, it is forwarded to the level's server.
- The capacity occupied by pilgrims inside the zone can be calculated by number of pilgrims has entered minus number of the pilgrims exit.

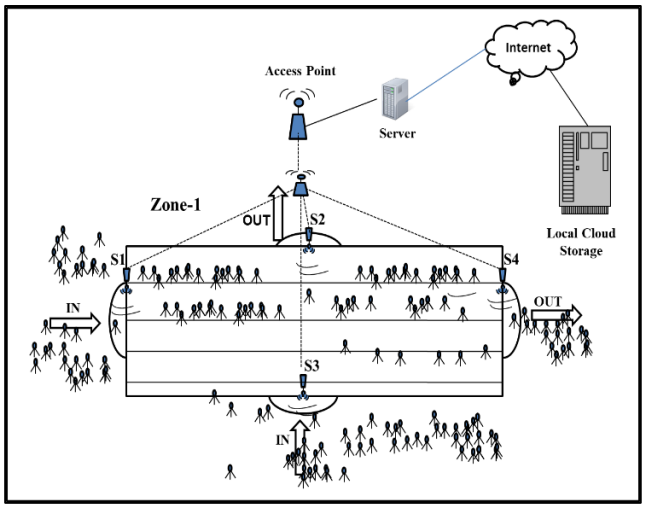


Figure 6: Smart capacity estimation of zone inside building [1]

Capacity occupied by pilgrims inside zone = number of pilgrims entered - number of the pilgrims exit

Z.IN=Entered to the zone
Z.OUT= Exit from the zone

$$Zone\ Capacity\ (Occupied) = \sum_{i=1}^{n}(Z.IN)i - \sum_{j=1}^{k}(Z.OUT)j \quad (1)$$

- The available capacity of the zone can be calculated by the maximum capacity of the zone minus total number of pilgrims inside.

The remaining capacity = maximum capacity of zone - (total number of pilgrims inside the zone).

Maximum Zone Capacity = Max (Z.C)

$$Zone\ Capacity\ (Available) = Max(Z.C) - \left(\sum_{i=1}^{n}(Z.IN)i - \sum_{j=1}^{k}(Z.OUT)j\right) \quad (2)$$

2- Proactive approach to estimate capacity at Ground Level is depicted in Figure 5.

- On each zone, sensor nodes will collect the data for pilgrims getting in or out. Afterward, it adds up or subtracts the number of the pilgrims got in and out from the head node or access point.
- The number of pilgrims getting in or out from all zones will be added up or subtracted by head nodes and then updated by the ground level server.
- If a pilgrim moves to the first level from ground level, then it will be subtracted from the ground level server but added into the first level server and vice versa.

L-IN=Entered to the level
L-OUT= Exit from the level

$$Level\ Capacity\ (Occupied) = \sum_{i=1}^{n}(L.IN)i - \sum_{j=1}^{k}(L.OUT)j \quad (3)$$

- From head node data can be forwarded to server to estimate the available capacity for ground level. If the ground level reaches to maximum capacity, then the ground level mentions no entry. Because ground level is full, it will be disturbance for the pilgrims if the pilgrims from outside entered into ground level and move to the first level. Therefore, security officers move the crowd to the first level from outside gate without entering into ground level and vice versa.

Maximum Level Capacity = Max(L.C)

$$Level\ Capacity(Available) = Max(L.C) - \left(\sum_{i=1}^{n}(L.IN)i - \sum_{j=1}^{k}(L.OUT)j\right) \quad (4)$$

3- Proactive approach to estimate capacity at First or at different levels is depicted in Figure 5:

- If the pilgrims move directly into the first level from outside gates, then there will be an addition to the ground level server and vice versa.
- If the pilgrims move into the First level via ground level, then it will be subtracted from the ground level but added into the first level and vice versa.
- From head node or access point data can be updated to the server to estimate the available capacity for first level. If the first level reaches to the maximum capacity, then at first level no entry appear at entrance. A message of maximum capacity is generated to the security officers by using IoT. The security officers move the crowd to second level from outside gate without entering into ground level or first level and vice versa.
- To avoid disturbance, pilgrims are allowed first to fill up top level then second and first level respectively. In this way first and second level pilgrims will not be disturbed.
- The data from each level forwarded to the local cloud storage to calculate the available capacity for the overall building. The data can be analyzed for different positions and events such as Ramadan, Pilgrimage, Jummah etc.

At present, most of prevailing wireless sensor devices are restricted because of memory, processing speed, efficiency and capabilities of communication. The above reasons are because of technology and economic limitations [36]. Low-power, low-cost, a multifunctional sensor device has been developed and designed by various industries for different applications, with high attention. Amongst the sensor devices one of the big issues in WSN is to classify and produce an organizational structure [37]. On behalf of clustering topologies, the processing of sensor devices can be improved which in turn improve the processing time. As a result, performance of capacity estimation will increase [38]. The collection of data will provide help to update information at zone, level, building and at crowded areas during Pilgrimage, Ramadan and Umrah for management.

The cloud storage, IoT and telecommunication play an important role in order to collect, assignment and process data locally and globally for smart capacity estimation [39]. To get results for different applications and different analysis, enormous amount of data can be sensed, stored and processed by the deployment of large number of sensor devices, cloud storage and IoT connections [40]. Detail of sensor topologies is given in section (2.4).

5. Use-Case for Smart Capacity Estimation

Data related to the pilgrims have been collected long time before by management and companies (travel agents) for pilgrimage, Umrah and visit. On the completion of e-visa process, management and companies receive information about pilgrim's residence in different sector or region. The information includes route of travel, reservation (city, hotel) and number of days to stay in different places. Pilgrims can be allocated the zones and levels according to their point of entrance according from their sector or region. As in Masjid Al-Haram, pilgrims have different positions therefore; capacities vary according to the different positions such as standing, sitting, walking and prayer positions. Capacity also depends on the events. Hence, it is important to measure the capacities according to each position and event.

5.1. Positions and Capacity

Capacity of the area is also depending on the position of the pilgrims such as standing, sitting, walking (performing Tawaf and Saee) and prayer position as given in Figure 7 and 8.In case of standing position space required for a pilgrim is less than as compared to the other position therefore, in standing position the available capacity will be maximum but in case of sitting, the pilgrim require more space than the pilgrim standing, therefore, in sitting position maximum available capacity will be reduced. In case of walking, the pilgrim required more space than sitting and standing position therefore, the maximum available capacity will be reduced as compared to sitting and standing maximum capacities.

In case of prayer position, the pilgrim needs more space than the standing, sitting, and walking position as the prayer position are dynamic such as standing (Qeyyam), sitting (Tashahud) and Sajood. Roughly, if we consider the standing position as the standard measurement then space ratio between standing and sitting position is measured as [5:9]. In case of walking position, the space ratio in between standing, and walking position is measured as [10:13]. In case of prayer position, the space ratio in between standing, and prayer position is as [1:3]. By calculating the capacities according to the crowd position, the crowd maximum capacity can be set according to standing, sitting, walking and performing prayer. At the time of prayer, the capacity will be minimum as compared to standing, sitting and walking position but in case of standing the capacity will be maximum as compared to sitting, walking and prayer position.

We can divide the sector for standing, sitting, walking and prayer. If crowd has different positions such as standing, sitting, walking and prayer within one sector or zone then we can calculate the average capacity. In case of The Holy Masjid Al-Haram, pilgrims want to see Baitullah (Kaaba), they can be managed on first, second and on top level near the balcony wall so that they can see the Kaaba in standing position. Sitting position is required when pilgrims want to recite the Holy Quran or Dhikr of Allah (Mention of Allah). They can be managed behind the Mataf area. Walking position is required in the Mataf or circular area on first, second and at top level. Prayer position can be at any sector but before prayer we need to calculate the capacity available for prayer and crowd inside the Masjid Al-Haram.

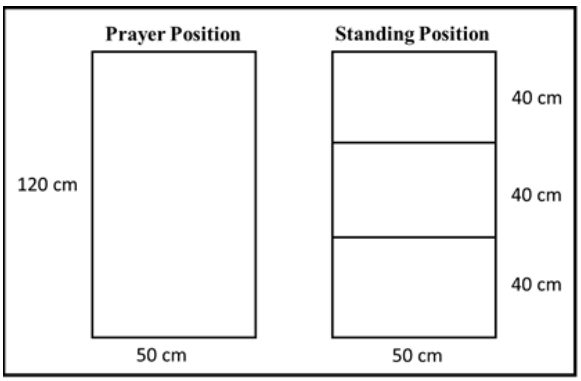


Figure 7: Capacity for prayer and standing positions

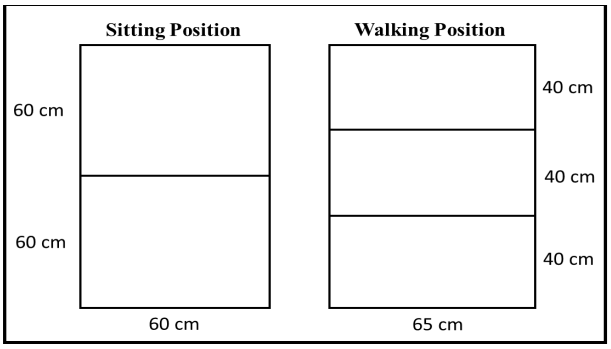


Figure 8: Capacity for sitting and walking positions

As we roughly calculated the space taken by the different position are given in Figures 7 and 8:

Standing = 40 x 50 = 2000 Sq CM

Sitting = 60 x 60 = 3600 Sq CM

Walking = 40 x 65 = 2600 Sq CM

Prayer = 120 x 50 = 6000 Sq CM

5.2. Stop and Wait

Calculate the average number of pilgrims entered in an area (Zone or sector) in unit time for example in one minute or in an hour. Average will provide help to calculate after how much time it is expected that the sector or zone will reach to its maximum limit. In this way we can control the crowd before it exceeds to its maximum limit. Usually it is observed that some time crowd exceeds the maximum limit and at prayer or Tawaf time pilgrims face a big problem which can lead to a stampede.

5.3. Move and Enter in the Zone

Rate of in and out to manage after maximum capacity, when zone and level reaches to its maximum capacity. Many pilgrims stay outside waiting for space availability. The space available for waiting pilgrims can be calculated by maximum capacity minus number of the pilgrims making the exit (going out). The space availability can be in a sector or zone or in whole area.

Table 1: Comparison between existing and proposed capacity estimation

S.No.	Existing Capacity Estimation	Proposed Capacity Estimation
i.	Initial capacity estimation is roughly calculated on the basis of previous experience and best guesses.	Initial capacity estimation will be calculated on the basis of data collected from central storage system.
ii.	Each zone capacity is calculated by the security officers or by computerized camera's system.	For each zone, automatically real time zone capacity will be estimated by zone sensors and sensor head.
iii.	Capacity available in a zone is controlled manually not automatically by security officers with the help of wireless system.	Capacity available in a zone will be calculated and automatically displayed. Zone security officer will manage the zone according to the available capacity.
iv.	There is no assurance about the exact maximum capacity estimation. Wrong estimation of maximum pilgrims' capacity can overcrowd the zone or level.	Maximum capacity for ach zone or level is already set therefore; maximum capacity alert will be generated automatically for no entry without any delay.
v.	According to dynamic positions and events there is no proper capacity records for zone and level.	Security officers can get records from local cloud storage system for future planning for each zone and level according to positions and events.
vi.	Real time capacity available cannot be calculated therefore capacity available can only be roughly estimated.	Real time capacity available will provide help to allow exact number of pilgrims to enter into a specific zone or level.
vii.	Difficult to estimate capacity available during different positions, events during prayer time and Tawaf time as each activity has different capacities that is number of pilgrims.	Capacity available during different positions, events during prayer time and Tawaf time is automatically estimated because prayer time and Tawaf time maximum capacity threshold can be set automatically according to activity.
viii.	There is no automatic way to trace out the lost pilgrim because existing system only support to manual capacity estimation and for security purpose.	Pilgrim lost can be traced out on the bases of the sensor device location with help of IoT.
ix.	CCTV cameras implementation provide help to observe and control the act of misconduct but there is no automatic way to identify the pilgrim (mis-conductor).	If the act of misconduct is done, then pilgrim (mis-conductor) will be identified on the bases of information stored on sensor device and on the cloud storage with the help of IoT.
x.	There is no such application to provide schedule, when the capacity is available for a specific building, sector and region.	Application can be developed on the basis of IoT to provide the schedule for visit to Masjid Al-Haram according to building, sector and region.

5.4. Evacuation or Exit from the Zone

In any case if the capacity exceeds to maximum limit in a zone or level, it can be managed in another zone or level. In worst case the pilgrims have to evacuate to the waiting place and put into high priority to get in when space will be available.

6. Comparison between Existing and Proposed Capacity Estimation

In the section 3, capacity estimation about existing system is described. It consists of manual capacity estimation, wireless system and computerized camera's systems. The proposed capacity estimation is discussed in the section 4. The comparison and preliminary analysis between existing and proposed smart capacity estimation systems are given in Table 1.

7. Algorithm for Proposed Capacity Estimation

We have taken the simple scenario to design the algorithms at different zones, different levels and in the whole building.

7.1. Algorithm for Zone Capacity Estimation

In zone capacity estimation, we simply check the available capacity in zone. If maximum zone capacity is greater than the number of pilgrims has entered in a zone minus number of pilgrims have exited from the zone. If the condition is true, then display the GREEN sign and also display the available capacity for pilgrims in waiting position. Otherwise display RED sign that zone is full and there is no entry.

Algorithm 1: Zone capacity estimation

```
Step-1   Begin
Step-2   GET value for Max(Z.C)
Step-3   If (Max(Z.C) > (Z.IN)-(Z.OUT))
Step-4      Display LED Turn(Green)
Step-5      Display Zone Available Capacity
Step-6   Else
Step-7      Display LED (RED)
Step-8   Endif
Step-9   End
```

7.2. Algorithm for different Level Capacity Estimation

In different levels capacity estimation, we simply check the available capacity in a level. If maximum level capacity is greater than the number of pilgrims has entered in a level minus number of pilgrims have exited from a level. If the condition is true, then display the GREEN sign for available capacity. Also display the zone number and level number where capacity is available for pilgrims in waiting position. Otherwise display RED sign that level is full and there is no entry.

Algorithm 2: Different level capacity estimation

```
Step-1   Begin
Step-2   GET value for Max(L.C)
Step-3   If (Max(L.C) > (L.IN)-(L.OUT))
Step-4      Display LED Turn(Green)
Step-5      Display Zone Number & Available Capacity
Step-6   Else
Step-7      Display LED (RED)
Step-8   Endif
Step-9   End
```

Cluster verification algorithm [31], represents the scenario of verification of cluster members (CMs) in cluster form. It discusses the number of verified and unverified CMs in a cluster. But in capacity estimation algorithm, it verifies and count the number of the pilgrims getting in or out from zone, level or in a building. All CMs remain within the cluster head transmission range for verification. But in capacity estimation algorithm, the pilgrims in the range of entry and exit sensors are verify when they make entry or exit from zone or level. CMs can enter to different clusters but in capacity estimation, it is not allowed to enter other than allocated zone or level. Zone and level are allocated so that pilgrims enter from the gate that is near to their zone and level. In this way crowd will be distributed in different zones, as a result a single path and zone will not be crowded. In the cluster verification, the number of the CM verify are listed out but in capacity estimation, it tells the available capacity for a zone and as well level and building. Cluster verification represents the number of CMs supported by a cluster head but in capacity estimation, the number of the pilgrims can be managed according to the available space in a zone or level or building. In cluster verification more than one CM can be verified. But in capacity estimation, the verification and counting of the pilgrims is done at entrance and exit one by one. In cluster verification, cluster head generate the signals for verification and in response CMs generate signals for their identification. But in capacity estimation, pilgrims (sensor devices) generate signals (that is an event driven approach) for verification and in response zone sensors verify and count the number of pilgrims getting in or out from zone.

8. Simulation and Results

Contiki OS is a lightweight open source operating system for sensor network. It was designed by Adam Dunkels, at the Swedish Institute of Computer Sciences. C programming language is used in both Contiki OS and its programs. Contiki is flexibly portable OS and it has been ported to different platforms. Texas Instruments MSP-430 in addition to Atmel AT mega series of microcontrollers are the most often used platforms. Contiki employs event-driven programming model to deal with concurrence. It is Contiki's main advantage that all processes share one stack, allowing savings of memory. In order to utilize this model Proto threads. Unconditional blocking wait is provided by Proto threads to save the state when it blocks. It jumps back to the next instruction on resumption of Proto thread. It also employs Rime stack that is a stack for sensor networks for lightweight communication. The main purpose of Rime stack is to support code reusing and simplify the implementation of sensor networks. Contiki is used in several systems, such as electric meters, alarm generating systems, remote control of house, radiation detection and streetlights control. Cooja is Java based Contiki network simulator with a graphical user interface (GUI). It is a trial system given by Contiki, and additionally empowers with original instrumentation before running it on the real target hardware. It integrates to simulate with the external tools to provide additional features to the large and small networks. Motes can be emulated, which is faster and allows simulation of larger networks, or at the hardware level. Two emulator software packages are contained by this tool: Avrora and MSPSim. Cooja can emulate multiple platforms like: TelosB/SkyMote, MicaZ mote, Zolertia Z1 mote, ESB, Wismote [41].

The Figure 9 shows full screen view of Contiki/Cooja simulator. Network simulator "Contiki/Cooja" is separated into small windows such as: Network Window, Notes Window, Simulation Control Window and Mote Output Window.

Network Window is employed to draw the WSN in random manner, linear, ellipse or manual positions. Notes Window is employed to write down the information or detail about the simulations and detail of its results. In order to start, stop, pause or reload the simulations, simulation control Window is used. The results for data packets have transmitted and received is provided by Mote Output Window. For our scenario, we use the Tmote Sky platform. Tmote Sky is one of the high data rate platforms work with low power consumption. It works well for many months to years without replacing the battery. Therefore, it is one of the most demanding wireless sensors. It is notable that options are best used in a selection manner, as the window can become disordered with too many selected options. The number for each mote is displayed by Mote IDs. Mote Output window can be used for analyzing and filtering results. On the simulated motes, LEDs are beneficial in order to observe the LED lights.

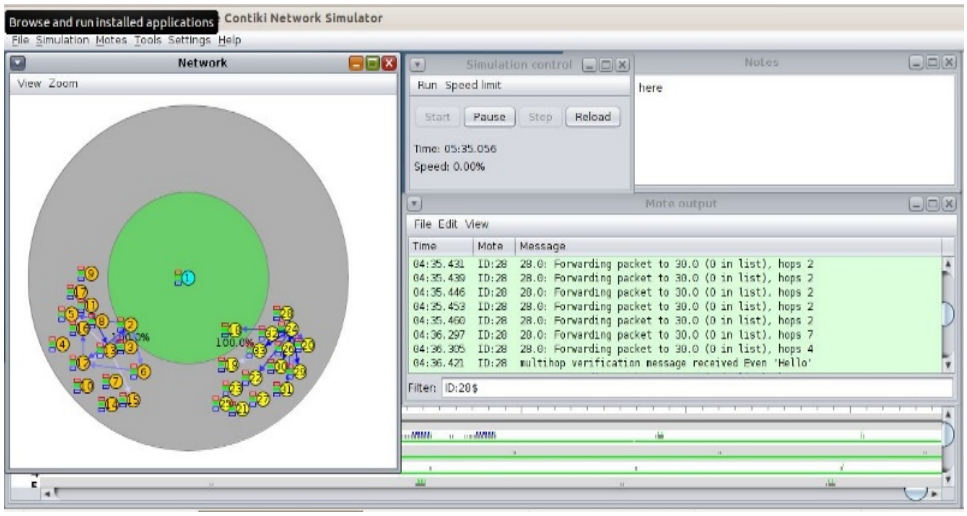


Figure 9: Full view of the Contiki/Cooja simulator

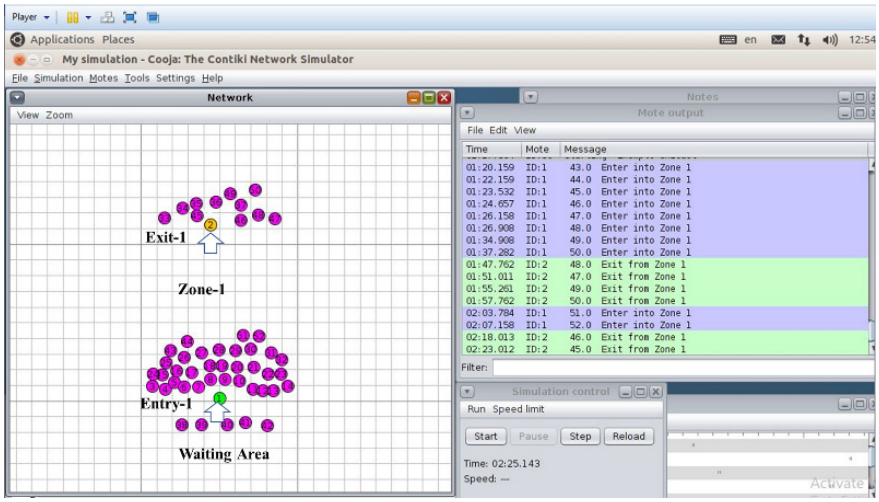


Figure 10: Single zone having maximum capacity 50

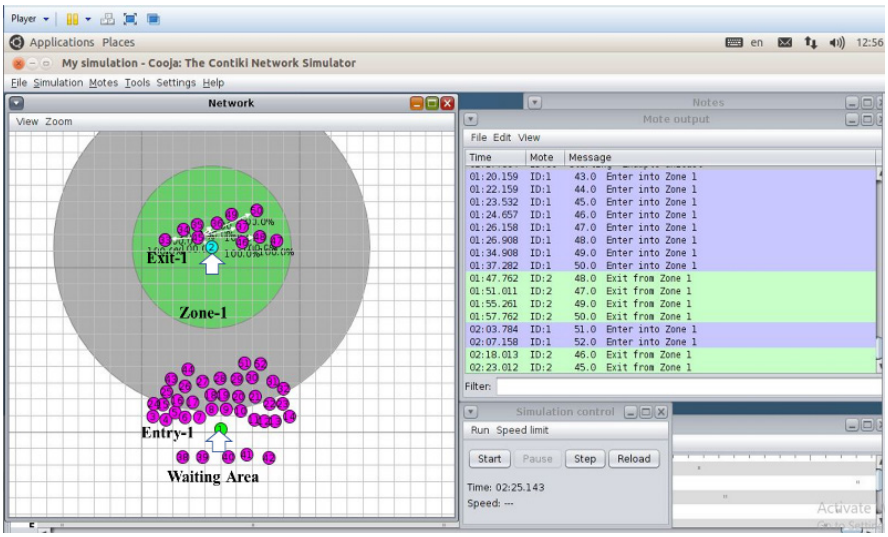


Figure 11: Single zone having maximum capacity 50 with coverage

We took the simple scenario to make understanding of the flow and function of each node as given in Figures 4 and 8. Head node function is to keep the record and status of CMs and forward record to the server for update. The status of the CM can be waiting to enter, enter and exit. Since the log file generated by the Contiki/Cooja simulation tool, we extracted the data for single and double zone in a level. For each zone we calculated the number of the pilgrims entered, exited, zone capacity occupied and available capacity for pilgrims. From this datum we can calculate the capacity for different positions and events. We also draw the table and graph for better understanding.

8.1. Single Zone Scenario

In single zone scenario, we consider one entry point and one exit point. We also set the maximum value for zone then we allow the specific number of the CMs to enter the zone and some of them allowed to exit from the zone. The simulation is shown in the Figures 10 to 18 and data is given in the Table 2. To make clarity of the figures, we consider the complete figure and then divided into its small parts (network, mote output and simulation control windows).

The simulation for maximum capacity 50 pilgrims is shown in the Figures from 10 to 14 and data is given in the Table 2.

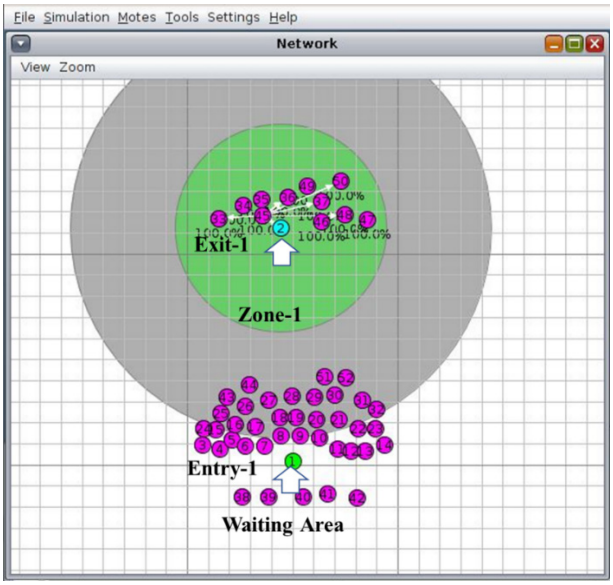


Figure 12: Network window from Figure 11

The Figures 10 and 11 represent the single zone having maximum capacity of 50 pilgrims. Figure 10 shows the capacity without coverage area while Figure 11 shows the capacity with coverage area in circular form.

The Figure 12 represents the network distribution of sensor nodes, entry point and exit point nodes in single zone having maximum capacity of 50 pilgrims. It also shows the number of nodes in waiting, occupy and exit position to calculate the available capacity.

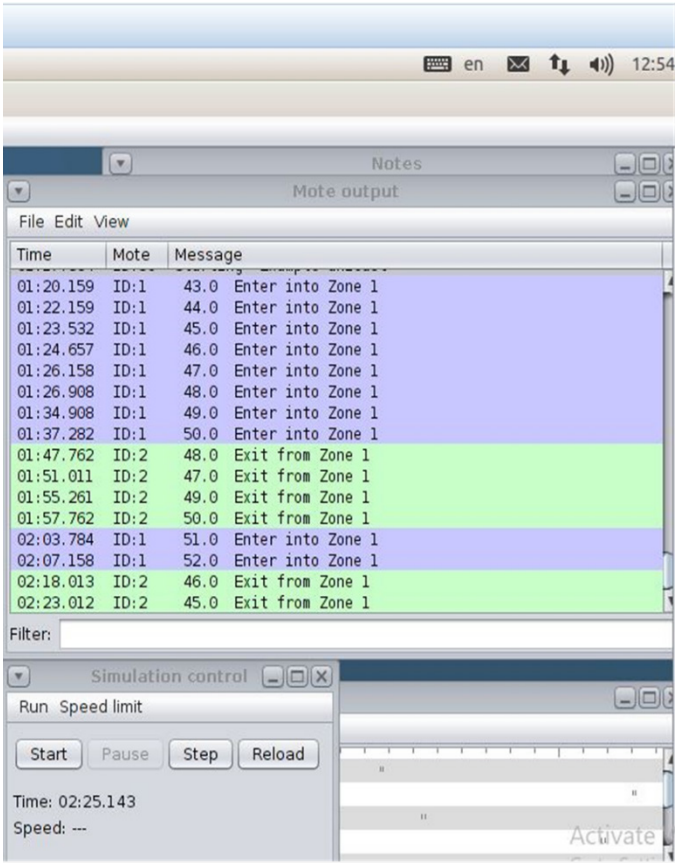


Figure 13: Mote output and simulation control windows from Figure 11

Figure 13 represents the most output with time, mote ID and its status enter or exit. The mote output is saved in the log file from where we can estimate the available capacity in a zone.

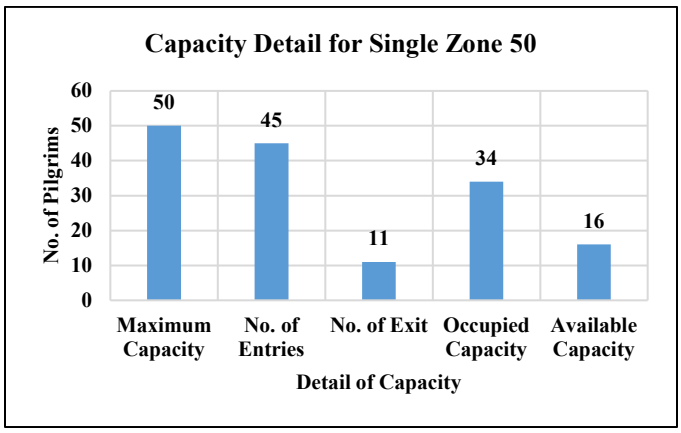


Figure 14: Capacity detail for single zone having maximum capacity 50

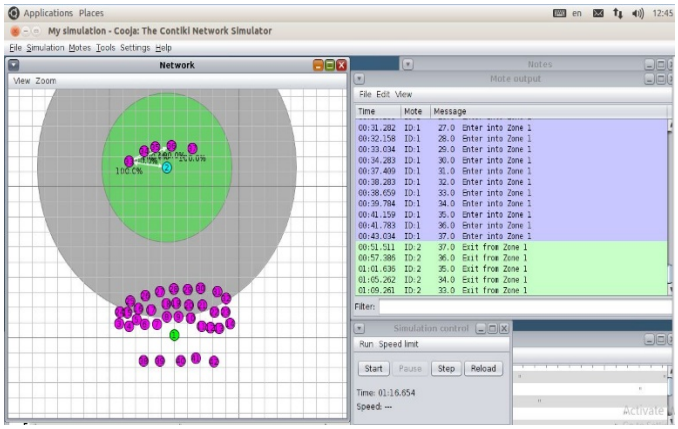


Figure 15: Single zone having maximum capacity 40 with coverage

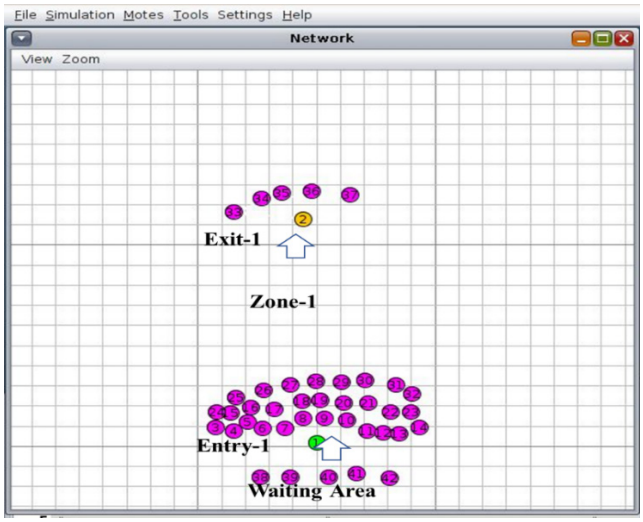


Figure 16: Network window from Figure 15

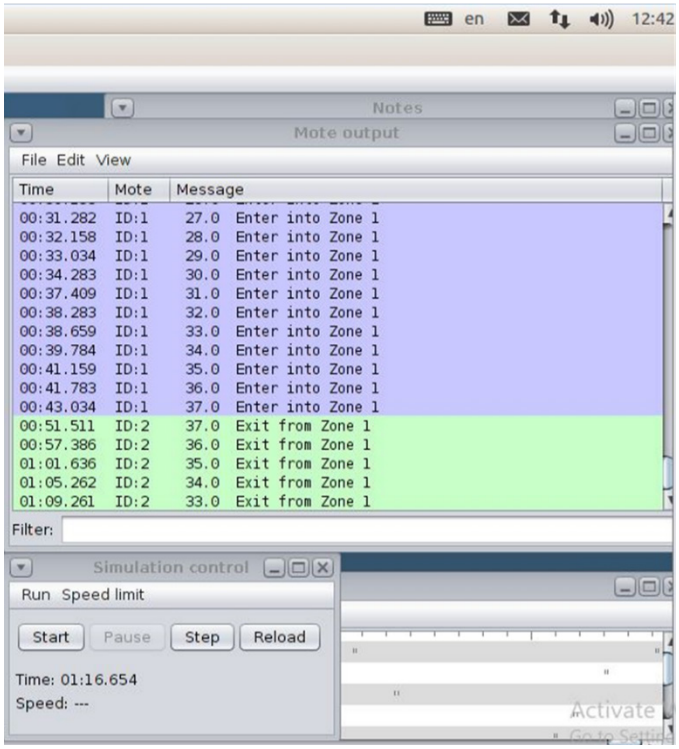


Figure 17: Mote output and simulation control windows from Figure 15

Figures from 10 to 14 considered that the maximum capacity for zone is 50 Pilgrims. Graph in Figure 14 shows that 45 of them have entered and 5 are in waiting position. 11 pilgrims have made the exit, occupied capacity in the zone is 34 and available capacity is 16.

The simulation for maximum capacity 40 is shown in the Figures from 15 to 18 and data is given in the Table 2. To make clarity of the figures, we consider the complete figure and then divided into its small parts (network, mote output and simulation control windows).

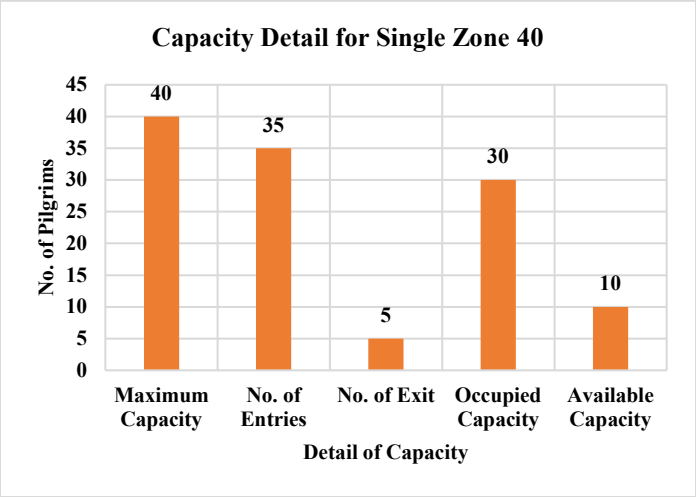


Figure 18: Capacity detail for single zone having maximum capacity 50

Figures from 15 to 18 show the maximum capacity for zone is 40 Pilgrims. Graph in Figure 18 shows that 35 out of 50 pilgrims have entered and 5 are in waiting position. 5 pilgrims have made the exit, occupied capacity in the zone is 30 and available capacity is 10.

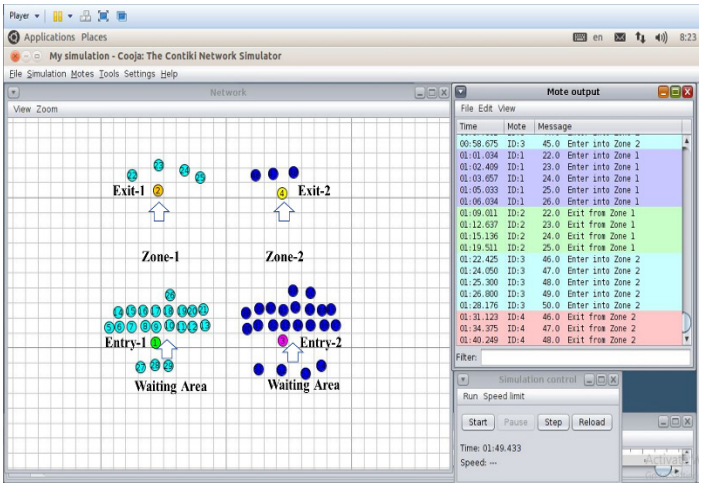


Figure 19: Each zone has maximum capacity 25 but total 50

8.2. Different Levels and Double Zone Scenarios

In different levels and double zone scenarios, we consider one entry point and one exit point in each zone and also consider that each level has double zone. We also set the maximum value for each zone and each level. We allowed the specific number of the nodes to enter in each zone and some of them allow to exit from each zone. The simulation is shown in the Figures 19-32 and data are given in the Table 2.

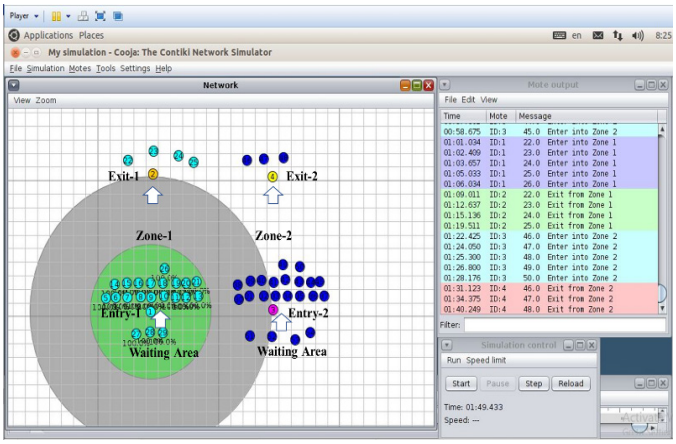


Figure 20: Each zone has maximum capacity 25 but total 50 with coverage

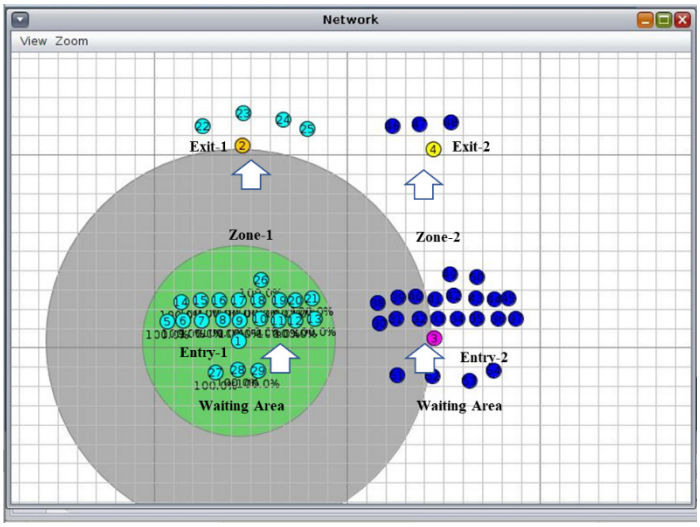


Figure 21: Network window from Figure 20

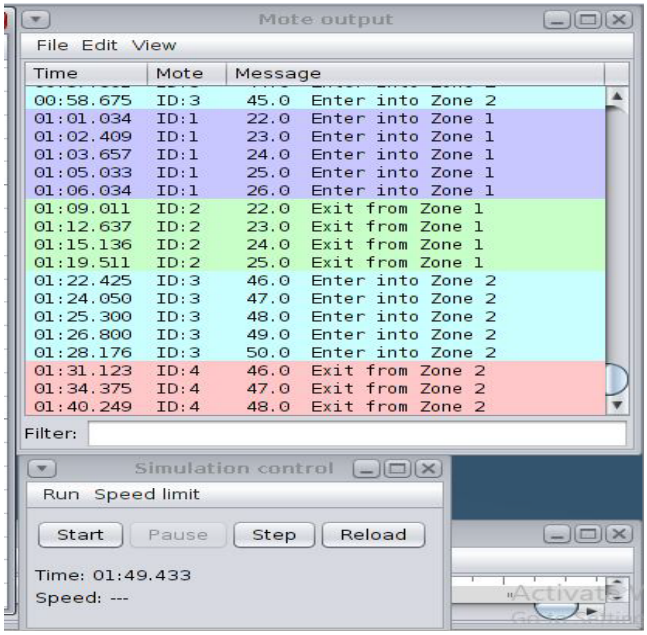


Figure 22: Mote output and simulation control windows from Figure 20

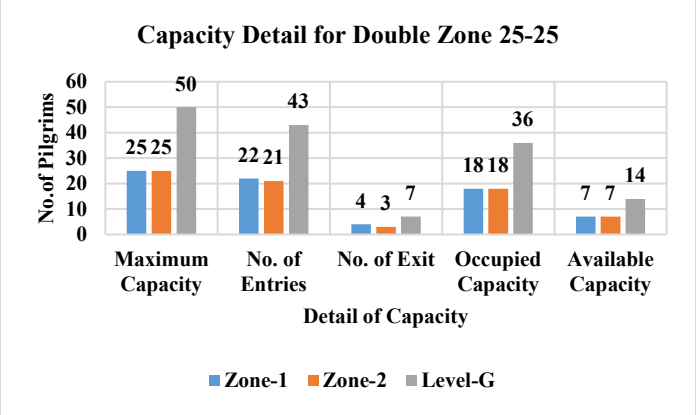


Figure 23: Capacity detail for double zones having maximum capacity 25 in each but Level-G has maximum capacity 50

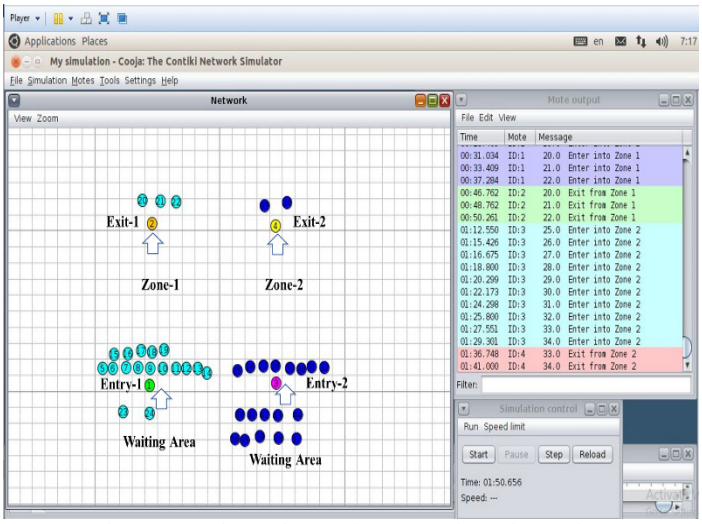


Figure 24: Each zone has maximum capacity 20 but total 40

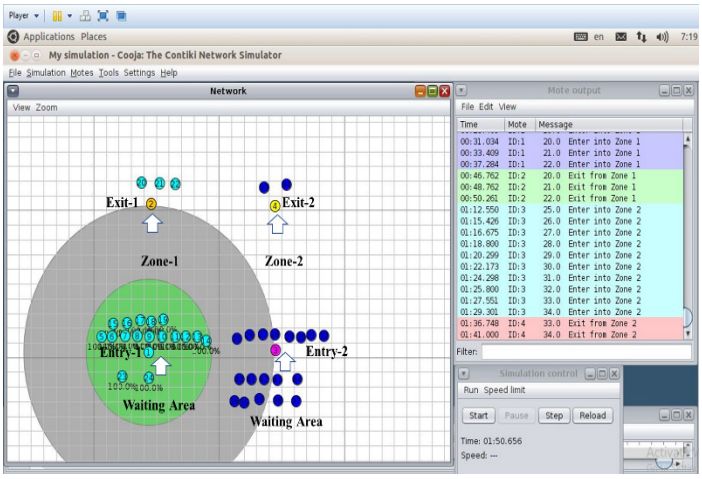


Figure 25: Each zone has maximum capacity 20 but total 40

In Figures 19 to 23, it is considered that the maximum capacity for zone-1 and zone-2 is 25. Zone-1 and zone-2 made the level-G and total capacity is 50 Pilgrims. In zone-1, 22 have entered and 3 are in waiting position. 4 pilgrims have made the exit, occupied capacity in the zone is 18 and available capacity is 7. In zone-2, 21 have entered and 4 are in waiting position. 3 pilgrims have made the exit, occupied capacity in the zone is 18 and available capacity is 7. From the log file or zone-1 and zone-2, we calculated the capacities for Level-G. Level-G has maximum capacity 50 pilgrims. 43 have entered and 7 are in waiting position. 7 pilgrims have made the exit, occupied capacity in the zone is 36 and available capacity is 14.

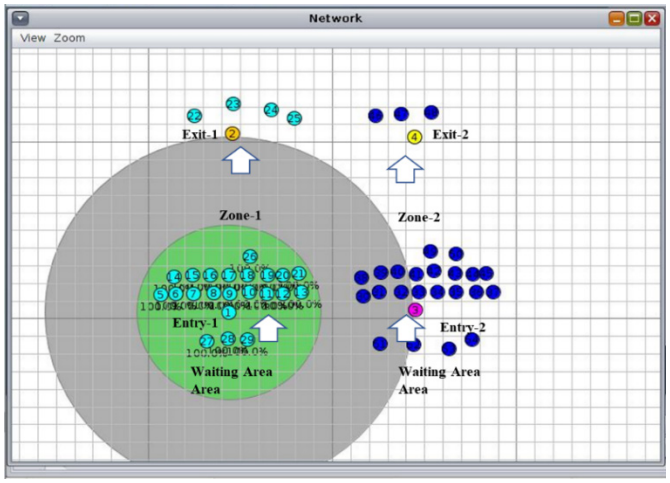


Figure 26: Network window from Figure 25

In Figures 24 to 28, it is considered that the maximum capacity for zone-1 and zone-2 is 20 pilgrims. Zone-1 and zone-2 made the level-1 and total capacity in level-1 is 50 Pilgrims. In zone-1, 18 have entered and 2 are in waiting position. 3 pilgrims have made the exit, occupied capacity in the zone is 15 and available capacity is 5. In zone-2, 10 have entered and 10 are in waiting position. 2 pilgrims have made the exit, occupied capacity in the zone is 8 and available capacity is 12. From the log file or zone-1 and zone-2, we calculated the capacities for Level-1. Level-1 has maximum capacity 40 pilgrims. 28 have entered and 12 are in waiting position. 5 pilgrims have made the exit, occupied capacity in the zone is 23 and available capacity is 17.

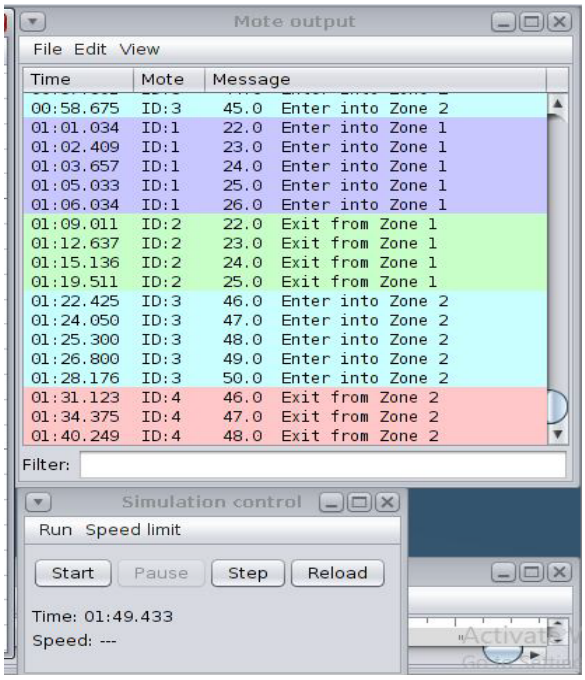


Figure 27: Mote output and simulation control windows from Figure 25

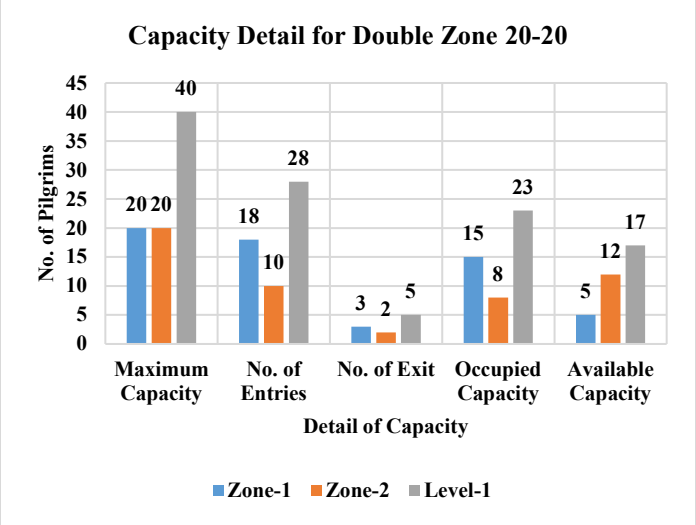


Figure 28: Capacity detail for double zones having maximum capacity 20 in each but Level-1 has maximum capacity 40

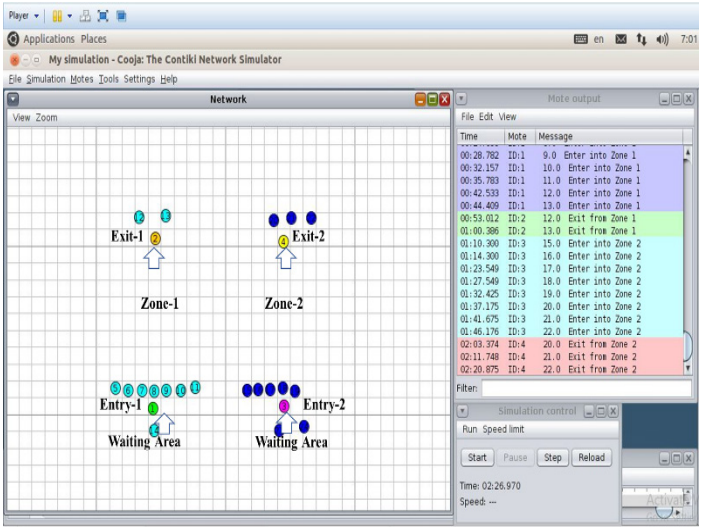


Figure 29: Each zone has maximum capacity 10 but total 20

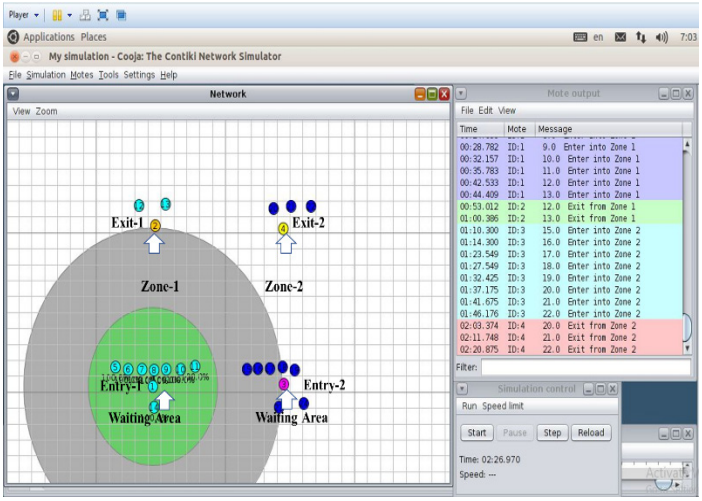


Figure 30: Each zone has maximum capacity 10 but total 20 with coverage

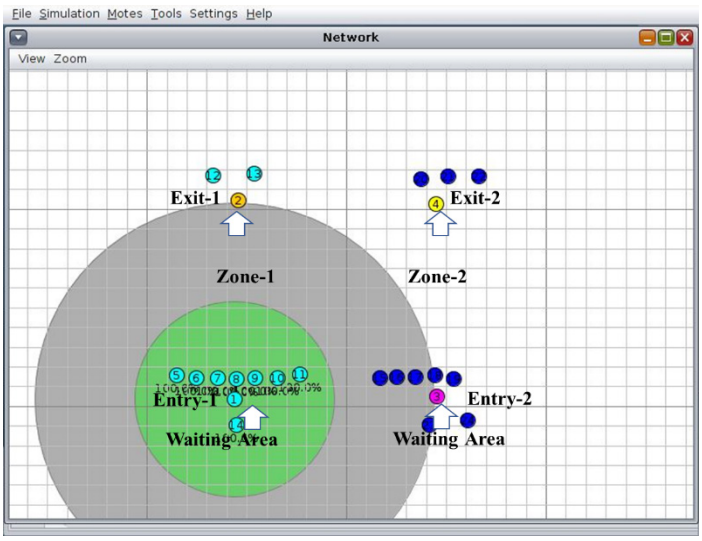


Figure 31: Network window from Figure 30

In Figures 29 to 32, it is considered that the maximum capacity for zone-1 and zone-2 is 10 pilgrims. Zone-1 and zone-2 made the level-2 and total capacity in level-2 is 20 Pilgrims. In zone-1, 9 have entered and 1is in waiting position. 2 pilgrims have made the exit, occupied capacity in the zone is 7 and available capacity is 3. In zone-2, 8 have entered and 2 are in waiting position. 3 pilgrims have made the exit, occupied capacity in the zone is 5 and available capacity is 5. From the log file or zone-1 and zone-2, we calculated the capacities for Level-2. Level-2 has maximum capacity 20 pilgrims. 17 have entered and 3 are in waiting position. 5 pilgrims have made the exit, occupied capacity in the zone is 12 and available capacity is 8.

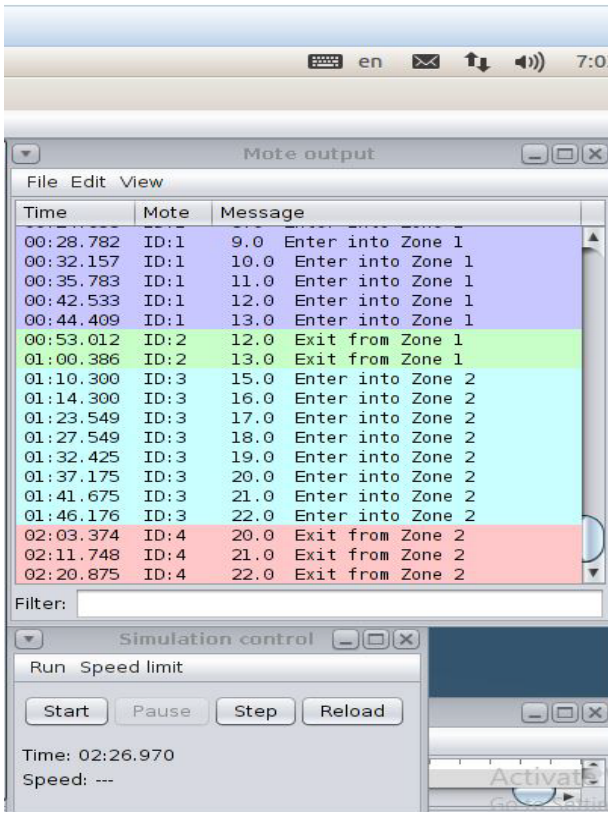


Figure 32: Mote output and simulation control windows from Figure 30

Table 2: Zones and Levels and capacities

Zone/Level	Maximum Capacity	No. of Entries	No. of Exit	Occupied Capacity	Available Capacity
Single Zone	50	45	11	34	16
	40	35	5	30	10
Level/Double Zone					
Zone-1	25	22	4	18	7
Zone-2	25	21	3	18	7
Level-G	50	43	7	36	14
Zone-1	20	18	3	15	5
Zone-2	20	10	2	8	12
Level-1	40	28	5	23	17
Zone-1	10	9	2	7	3
Zone-2	10	8	3	5	5
Level-2	20	17	5	12	8

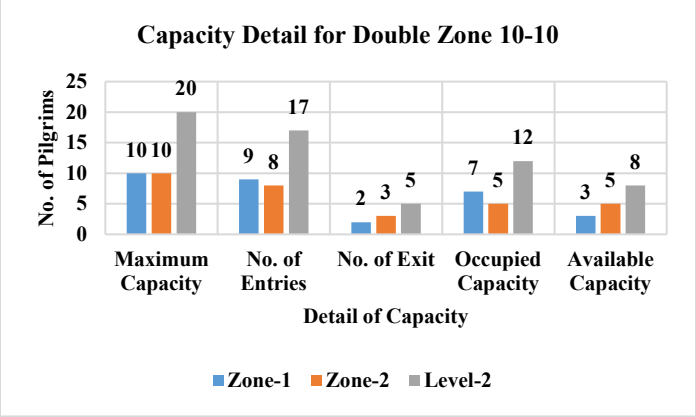


Figure 33: Capacity detail for double zones having maximum capacity 10 in each but Level-2 has maximum capacity 20

9. Discussion

The present analysis deals with capacity estimation for pilgrims at crowded place. The prime focus is on the allotment of zone and level number. This is performed according to moving status and events, single or multi sink, entry and exit point.

The concern of discussion is the allotment of zone and level number to pilgrims. According to discussion allotment of zone and level number to pilgrims should be specific (fixed) or flexible for dynamic position and events accordingly. Specific zone and level number according to dynamic positions and events is preferred because it is noticed if allotment is flexible then some zone and level number are more crowded than available capacity while others remain empty or less occupied. Furthermore, prayer position requires more area than standing and sitting position comparatively as given in section (5.1). As compared to routine days, pilgrimage, Ramadan and Jummah needs more area. The above mention reasons lead to non-uniform distribution of crowd. What is more, pilgrims having flexible zone and level move in different directions. It can result in congestions (bottleneck) that can lead to stampede. But in particular case, pilgrims are allotted the specific tracks to approach their zone and level number, that is why, chances of bottleneck and stampede are lessened or minimized. In flexible case, pilgrims are unable to move flexibly from overcrowded to less crowded zone or level (where capacity is available). This is because they have right to enter any zone and level number. But in particular case, pilgrims are allotted by the zone and level number. They can only move into allotted zone and level number. As allotment is performed considering maximum capacity of zone or level (a proactive approach), the zone or level does not get overcrowded. Place can be occupied within the zone on the basis of first come and first serve.

Entry point and exit point is the second point of discussion. In case of flexible, the pilgrims can enter and exit from any one of the points. Again, some entry points become more crowded than the other and cause bottleneck and there are chances of stampede. Moreover, pilgrims' attempt of entry and making exit from same location also causes bottleneck and stampede. But in specific case, each zone and level have particular point of entry and exit. In this way, chances of problem of bottleneck and stampeded can be overcome. Furthermore, pilgrims can conveniently access the transport of their residential area.

Third point of discussion is that there should be a single or multiple sink for WSN system. Massive crowd, scalability, latency, reliability, and network lifetime are the numerous challenges that may be created by the design of WSNs. The question whether sink should be single or multi sink is dependent upon its application. In the case of single sink and massive crowd, single sink has to control all traffic. The streaming of data traffic will be in one direction which causes congestion. There are probabilities that single sink is ineffective to control massive amount of data which causes the drop of data. This drop of data will cause the high latency rate. However, there are separate solutions to control the massive crowd and data in case the multi sinks are used. In the case of multi sink, data flow will be in multiple directions rather in one direction. Probability of congestion can be minimized in this way. In order to avoid the congestion and to distribute the task among sinks different type of data traffic or task can be allotted to different sinks. If the data is dropped on one path, it can be received on the alternative path. In this way latency rate can be minimized. When

data is received from different paths and sinks, it will provide the data redundancy. Single sink solution is not appropriate in case of long-distance communication because transferring data to long distance with the support of sensor nodes means consumption of energy to a lot degree. This consumption of energy will lead the sensor node die early and hence it will shorten the network life span. But in case of multi sink, the sensor nodes will get the shortest path as multi sink will divide the path and will consume less energy by increasing the network life span. Increasing of lifetime of network and receiving of data from multiple courses increase the reliability of the network. In our case, as data is massive, crowd is large, dynamic, and on large area therefore WSN required the multi sinks solution.

10. Conclusion and Future Directions

This paper provides an overview of a dynamic positions and events, proactive approach with framework and its implications towards smart capacity estimation using WSN and cloud computing in the IoT paradigm.

We discussed the framework and proactive approach in the context of dynamic position and events for smart capacity estimation. We examined how smart capacity estimation automate and minimizes time to manage available capacity in a zone or level by using framework and proactive. Explanation is done by use-case of masjid Al-Haram for smart capacity estimation for crowd management. We also compared the existing capacity estimation and smart capacity estimation systems. We also defined the algorithms with respect to Zone and Level. We also ran different simulation scenarios and also discussed the results. In future we will calculate the capacity according to the different position. The dynamic capacity can be calculated on the basis of time frame and capacity can be maximized and minimized according to time frame. In future algorithms can be defined how to maintain social distancing, how to calculate the capacity with social distancing and find the empty space with direction? Algorithms devised how to define the exit plan to move out without facing the congested path. Smart capacity estimation with Sensor and without sensor. Without sensor, capacity of pilgrims can be counted by putting the manual counter. If proposed proactive approach and framework for smart capacity estimation is commercially applied, it can provide help in different application developments, such as: 1. Capacity available in a zone or at a level can be calculated precisely. 2. To estimate the occupied and remaining capacity in a building or in an area. 3. Alert generation before the zone or level reached to its maximum threshold. 4. Positions and Events based capacity estimation. 5. Capacity estimation on the basis of social distancing. 6. Calculation of the social distance. 7. Drinking water estimation to supply the water in different zones and levels. 8. On the basis of wheelchairs' requests, wheelchairs supply plan according to demand in different zones and levels. 9. Estimation of the Holy Quran scripts according to different languages. 10. If water, chairs or the Holy Quran scripts are not enough in a zone or level then generate the alert to workers for supply. 11. Sharing of zone capacity data with different zone security officers to manage the crowd capacity. 12. In case of emergency, provide path for evacuation from affected zone or level to outside.

References

[1] N.A. Nawaz, H.R. Malik, A.J. Alshaor, "A framework for smart capacity estimation at crowded area using WSN," in Proceedings - 2019 International Arab Conference on Information Technology, ACIT 2019, 159–164, 2019, doi:10.1109/ACIT47987.2019.8991037.

[2] H. Paloheimo, M. Lettenmeier, H. Waris, "Transport reduction by crowdsourced deliveries – a library case in Finland," Journal of Cleaner Production, **132**(2016), 240–251, 2016, doi:10.1016/j.jclepro.2015.04.103.

[3] W. Xi, J. Zhao, X.Y. Li, K. Zhao, S. Tang, X. Liu, Z. Jiang, "Electronic frog eye: Counting crowd using WiFi," in Proceedings - IEEE INFOCOM, 361–369, 2014, doi:10.1109/INFOCOM.2014.6847958.

[4] C.C. Loy, K. Chen, S. Gong, T. Xiang, Crowd Counting and Profiling: Methodology and Evaluation, 347–382, 2013, doi:10.1007/978-1-4614-8483-7_14.

[5] B. Guo, Z. Wang, Z. Yu, Y. Wang, N.Y. Yen, R. Huang, X. Zhou, "Mobile Crowd Sensing and Computing: The review of an emerging human-powered sensing paradigm," ACM Computing Surveys, **48**(1), 1–31, 2015, doi:10.1145/2794400.

[6] S. Hashish, M. Ahmed, "Efficient wireless sensor network rings overlay for crowd management in Arafat area of Makkah," in 2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems, SPICES 2015, 1–6, 2015, doi:10.1109/SPICES.2015.7091437.

[7] H. Elazhary, "Cloud-based Context-aware Mobile Applications and Framework for Hajj and Umrah Management," International Journal of Computer Trends and Technology, **47**(2), 106–115, 2017, doi:10.14445/22312803/ijctt-v47p115.

[8] F. Tao, J. Panneerselvam, T. Holding, L. Liu, "A cloud-based sustainable business model for effective ICT provision in higher education," in Proceedings - 9th IEEE International Symposium on Service-Oriented System Engineering, IEEE SOSE 2015, 222–228, 2015, doi:10.1109/SOSE.2015.9.

[9] T. Gamage, D. Anderson, D. Bakken, K. Birman, A. Bose, C. Hauser, K. Maheshwari, R. Van Renesse, Mission-critical cloud computing for critical infrastructures, 1–16, 2017, doi:10.1201/b16908.

[10] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, C.H. Youn, "Wearable 2.0: Enabling Human-Cloud Integration in Next Generation Healthcare Systems," IEEE Communications Magazine, **55**(1), 54–61, 2017, doi:10.1109/MCOM.2017.1600410CM.

[11] K. Nowicka, "Competitiveness of Polish enterprises in relation to the potential of cloud computing," Kwartalnik Nauk o Przedsiębiorstwie, **42**(1), 85–99, 2017, doi:10.5604/01.3001.0010.0158.

[12] N.A. Nawaz, A. Waqas, Z.M. Yusof, A. Shah, "A framework for smart estimation of demand-supply for crowdsource management using WSN," in ACM International Conference Proceeding Series, 1–5, 2017, doi:10.1145/3018896.3025140.

[13] Y. Mao, C. You, J. Zhang, K. Huang, K.B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," IEEE Communications Surveys and Tutorials, **19**(4), 2322–2358, 2017, doi:10.1109/COMST.2017.2745201.

[14] S.H.H. Madni, M.S.A. Latiff, Y. Coulibaly, S.M. Abdulhamid, "Resource scheduling for infrastructure as a service (IaaS) in cloud computing: Challenges and opportunities," Journal of Network and Computer Applications, **68**(2016), 173–200, 2016, doi:10.1016/j.jnca.2016.04.016.

[15] S. Goyal, "Public vs Private vs Hybrid vs Community - Cloud Computing: A Critical Review," International Journal of Computer Network and Information Security, **6**(3), 20–29, 2014, doi:10.5815/ijcnis.2014.03.03.

[16] C. Stergiou, K.E. Psannis, B.G. Kim, B. Gupta, "Secure integration of IoT and Cloud Computing," Future Generation Computer Systems, **78**(2018), 964–975, 2018, doi:10.1016/j.future.2016.11.031.

[17] N. Pavón-Pulido, J.A. López-Riquelme, R. Torres, R. Morais, J.A. Pastor, "New trends in precision agriculture: a novel cloud-based system for enabling data storage and agricultural task planning and automation," Precision Agriculture, **18**(16), 1038–1068, 2017, doi:10.1007/s11119-017-9532-7.

[18] M.A. Zamora-Izquierdo, J. Santa, J.A. Martínez, V. Martínez, A.F. Skarmeta, "Smart farming IoT platform based on edge and cloud computing," Biosystems Engineering, **177**(2019), 4–17, 2019, doi:10.1016/j.biosystemseng.2018.10.014.

[19] E.A. Shammar, A.T. Zahary, "The Internet of Things (IoT): a survey of techniques, operating systems, and trends," Library Hi Tech, **38**(1), 5–66, 2019, doi:10.1108/LHT-12-2018-0200.

[20] J. Qiu, Z. Tian, C. Du, Q. Zuo, S. Su, B. Fang, "A survey on access control in the age of internet of things," IEEE Internet of Things Journal, **7**(6), 4682–4696, 2020, doi:10.1109/JIOT.2020.2969326.

[21] R. Zheng, T. Zhang, Z. Liu, H. Wang, "An EIoT system designed for ecological and environmental management of the Xianghe Segment of China's Grand Canal," International Journal of Sustainable Development and World Ecology, **23**(4), 372–380, 2016, doi:10.1080/13504509.2015.1124470.

[22] X. Li, D. Li, J. Wan, A. V. Vasilakos, C.F. Lai, S. Wang, "A review of industrial wireless networks in the context of Industry 4.0," Wireless Networks, **23**(1), 23–41, 2017, doi:10.1007/s11276-015-1133-7.

[23] M. Torres-Ruiz, J.H. Juárez-Hipólito, M.D. Lytras, M. Moreno-Ibarra, "Environmental noise sensing approach based on volunteered geographic information and spatio-temporal analysis with machine learning," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 95–110, 2016, doi:10.1007/978-3-319-42089-9_7.

[24] J.H.J. Hipólito, M.A.M. Ibarra, M. Torres-Ruiz, G. Guzmán, R. Quintero, Innovation on user-generated content for environmental noise monitoring and analysis in the context of smart cities, 490–519, 2018, doi:10.4018/978-1-5225-7033-2.ch021.

[25] A. Kumar, M. Zhao, K.J. Wong, Y.L. Guan, P.H.J. Chong, "A comprehensive study of IoT and WSN MAC protocols: Research issues, challenges and opportunities," IEEE Access, **6**(2018), 76228–76262, 2018, doi:10.1109/ACCESS.2018.2883391.

[26] K. Zkik, G. Orhanou, S. El Hajji, "A new secure framework in WSNs using ECC: Medical application," in Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017, 1–7, 2018, doi:10.1109/ICEngTechnol.2017.8308144.

[27] T. hoon Kim, C. Ramos, S. Mohammed, Smart City and IoT, Future Generation Computer Systems, 159–162, 2017, doi:10.1016/j.future.2017.03.034.

[28] F. Xiao, Q. Miao, X. Xie, L. Sun, R. Wang, "Indoor Anti-Collision Alarm System Based on Wearable Internet of Things for Smart Healthcare," IEEE Communications Magazine, **56**(4), 53–59, 2018, doi:10.1109/MCOM.2018.1700706.

[29] M. Al-Kuwari, A. Ramadan, Y. Ismael, L. Al-Sughair, A. Gastli, M. Benammar, "Smart-home automation using IoT-based sensing and monitoring platform," in Proceedings - 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering, CPE-POWERENG 2018, 1–6, 2018, doi:10.1109/CPE.2018.8372548.

[30] W. He, G. Yan, L. Da Xu, "Developing vehicular data cloud services in the IoT environment," IEEE Transactions on Industrial Informatics, **10**(2), 1587–1595, 2014, doi:10.1109/TII.2014.2299233.

[31] N.A. Nawaz, A. Waqas, Z.M. Yusof, A.W. Mahesar, A. Shah, "WSN based sensing model for smart crowd movement with identification: An extended study," Journal of Theoretical and Applied Information Technology, **95**(5), 975–988, 2017.

[32] X. Li, J. Niu, S. Kumari, F. Wu, A.K. Sangaiah, K.K.R. Choo, "A three-factor anonymous authentication scheme for wireless sensor networks in internet of things environments," Journal of Network and Computer Applications, **103**(2018), 194–204, 2018, doi:10.1016/j.jnca.2017.07.001.

[33] M.J. McGrath, C.N. Scanaill, M.J. McGrath, C.N. Scanaill, Sensor Network Topologies and Design Considerations, 79–95, 2013, doi:10.1007/978-1-4302-6014-1_4.

[34] D.L. Guidoni, F.S.H. Souza, J. Ueyama, L.A. Villas, "RouT: A routing protocol based on topologies for heterogeneous wireless sensor networks," IEEE Latin America Transactions, **12**(4), 812–817, 2014, doi:10.1109/TLA.2014.6868887.

[35] H.R. Faragardi, H. Fotouhi, T. Nolte, R. Rahmani, "A Cost Efficient Design of a Multi-sink Multi-controller WSN in a Smart Factory," in Proceedings - 2017 IEEE 19th Intl Conference on High Performance Computing and Communications, HPCC 2017, 594–602, 2018, doi:10.1109/HPCC-SmartCity-DSS.2017.77.

[36] E. Fadel, V.C. Gungor, L. Nassef, N. Akkari, M.G. Abbas Malik, S. Almasri, I.F. Akyildiz, "A survey on wireless sensor networks for smart grid," Computer Communications, **71**(2015), 22–33, 2015, doi:10.1016/j.comcom.2015.09.006.

[37] H. Toral-Cruz, F. Hidoussi, D.E. Boubiche, R. Barbosa, M. Voznak, K.I. Lakhtaria, A survey on wireless sensor networks, 171–210, 2015, doi:10.4018/978-1-4666-8687-8.ch006.

[38] R. Zhu, Y. Qin, J. Wang, "Energy-aware distributed intelligent data gathering algorithm in wireless sensor networks," International Journal of Distributed Sensor Networks, **7**(1), 235724, 2011, doi:10.1155/2011/235724.

[39] A. Waheed Khan, A.H. Abdullah, M.H. Anisi, J. Iqbal Bangash, "A comprehensive study of data collection schemes using mobile sinks in wireless sensor networks," Sensors (Switzerland), **4**(2), 2510–2548, 2014, doi:10.3390/s140202510.

[40] I.A.T. Hashem, I. Yaqoob, N.B. Anuar, S. Mokhtar, A. Gani, S. Ullah Khan, "The rise of 'big data' on cloud computing: Review and open research issues," Information Systems, **47**(2015), 98–115, 2015, doi:10.1016/j.is.2014.07.006.

[41] E.N. Sharad, I.K. Aulakh, "Evaluation and implementation of cluster head selection in WSN using Contiki/Cooja simulator," Journal of Statistics and Management Systems, **23**(2), 407–418, 2020, doi:10.1080/09720510.2020.1736324.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 439-446 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Effective Learning of Tax Regulations using Different Chatbot Techniques

Rafael Mellado-Silva*, Antonio Faúndez-Ugalde, María Blanco-Lobos

School of Commerce, Faculty of Economic and Administrative Sciences, Pontificia Universidad Católica de Valparaíso, Valparaíso, 2540000, Chile

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 01 November, 2020
Online: 20 November, 2020

Keywords:
Chatbot
Tax Learning
Automation
Decision tree

ABSTRACT

Teaching tax-related regulations have always been a challenge due to the inclusion of external variables that hinder the learning process, such as the high complexity of tax systems and legislation variability. Universities have sought different alternatives to support the teaching process both outside and inside the classroom, ranging from recreational activities to active learning. This article will show the experience resulting from using a chatbot to support learning in accounting students for the teaching of tax regulations related to the Chilean tax system, comparing two types of tools, on the one hand, a free conversation chatbot using natural language processing versus a rule-based chatbot driven by a decision tree. The experimentation process was carried out with 50 higher education students, divided into an experimental group and a control group, in two different courses. The results obtained demonstrated in both cases greater effectiveness of the use of the chatbot in learning the tax matter, both in the free conversation chatbot where the experimental group obtained a 15.7% improvement versus the control group that obtained a 1.05% improvement, as in the chatbot that applied decision tree where the experimental group obtained a 32% improvement versus the control group with 5.2%. Considering the complexity of the content in tax matters and the little innovation in the existing teaching subjects in the area and that the students improve learning using both chatbot tools compared to other learning techniques, is considered a relevant contribution to teaching innovation.

1. Introduction

This article is the continuation of a lengthy investigation developed in courses on the subject of taxation [1], which seeks to generate support for learning the Chilean tax system through the use of a chatbot with an automated conversational system technique; these technological systems aim to establish new mechanisms that have a direct impact on student learning, especially on issues such as the tax regime of a country, an area characterized by the complexity in the application of tax regulations.

Significant studies have taken place about the behavior and learning of students [2], [3], where motivation has been one of the key factors due to its direct impact on the perception and predisposition that students have towards the teaching activities delivered by teachers [4]. In this sense, the lack of motivation is one of the most critical causes when studying the failure and dropout of educational systems [5]. Socioeconomic and health factors also influence learning processes [6], [7], as well as the design problems of the different learning resources created by teachers [8], [9]. However, what research mostly agrees with is that this is a cross-cutting problem in society [3], [10], [11].

One technique that has become a common factor is the use of technology and software to support the teaching process and enhance the students learning [12], [13]. The integration of technologies in the classroom is a constant challenge that teachers and educational communities must face, but it is necessary based on the global demands on students' development of skills [14]. Based on this, there are numerous innovations, but the most outstanding and most attractive to students are those related to digital credentials, virtual assistants, and blockchain [15], [16].

This work puts into application in educational innovation, the use of a chatbot for learning theory, and application of Chilean tax regulations. The learning processes of tax regulations are always a

*Corresponding Author: Rafael Mellado-Silva, Pontificia Universidad Católica de Valparaíso, +56 32 2273340, Email: rafael.mellado@pucv.cl

great challenge for education and tax entities [17], [18]. In the case of Chile, the last ten years have seen critical structural changes in the tax system related to income tax, the reforms of which have been analyzed and studied [19], both due to the complexity that the tax system has acquired well as the particular characteristics defined for corporate taxes, which have direct effects on the economies of the countries [20]. Because of the above, different investigations have revealed the importance of teaching correctly and innovatively the subject of taxes [21]. With this, it is that innovative teaching mechanisms must be sought to produce effective learning in students.

The research objective is to compare the effectiveness in the learning of tax regulations in students of accounting careers through the use of two chatbot techniques, that is, one, through an automated conversational system and, two, based on a tree decision. This article presents the result of applying chatbots linked to the study of the tax system regulations, which cannot be exempt from applying, integrating, and using information technologies in their learning process [22].

Considering the complexity of the tax systems and the constant legislative variability produce an effect of obsolescence in the subjects, affecting the students' rejection towards these subjects and, consequently, very high rates of failure and dropout [23]. To address this problem, there were two chatbots with knowledge of tax matters. The first with a free communication with the students as applied in [1] and the second using a decision tree. The study was carried out in an N=50, dividing the students into an experimental and control group, in the accounting career in a Chilean higher education institution. The results obtained are promising, where, in the case of the free communication chatbot there is a difference of 14.65% in favor of the experimental group and in the case of the decision tree chatbot there was a difference of 26.8% in the results in favor of the experimental group.

For this article, we will consider a chatbot as a software tool that allows users to have a conversation between a human being and an artificial entity [24]. One of these technology precursors is Joseph Weizenbaum, with Eliza [25] applied to psychoanalysis with promising results [26]. As the winners of the Loebner [27] contest point out, chatbots have evolved from simple pattern matching systems to increasingly complicated patterns of computer interaction and reasoning, being applied in education [28], entertainment [29], health [30], among others. The rise of chatbots has occurred in customer service, such as Hennes & Mauritz AB, which seeks to replicate a human conversation to facilitate information collection by a customer [31]. The purpose of chatbots is to automate the process of communication and support for people [32].

At an educational level, there have been various applications to chatbots with mixed results, but most of them agree that an improvement in learning and student satisfaction is evident [33], [34], generating greater interest in the subjects test [35]. For this reason, three main factors stand out to determine their participation, the performance expectation, the effort expectation, and the habit [36], which indicates that the commitment is linked to the benefits versus the effort to which the students are subjected by supplementing your learning with a chatbot system. This is why it is critical to innovate in non-traditional areas and that the effect has a direct application in business, accounting, and auditing matters.

The research questions posed are in line with determining the effectiveness of the two tools applied and their impact on effective learning:

- Does the use of the two chatbot techniques positively contribute to learning tax regulations?
- Is learning through a chatbot more effective than through the use of gamified activities and e-learning videos?

Based on the above, we established as a research hypothesis:

Incorporating a chatbot, independent of the applied technique, for the learning process of tax regulations in higher education provides better results due to the students' academic performance.

Due to the above, a null hypothesis is also established:

The inclusion of technological tools, particularly chatbots, does not improve effective learning on tax matters compared to traditional methodologies or e-learning.

Concerning the structure of this paper, the theoretical framework that will deliver the definitions and general aspects that must be taken into account is presented below; later, other experiences applied in the matter will be seen, and then proceed with the experimental design and case study for the two techniques to be compared. This article ends with the results obtained and their corresponding discussion.

2. Related Works

Australian work by [46] shows that a significant number of accounting schools resist the adoption of new technologies in their training processes. The reasons are the lack of interest in implementing technologies with an extra workload, the little support, and the lack of resources, but the most relevant is the lack of educators' time to learn and develop skills and competencies to adopt new technologies in education systems efficiently. The work recommends considering a new model that reflects the current innovative technological nature to provide education, anywhere, anytime, and for anyone.

As far as tax learning is concerned, the work of [47] mentions how tax education faces significant challenges in the 21st century and delivers results to change current educational paradigms, prompting a more effective teaching practice based on conceptual knowledge over technical ability. In [48] it is explained how the subject of taxes is elementary to complement other subjects within accounting programs, and it seeks to find effective teaching methods using computer software, which, although it presents good results, its use is not every day. In [49] the work to support the understanding of tax systems and the effect on the systemic change that arises when different organizations are willing to learn from each other is presented. For this, real case studies are used in the companies Oxfam and Unilever. In Spain, there is the work of [50], which analyzes part of the challenges that exist for teaching tax law, considering students who are studying in a highly multidisciplinary degree; and in [51] whose economic concept is the distribution of the tax burden between buyers and sellers, it is

suggested that collaboration between teachers is more effective than the lesson study model. The work of [52] demonstrates how tax learning is necessary even in non-accounting programs, relating the interest of professionals in learning tax matters and the needs of companies.

As mentioned in [53], chatbots as learning tools are still in their infancy. The study carried out in [54] suggests that although initially arousing considerable interest in behavior, some tasks may not be interesting enough to impact subsequent interest in the broader domain of the study; for this purpose, in [55] a four-phase model of lasting interest development. In [56], it is stated that digital tools can provoke feelings of isolation and detachment caused by the lack of interaction with teachers, which is approached through the design of a hybrid-chatbot. Finally, it must be guaranteed that a chatbot's novelty does not define the end-user experience, although it is undoubtedly part of the initial interactions with these types of tools [54].

3. Solution Design

The solution was thought of in accounting students from a Chilean university, where the learning approach is related to tax regulations. In this context, it is essential to point out the Chilean tax system's difficulty due to its complexity, variability, and a very extensive law [57].

The proposed solution's objective was the implementation of two chatbots with pedagogical resources, which, using an automated conversational system and decision trees, could be applied to the context of tax regulations at the university level.

To respond to the proposed objective, two chatbots were designed, one through an automated conversational system and the other based on a decision tree, both focused on knowledge of tax regulations under the Chilean legal context. Also, the different tests and activities were designed to demonstrate the interaction and verify the effectiveness of the tools in the learning process.

The chatbot architecture is fundamental to explain its operation, and, as is well known, there are several architectures [37], [38]. In general, there is a standard operation that can be divided into three stages, as seen in Figure 1, where the user interacts with the voice manager through a device, where the analysis and generation of the query are carried out under the work context, considering that the bot must generate a response without considering the sense of intention. Finally, the generation of messages includes the planning of words, sentences, and coherence to deliver. With these aspects defined, we can establish that the user writes a question in his application, which must be answered by the chatbot and with greater complexity and volume of knowledge, then a more significant number of interactions between a bot and user [40], [41].

Decision trees are highly complex data structures sharing binary trees [42], [43]. This type of tree is characterized by storing its different nodes in a hierarchical and orderly manner under a defined criterion based on the problem addressed. Compared to linked lists, queues, or stacks, in this case, the data is not stored linearly [44].

In this case, the constraints to be applied to the decision tree have characteristics of an N-ary tree with single essential control used in binary search trees. The walkthroughs and handling of the decision tree structure are handed over to the natural language processing engine. The algorithm used by the classification and interpretation engine conforms to a conditional inference tree; consider that the Recast.ai engine handles the classification of intentions automatically, so no changes were made to it in this research.

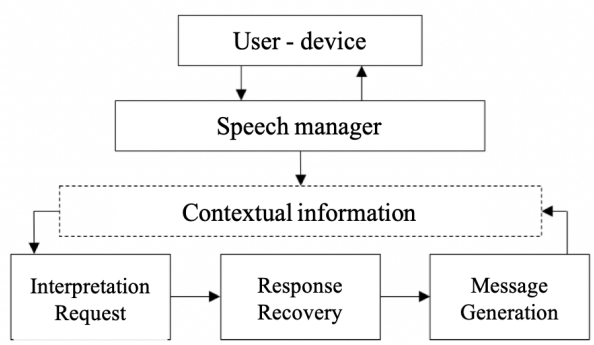


Figure 1: General chatbot architecture [39]

In this case, we start with a single node and then branch out into possible results based on the intent analysis of the natural language processing engine [45]. The decision node will represent the intention in which the user's interaction will be classified, giving rise to the existence of a terminal node that will show the result of the intention. The first node will correspond to the concept of tax and the last to specific answers.

4. Method

4.1. Pedagogical context

The learning context of this research's tax regulations focuses on the career of auditor accountant (public accountant) of a Chilean university. Thus, teaching and learning the rules that regulate tax regulations is a constant challenge, especially to achieve effective learning. The different indicators of the courses approval do not generate significant delays in the students' curricular progress. The professional profile for graduating the auditor accountant career (public accountant) considers the development of specific competencies that are directly related to the work performed:

- Generic fundamental training competencies: proactivity and responsibility; carries out its work with professional ethics.
- Specific disciplinary competencies: understand the conceptual framework of the current internal fiscal tax legislation.
- Specific professional competencies: apply the conceptual framework of the current internal fiscal tax legislation; apply the general, and special regulations of the tax legislation of the taxes levied on business income.

4.2. Learning Objectives

The Taxation courses of the auditor accountant career (public accountant) have a content structure that contributes to the learning results; in this way, the contents associated with this experiment are: (1) Understanding of the examination regulations, their practical application in the taxpayer's life cycle; (2) Understanding of the corporate tax structure.

For the technique through an automated conversational system, the learning outcomes associated with the experiment and the course are [1]:

- OL1: Recognizes basic concepts related to corporation tax to apply them to the company.
- OL2: Identify basic concepts of the different corporate taxes to compose tax returns.
- OL3: Applies basic concepts of the tax object to calculate taxes.

For the decision tree-based technique, the learning outcomes associated with the experiment and the course are:

- OL4: Defines the types of tax examination to identify the procedure applied in each of them.
- OL5: It includes the indirect control rules for their correct application in the taxpayers' life cycle.
- OL6: It includes direct inspection rules to identify the rights that assist taxpayers.

4.3. Methodology

For methodological purposes, we worked with the Taxation 1 and Taxation 3 courses of the auditor-accountant career, where the main subject of these courses is the regulations on audit and corporate taxes, respectively. The experimentation process for both chatbots is considered the same methodological steps are shown in the BPMN diagram in Figure 2.

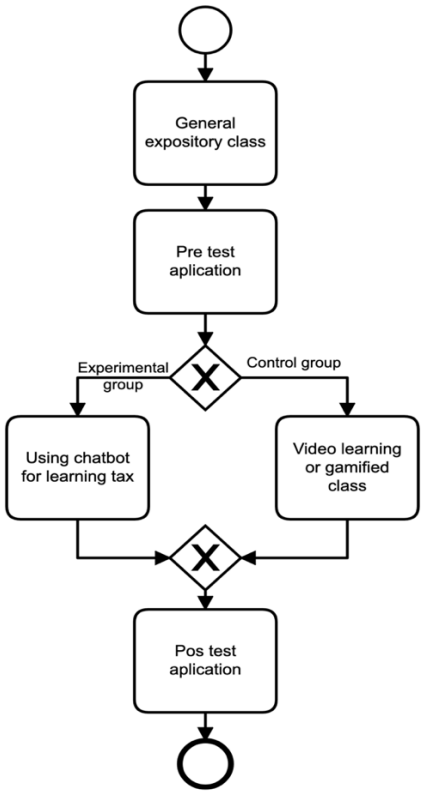


Figure 2: Stages of the methodology

First, a class of theoretical contents is held in both courses with all students without differentiating or separating individuals. This class is face-to-face and in a classroom with a projection. After that, a pre-test is applied to all students based on the contents taught in the face-to-face class of the Tax 1 and Tax 3 courses. Both courses are then divided into two groups, which we will call the experimental group and the control group. In the case of the Taxation 1 course, the use of videos through the *Moodle* platform was applied to the control group, and in the case of the experimental group, the use of the chatbot based on the decision tree technique was applied. In the case of the Taxation 3 course, the control group was given playful activities in the classroom led by a teacher, and in the case of the experimental group, the use of the chatbot with the technique was applied through an automated conversational system. A post-test is applied to both courses to measure the difference generated in the learning process. In this way, the results obtained in the tests by each of the groups of both courses will be compared, and the effectiveness of using the chatbot with the two techniques will be shown. Finally, it should be noted that, in the Taxation course, 1 of the of 16 students, eight remained in the experimental group and 8 in the control group. In the case of the Taxation course, 3 of 34 students, 18 remained in the experimental group, and 16 in the control group.

As mentioned above, two tests were applied to each course, one that was developed before dividing the group and another afterward to measure effective learning. An example of interaction in the evaluations is presented in Figure 3, using Google Forms.

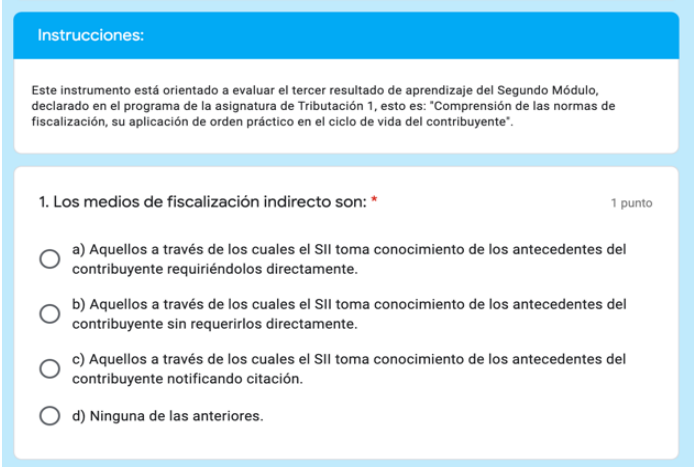


Figure 3: Example evaluation rendered by the student

4.4. Mechanics

In the Tax 1 course, the students belonging to the control group saw explanatory videos with the contents. For the experimental group, a guide to questions to be solved was delivered with the bot's support. The chatbot used was called Tribuchat, and it was linked to a Telegram chat, with which the students who made use of their mobile devices interacted under the guidance of the chatbot. Figure 4 shows an example of interaction with the bot.

In the case of the Tax 3 course, the students belonging to the control group did playful activities that included the themes addressed in the planning. The teacher-guided these activities. For the experimental group, a guide to questions to be solved was delivered with the bot's support. The chatbot was called Tribubot, and it was linked to a Facebook Messenger chat, with which the students, using their mobile devices, freely interacted with the chatbot. Figure 5 shows an example of interaction with the bot.

It is important to note that feeding both bots was carried out, in which the teachers in charge of the subjects generated the expert knowledge base on the subject and thinking about the interaction with the students, which was loaded into the tool SAP Recast.ai. Finally, interaction actions were included to measure a formative nature's learning through questions asked by the same bot to the students.

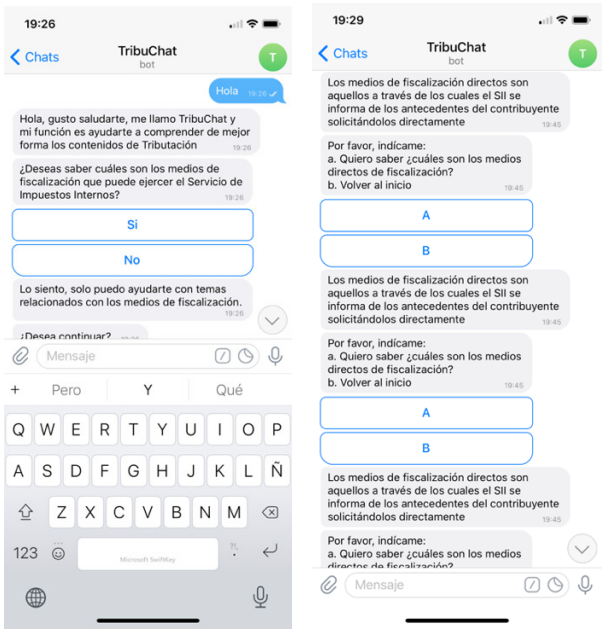


Figure 4: Chatbot group 1 interaction example

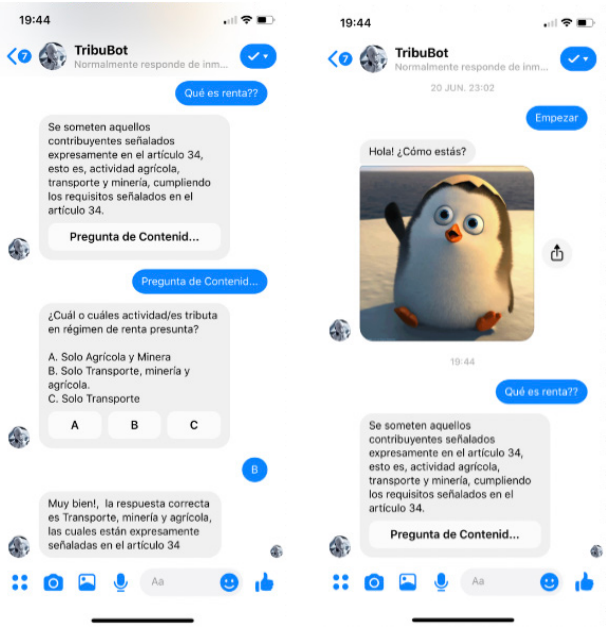


Figure 5: Chatbot group 2 interaction example

5. Results

The display of results obtained will be classified in the general results, later categorizing by gender, and finally according to performance. Table 1 shows the results obtained in the tests applied to the Taxation 1 and 3 courses, and the averages of correct, incorrect, omitted answers, and the percentage of correct answers that were.

Table 1: General Results

Course	Group	Clasification	Pretest	Postest
Tax 1 course	Control	Correct	7	7,125
		Wrong	5	4,875
		% correct	58,33%	59,38%
	Experimental	Correct	6,125	8
		Wrong	5,875	4
		% correct	51,0%	66,7%
Tax 3 course	Control	Correct	6,88	7,5
		Wrong	4,01	4
		% correct	57,3%	62,5%
	Experimental	Correct	6,56	10,39
		Wrong	4	1,87
		% correct	54,6%	86,6%

In Table 2, we can see the characterization and results in the test of the groups of Taxation 1 and 3 courses based on composition according to gender.

Table 2: Characterization by gender

Course	Tax 1 course		Tax 1 course	
Group	Control	Exp.	Control	Exp.
N	8	8	16	18
Male	25%	37,5%	43,75%	50%
% Improvement	0%	27,8%	5,7%	60,32%
Female	75%	62,5%	56,25%	50%
% Improvement	2,3%	32,3%	13,4%	58,21%

Table 3 shows groups' behavior according to the average of the Tax 1 and 3 course; for this, those below the average, equal and above, are grouped in their categories of the experimental and control groups.

Table 3: Behavior of groups according to the average

	Test	Group	Ref. $\bar{x}$	%<$\bar{x}$	%=$\bar{x}$	%>$\bar{x}$
Tax 1	Pretest	Control	7	12,5%	75%	12,5%
		Exp.	6,125	25%	62,5%	12,5%
	Postest	Control	7	25%	37,5%	37,5%
		Exp.	6,125	0%	25%	75%
Tax 3	Pretest	Control	6,88	6,25%	68,75%	125%
		Exp.	6,56	5,56%	77,78%	16,66%
	Postest	Control	6,88	0%	81,25%	18,75%
		Exp.	6,56	0%	0%	100%

In addition, to statistically validate the results, some tests were applied. Because the new experiment points directly to tax course 1, then as the results were analyzed, we begin with Table 4, where the samples' normality is appreciated through a Shapiro-Wilks Test. As can be seen, for the sample of the experimental group in the post-test, normality was rejected with a high level of significance, so that we will use non-parametric tests for the comparison of means.

Table 4: Results of Shapiro-Wilks test.

Group	Test	P-value
Control	Pretest	0,6877
	Postest	0,2993
Experimental	Pretest	0,1842
	Postest	**0,02066**

Table 5 shows the result of applying the Wilcoxon test for paired data. In this case, the difference in means for the control group between each test is not statistically significant; on the contrary, in the experimental group, it is observed that the mean of the post-test is higher with a high level of statistical significance (0.0001526 < <0.05).

Table 5: Wilcoxon test for paired data

Samples/Group		P-value
Control Pretest	Control Postest	0,08052
Experimental Pretest	Experimental Postest	**0,0001526**

In this first investigation, the data concludes that the research hypothesis is accepted, giving a difference between groups of p=9,722e-06. Specifically, the experimental group results were statistically much higher than those shown by the control group; therefore, the chatbot was more effective than the other method used.

Regarding the comparison of the two tools, only the sample data of each experimental group were taken. The research hypothesis is addressed: "The students' performance is different given the applied chatbot". To do this, first, a difference in proportion test was applied, where this test analyzes whether the percentage or proportion of a sample is statistically significance concerning another. Table 6 represents the data of the experimental group students who obtained a performance higher than the global average (control + experimental) in each post-test and the total sample of students of the experimental group.

Table 6: Proportion difference test by course

Students	Tax 1 course	Tax 3 course
Above average	7	15
Totals	8	18

Furthermore, with a p-value = 1, it can be concluded that this proportion is the same in both groups, and we cannot find a significance difference.

Finally, the result changes since the hypothesis focus only on a one-sided inequality and the test; therefore, a significance mean difference can be accepted and affirmed that the first chatbot students had a better performance (p = 0.04418). Also, it is noticeable that course 1 was much higher than those of course 2, but there was also a larger sample in that case (a little more than twice as many students), so that although the inequality between chatbots could be maintained. With the above, we have ruled out the null hypothesis.

6. Discussions

Due to what is stated in Table 1, it is evident that the control group of the Tax 1 course presents an advantage of 12.56% over the experimental group in the pretest, which was reversed in the posttest, leaving a 10.97% in favor of the experimental group; which leads us to deduce in the first instance that if the performance of the students in the experimental group is compared with the control group, the chatbot tool produced an evident improvement in learning to the detriment of the control group that took explanatory videos. This justifies the first research question about contributing positively to learning the contents related to the Chilean tax system, which is characterized mainly by the complexity of its normative structure.

The above also allows supporting the second research question on the effectiveness in the use of the chatbot. Indeed, the differences between incorrect answers and errors made are evident, where the control group presents discrete performance compared to the experimental group.

In the case of Table 1 related to the Tax 3 course, it is evidenced that the control group also presented an advantage of 4.94% over the experimental group in the pretest, reversing the situation in the posttest by 27, 82% in favor of the experimental group. The evidence is also significance when comparing the wrong questions that did not decrease.

If the behavior by gender in Table 2 is studied, it is essential to note that in the Tax 1 course, the men belonging to the control group did not show improvement (0%) versus the men in the experimental group (27.8%). In women's case, the improvement in those belonging to the control group is only 2.3% versus 32.3% in the experimental group. In the case of the Tax 3 course, the result in Table 2 is not so different compared to the Tax 1 course. Indeed, it can be found that men who make use of the chatbot show more than a 60% improvement in its results; on the other hand, those who do not use the chatbot improve by 5.7%; in the case of women, the difference is also greater, because those women who use bot the percentage of improvement is greater than 58% but those who do not use it improve 13.4%.

If the behavior is analyzed according to the means in the Tax 1 course, as shown in Table 3, the control group goes from having 75% in the average to 37.5%, leaving 37.5% above the average versus the initial 12.5%. In the case of the experimental group of 62.5% who were in the mean, only 25% remained, and those who were above the mean grew from 12.5% to 75% of the group, which together with finally remaining 0 % below the mean, then the effect of the tool applied in the experimental group is greater than that of the control group since many students being below the mean or within it rose to be at or above it. The above is a critical case to highlight and take into account for later statistical analysis. Finally, as shown in Table 3 about the Taxation 3 course, initially, 5.56% of the students who use chatbot were below the average; after using the chatbot, that number decreases to 0%, which shows a positive impact of the tool. It is not minor to identify that the control group also improves the results after applying the gamified activities.

7. Conclusions

It is already known that the teaching of tax matters is a constant challenge that expert teachers in the matter must face daily, especially considering the complexity of the contents and the repeated modifications of the legislation in the last ten years.

Throughout this paper, we have been able to see different experiences that try to improve effective learning, comparing two chatbot techniques, that is, through an automated conversational

system and another based on a decision tree, showing that independent of the technique of the chatbot used has better results than guided learning through videos or gamification techniques. The results showed that, in all cases, students who use a chatbot, regardless of the technique used, perform better than other tools, justifying their effectiveness according to what was raised in the research questions. The results show that the performance of the students of the experimental group produced an evident improvement in the learning of the contents related to the Chilean tax system, which is characterized by its complexity, derived mainly from the withdrawn legislative reforms in the last ten years.

As future work, the perspective should be taken to analyze and quantify the impact of the use of this type of tool and the cognitive load that it generates in students. Besides, it would be essential to study the change suffered by the levels of frustration of students compared to the different techniques that exist in novel and impact studies, directly related to emotionality and the application of algorithms. In this sense, an artificial emotion can be implemented as a synthetic abstraction derived from the observation of human emotions, which allows improving the effectiveness of the decision made [58]. It is also relevant to test other techniques and tools such as ROL-type games [59], gamification techniques [60], or the application of multilingual verbal communication with a focus on human-robot interaction.

Finally, it should be noted that through this research, it is shown that a chatbot improves the results in learning subjects related to tax regulations and, therefore, it is necessary to continue with studies that cover other aspects of effective teaching and learning.

Acknowledgment

Tomás Vidal Maulén, Xavier Diaz and Roberto Rojas: PUCV Research Assistants. Rafael Mellado-Silva is a beneficiary of the INF-PUCV doctoral scholarship.

References

[1] R. Mellado-Silva, M. Blanco, A. Faundez, H. De La Fuente, "Support to the learning of the Chilean tax system using artificial intelligence through a chatbot," in 2019 38th International Conference of the Chilean Computer Science Society (SCCC), 1–8, 2019, doi:10.1109/SCCC49216.2019.8966410.

[2] R. Goodman, "Problem-based learning: merging of economics and mathematics", Journal of Economics and Finance, **34**(4), 477-483, 2010, https://doi.org/10.1007/s12197-010-9154-7

[3] M. Prince, R. Felder, "Inductive teaching and learning methods: Definitions, comparisons, and research bases", Journal of Engineering Educaction, **95**(2), 123-138, 2006, https://doi.org/10.1002/j.2168-9830.2006.tb00884.x

[4] D. Schunk, B Zimmerman, "Motivation and Self-Regulated Learning: Theory, Research and Applications", Routledge, **1**, 38-56, 2012, https://books.google.cl/books?id=8VCLdg2DmoQC&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

[5] C. Cofer, M. Appley, "Motivation: Theory and research", John Wiley, 1, 958, 1964, https://books.google.cl/books/about/Motivation.html?id=O7x_KgAACAAJ&redir_esc=y

[6] Guerrero, E. Hishinuma, N. Andrade, S. Nishimura, V. Cunanan, "Correlations among Socioeconomics and Family Factors and Academic, Behavioral, and Emotional Difficulties inn Filipino Adolescents in Hawai", International Journal of Social Psychiatry, **52**(4), 343-359,2006, doi:10.1177/0020764006065146

[7] P. Lantz, J. House, J. Lepkowski, D. Williams, R. Mero, J. Chen, "Socioeconomics Factors Health Behaviors, and Mortality: Results From a Nationally Representative Prospective Study of US Adults", JAMA, **279**(21), 1703-1708, 1998, doi:10.1001/jama.279.21.1703

[8] V. Aleyen, E. Stahl, S. Schworm, F. Fischer, R. Wallace, "Help Seeking and Help Design in Interactive Learning Environments", Review of Educational Research, **73**(3), 277-320, 2003, https://doi.org/10.3102/00346543073003277

[9] D. Gjibels, F. Dochy, P. Van den Bossche, M. Segers, "Effects of Problem-Based Learning: A Meta-Analysis From the Angle of Assessment", Review of Educational Research, **75**(1), 27-61, 2005, https://doi.org/10.3102/00346543075001027

[10] M. Chi, M. Bassok, M. Lewis, P. Reimann, R. Glaser, "Self-Explanations – How Students Study and Use Examples in Learning to Solve Problems", Cognitive Science, **13**(2), 145-182, 1989, https://doi.org/10.1016/0364-0213(89)90002-5

[11] C. Hmelo-Silver, "Problem-Based Learning: What and How Do Students Learn?", Educational Psychology Review, **16**(3), 235-266, 2004, https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

[12] Y. Ye, K. Kishida, "Toward an understanding of the motivation of open source software developers", in 25th International Conference on Software Engineering, 419-429, 2003, doi:10.1109/ICSE.2003.1201220.

[13] W. Nuankaew, K. Phanniphong, S. Bussaman, D. Teeraputon, P. Nuankaew, "Mentoring Model in an Active Learning Culture for Undergraduate Projects, Astes Journal, **5**(4), 805-815, 2020, http://www.astesj.com/publications/ASTESJ_050495.pdf

[14] W. Lambrechts, I. Mulà, K. Ceulemans, I. Molderez, V. Gaeremynck, "The integration of competences for sustainable development in higher education: an analysis of bachelor programs in management", Journal of Cleaner Production, **48**, 65-73, 2013, doi:10.1016/j.jclepro.2011.12.034.

[15] R. Sharif, "The relations between acculturation and creativity and innovation in higher education: A systematic literature review", Educational Research Review, 28, 1-26, 2019, doi:10.1016/j.edurev.2019.100287.

[16] L. Carvalho, P. Goodyear, "Design, learning networks and service innovation", Design Studies, **55**, 27-53, 2018, doi:10.1016/j.destud.2017.09.003.

[17] E. Engelbrecht, "Adapting to changing expectations: Post-graduate students experience of an e-learning tax program", Computers & Education, **45**(2), 217.229. 2005, doi:10.1016/j.compedu.2004.08.001.

[18] Faúndez-Ugalde, R. Mellado-Silva, E. Aldunate-Lizana, "Use of artificial intelligence by tax administrations: An analysis regarding taxpayers rights in Latin American countries", Computer Law & Security Review, 38, 1-13, 2020. 2020, doi:10.1016/j.clsr.2020.105441.

[19] T. Fairfield, "The political economy of progressive tax reform in Chile", Revista de Economia Institucional, **17**(32), 129-156, 2015, https://ssrn.com/abstract=2625309

[20] T. Derya-Baskan, E. Balikcioglu, "The Effects of Firm Components of Open Joint Manufacturing Companies on Corporate Taxes", Sosyoekonomi, **26**(37), 219-226, 2018, doi: 10.17233/sosyoekonomi.2018.03.12

[21] M. Crandall-Hollick, "Higher Education Tax Benefits: Brief Overview and Budgetary Effects, Congressional Research Service **1**, 1-17, 2018, https://files.eric.ed.gov/fulltext/ED593609.pdf

[22] L. Lee, J. Hung, "Effects of blended e-Learning: a case study in higher education tax learning setting", Human-Centric Computing and Information Sciences, **5**(1), 13, 2015, doi:10.1186/s13673-015-0024-3.

[23] J. Walmsley, "Talking to Top people: Some Issues Relating to the Citizenship of People with Learning Difficulties", Disability, Handicap & Society, **6**(3), 219-231, 1991, doi:10.1080/02674649166780261.

[24] Sheehan, H. Jin, U. Gottlieb, "Customer service chatbots: Anthropomorphism and adoption", Journal of Business Research, **115**, 14-24, 2020, doi:10.1016/j.jbusres.2020.04.030.

[25] J. Weizenbraum, "ELIZA—a computer program for the study of natural language communication between man and machine", Communications of the ACM, 9(1), 36-45, 1966, https://dl.acm.org/doi/10.1145/365153.365168

[26] N. Block, "Psychologism and Beehaviorism", The Philosophical Review, **90**(1), 5-43, 1981 https://www.researchgate.net/publication/265578700_Psychologism_and_Behaviorism

[27] L. Bradesko, D. Mladenic, "A Survey of Chabot Systems through a Loebner Prize Competition", in Slovenian Language Technologies Society 8th Conference of Language Technologies, 2012, https://www.researchgate.net/publication/235664166_A_Survey_of_Chatbot_Systems_through_a_Loebner_Prize_Competition

[28] J. Jia, "CSIEC: A computer assisted English Learning chatbot based on textual knowledge and reasoning", Knowledge-Based Systems, **22**(4), 249-255, 2009, https://doi.org/10.1016/j.knosys.2008.09.001

[29] Lee, S. Jung, S. Kim, G. Lee, "Example-based dialog modeling for practical

multi-domain dialog system", Speech Communication, **51**(5), 466-484, 2009, https://doi.org/10.1016/j.specom.2009.01.008
[30] R. Crutzen, G. Peters, S. Portugal, E. Fisser, J. Grolleman, "An Artificially Intelligent Chat Agent That Answer Adolescents Questions Related to Sex, Drugs and Alcohol: An Exploratory Study", Journal of Adolescent Health, **48**(5), 514-519, 2011, https://doi.org/10.1016/j.jadohealth.2010.09.002
[31] P. Brandttzaeg, A. Følstad, "Why People Use Chatbots", Internationnal Conference on Internet Science, 377-392, 2017, doi:10.1007/978-3-319-70284-1_30.
[32] H. Shah, K. Warwick, J. Vallverdú, D. Wu, "Can machines talk? Comparison of Eliza with modern dialogue systems", Computers in Human Behaviors, **58**, 278-295, 2016, doi:10.1016/j.chb.2016.01.004.
[33] O. Simpson, "Supporting Students in Online, Open and Distance Learning", 1, 348-362, Routledge, 2018, https://doi.org/10.4324/9780203417003
[34] J. Lee, "Online support service quality, online learning acceptance, and student satisfaction", The Internet and Higher Education, **13**(4), 277-283, 2010, https://doi.org/10.1016/j.iheduc.2010.08.002
[35] L. Fryer, K. Nakao, A. Thompson, "Chatbot learning partners: Connecting learning experiences, interest and competence", Computers in Human Behavior, 93, 279-289, 2019, https://doi.org/10.1016/j.chb.2018.12.023
[36] F. Almabri, D. Bell, M. Merhi, "Understanding Student Acceptance and Use of Chatbots in the United Kingdom Universities: A Structural Equiation Modelling Approach", in 2020 6th International Conference on Information Management (ICIM), IEEE, London, United Kingdom, 284-288, 2020, doi:10.1109/ICIM49319.2020.244712.
[37] S. Ghose, J. Barua, "Toward the implementation of a topic specific dialogue based natural language chatbot as an undergraduate advisor", in 2013 International Conference eon Informatics, Electronics and Vision (ICIEV), 1-5, 2013, doi: 10.1109/ICIEV.2013.6572650
[38] M. Yan, P. Castro, P. Cheng, V. Ishakian, "Building a Chatbot with Serverless Computinng", in Proceedings of the 1st International Workshop on Mashups of Things and APIs, **5**, 1-5, 2016, doi: 10.1145/3007203.3007217
[39] Braun, A. Hernandez, F. Matthes, M. Langen, "Evaluating Natural Language Understanding Services for Conversational Question Answering Systems", in Proceedings of the 18th Annual SIGdial Meeting of Discourse and Dialogue, 174-185, 2017, doi: 10.18653/v1/W17-5522
[40] M. Dahiya, "A Tool of Conversation: Chatbot", International Journal of Computer Sciences and Engineering, **5**, 158-161, 2017, https://www.researchgate.net/publication/321864990_A_Tool_of_Conversation_Chatbot
[41] S. Abdul-Kader, J. Woods, "Survey on Chatbot Design Techniques in Speech Conversation Systems", International Journal of Advanced Computer Science and Applications, **6**(7), 2015, http://dx.doi.org/10.14569/IJACSA.2015.060712
[42] C. Wang, A. Wang, J. Xu, Q. Wang, F. Zhou, "Outsourced privacy-preserving decision tree classification service over encrypted data", Journal of Information Security and Applications, **53**, 788-813, https://doi.org/10.1016/j.jisa.2020.102517
[43] Wang, Q. Wang, F. Nie, Z. Li, W. Yu, F. Ren, "A linear multivariate binary decision tree classifier based on K-means splitting", Pattern Recognition, 107, 423-437, 2020, doi:10.1016/j.patcog.2020.107521.
[44] M. Van den Boomen, M. Spaan, R. Schoenmaker, A. Wolfert, "Untangling decision tree and real options analyses: a public infrastructure case study dealing with political decisions, structural integrity and price uncertainty", Construction Management and Economics, **37**(1), 24-43, 2019, doi:10.1080/01446193.2018.1486510.
[45] S. Garg, "An evaluation of investor acceptability for physical gold using classification (Decision Tree)", Materials Today: Proceedings, FALTA VOLUMEN, FALTAN PAGINAS, 2020, doi:10.1016/j.matpr.2020.06.177.
[46] K. Watty, J. McKay, L. Ngo, "Innovators or inhibitors? Accounting faculty resistance to new educational technologies in higher education", Journal of Accounting Education, **36**, 1-15, 2016, doi:10.1016/j.jaccedu.2016.03.003.
[47] N. Ahamad, M. Lai, "Undergraduate tax education in the new millennium: Technical vs. conceptual", in 2010 International Conference on Science and Social Research, 602-606, 2010, doi:10.1109/CSSR.2010.5773852.
[48] J. Craner, A. Lymer, "Tax education in the UK: a survey of tax courses in undergraduate accounting deegrees", Accounting Education, **8**(2), 127-156, 1999, doi:10.1080/096392899330973.
[49] P. Senge, M. Dow, G. Neath, "Learning together: new partnerships for new times", Corporate Governance: The international journal of business in society, **6**(4), 420-430, 2006, doi: 10.1108/14720700610689531
[50] P. Marín, "Enseñanza y aprendizaje del derecho tributario, y ejercicio profesional, en ADE", Revista de Educación y Derecho, **12**, 129-150, 2015,
[51] M. Pang, F. Marton, "Beyond 'lesson study': Comparing two ways of facilitating the grasp of some economic concepts", Instructional Science, **31**(3), 175-194, 2003, https://doi.org/10.1023/A:1023280619632
[52] M. Lai, Y. Zalilawati, M. Amran, K. Choong, "Quest for Tax Education in Non-Accounting Curriculum: A Malaysian Study", Asian Social Science, **9**(2), 154, 2013, doi:10.5539/ass.v9n2p154.
[53] D. Gonda, J. Luo, Y. Wong, C. Lei, "Evaluation of Deeveloping Educational Chatbots Based on the Seven Principles for Good Teaching", in 2018 IEEE Internationnal Conference on Teachingn, Assessment, and Learning for Engineering (TALE), 446-453, 2018, doi:10.1109/TALE.2018.8615175.
[54] L. Fryer, M. Ainley, A. Thompson, A. Gibson, Z. Sherlock, "Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners", Computers in Human Behavarior, **75**, 461-468, 2017, https://doi.org/10.1016/j.chb.2017.05.045
[55] S. Hidi, K. Renninger, "The Four-Phase Model of Interest Development", Educational Psychologist, **41**(2), 111-127, 2006, https://doi.org/10.1207/s15326985ep4102_4
[56] Wu, C. Lin, Y. Ou, C. Liu, W. Wang, C. Chao, "Advantages and Constraints of a Hybrid Model K-12 E-Learning Assistant Chatbot", IEEE Access, 8, 77788-77801, 2020, doi:10.1109/ACCESS.2020.2988252.
[57] G. Campos, "Complejidad del Sistema Tributario", PriceWaterhouseCoopers, 2017. https://www.pwc.com/cl/es/prensa/columnas-de-opinion/complejidad-del-sistema-tributario.html
[58] D. Cabrera, C. Cubillos, A. Cubillos, E. Urra, R. Mellado, "Affective Algorithm for Controlling Emotional Fluctuation of Artificial Investors inn Stock Markets", IEEE Access, **6**, 7610-7624, 2018, doi:10.1109/ACCESS.2018.2802781.
[59] R. Mellado, B. Melgarejo, C. Velasquez, C. Cubillos, S. Roncagliolo, N. Gonzalez, "ROLE video game tool for teaching myths and legends to school basic students", in 2018 37th International Conference of the Chilean Computer Science Society, 1-8, 2018, doi:10.1109/SCCC.2018.8705258.
[60] R. Mellado, M. Blanco, C. Cubillos, H. Fuente, A. Faúndez, "Learning of Information Technologies in Administration Science through gamification techniques" in 2019 IEEE Chilean Conference on Electrical, Electronics, Engineering, Information and Communication Technologies, 1-7, 2019, doi:10.1109/CHILECON47746.2019.8988017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 447-454 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Electronic Warfare Methods Combatting UAVs

Miroslav Kratky[*,1], Vaclav Minarik[1], Michal Sustr[2], Jan Ivan[2]

[1]*University of Defence, Department of Air Defence, 662 10, Czech Republic*

[2]*University of Defence, Department of Fire Support, 662 10, Czech Republic*

ARTICLE INFO

Article history:
Received: 27 August, 2020
Accepted: 15 October, 2020
Online: 20 November, 2020

Keywords:
Unmanned Aerial System – UAS
Unmanned Aerial Vehicle – UAV
Air Defence
Electronic War-fare
Cybernetics
Neural Network

ABSTRACT

This paper describes methods of eliminating Unmanned Aerial Vehicles (UAV) non-destructively, using Electronic Warfare Methods. The aim is to introduce certain methods of UAV detection and elimination in a complex environment and terrain, e.g., in an urban and battlefield environment, that will result in finding the control device position and the UAV itself. Neural networks, cyber penetration elements, and wireless network scanning programs are all used to address this issue. The output of this article is a new concept of a comprehensive solution, which can be implemented into the existing complex system of electronic defence against UAVs, e.g., within the allied base. Conclusions will be also used to further improve the above-mentioned topics at the authors' workplace, within the frame of long-term projects and specifically as a part of solutions applicable to the force protection of combat support units, namely field artillery, which is described here in detail.

1. Introduction

This paper is an extension of the work originally presented in proceedings of the "2019 International Conference on Military Technologies" (ICMT), Brno, Czech Republic [1]. The original material was enriched by the proposed concept applied in the urban area. The whole section proposing a possible concept applicable to artillery units was added.

A very dynamic development of Unmanned Aerial Systems (UASs) is becoming highly visible as they are used in all areas of human activities [2].

With the gradual widening of the UASs complexity and complementing other features, there is a demand for more complex security. The reason is that both military forces and terrorists have increasingly used Unmanned Aircraft Systems to plan, prepare, and execute attacks on ally and partner's forces and on "soft targets" in the civilian sector. Preventing, protecting, and recovering from such attacks require a cross-governmental approach, bridging the different efforts that Allies, partners, NATO, other international organizations, industry, and academia are making on this topic, both in the military and in the civilian domain. The threat description, probable scenarios, and protection models concepts are well described now by several trustworthy documents, e.g., in [3].

In another word, simultaneously with UASs development, we have been witnessing the counter-UAS (C-UAS) systems rapid improvement as well. Moreover, the C-UAS systems development concerns not only the direct defence against flying apparatuses themselves, but this also applies to the whole loop of the air defence system: detection, command-control system and elimination itself. The whole engagement process is now a highly complex, sophisticated, and, from a scientific point of view, multidisciplinary one.

Thus, after transforming from simple radio-controlled machines into sophisticated, "smart", digitally-controlled UAVs, an opportunity was developed to combat the whole UAS not only with standard methods, but – apart from another ones - also with cybernetic methods. Contemporary modern UAV can be seen as a small computer or a mobile phone with the ability to fly. This connects the current IT security issues and air defence issues together with the UAVs.

During the next development, it will be necessary to take into account the possibility of anti-aircraft defence congestion caused by the high availability of micro UAVs at a very low affordable price. Therefore, the developed means should be able to operate a large number of UAVs at the same time and at a minimum cost.

[*]Corresponding Author: Miroslav Kratky, Email: miroslav.kratky@unob.cz

2. Methods of Cyberattack Applied to UAVs

In today's digital age, the growth of cyberattacks can be seen not only on personal computers, servers, and mobile phones, but also with the introduction of "internet of things", on any device that is able to connect and communicate over the network. This opens the possibility of cyberattacks on most types of UAVs too. Cybernetic methods of attack, like conventional methods, can be divided into two basic groups according to the expected effect. These methods are non-destructive and destructive.

2.1. Non-destructive Methods of Cyberattack

Non-destructive cybernetic methods of combat are understood as methods in which there is no direct destruction of any particular component of the affected system. This group includes the most contemporary cyberattacks. We can further divide these attacks into several subareas (in detail: [4]). Graphical representation of non-destructive cyberattack methods on UASs is depicted in Fig. 1.

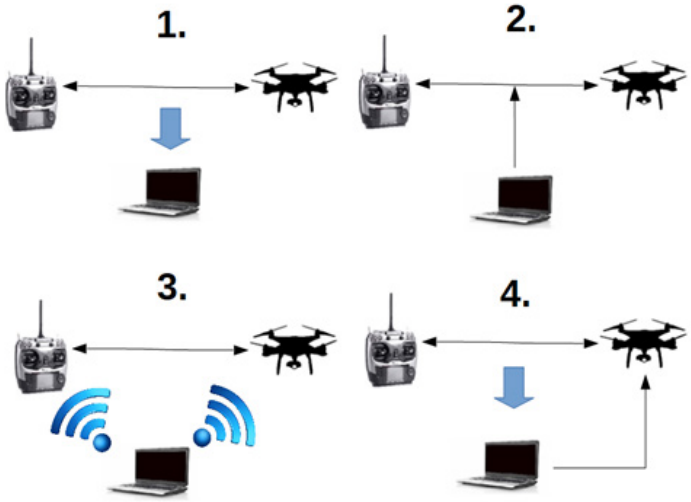


Figure 1: Graphical representation of non-destructive cyberattack methods applied to UAV´s

1) Leakage of Information

This type of attack results in disclosure or leakage of protected information. The advantage is its detection difficulty and in most cases the speed of attack. In UAVs, this attack is about getting the downlink channel information providing the UAV's mission or about getting the data to find the password for Wi-Fi communication.

2) Disturbance of Integrity

This subgroup includes attacks where the UAV's data are destroyed, damaged, or changed. This type of attack is very well detectable, but in most cases until after its accomplishment.

3) Denial of Service

Denial of Service (DOS) attacks make it impossible to use a particular service or system of UAS. The C-UAS defender will focus on a particular access channel or a specific service and will disable real-time activity by systematically sending requests (see DoS attack below). This attack is very visible and is usually suppressed quite quickly.

4) Unlawful Use of Information

These attacks focus on using the obtained information to access non-public parts of the system or to use certain services without authorization. In UAVs, for example, the acquired password can be used to decrypt the intercepted communication or to take over its control.

2.2. Destructive Methods of Cyberattack

The destructive C-UAS cybernetic methods of combat have a direct impact on the part of the attacked system that is physically irreversibly damaged as a result of this activity. These methods mainly use the vulnerability in the lower layers of the OSI (Open Systems Inter-connection) model and focus primarily on individual hardware components that are used in multiple systems. The operation of these methods is primarily based on a mechanical damage. For example: the cybernetic attack induces a collision of mechanically moving components or the cybernetic attack forces a battery or some other component to overheat and thus damage themselves.

3. Signal Detection of the UAV

When combatting mini and micro UAVs, one of the biggest problems is the detection and identification of the UAS itself, especially in an urban densely built area. Using radar or other detection methods with optical equipment (in the visible or IR optical band) is often considerably complicated by many fixed obstacles [5]. Methods using specific acoustic characteristics of the UAVs are also limited due to the interfering ambient noise [6].

The best choice in such environments is, therefore, the detection and localization of signals transmitted by the UAV itself or its control station.

These signals can be relatively well detected and identified due to their known specific transmission frequencies and known encoding.

However, this detection becomes considerably harder if the UAV is managed only by Wi-Fi in an electronically complex environment. Contemporary modern cities are full of devices that use the same Wi-Fi standards, and that hides the UAV' control signal among the other Wi-Fi networks within its range [7].

3.1. Localization Position of the remote control station or UAV

An UAV recognition method can be used, utilizing the MAC address to locate the position of the control station or the Wi-Fi standard to locate the position of the connected UAV itself. Each producer has a certain range of initial MAC address characters, which makes it uniquely identifiable. However, a problem occurs when the Wi-Fi module is modified or if the UAV MAC address is changed by software. Both of these methods are realizable by at least an average capable IT man.

3.2. Analysis of Data Flow Using Neural Network

One of the progressively evolving technologies is the technology of neural networks (NN). The main domain of these networks is their relatively rapid analysis of a large volume of data and the data subsequent evaluation. Neural networks, unlike algorithmic solutions, do not use serial computations. The task is solved simultaneously with several layers of neurons that interact

with each other. Neural network inputs can also be parameters of the UAS Wi-Fi traffic, such as the number of frames, their size, and generally the data flow over time.

Thanks to NN advantages, especially in the field of data processing, i.e., their ability to take in a lot of inputs, process them to infer hidden as well as complex, nonlinear relationships, NNs are playing a big role in signal characteristic recognition. Based on these features, the NN should be able to recognize which device it is both the UAV itself and the control station as well. Thus, the next part is devoted the NN application to the UAS detection and identification.

4. The UAS Signal Detection Using Neural Network

Neural networks can be used in the process of data mining. This is currently an increasingly inflected term. In general, it can be understood as a process aimed at discovering dependencies or finding the required information in a large volume of usually experimentally obtained data. The output is a certain knowledge that can be used in solving a decision problem, predicting values for other new data, or simply understanding a certain phenomenon or context. [8]

In principle, three methods can be used to identify UASs using 802.11 standards. UAS can be identified by the SSID, which is the name of the access point. However, this can be changed very easily. Furthermore, we can use the identification according to the MAC (Media Access Control) address, because in the MAC address the first six characters identify the manufacturer. This method is already suitable for use, but the MAC address can also be changed. The third method is based on the analysis of the data flow and the creation of the so-called fingerprint. For this purpose, a neural network can be successfully used, which has the task of analysing the data flow to decide whether it is intercepted traffic originating from UAS communication or not.

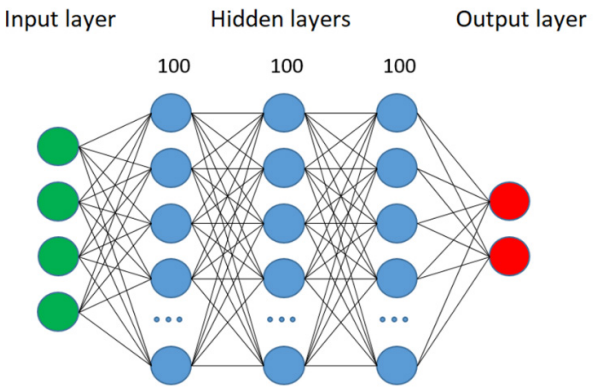


Figure 2: Topology of the created neural network

Input data is a key element in training neural networks. The ability of the network to learn and evaluate, or rather to evaluate with sufficient accuracy, depends on the appropriate selection of individual parameters and the input data set.

The selection of input parameters was made on the basis of knowledge obtained by analysis of data flows originating from the tested UAS and several other applications (e.g. Skype, YouTube). Analysis of data streams revealed that one of the determining parameters is the number of unique frame sizes in a given sample. This is due to the nature of the transmitted stream, where certain activities or communications create a larger number of unique frame sizes than others. The second determining parameter is the number of frames transmitted per second. This is mainly related to the overall data flow, but it also in some way represents the disposition of the intercepted communication. The average frame size was chosen as the third parameter. It expresses whether the captured traffic contains rather smaller or larger frames. The size of the frames is affected by the disposition of the transmitted data. The last parameter was the ratio of the number of frames with a size greater than 100 B and less than 100 B. This value was chosen mainly based on the analysis of all data streams, where different operations have this ratio different. The topology of created neural network is depicted in the Fig. 2.

Based on these parameters, the neural network can distinguish whether it is UAS operation even when masking the SSID or changing the MAC address of the transmitted frames.

5. Possibilities for UAS Elimination Using EW Methods

5.1. Possibilities of Penetration into the UAV Communication Channel

In general, the resilience of security depends on several factors, of which the quality of the encryption used and the length of the password usually have the greatest influence. Specifically, WPA2 security uses the CCMP (Counter Cipher Mode Protocol) algorithm based on the Advanced Encryption Standard (AES) encryption algorithm. The length of the password that can be used with WPA2 is between 8-63 characters. Security resilience depends not only on the length of the password, but also on the character set used. It is also advisable to choose passwords that are not contained in dictionaries, or that do not contain word forms or diminutives [7].

Brute force Attack and Dictionary Attack by Cloud Computing

Brute force Attack and Dictionary Attack can be amplified by using Cloud Computing. In the case of a brute force attack, it is necessary to test all possible combinations in the set given by the specified parameters. Any known information about password parameters significantly speeds up the successful use of this attack. Brute force Attack will certainly find a password in the future, but it is better to use a dictionary attack first. Dictionary attack is based on creating as big as possible a dictionary of known words, which increases the chance of an earlier password being discovered. Commonly used high-performance computer sets can test even with a GPU (Graphic Processor Unit) of only 400 KHps (Kilo Hash per second) when used for WPA2 encryption. The way to amplify computing power is called Cloud Computing. With these services, it is possible under certain circumstances to break relatively long passwords in short or real time. These services can be rented from technology companies such as Google or Amazon, from 10 minutes for the duration of the rental [9].

Rainbow Table

In some cases, the "Hashing Function" format is used for encrypted communication. This function recalculates the password entered by the user and the output is "Hash", which is a transformation of a string to a given number of characters. These

hashes are precalculated within the "Rainbow Table" to make it easier to crack the password [10].

DoS Attack

Denial of Service attack is an attack which transmits recurring deauthentication frames that result in disconnection of communicating devices on 802.11 [11].

KRACK Method

The Key Reinstallation Attack (KRACK) method is one of the relatively newly discovered methods focusing on a certain vulnerability of the Wi-Fi standard. It uses a four-way authentication process, in which communication is established between the AP (Access point) and the client. Simply put, it works on the principle of delaying the response to the third message during the four-way authentication sent to the access point. As a result, the AP sends a new 3rd message with an increased "counter number". When it receives the 3rd message for the first time, the client installs the GTK (Group Transient key) and PTK (Pairwise Transient Key) keys and sends the 4th message, which is then held. Upon receiving a retransmitted 3rd message with an increased counter number, the client reinstalls the previously installed GTK and PTK keys and resends the 4th message. Subsequently, both 4th messages are left to pass to the AP. Knowledge of changing the original and later reinstalled GTK and PTK keys can then be used to decrypt the communication [12].

5.2. Options after Successful Penetration into the UAV System

From the attacker's point of view, there are several ways to penetrate the system, differing in danger and enabling different goals to be met [1].

Secret Observation

The most inconspicuous way to start a penetration is a simple observation of the video stream or telemetry data, which are received by the UAS control station. This makes it possible to identify the equipment of the UAS with specific sensor (measuring) devices or to identify the operator' area of interest.

Sending Unobtrusive Commands

The operator' control can be affected by sending confusing or conflicting commands. Such an effect on the UAS may lead to the termination of the UAS task, as the operator will think it is a technical fault.

Complete take over the UAS

For the complete takeover of the UAS, it is necessary to disconnect the original RC and prevent it from being reconnected. To do this, you need to know the password and the ability to change it in real time as well. Successful takeover largely depends on the manufacturer of the individual components or the UAS model itself.

6. The Proposed Concept of the EW C-UAS System

6.1. Basic principles and functioning of the proposed concept

The basic principles of system operation are clearly shown in the Figure 3. The whole system is divided into one main and two support sections, each of which brings a different ability, or added value. The proposed concept [6] may represent the addition of another element of complex defence against micro and mini UAS.

The concept includes devices to detect signals and record Wi-Fi data in passive mode and send deauthentication frames in active mode. Within one computer and control centre, information are collected from one or more devices.

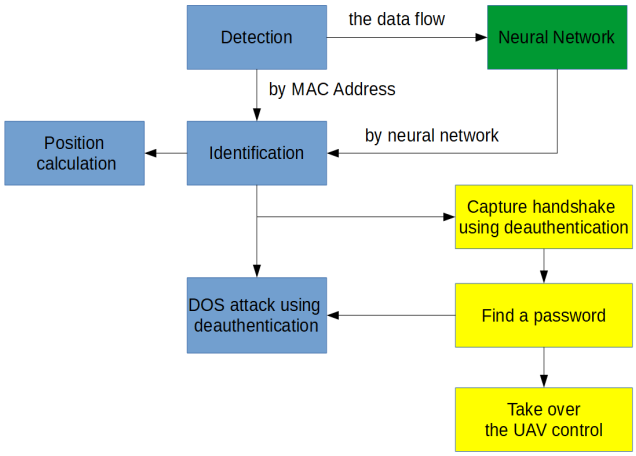


Figure 3: The proposed neural network

Detection, Identification, Position calculation and DoS Attack Using Deauthentication

The main section (shown in blue) provides the basic capabilities of the system and is the only one that can act automatically in a limited way. It contains 4 blocks providing initial signal detection (capture), device identification according to MAC address, position calculation based on information from one or more sensors and, if necessary, the DOS attack itself. Upon successful capture and identification of an unwanted UAS, we are able to send a deautentication frame to the nearest sensor to disconnect communication between the RC and the UAS or between the RC and the display device. After sending only one or a few deauthentication frames, the connection is established automatically, but if the deauthentication frames are still sent, the connection will not be established. [9]

Neural Network

The first support section (shown in green) evaluates the intercepted data stream using a neural network (explained in more detail in Section 4) and, based on the results, identifies the type of device or the type of traffic in the intercepted communication. This information is passed on to the main section for subsequent specification [6].

Possibilities to Find Password

The second support section (shown in yellow) aims to bring the ability to take control of the UAV. The process usually begins by sending a deauthentication frame and then capturing a 4-way handshake. This provides a signal pattern to perform the password retrieval process. After successfully finding the password, it is possible to take over the control, either completely or partially, or penetrate into the communication system. [6]

6.2. The proposed concept applied in the urban area

For practical use in urban areas, e.g., the device *Alfa AWUS036ACM* (see Fig. 4) can be used as a sensor and

transmitter, providing sufficient omnidirectional range, supporting 802.11 a / b / g / n / and ac standards, and operating in the 2.4 GHz and 5 GHz frequency bands, respectively. Another indisputable advantage of this device is its low price in combination with normal commercial availability [6].

Figure 4: Device *Alfa AWUS036AC*

Figure 5 shows an object with a marked critical defence point (red circle), access roads (3, 5), high-rise areas (1,2,4), and a low building (6), which poses the greatest risk due to the possibility of direct visibility to defended point

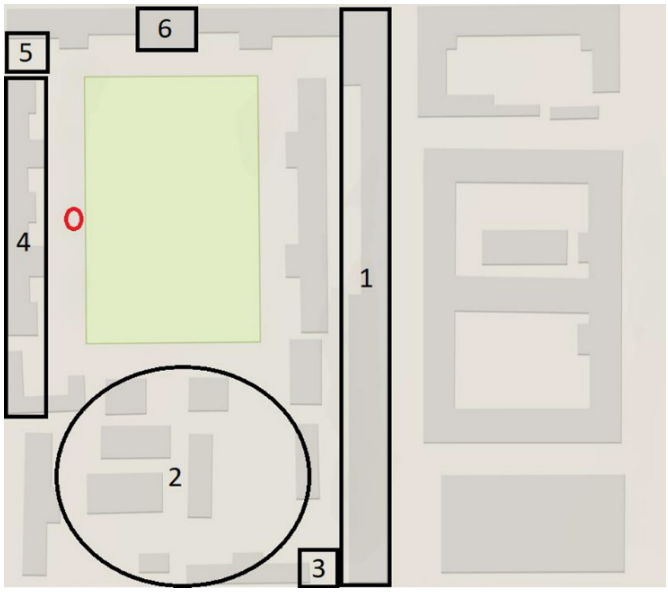


Figure 5: Defended critical point

It is advisable to place the sensors so that they cover most of the area and selected places outside the area from which the UAS control could be performed. The Alfa AWUS036AC device, which is connected to a certain small computer, e.g. *Raspberry Pi*, is taken as a sensor. All sensors would then be connected to a central point where evaluation and control would take place.

Despite of this that the quantity of sensors seems to be a little considerable, by using low-price hardware mentioned above, we are able to cover the protected area completely. Connection to the EW C-UAS system proposed in subsection 6.1 can sustain full control of this defended critical area.

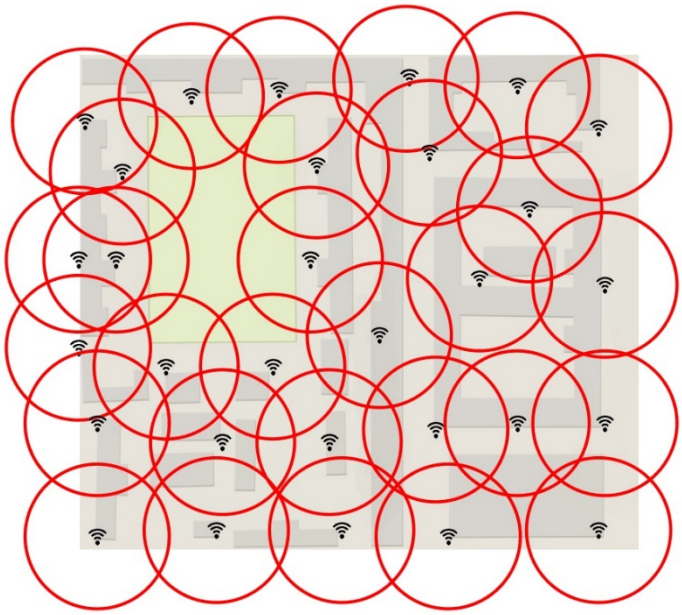


Figure 6: Defended critical point

6.3. Partial conclusion of Section 6

The previous parts described the general principles and division of cyber-attacks and the operation of the 802.11 standard of interest. To detect UAS using this standard was developed by applying the neural network and the concept in which it is embedded. However, instead of a neural network, it would be possible to use other methods of genetic programming, but the use of a neural network is suitable for this task.

Within this theoretical concept, the use of an existing HW and its deployment on the example of a defended object in a non-war area was also proposed. The use of the system in the military field is dealt with in the following section.

7. The proposed concept applied to artillery units

Based on knowledge from current conflicts, especially from the war in Ukraine and Syria, it is clear that UAVs - small, light, and cheap unmanned aerial vehicles pose a great risk to all combat units. The availability of these tools, which are available for small amounts of money, together with their capabilities makes them an ideal means of conducting aerial reconnaissance.

Findings from the conflicts in Ukraine and the Middle East clearly show that these instruments, when applied to fire support units, specifically artillery, can be used very effectively, especially for:

- reconnaissance (uncovering the battle group),
- target acquisition,
- artillery fire control,
- evaluation of the effectiveness of fires,
- perform attacks using explosives,
- perform attacks, using weapons s of mass destruction (chemical, biological, and possibly also radioactive "dirty" bombs).

In recent years, we can find countless combat situations, where small UAVs have successfully performed the tasks mentioned above. To illustrate this, several examples of the use of small UAVs are given in the table No 1.

Table 1: Examples of small UAV attacks

Year	Example
2005	Al-Qaida used Chinese made remote control model airplanes to recon Pakistan security forces prior attacks. It was also weaponized with IED.[13]
2014	Islamic state used small quadcopters for recon missions of Syrian military bases prior to ground attacks. [13]
2015	Drone attacks on military bases in Ukraine. Each drone was equipped with one grenade (ZMG-1 type). [14]
2015	Several attacks on ammunition depots at Svatovo, Ukraine. Drones were equipped with grenades. [14]
2016	Two French Special Forces soldiers were injured and an exploding ISIS drone killed two Kurdish fighters. [14]
2019	Drone attacks struck two key oil installations inside Saudi Arabia [15].
2020	Pro-Kremlin mercenaries use an unmanned aerial vehicle to drop a rocket-propelled grenade on Ukrainian emplacements. [16]

These examples clearly show the possibilities of using small UAVs and their ability to cause significant losses. The use of small commercial UAVs in combat operations can provide the enemy with key information about the positions of our own troops and their manoeuvres. In combination with explosives or weapons of mass destruction, it is possible to cause significant losses in technique and manpower. For this reason, it is necessary to be prepared for this variant and be able to defend effectively against these types of attacks.

7.1. Artillery units in the regular operations and proposed concept of application

In a peer-to-peer war, all kinds of military forces and resources are involved in combat operations. For simplicity, examples of the application of the proposed concept of protection against small UAVs are reduced only to artillery units. However, a similar approach can be used for all types of units and their specializations wherever there is a real possibility of attack by these UAVs.

The artillery battalion (battery) is conducting its operations within a given position area of artillery (PAA) within the zone of attack (defence) of the brigade (task force). Part of this PAA are also battalion (battery) Fire Direction Centres (FDC). Fig. 7 gives an example of PAA [17].

The danger emerging from the use of small UAVs is, in particular, the possibility of uncovering a combat formation and subsequent directing fire on firing units and command posts, and attacks using explosives placed on the UAV or attacks using weapons of mass destruction. Artillery units are usually deployed in a hidden position, outside the firing positions, when they do not conduct fires. Weapon systems occupy firing positions only when they are firing. Firing positions and firing points are usually in the open space. At this point, they are vulnerable and easy targets.

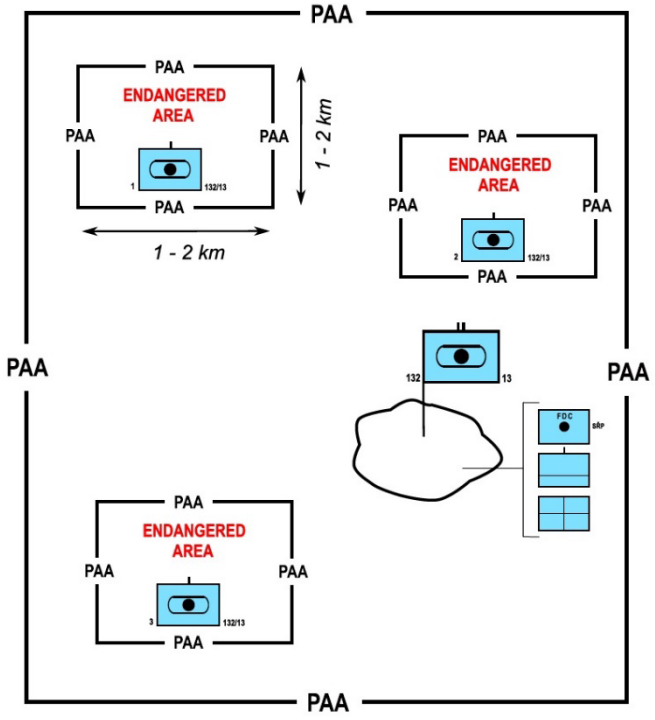


Figure 7: The position area of artillery [18]

One of the main risks the artillery is facing is the counterbattery fires. Artillery units are always priority targets and it is necessary to avoid the risk of detection, which precedes the implementation of counterbattery fire. The tactic of using small UAVs by enemy reconnaissance and diversion groups has always been a danger. In connection with the development of capabilities and possibilities of using not only military, but also small commercial UAVs, new risks are emerging and their frequency may be higher due to their massive spread and availability.

With a standard time of one fire mission and time to leave the firing position, the probability of successful counterbattery fire is low. In the case of proactive counterbattery activity, where artillery units are actively searched for and neutralized without unmasking themselves by firing, the probability of success is much greater than in the case of reactive counterbattery fire when it is reacted to firing artillery. The enemy gains an advantage by knowing the coordinates of the firing positions and starts the fire mission when the cannons are taking up firing points to conduct fire. It is this key information that can be easily obtained using small UAVs. They are able to uncover a combat formation, detect individual artillery weapon sets and already in the phase of the manoeuvre into the firing position, warn the enemy, and provide sufficient information for the preparation and execution of fire. At the same time, they can provide a detailed evaluation of battle damage assessment (BDA) after firing.

Another variant is the use of UAVs as carriers of explosive devices, which can be used to destroy targets. In the case of artillery, a suitable target are FDCs. Destruction of them will

eliminate the entire firing battery from combat for a period of time. Alternatively, it is possible to destroy individual artillery weapon sets, which cause a deficiency in fire support.

Based on these findings, the proposed concept of detection and neutralization of UAVs is a very effective tool for reducing the risk of detection by these means and the subsequent elimination of artillery units by enemy counterbattery fire, or explosive devices. From the point of view of the effective use of the proposed concept, it is necessary that the established measures reflect the tactics and procedures of artillery fire units. The biggest restrictions will be mainly area requirements. The firing battery of self-propelled cannon howitzers is deployed in the area of firing positions, usually 1 - 2 x 1 - 2 km, depending on the position conditions and the combat task [19]. This space must be covered with the ability to neutralize the UAV - this can be achieved by using a directional antenna with sufficient gain or power, which will ensure the required range. The individual antennas should be connected to an automated system, from which their focus and modes of operation will be controlled.

Firing units are most exposed to observation at the moment of performing fire missions, when they are deployed in firing positions. A suitable variant is the placement of equipment for neutralization of UAVs on vehicles of artillery reconnaissance/survey units, or on artillery weapons sets [20]. To a minimal extent, it would be sufficient to place this device on the lead gun of fire platoons, or vehicle of the platoon commander. However, this depends on the possibility of increasing the scope of the opportunity of neutralizing the enemy UAV within the proposed solution. At command points (FDCs), it would be sufficient to place the equipment on one vehicle of FDC.

7.2. Artillery units in asymmetric operations and proposed concept of application

The use of artillery in an asymmetric conflict has its specifics. Firing units are usually located on permanent or forward operational bases and provide fire support to units performing framework operations in the area of responsibility. They do not manoeuvre and are constantly at the firing points. In an asymmetrical operation, the risk of counterbattery fire is minimal according to current experience. However, commercially available UAVs can radically change this state if used correctly [21]. The danger of the UAV being used by the enemy lies mainly in the possibility of uncovering the combat formation, finding out the position of the firing point of individual artillery weapon sets, and possibly leading an attack on them using explosives placed on the UAV. In an asymmetrical environment, this way of conducting combat is one of the few ways to put the fire units located in the base area in danger. A coordinated attack can cause significant technical losses and limit the ability to provide artillery fire support in the area of operations. It is therefore appropriate to consider how to defend against such an attack and to provide protection for the artillery fire support units and other base personnel. Placing the defence EW device on the base in such a way, which achieves coverage of the base perimeter, is a suitable variant of solving the problem. In the event of a breach of the base perimeter, the operator is able to ensure the protection of base members and prevent the UAV from flying over the base within the proposed method.

7.3. Partial conclusion of Section 7

This concept of protection against small UAVs will ensure that units are not attacked or uncovered and the risk of losing artillery fire support will not increase. At the same time, the simplicity of the proposed solution will not place excessive demands on interventions in the construction of vehicles or buildings. The solution can be applied to all types of troops and its specialization, wherever the risk of using small UAVs in the combat activities of the enemy can be assumed.

8. Conclusion

UAV defence in a cyber environment combines elements of cyber security and air defence. The development of new methods for attack and defence is necessary due to the dynamics of the development of these areas, which is confirmed by current knowledge from the fighting in Nagorno-Karabakh, where unmanned aerial vehicles and artillery play a major role. One of the possible ways is the proposed solution for penetration into UAS control systems.

The methods described in this article focus on supplementing and extending existing complex solutions of UAV defence models. Their application clearly achieves an increase in capability in the implementation of defensive measures and the fight against the UAV opponent. Simplicity and ease of materials and new inexpensive technical means represent an advantage for application in practice, e.g. in the military environment, as described and suggested by the examples of artillery units.

As the main achievement, authors consider a practical demonstration of the possibility for identifying the device based on the characteristics of the data frames. For this purpose, a neural network was developed which, based on the entered parameters, can evaluate whether the intercepted traffic comes from UAV communication.

As all the UAS defence systems are usually aimed at a certain type of opponent, this concept is not universal. It focuses on complementing the existing comprehensive defence model with more options, thus creating the ability for the defence institutions to fight more effectively with highly sophisticated adversary means.

Subsequent research in this area could be aimed at expanding the input parameters of the neural network and output neurons. The neural network should then be able to identify not only the device or the type of data stream transmitted during UAV operation, but also other types of operation. The whole system could then be used for overall monitoring - for the purpose of using Wi-Fi networks in the area of interest. However, extending this system to such a level could mean a significant intrusion on the privacy of individual users. If implemented in practice, this problem would be forced by the author to solve both at the technical and legal level.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was performed as a part of the project "PROKVES" for Development of Electro – Military Departments at Faculty of Military Technology and "SPECIFIC RESEARCH" at Faculty of Military Leadership, University of Defence in Brno. They have been financed from the Institutional Support Department funded by the Ministry of Defence Czech Republic.

References

[1] V. Minařík a M. Krátký, "Cybernetics fight against the UAV". 2019 International Conference on Military Technologies (ICMT), Brno, Czech Republic: Institute of Electrical and Electronics Engineers Inc., 2019, 8870103.

[2] M. Kratky and J. Farlik "Countering UAVs – the Mover of Research in Military Technology". Defence Science Journal 2018, **68**(5), 460-466. https://doi.org/10.14429/dsj.68.12442.

[3] NATO C-UAS Working Group, "NATO Countering Class I Unmanned Aircraft Systems (C-UAS) Handbook". Version 1.8. Brussels, BEL 2020.

[4] J. Kolouch, "Kybernetické útkoy" 2019. https://csirt.cesnet.cz/_media/cs/documents/kyberneticke_utoky.pdf

[5] T. Kornelly, J. Casar, V. Stary and J. Farlik, "Mobile phone optical sensor usage for navigation tasks" 2017 International Conference on Military Technologies (ICMT), Brno, 2017, 671-675. doi: 10.1109/MILTECHS.2017.7988842.

[6] V. Minařík, "Elektronický boj v obraně proti bezpilotním prostředkům". Brno, 2019. Ph.D. exam thesis. University of Defence.

[7] V. Minařík a M. Krátký, "The Non-destructive Methods of Fight Against UAVs". 2017 International Conference on Military Technologies (ICMT), Brno, 2017, 690-694. doi: 10.1109/MILTECHS.2017.7988845.

[8] M. Lastovka, "Data mining jako moderní prostředek analýzy dat a možnosti jeho uplatnění v podmínkách AČR". Brno, 2015. Diploma thesis. University of Defence.

[9] "Ethical Hacking and Countermeasures v10". EC-council, 2019. https://iclass.eccouncil.org/.

[10] "Rainbow tables tajemství zbavené" [Online article]. 2015 [cit. 2019-01-5]. Available from: https://www.soom.cz/clanky/1165--Rainbow-tables-tajemstvi-zbavene.

[11] J. Kolouch, P. Bašta, "CYBERSECURITY". Praha: CZ.NIC, z. s. p. o, 2019. ISBN 978-80-88168-34-8.

[12] M. Vanhoef, F. Piessens, „Key Reinstallation Attacks: Forcing Nonce Reuse in WPA2" [Online article]. 2017 [cit. 2019-01-5].

[13] J. R. Bunker, Terrorist and insurgent unmanned aerial vehicles: use, potentials, and military implications. Carlisle, Pennsylvania: US Army war college press, 2005.

[14] Russian Drone With Thermite Grenade Blows Up a Billion Dollars of Ukrainian Ammo. Popularmechanic [online]. 2017, 27. 7. 2017, [cit. 2020-08-13]. Dostupné z: https://www.popularmechanics.com/military/weapons/news/a27511/russia-drone-thermite-grenade-ukraine-ammo/.

[15] Donbas: Russian militants use drone to attack Ukrainian positions. 112UA [online]. 1. 6. 2020, [cit. 2020-08-13]. Dostupné z: https://112.international/conflict-in-eastern-ukraine/donbas-russian-militants-use-drone-to-attack-ukrainian-positions-51819.html

[16] J. Ivan, K. Šilinger, L. Potužák. "Target acquisition systems suitability assessment based on joint fires observer mission criteria determination". In: Madani K., Gusikhin O. ICINCO 2018 - Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics. Portugal: SciTePress, 2018, roč. 1, 397-404. ISBN 978-989-758-321-6.

[17] B. Čermák, R. Hegyi, L. Potužák, M. Kalina, "Bojové použití dělostřelectva", Military publication Pub-35-14-01, TRAIDOC Vyškov, 2013.

[18] J. Ivan, "Situační značky a zkratky pro dělostřelectvo: (vojenská symbolika a taktické značky pro dělostřelectvo dle APP-6)". University of Defence in Brno. ISBN 978-80-7582-122-5.

[19] M. Bláha, K. Šilinger, "Application support for topographical-geodetic issues for tactical and technical control of artillery fire". In: International Journal of Circuits, Systems and Signal Processing, **12**, 48-57. ISSN 1998-4464.

[20] IVAN, Jan and Jan POTUŽÁK, Jiří ŠOTNAR. Dělostřelecká rekognoskace pro zabezpečení činnosti autonomních zbraňových systémů a základní požadavky na rekognoskační jednotky. *Vojenské rozhledy*. 2019, **28** (4), 063-077. ISSN 1210-3292 (print), 2336-2995 (on-line). Available at: www.vojenskerozhledy.cz.

[21] KARBER, Phillip A. Lessons Learned from the Russo-Ukrainian War: Personal Observations [online]. 8 July 2015. Available at: https://www.researchgate.net/publication/316122469_Karber_RUS-UKR_War_Lessons_Learned.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 455-463 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Fault-Tolerant Control of Permanent Magnet Synchronous Motor Drive under Open-Phase Fault

Amr Saleh[1,*], Nada Sayed[2], Ghada Ahmed Abdel Aziz[3], Mona Nagieb Eskander[3]

[1]*Electrical Engineering Department, Faculty of Engineering, Fayoum University, Fayoum, 63514, Egypt*

[2]*Engineering Mathematics and Physics Department, Faculty of Engineering, Fayoum University, Fayoum, 63514, Egypt*

[3]*Power Electronics and Energy Conversion Department, Electronics Research Institute, Cairo, 12622, Egypt*

ARTICLE INFO

Article history:
Received: 27 August, 2020
Accepted: 15 October, 2020
Online: 20 November, 2020

Keywords:
Fault Diagnostic
Fault Tolerant Control
Field Oriented Control
Open Phase Fault
PMSM

ABSTRACT

This paper presents an integrated solution for a fault-tolerant three-phase permanent magnet synchronous motor (PMSM) field-oriented control drive subjected to an open-phase fault (OPF) integrated with effective fault-tolerant detection methodology. The fault detection methodology is based on model predictive current control (MPCC), which is easy to apply, detect OPF in a range of microseconds, and robust under-speed or load transient. On the other hand, the fault-tolerant compensation is based on a neutral point connection together with stator current regulation to maintain the magneto-motive force (MMF) unchanged under open-phase failure. Controlling the motor phase currents in the post-fault condition ensures a rotating magnetic field similar to that produced during healthy conditions thus, reducing the saturation impact and ensuring the reliability of the control operation. The proposed integrated fault-tolerant drive is validated using MATLAB simulation that ensure the effectiveness of the proposed solution in steady state and during transients under different loading and rotor speed condition.

Nomenclature

v_a, v_b, v_c	The *abc* stator voltages
v_d, v_q	The *dq*-axis stator voltages
i_a, i_b, i_c	The *abc* stator currents
i_{ds}, i_{qs}	The *dq*-axis stator currents
P	The number of pole pairs
R_s	The stator resistance
L_d, L_q	The *dq*-axis stator inductances
L_{ls}	The leakage inductance
ω_e	The rotor electrical speed
λ_m	The permanent magnet flux linkage
T_e	The electromagnetic torque
N	The winding turns
θ_r	The electrical rotor position angle

1. Introduction

Permanent magnet synchronous motor (PMSM) is employed when high torque density/volume is needed, including systems with high safety requirements [1], [2]. As the more complexity of the electrical drives become, the possibility of system faults and failures increases which are hard to be prevented and can lead to severe effects. Fault-tolerant (FT) capability is the main requirement of motor drives in various applications such as electric vehicles, aerospace and spacecrafts, as the reliability of the motor drive system is very critical [3]. Such a system should identify faults, locate their origin, and, if possible, continue to operate after the fault happens. The Fault-tolerant control scheme involves fault detection and compensation. The open-circuit fault may not cause instantaneous damage and can even remain undetected for long periods thus leading to secondary failures [4]. The open-phase fault (OPF) happens in electrical machine drives when a winding phase is disconnected or one inverter leg fails or even due to a protection system isolating a short-circuit condition. In this situation, a large ripple of electromagnetic torque is created and serious mechanical vibration may happen. Fast and accurate methods of detection and fault-tolerance are therefore urgently needed to identify, isolate the fault early, and prevent the destruction of the whole system [5].

Among the numerous techniques for identifying faults by sensing waveforms is the Machine Current Signature Analysis (MCSA) which is based on Fast Fourier transform (FFT) stator

*Corresponding Author: Amr Saleh, Faculty of Engineering, Fayoum University, +201205555699, Email: aae00@fayoum.edu.eg

current analysis but is limited to steady-state process and requires stator current monitoring for at least one or more electrical post-fault periods before detecting the fault successfully. It also needs substantial processing energy and memory resources. Another technique known as Park's vector approach relies on the analysis of stator current in the rotor reference frame [6]. The current trajectories have a circular shape in normal conditions, the trajectories shape changes to the ellipse in the case of open phase fault (OPF). Modifications and developments of this technique are presented in [7]. These methods are much simpler and faster than MCSA, but during transient speed and load, they are subject to false alarms. In [8], a model reference adaptive system (MRAS) detection technique has been presented which relies on voltage monitoring and model current calculation when a fault happens in the steady-state; the threshold value of the detection was estimated empirically. The model predictive current control (MPCC) that has been presented for fault diagnostic in PMSM [9], [10] is employed here, where it relies on the stator current prediction and estimated error. MPCC scheme for the fault detection is found to be fast, requires minimum processing effort and robust under transient conditions.

Fault-tolerant control (FTC) compensation techniques for OPF PMSM drives have been introduced in two major forms depending on whether reconfiguration in hardware is needed or not. Methods with reconfiguration [11] are most applicable due to the fact that single-phase operation of the method without reconfiguration [9] even with modifications applied has the probability of matching the system mechanical resonance frequency at low speeds. One of the reconfiguration methods with extra inverter leg connected to the neutral point of PMSM is presented in [12], however still the main drawback of this method is the extra complexity of the power converter and additional cost and size. In [13] the authors employed a fourth leg in a manner that it replaces the faulty leg, yet the same drawback of extra complexity and additional cost and size remains the same. In [14] fault-tolerant systems of five-phase permanent magnet motor drive was studied. However, the multiphase machines are not a suitable solution in some application owing to the extra weight and costs. In this paper, the FTC compensation methodology implemented is targeting a lower cost compensation methodology [15], however, the magnitude of the phase current is limited to its permissible limit in order to minimize saturation impacts and ensure a sustainable motor drive fault-tolerant operation [16].

It is commonly found in literature research efforts that are dedicated for the fault detection [6]- [9] and others that present the fault compensation methodology [11]-[15], while it is rarely presented in the literature papers that provide complete solution from the instant of open phase fault, it is detection until applying the fault compensation methodology

In this paper the authors are introducing an integrated solution for a field oriented controlled PMSM drive that incorporates the benefits of fast open phase fault detection using a methodology based on MPCC and a reconfiguration fault compensation method that utilizes minimum additional hardware achieving a balanced operation with balanced current in magnitude thus dramatically reduces the torque ripples and consequently the ripples in the rotor speed as presented in [15]. In contrary to using a pulse width modulator in [15], the authors here deduced the required transformation to apply a space vector modulator which provides less harmonic distortion.

The paper is arranged as follows, Section 2 presents the modeling of PMSM, Section 3 presents the concept of the field-oriented control (FOC) of PMSM, Section 4 presents the concept of the MPCC used for OPF detection, Section 5 describes the proposed FTC compensation strategy and Section 6 shows the simulation results. Finally, Section 7 provides the conclusion of this work.

2. PMSM Mathematical Model

The electrical dynamic equations of the PMSM in the abc stationary reference frame can be written as [15], [17]:

$$\mathbf{V}_{\mathrm{abc}} = \mathbf{R}_{\mathrm{s}}\mathbf{I}_{\mathrm{abc}} + \mathbf{L}_{\mathrm{abc}}\frac{\mathrm{d}\mathbf{I}_{\mathrm{abc}}}{\mathrm{dt}} + \omega_e\,\mathbf{I}_{\mathrm{abc}}\frac{\partial\mathbf{L}_{\mathrm{abc}}}{\partial\theta_{\mathrm{r}}} + \omega_e\,\boldsymbol{\psi}_{\mathrm{abc}}$$

$$\boldsymbol{\psi}_{\mathrm{abc}} = \lambda_m\begin{bmatrix}\cos\theta_r\\ \cos\left(\theta_r-\frac{2\pi}{3}\right)\\ \cos\left(\theta_r+\frac{2\pi}{3}\right)\end{bmatrix} \tag{1}$$

Where $\mathbf{R}_{\mathrm{s}} = \mathrm{diag}(R_s)$, $\mathbf{V}_{\mathrm{abc}} = [v_a \quad v_b \quad v_c]^T$,

$\mathbf{I}_{\mathrm{abc}} = [i_a \quad i_b \quad i_c]^T$ and the terms of the matrix $\mathbf{L}_{\mathrm{abc}}$ are

$$L_{hh} = L_{ls} + \left(\frac{L_d+L_q-2L_{ls}}{3}\right) - \left(\frac{L_d-L_q}{3}\right)\cos[2(\theta_r+\alpha)]$$

$$L_{kh} = L_{hk} = \frac{-1}{2}\left(\frac{L_d+L_q-2L_{ls}}{3}\right) - \left(\frac{L_d-L_q}{3}\right)\cos[2(\theta_r+\gamma)]$$

$$\gamma = \begin{cases}-\frac{\pi}{3} & (h,k)=(a,b)\\ \pi & (h,k)=(b,c)\\ \frac{\pi}{3} & (h,k)=(a,c)\end{cases},\quad \alpha = \begin{cases}0 & h=a\\ -\frac{2\pi}{3} & h=b\\ \frac{2\pi}{3} & h=c\end{cases} \tag{2}$$

For the mechanical part:

$$T_e = J\frac{d\omega_r}{dt} + B\omega_r + T_L \tag{3}$$

T_L is the load torque, B is the damping coefficient, ω_r is the rotor mechanical speed and J is the moment of inertia.
The electromagnetic torque can be represented as in (4):

$$T_e = \lambda_m P[\,(i_a - \tfrac{i_b}{2} - \tfrac{i_c}{2})\cos\theta_r + \tfrac{\sqrt{3}}{2}(i_b - i_c)\sin\theta_r] \tag{4}$$

For the loss of one phase, the current in the lost phase becomes equal to zero keeping a single-phase current circulating in the remaining two phases.
For the post fault operation with the neutral of the star-connected three-phase windings of the PMSM connected to the center point of the DC link, the PMSM equations can be re-written as in [15]:

$$\mathbf{V}_{\mathrm{ij}}(\mathrm{t}) = \mathbf{R}_{\mathrm{s}}\mathbf{I}_{\mathrm{ij}}(t) + \frac{\mathrm{d}\boldsymbol{\lambda}_{\mathrm{ij}}(\mathrm{t})}{\mathrm{dt}} \tag{5}$$

$$\boldsymbol{\lambda}_{\mathrm{ij}}(\mathrm{t}) = (\mathbf{L}_{\mathrm{S}} + \mathbf{L}_{\mathrm{Mij}})\mathbf{I}_{\mathrm{ij}}(\mathrm{t}) + \mathbf{M}_{\mathrm{ij}}\lambda_m \tag{6}$$

where, (i,j)=$\begin{cases}(b,c)\ phase\ a\ open\\ (a,c)\ phase\ b\ open\\ (a,b)\ phase\ c\ open\end{cases}$

$$\mathbf{V}_{ij} = [v_i \quad v_j]^T \qquad \mathbf{I}_{ij} = [i_i \quad i_j]^T \qquad \boldsymbol{\lambda}_{ij} = [\lambda_i \quad \lambda_j]^T$$

$$\mathbf{L}_S = \begin{bmatrix} L_{ls} + \left(\frac{L_d + L_q - 2L_{ls}}{3}\right) & \frac{-1}{2}\left(\frac{L_d + L_q - 2L_{ls}}{3}\right) \\ \frac{-1}{2}\left(\frac{L_d + L_q - 2L_{ls}}{3}\right) & L_{ls} + \left(\frac{L_d + L_q - 2L_{ls}}{3}\right) \end{bmatrix}$$

$$\mathbf{L}_{Mij} = -\left(\frac{L_d - L_q}{3}\right)\begin{bmatrix} \cos[2(\theta_r + \beta_1)] & \cos[2(\theta_r + \beta_2)] \\ \cos[2(\theta_r + \beta_2)] & \cos[2(\theta_r + \beta_3)] \end{bmatrix}$$

$$\mathbf{M}_{ij} = \begin{bmatrix} \sin(\theta_r + \beta_1) \\ \sin(\theta_r + \beta_3) \end{bmatrix} \tag{7}$$

β_1, β_2 and β_3 are selected under the condition of open phase fault, as defined in (8):

$$(\beta_1, \beta_2, \beta_3) = \begin{cases} \left(\frac{-2\pi}{3}, 0, \frac{2\pi}{3}\right) & \text{phase } a \text{ open} \\ \left(0, \frac{\pi}{3}, \frac{2\pi}{3}\right) & \text{phase } b \text{ open} \\ \left(0, \frac{-\pi}{3}, \frac{-2\pi}{3}\right) & \text{phase } c \text{ open} \end{cases} \tag{8}$$

The torque consists of two-component: reluctance torque (T_{er}) and excitation torque (T_{em}) that is expressed in (9) and (10):

$$T_e = T_{er} + T_{em} \tag{9}$$

$$T_e = \frac{P}{2}(\mathbf{I}_{ij}(t))^T \frac{\partial \mathbf{L}_{Mij}}{\partial \theta_r}\mathbf{I}_{ij}(t) + P\lambda_m(\mathbf{I}_{ij}(t))^T \frac{\partial \mathbf{M}_{ij}}{\partial \theta_r} \tag{10}$$

3. Field oriented control (FOC) of PMSM

The FOC method used for AC electrical machine is concerned to let the AC machine behaves like the dc machine in terms of decoupled torque and flux control. A FOC control algorithm is manipulating the stator current vector in a d-q reference frame attached to the rotor. The torque equation of the PMSM in a d-q reference is given as:

$$T_e = 3P(\lambda_m i_{qs} + (L_q - L_d)\, i_{ds}\, i_{qs})/2 \tag{11}$$

Keeping the direct component of the stator current vector equals to zero results in a torque that is controlled through the manipulation of the quadrate component of the stator current vector. λ_m is the permanent magnet flux linkage resembling the flux in the DC machine so the torque can be expressed as:

$$Te = \frac{3P\lambda_m i_{qs}}{2} = K_t i_{qs} \qquad , \; K_t = \frac{3P\lambda_m}{2} \tag{12}$$

FOC has the advantage of having a fast response and little torque ripple.

4. MPCC method for OPF detection

The model predictive current control (MPCC) applied for fault diagnostic relies on the stator current prediction and estimated error to detect the abnormal situation. The current prediction is derived from the stator voltage equations in the $d-q$ reference frame of the machine attached to the rotor (rotor field) as given in (13) and (14) [9], [10].

$$v_d = R_s i_{ds} + L_d \frac{di_{ds}}{dt} - \omega_e L_q\, i_{qs} \tag{13}$$

$$v_q = R_s i_{qs} + L_q \frac{di_{qs}}{dt} + \omega_e L_d\, i_{ds} + \omega_e \lambda_m \tag{14}$$

By simplifying the derivation of the predicted current from (13) and (14), justified by the short sampling interval of Δt, the rotor speed ω_e can be considered to be quasi-constant, and linear change of stator currents during the short sampling interval [9].The PMSM equations can, therefore, be discretized with Euler approximation, resulting in the prediction of stator current at the sample t_{i+1} as given below:

$$i_{ds(t_{i+1})} = i_{ds(t_i)} + \frac{\Delta t}{L_d}(v_{d(t_i)} - R_s i_{ds(t_i)} + \omega_{e(t_i)} L_q\, i_{qs(t_i)}) \tag{15}$$

$$i_{qs(t_{i+1})} = i_{qs(t_i)} + \frac{\Delta t}{L_q}(v_{q(t_i)} - R_s i_{qs(t_i)} - \omega_{e(t_i)} L_d\, i_{ds(t_i)} - \omega_{e(t_i)} \lambda_m) \tag{16}$$

The objective function is chosen at the sampling time t_i, as the sum of the squared errors between the predictive values of d- and q-axis currents and their corresponding actual values as expressed in (17) [10]:

$$E = (i^*_{ds(t_i)} - i_{ds(t_{i+1})})^2 + (i^*_{qs(t_i)} - i_{qs(t_{i+1})})^2 \tag{17}$$

where E is considered to be the fault indicator. Under normal conditions and without system failure, E is relatively small within the expected range, once a fault occurs, the value of E will increase in magnitude exceeding the error threshold E_{th}. The chart in Figure 1 shows how the proposed algorithm differentiates between actual fault condition and transient operation condition. In this algorithm, the condition of having one of the phases with zero measured current and zero predicted current fulfills the condition of OPF. Few sampling intervals are sufficient to detect the fault. The proposed solution has the following advantages:

1) The OPF can be identified rapidly, in a range of few sampling intervals that are in the order of µsec.
2) No considerable processing power is needed.
3) Effective during transient speed and load.

5. The fault-tolerant compensation topology

Between many configurations, the minimum level of fault-tolerance in a three-phase drive for an OPF can be accomplished by isolating the faulty element, opening the corresponding motor phase, and connecting the motor stator neutral point to the mid-point of the inverter dc bus [15], as shown in Figure 2. During the OPF, after the system detects the abnormal condition (OPF) the system firstly disconnects the faulty phase leg by disabling the PWM signals to the corresponding switches of the inverter to avoid its impact on drive behavior after applying the post fault strategy, then fault compensation technique should begin and can be split into two steps as follows:

a) Connect the PMSM neutral to the mid-point of DC bus by firing TRn to continue the machine operation as a two-phase motor as illustrated in Figure 2.
b) Apply the post-fault current regulation technique as will follow in this section.

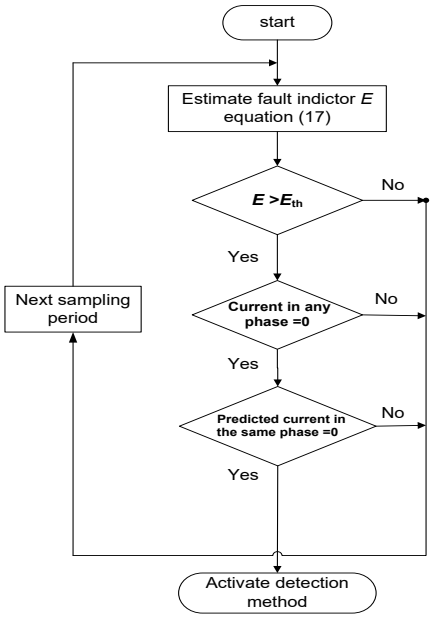


Figure 1: Flow chart of MPCC

For the PMSM FOC drive, in the healthy operation, the stator phase current has two components i_{qs}, i_{ds}, representing the torque and the flux components, respectively. During the post-fault stage, the motor neutral to the mid-point of the DC link split the machine phase currents into two currents where the zero-sequence component of the phase current becomes active, i.e. $\Sigma I_n \neq 0$. The fault-tolerant control strategy [11], [15] targets to maintain the same magneto-motive force (MMF) of the three-phase healthy operation and balance the unbalanced open-phase fault operation. This can be achieved by the following transformations to obtain the set of reference voltages for the remaining two healthy phases as follows:

$$\mathbf{V}^{u}_{\alpha\beta} = \mathbf{A}\mathbf{V}_{ij}(t) \quad (18)$$

$$\mathbf{V}_{\alpha\beta} = \mathbf{B}\,\mathbf{V}^{u}_{\alpha\beta} \quad (19)$$

$$\mathbf{V}^{r}_{qd} = \mathbf{R}_{v}\mathbf{V}_{\alpha\beta} \quad (20)$$

Also, similar matrices **A**, **B**, and $\mathbf{R_I}$ instead of $\mathbf{R_v}$ are used to transform the phase currents.

Where, $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ $\quad (\mathbf{A})^{-1} = \frac{1}{2}\mathbf{A}$

$$\mathbf{B} = \begin{bmatrix} \left[2\cos\left(\frac{\beta_1-\beta_3}{2}\right)\right]^{-1} & 0 \\ 0 & \left[2\sin\left(\frac{\beta_1-\beta_3}{2}\right)\right]^{-1} \end{bmatrix}$$

$$(\mathbf{B})^{-1} = \begin{bmatrix} 2\cos\left(\frac{\beta_1-\beta_3}{2}\right) & 0 \\ 0 & 2\sin\left(\frac{\beta_1-\beta_3}{2}\right) \end{bmatrix}$$

$$\mathbf{R_V} = \begin{bmatrix} \cos(\theta_r+\beta_2) & -\sin(\theta_r+\beta_2) \\ \sin(\theta_r+\beta_2) & \cos(\theta_r+\beta_2) \end{bmatrix} = ((\mathbf{R_V})^{-1})^{T}$$

$$\mathbf{R_I} = \frac{1}{3}\begin{bmatrix} \cos(\theta_r+\beta_2) & -3\sin(\theta_r+\beta_2) \\ \sin(\theta_r+\beta_2) & 3\cos(\theta_r+\beta_2) \end{bmatrix}$$

$$(\mathbf{R_I})^{-1} = \begin{bmatrix} 3\cos(\theta_r+\beta_2) & 3\sin(\theta_r+\beta_2) \\ -\sin(\theta_r+\beta_2) & \cos(\theta_r+\beta_2) \end{bmatrix}$$

And **S** is the transformation between the balanced alpha-beta frame and the alpha-beta reference frame of the SVPWM given as:

$$\mathbf{S} = \begin{bmatrix} \cos(\beta_2) & -\sin(\beta_2) \\ \sin(\beta_2) & \cos(\beta_2) \end{bmatrix}$$

Then the resulting two-phase post fault currents will increase to $\sqrt{3}$ of the healthy phase current with phase angle shifted 60° away from each other [11], [12]. This is illustrated by the phasor diagram in Figure 3 for an OPF in phase (a).

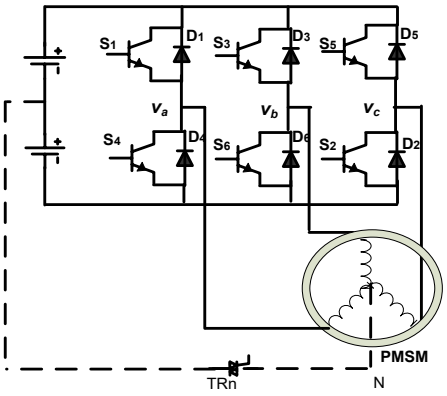


Figure 2: FTC motor drive configuration

This ensures the same rotating magnetic field and thus the same stator flux linkage by maintaining the same MMF between healthy and fault-tolerant conditions. The compensation of voltage drops due to the variation of stator resistance [11] was also applied to remove their effect during the fault-tolerant operation as shown in Figure 4. The flow chart in Figure 5 summarize applied post fault compensation methodology.

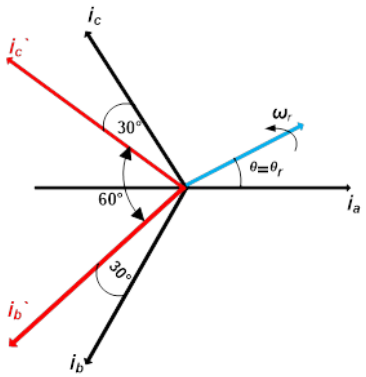


Figure 3: The phasor diagram of the PMSM under OPF

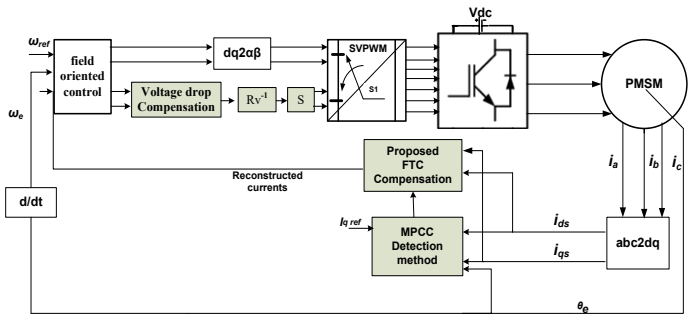


Figure 4: The proposed FTC system for the PMSM drive

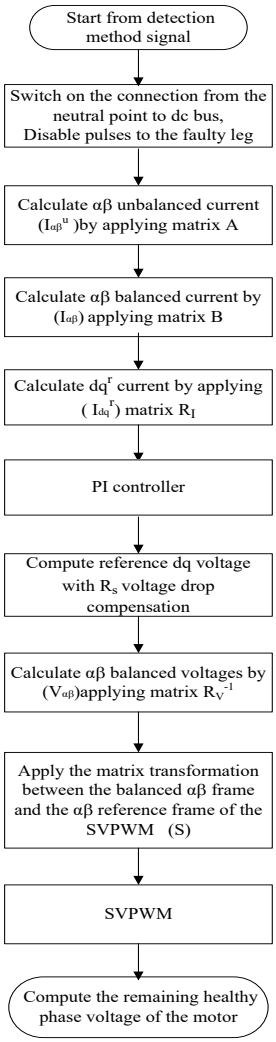


Figure 5: Flow chart of fault compensation control steps

6. Simulation results

In this section, MATLAB simulation results are presented from a MATLAB Simulink model developed for the PMSM drive system in Figure 4 with proposed fault detection and fault-tolerant techniques under different modes of operation (healthy, during an open-phase fault, fault-tolerant compensation). Four cases under the different conditions and with different speed and load conditions as presented in Table 2. The PMSM parameters are listed in Table 1.

Table 1: Three Phase PMSM Parameters

Parameter	Value
Rated power	0.47 kw
Rated torque	2.2 N.m.
Rated current	2 A
Rated speed	1000 rpm
The number of pole pairs (p)	3
Moment of inertia coefficient	0.000198 kg.m^2
Stator Resistance (R_s)	5.95 Ω
d-axis Inductance (L_d)	12.5 mH
q-axis Inductance (L_q)	12.5 mH
The PM flux linkage (λ_m)	0.2955556 Wb

Table 2: Test cases

Cases	Description
A	In this case phase b is disconnected, detection method (MPCC) is disabled; reference speed 1000 rpm, and load torque 0.5 N.m.
B	In this case phase b is disconnected, detection method (MPCC) is disabled, reference speed 500 rpm, and load torque 1 N.m.
C	In this case phase b is disconnected, the detection method (MPCC) enabled, reference speed 400 rpm and load torque 1 N.m.
D	In this case of phase, b is disconnected, detection method (MPCC) is enabled, with transient speed change from 400 rpm to 800 rpm and load torque 1 N.m.

A. phase b is disconnected, detection method (MPCC) is disabled; reference speed 1000 rpm, and load torque 0.5 N.m.

The PMSM is first driven at healthy operating conditions. An OPF at phase *b* is then applied at t=0.1 sec with the fault detection subsystem being disabled for illustrating the error signal that is used for detecting the fault when the integrated fault detection, fault-tolerant compensation technique is fully applied. At t = 0.2 sec, The FTC current regulation technique is allowed to start thus, connecting the neutral point to the mid-point of the DC link and applying the required modification for the current control to return back the magneto-motive force to its value during the healthy operation as explained in section 5. Figure 6 shows the stator currents during the OPF in phase b at t = 0.1 sec, the phase *b* current falls to zero while the other two-phase currents have the same value of the current, but with opposite polarity (phase-shifted 180°). When the FTC compensation is applied at t = 0.2 sec, the currents of the remaining healthy phases is increased in magnitude by 1.73 and shifted 60° apart from each other.

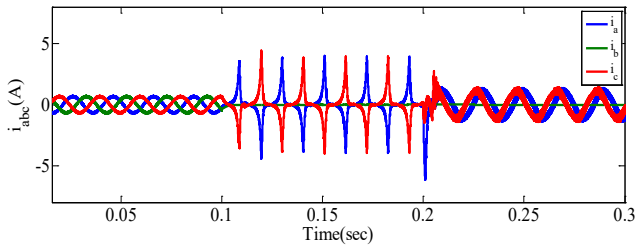


Figure 6: Stator currents before, during the fault and after FTC compensation (case A)

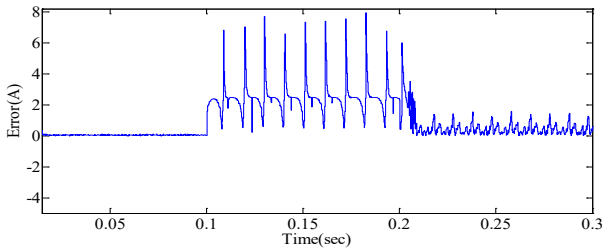


Figure 7: Fault detection error magnitude before, during the fault and after FTC compensation (case A)

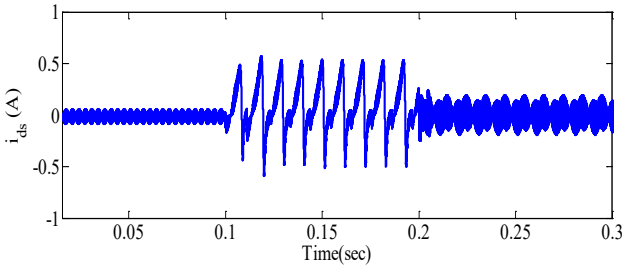


Figure 8: d-axis current before, during the fault and after FTC compensation (case A)

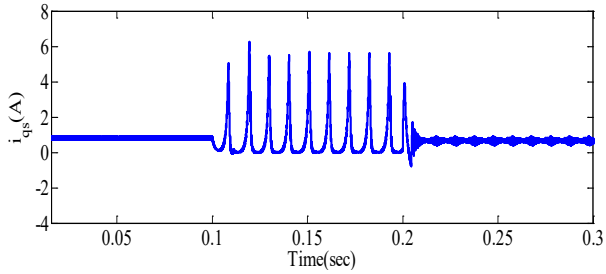


Figure 9: q-axis current before, during the fault and after FTC compensation (case A)

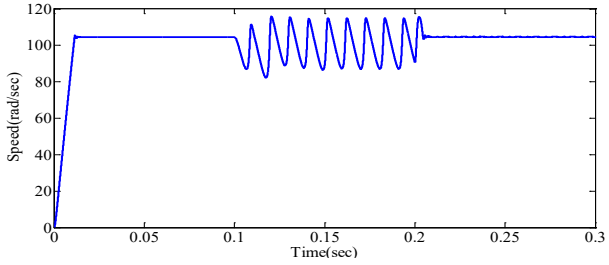


Figure 10: Rotor speed before, during the fault and after FTC compensation (case A)

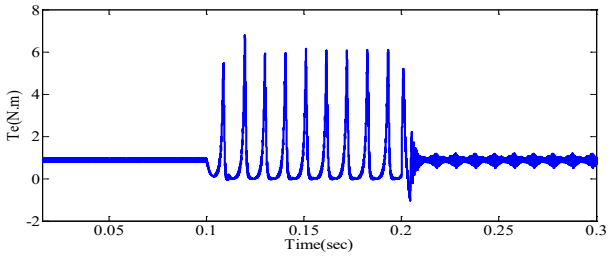


Figure 11: Electromagnetic torque before, during the fault and after FTC compensation (case A)

During the OPF the error magnitude increases due to the OPF (t = 0.1sec – 0.2 sec) then back to the acceptable range after the FTC technique is applied (t = 0.2sec - 0.3sec) as presented in Figure 7. The dq currents, speed, and torque ripples severely increase during the fault due to the OPF operation, then the FTC technique constrained it as shown in Figure 8, Figure 9, Figure 10, and Figure 11.

B. Phase, b is disconnected, detection method (MPCC) is disabled, reference speed 500 rpm, and load torque 1 N.m.

During single-phase operation (OPF occurs at t= 0.1 sec) without enabling the detection technique (MPCC), the system cannot identify the fault and this leads to the PMSM stop and cannot continue its operation. Until the FTC compensation is applied at (t=0.2 sec), PMSM can return back to work with the remaining two phases at t=0.2 sec a Shown in Figures 12, 13 and14.

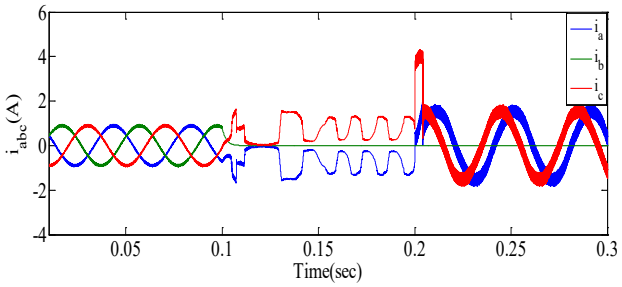


Figure 12: Stator currents before, during the fault and after FTC compensation (case B)

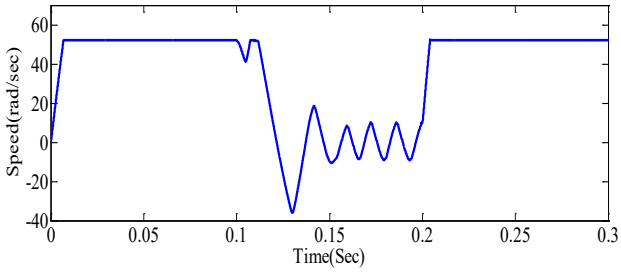


Figure 13: Rotor speed before, during the fault and after FTC compensation (case B)

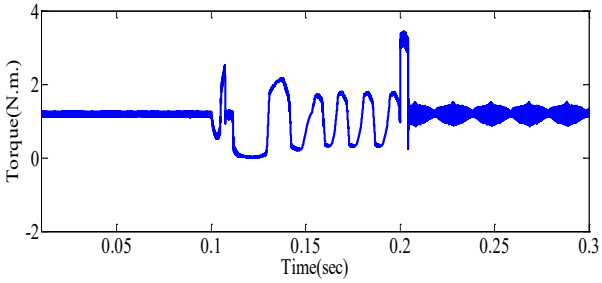


Figure 14: Electromagnetic torque before, during the fault and after FTC compensation (case B)

C. phase b is disconnected, the detection method (MPCC) enabled, reference speed 400 rpm and load torque 1 N.m.

Phase b is opened at 0.1 sec with the fault detection technique (MPCC) enabled. The function of the fault detection method is to detect the OPF and trigger the FTC-compensation for operation. It is obvious from Figures 15 to 20 that the detection of the OPF is only in a few µsec depending on the error indicator. In normal operation (t = 0 - 0.1sec) the error as given in (17) is close to zero, at the instant of fault at (t= 0.1 sec) the error magnitude is increased and the MPCC detects it quickly before the error value becomes high. Once the fault detection takes place it activates the FTC which returns back the error value to an expected range (t=

0.1 - 0.3 sec) as shown in Figure 16. The stator currents are depicted in Figure 15. As shown in Figure 17 and Figure 18, the balance in the resultant MMF can be easily observed by obtaining the same average torque and speed as in the healthy operation at the instant the fault is detected and FTC is activated. i_{ds} and i_{qs} are depicted in Figure 19 and Figure 20, respectively. Comparing case C to case B, it can be concluded that the fast response of the fault detection prevents the drive from halting at low-speed operation.

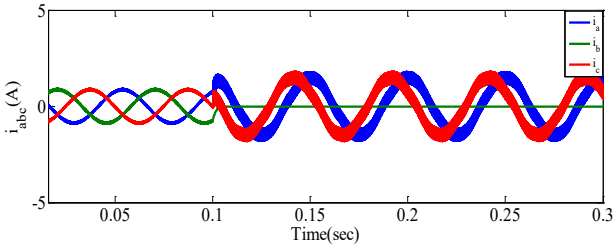


Figure 15: Stator currents in the healthy and post-fault operation

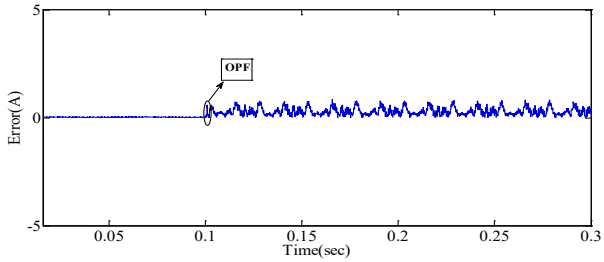


Figure 16: Fault detection error magnitude in the healthy and post-fault operation

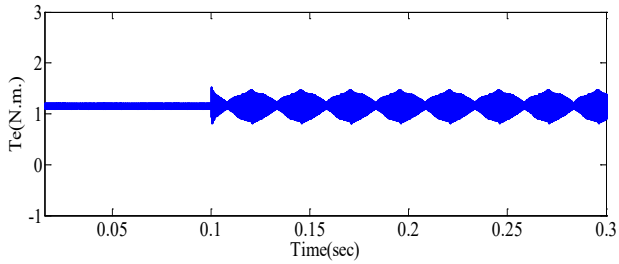


Figure 17: Electromagnetic torque in the healthy and post-fault operation

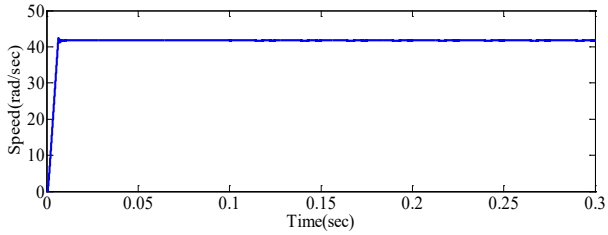


Figure 18: Rotor speed in the healthy and post-fault operation

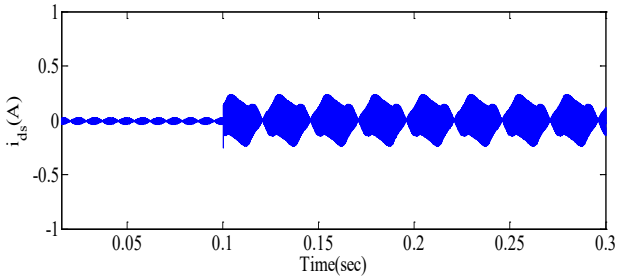


Figure 19: d-axis current in the healthy and post-fault operation

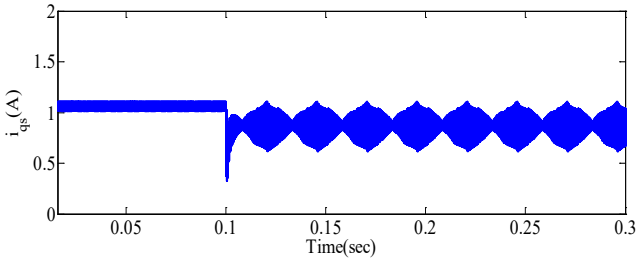


Figure20.: q-axis current in the healthy and post-fault operation

D. phase, b is disconnected, detection method (MPCC) is enabled, with transient speed change from 400 rpm to 800 rpm and load torque 1 N.m.

In this test case, the selective and discriminative performance of the fault detection method is examined under healthy, fault during transient and transients while operating in post-fault conditions. Figures 21-23 show the case where during the healthy operation the speed was changed from 400 rpm to 800 rpm while the detection algorithm being activated. A large error is observed by the fault detector similar to a fault condition however due to the presence of actual and predicted current, the algorithm takes no action. On the other hand, Figures 24-25 illustrate the same case for Figures 21-23 however during the transient change, an OPF of phase b is introduced (at t= 0.053 sec). The fault detection method observed the fault condition and takes the proper action for transferring the operation to the post-fault algorithm allowed to a balanced two-phase maintaining the same MMF as for the healthy condition with minimum torque ripples.

Moreover, the FTC-compensation strategy is tested to prove the robustness of the system during post-fault by changing the reference speed from 400 rpm to 800 rpm as indicated in Figure 26. It is obvious that the PMSM speed is totally able to follow the reference speed. Figure 27 and 28 illustrate the two-phase currents and the electromagnetic torque in the post-fault operation during the steady-state and transient conditions of the step change in the speed.

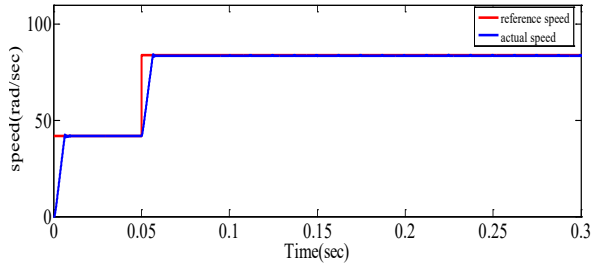


Figure 21: Rotor speed change in the healthy operation

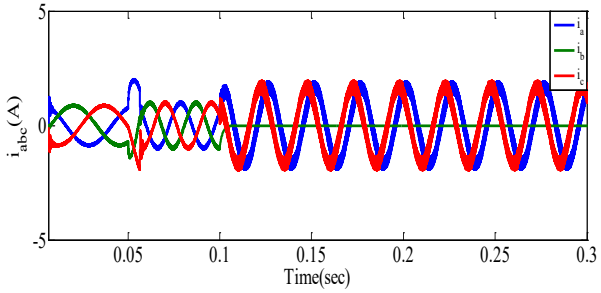


Figure 22: Stator currents in the healthy operation under-speed change

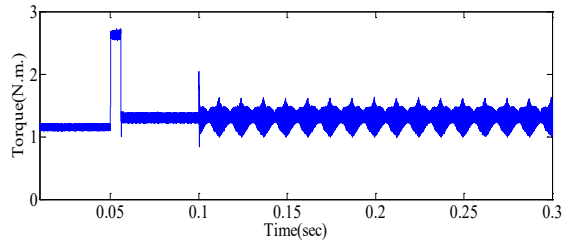


Figure 23: Torque in the healthy operation under-speed change

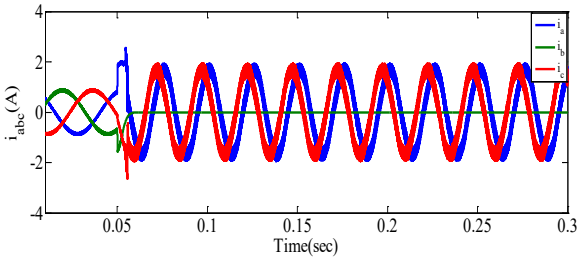


Figure 24: Stator currents under- transient speed change

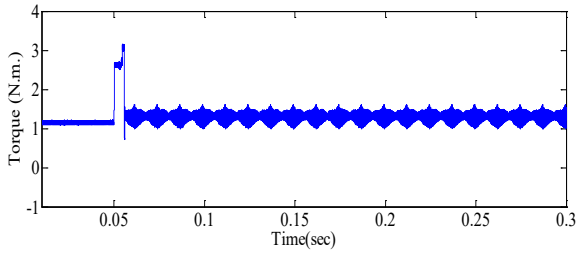


Figure 25: Torque under transient-speed change

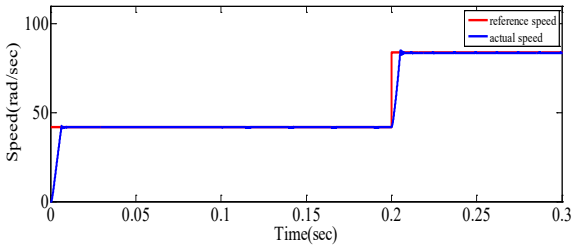


Figure 26: Rotor speed change in the post-fault operation

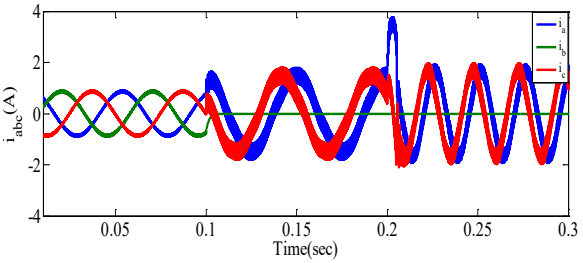


Figure 27: Stator currents in the post-fault operation under speed change

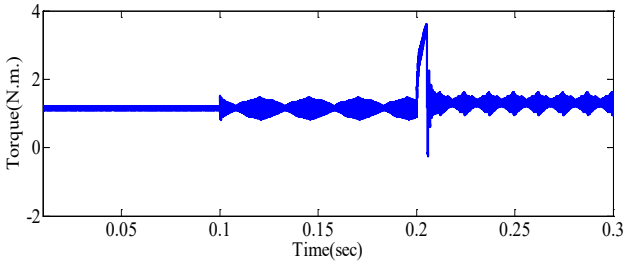


Figure 28: Torque in the post-fault operation under-speed change

7. Conclusion

An integrated fault-tolerant system for open-phase fault in PMSM field-oriented controlled drive has been presented. This drive system consists of a MPCC technique for the fault detection, which is reliable, fast as only a few sampling intervals are sufficient to identify the OPF during steady-state and effectively under transient conditions, together with a fault-tolerant control (FTC) compensation method which utilizes the neutral point to split the single-phase current into two-phase currents that are 60° phase-shifted and equal in magnitude (increased to $\sqrt{3}$ of the healthy phase current) with minimum hardware. This guarantees the balance in the resultant magneto-motive force allowing for the current vectors regulation returning the operation with almost the same average torque.

The simulation results show the smooth transition from the healthy operation to the post fault operation immediately at the time the fault occurs without loss of machine torque and consequently the speed control. Moreover, the fault detection method is robust enough to discriminate between a real OPF and a normal operation transient condition.

The results obtained under different testing conditions show the robustness and precision of the proposed fault detection and fault compensation methods which makes the proposed integrated fault-tolerant is suitable for PMSM drives used in industrial processes to ensure the reliability of the whole system operation even in the event of OPF.

Conflict of Interest

The authors do declare that there is no conflict of interest.

References

[1] F. Betin et al., "Trends in electrical machines control: samples for classical, sensorless, and fault-tolerant techniques," IEE Ind. Electron. Mag., **8**(2), 43–55, 2014, doi:10.1109/MIE.2014.2313752.

[2] H. Guo, J. Xu, and Y.-H. Chen, "Robust control of fault-tolerant permanent-magnet synchronous motor for aerospace application with guaranteed fault switch process," IEEE Trans. Ind. Electron., **62**(12), 7309–7321, 2015, doi:10.1109/TIE.2015.2453935.

[3] Y. Wu, B. Jiang, and N. Lu, "A descriptor system approach for estimation of incipient faults with application to high-speed railway traction devices," IEEE Trans. Syst., Man, Cybern. Syst., **49**(10), 2108-2118, 2019, doi: 10.1109/TSMC.2017.2757264.

[4] W. Wang, J. Zhang, M. Cheng, and S. Li, "Fault-tolerant control of dual three-phase permanent-magnet synchronous machine drives under open-phase faults," IEEE Transactions on Power Electronics, **32**(3), 2052-2063, 2017, doi:10.1109/TPEL.2016.2559498.

[5] J. Hang, J. Zhang, M. Cheng, and S. Ding, "Detection and discrimination of open-phase fault in permanent magnet synchronous motor drive system," IEEE Transactions on Power Electronics, **31**(7), 4697-4709, 2016, doi:10.1109/TPEL.2015.2479399.

[6] N. M. A. Freire, J. O. Estima, and A. J. M. Cardoso, "Open-circuit fault diagnosis in PMSG drives for wind turbine applications," IEEE Trans.Ind. Electron., **60**(9), 3957–3967, 2013, doi: 10.1109/TIE.2012.2207655.

[7] J. O. Estima, N. M. A. Freire, and A. J. M. Cardoso, "Recent Advances in Fault Diagnosis by Park's Vector Approach," in Proc. IEEE WEMDCD, 279–288, 2013, doi: 10.1109/WEMDCD.2013.6525187.

[8] S.-M. Jung, J.-S. Park, H.-W. Kim, K.-Y. Cho, and M.-J. Youn, "An MRAS-based diagnosis of open-circuit fault in PWM voltage-source inverters for PM synchronous motor drive systems," IEEE Trans. Power Electron., **28**(5), 2514–2526, 2013, doi: 10.1109/TPEL.2012.2212916.

[9] A. Kontarček, P. Bajec, M. Nemec, V. Ambrožič and D. Nedeljković, "Cost-effective three-phase pmsm drive tolerant to open-phase fault," IEEE

Transactions on Industrial Electronics, **62**(11), 6708-6718, 2015, doi:10.1109/TIE.2015.2437357.

[10] L. Wang, Shan Chai, Dae Yoo, Lu Gan and Ki Ng., PID and predictive control of electrical drives and power converters using MATLAB/Simulink, IEEE Wiley, 2015.

[11] G. Scarcella, G. Scelba, M. Pulvirenti and R. D. Lorenz, "Fault-tolerant capability of deadbeat-direct torque and flux control for three-phase PMSM drives," IEEE Transactions on Industry Applications, **53**, 2017, doi: 10.1109/TIA.2017.2743070.

[12] X. Zhou, J. Sun, H. Li and X. Song, "High performance three-phase pmsm open-phase fault-tolerant method based on reference frame transformation," IEEE Transactions on Industrial Electronics, **66**(10), 7571-7580, 2019, doi:10.1109/TIE.2018.2877197.

[13] X. Zhou, S. Li, M. Lu, F. Zeng, M. Zhu and Y. Yu, "New fault tolerance method for open-phase PMSM," IEEE Access, **7**, 146416-146427, 2019, doi:10.1109/ACCESS.2019.2946183.

[14] A. Akay, P. Lefley and M. Kansara, "Open-Circuit Fault-Tolerant Control for a Five-Phase Permanent Magnet Synchronous Machine Drive," in 2020 7th International Conference on Electrical and Electronics Engineering (ICEEE), 150-154, 2020, doi:10.1109/ICEEE49618.2020.9102486.

[15] A. Gaeta, G. Scelba, and A. Consoli, "Modeling and control of three-phase PMSMs under open-phase fault," IEEE Trans. Ind. Appl., **49**(1), 74–83, 2013, doi: 10.1109/TIA.2012.2228614.

[16] M. Amin, G. A. A. Aziz, M. N. Ibrahim and P. Sergeant, "Open-Phase Fault-Tolerant Current Reconstruction Control of Three-Phase Permanent Magnet Assisted Synchronous Reluctance Motors," 2019 IEEE Industry Applications Society Annual Meeting, Baltimore, 1-8, 2019, doi: 10.1109/IAS.2019.8912389.

[17] Ho, Ping-Kwong. An exact model and simulation of a permanent magnet synchronous motor system when driving a light load at low speed, Diss. The Hong Kong Polytechnic University, 2000.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 464-471 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Extending the Classifier Algorithms in Machine Learning to Improve the Performance in Spoken Language Understanding Systems Under Deficient Training Data

Sheetal Jagdale[1,*], Milind Shah[2]

[1]*Electronic Engineering Department, Shri. L.R Tiwari College of Engineering, Kanakia Park Mira Road (E) Thane Maharashtra, 401107, India*

[2]*Electronics and Telecommunication Engineering Department, Fr. C. Rodrigues Institute of Technology, Sector 9A, Vashi, Navi Mumbai, 400703, India*

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 31 October, 2020
Online: 20 November, 2020

Keywords:
Spoken Dialogue System (SDS)
Spoken Language Understanding (SLU)
Deficient Training Data
Rarely Visited Domain
Classifier

ABSTRACT

One of the open domain challenges for Spoken Dialogue System (SDS) is to maintain a natural conversation for rarely visited domain i.e. domain with fewer data. Spoken Language Understanding (SLU) is a component of SDS that converts user utterance into a semantic form that a computer can understand. If we scale SDS open domain challenge to SLU then it should be able to convert user utterance to a semantic form even if less data is available for the rarest visited domain. The SLU reported in literature incorporate classifiers for the task of identifying the domain of user utterance, understanding the intent of the user, and filling slots-value pair. Thus, to address open domain challenges, classifiers in SLU must be robust to scarce training data. This paper presents investigations to improve the performance of SLU to convert user utterance into semantic form even if less training data is available. Eleven classification algorithms from machine learning have experimented under deficient data. The evaluation matrices used are accuracy, f-score, and inter cross-entropy. Comprehensive experimentation is carried out on the two publicly available datasets DSTC2 and DSTC3 were carried out.The accuracy for Support Vector Machine (SVM) , Stochastic Gradient Descent (SGD) and Decision tree are 0.940, 0.960 , 0.955 for DSTC2 and 0.916, 0.900, 0.901 for DSTC3 database respectively. The F-score for SVM, SGD and Decision tree are 0.855, 0.868, 0.849 for DSTC2 dataset and 0.725, 0.715, 0.700 for DSTC3 database, respectively. The ICE for SVM and SGD are 1.191,1.100 for DSTC2 dataset and 3.180,2.999 for DSTC3 database, respectively. The performance of SLU based on SVM and SGD was found to be the best among all. The worst performance in terms of all three evaluation metrics was displayed by SLU incorporating Automatic Relevance Determination (ARD) and Relevance Vector Machine (RVC) classifier.

1. Introduction

Spoken Dialogue System (SDS) interacts with the user in the most natural form of communication. It helps the user to achieve a goal through a series of interactions. SDS is moving to the open domain now. In the open domain, SDS has to maintain natural conversation even if data are scarce. Recently a lot of work is reported in the literature for open-domain Spoken Language Understanding (SLU) where it should be able to convert user utterance to semantic form even if less data is available. SLU reported in literature incorporates classifiers at different levels. The first level classifier in SLU will classify utterance to a domain example restaurant, flight booking, and tourist information. The next level classifier will classify user intend within domain example for example finding a restaurant or flight.

Subsequent levels in SLU should fill the slots such as time, date, etc. Filling slots will allow the flow of conversation to achieve the user goal. Thus, the robustness of the classifier to infer the output even if training data is less is important. There is a need for investigation of the robust classifier for SLU, which can convert user utterance to semantic form even if less training data is available. The work in the paper investigates a robust classifier. This research is broadening of work initially reported in the conference paper [1] The work is extended by incorporating CNN

*Corresponding Author: Sheetal Jagdale, Email: sheetal.jagdale2@gmail.com

for feature extraction and investigating eleven classifiers from literature. Comprehensive experimentation was done on the DSTC2 dataset and the DSTC3 dataset. The evaluation was done on two domain datasets. DSTC2 datasets for the restaurant domain and DSTC3 datasets for the tourist domain were used for evaluation.

Classification is an important task in many applications as image processing, spoken dialogue systems, etc. Choosing the right classifier for an application is a very important task. A lot of work is reported in the literature for investigating a robust classifier for a particular application. The work [2] has predicted error rates of k-neighborhood, SVM, and decision tree classifiers for the application of computer-aided diagnostic (CAD). The evaluation of the classifiers is done on a small training set. Later the work was extended to three tasks in CAD. The task was detecting cancer, identifying high or low-grade and-later the work was extended by adding more features [3] Another work investigates classifiers for text classification [4] The Paper investigates Rocchio Classification, K Nearest Neighbor, Naïve Bayes Classification, SVM Classification, and Regression-Based Classification for the task of classifying text. The paper concludes with a study stating that the right choice of dimension reduction algorithm and classifier can improve the performance of text classification. One more paper [5] refers to the classifier investigation for the task of text classification under limited data. A naive Bayes classifier is also investigated for text classification [6] The work combines the dimension reduction method with a classification algorithm for text classification under limited training data. The study was reported in [7] Paper [8] tackles the issue of imbalance data under training data. They tackle the issue by a stable space for testing data proposed using a metric algorithm. The effect of training size on the neural network classifier is also demonstrated in the literature [9] Recently a work [10] reported classifiers for clinical treatment. The study also includes data augmentation techniques along with classifiers Thus, a lot of classification algorithm investigation for different applications is done. This paper investigates a robust classifier for the slot filling task of SDS.

To improve SLU's performance for classifying utterance to predefined slots, different classification algorithms from machine learning were investigated. The purpose of the work is to:

1) Extend a robust classifier algorithm in machine learning to the semantic decoder in spoken language understanding for the task of converting user utterance into a semantic form of dialogue act and slot-value pair.
2) To demonstrate the impact of data scarcity on SLU incorporating different classifiers.
3) Comprehensive experiment on SLU incorporating eleven classifiers on two public datasets DSTC2 and DSTC3. The metrics used for evaluation are accuracy, f-score, and item cross-entropy.

2. Related Background

2.1. Decision Tree [DT]

DT is a popular classifier and is used in various machine learning applications such as data mining, image processing, statistics, etc. It uses if and then rules and follow the top-down method. In paper [11] decision tree is used for mining applications. A decision tree is used to extract information from huge data. Another work discusses the framework for decision tree algorithm, splitting criteria, and pruning methods [12] The paper also discusses extended decision trees such as obvious tree s decision trees, incremental tree, tree inducers, and fuzzy trees. A decision is also used as an inductive inference system [13] The decision tree algorithm consists of a set of rules called a decision. It extracts knowledge from example. Thus, it resembles the machine learning principle of learning from examples. In a research work presented in paper [14], the IDA decision algorithm is proposed. The ID3 decision algorithm is also discussed and compared with IDA.ID3 algorithm uses entropy for finding the uncertainty of classification output. While IDA uses divergence for classification. As reflected in the results, the IDA algorithm performance displays its effectiveness over ID3. Another paper [15] proposes an improved ID3 algorithm. It incorporates the Taylor series in ID3 to improve the classifier. One of the popular applications of the decision tree is data mining [16]

2.2. Multilayer Perceptron [MLP]

One of the most powerful classifiers used in machine learning is the MLP classifier. It uses a supervised learning and feedforward network. One of the applications reported for MLP is for recognizing a number in the Bengali language [17] For numbers recognizing pre-processing techniques such as feature, the selection is also incorporated. Another application reported is for the inference of liver cirrhosis [18] However, it is very difficult to tune the parameter of a multilayer perceptron. To overcome this difficulty a methodology is proposed in the paper [19] The methodology can find the number of epochs to get the best performance which is optimal.

2.3. Gaussian naïve Bayes

Gaussian naive Bayes classifier is a probabilistic classifier applied to many applications. Bayes theorem is used for classification. One application is discussed in [20] for breast cancer classification. A framework consisting of data pre-processing, feature selection, and classifier is proposed for breast cancer classification.

2.4. Gaussian classifier

Gaussian classifiers are used in many applications such as language identification [21], spectrum recognition [22] and dialogue representation [23] The work in paper [24] discusses the parametric Gaussian classifier. In this work, a Gaussian classifier incorporating a neural network is discussed. The results demonstrate the effectiveness of the parametric approach applied to Gaussian distribution. The Gaussian classifier is also analyzed for limited training data [25] The work discusses the impact of limited data on the decision boundary of a classifier.

2.5. AdaBoost classifier

In literature AdaBoost are used for detecting objects[26] ,detecting faults in gearbox [27] and detecting hands [28] In another paper, AdaBoost is used in mobile intelligent terminals for security level classification [29] Various modified AdaBoost algorithms are were analyzed for classification margin[30]

Further, in the same paper, generalized error for eight variants of AdaBoost was reported. Personalized AdaBoost which is a variant of AdaBoost as reported in the paper has the least generalization error

2.6. Relevance vector machine

Relevance vector machine (RVM) is a probabilistic model used for classification and regression. The prediction of the output of RVM is probabilistic. RVM is used for the classification of hyperspectral images using wavelet kernel [31] As reported in the paper, training speed was improved by RVC incorporating wavelet kernel. Another application was reported for RVM for the detection of clustering in mammogram [32] For mammogram classification, RVM was able to maintain better classification with less computation complexity

2.7. Stochastic gradient (SGD)

It is a linear classifier that incorporates SGD learning. These classifiers are used in classifying emotions in Hindi text [33] In the Hindi emotion classification model, first pos tagging is done on the text and then features are extracted from the text. Extracted features are used to train the SGD classifier or random forest classifier. The results were improved by incorporating reduced features to SGD. Another application for SGD in literature is for the categorization of Bangla text [34] The results for Bangla text classification was improved using SGD.

2.8. Support Vector Machine [SVM]

It is a linear non-probabilistic classifier. It uses Platt's model to predict the output. One of the applications in literature is big data automated selection [35] The quality of data affects the classification performance of SVM. Results in the paper showed that optimizing the SVM parameter can improve the performance. SVM is also used for skin disease classification [36] In this work, the Image was flattened to get an array of pixel intensities. This array of labeled data was used to train SVM. Recently SVM using field programmable gate array is used for knowledge discovery.

2.9. K-nearest neighborhood [KNN]

It is a supervised classifier based on distance measurement. It is widely applied to data mining for the classification task.In recent work[37], KNN is used for saving energy for the home. KNN is used here for appliance classification. The KNN classifiers were able to detect home appliances in a short time. Lot variants of KNN such as dynamic KNN, weight-adjusted KNN and distance weighted KNN are proposed to speed up the learning process [38]

2.10. Random Forest [RF]

It is an algorithm based on a decision tree. In a paper [39], RF is used for Android malware detection applications. The framework proposed in the paper was evaluated on the android feature dataset. RF was used for the classification task of identifying applications as benign or malicious. Misclassification error for classifying malware was least for the random forest. Recently in a paper, RF is used for identifying and classifying applications for network traffic detection. The algorithm used for comparison were random forest, SVM, and logistic regression. RF outperformed all the classifiers in terms of accuracy and f-score.

2.11. Variational Relevance Vector Machine [VRVM]

A variant of RVM is proposed in paper [40] known as variational relevance vector machine. This method is RVM with variational inference. VRVM is applied for RVM multi-dimension array classification. Another application for VRVM is discussed in the literature for the classification of table data such as blocks of image [41] VRVM was found more robust then RVM.

3. Block diagram of SLU

SLU is an important component of the SDS system which converts user utterance to a semantic form consisting of dialogue act and slot value pair. Figure 1 shows the block diagram of SLU.

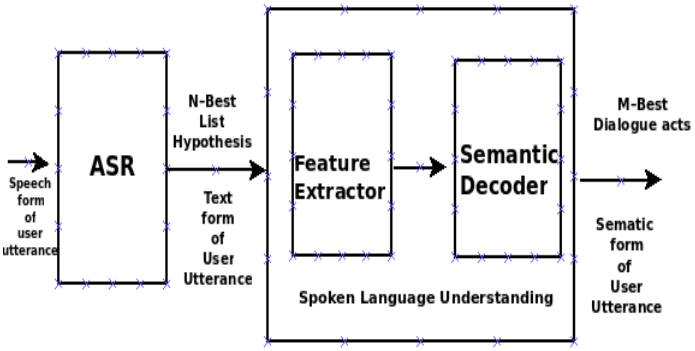


Figure 1: Structure of SLU

SLU plays a very important role in understanding user intent. User utterance is given to the Automatic Speech Recognition (ASR) system which converts speech into text form. The text input is given to SLU. It extracts features from text utterance. The features are fed to a set of classifiers to extract dialogue act and slot-value pair from features. The dialogue act and slot-value pair are combined to form a semantic representation of user utterance. Semantic representation is further analyzed to understand user utterance.

4. Implementation of SLU

Figure 2 represents a semantic tuple classifier for the task of semantic classification which is described in the paper [42]

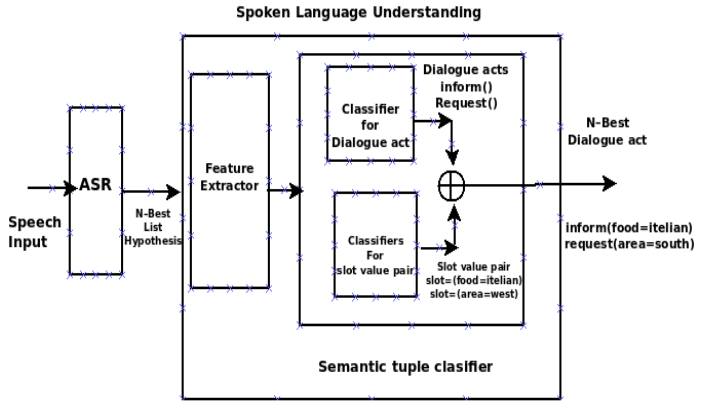


Figure 2: SLU incorporating classifiers for dialogue act and slot-value pair

The speech input is given to ASR. The ASR converts speech input to text form which is the N-best list. The features are extracted from the text. The feature extraction method in paper [42] is extended by incorporating CNN. The deep extracted feature is fed to the classifier. The first classifier classifies user utterance to dialogue act such as to inform (), request (), and bye (). This will help to understand the goal of the user. The next set of classifier identifies the slot-value pair for user utterance. The semantic form of user utterance consists of dialogue act and slot- value pair. Next, eleven classifiers from machine learning are investigated for dialogue act classification and to find slot–value pair. The classifiers investigated on SLU are support vector machine, stochastic gradient descent classifier, AdaBoost classifier, decision tree, relevance vector classifier, random forest, *k*-nearest neighbor classifier, multi-layer perceptron classifier, automatic relevance determination classifier, Gaussian naive Bayes classifier, and variational relevance vector machine. The experimentation was performed with a different amount of training data. The performance of SLU was studied and compared incorporating different classifiers under limited training data on DSTC2 and DSTC3 databases. The experimentation for measuring robustness on limited training data was conducted by training the model on 25% training data and then testing the model on entire testing data. Later experimentation was repeated for 50% training data, 75% of training data, and 100% training data. The evaluation was done on metric Inter Cross-Entropy (ICE), accuracy, and F-score.

5. Results and Discussion

5.1. Datasets

Experimentation is done on two public datasets DSTC2 and DSTC3. The DSTC2 dataset for the restaurant domain. It consists of 1765 dialogues for training data, 441 validation data and 1117 dialogue for testing data. The DSTC3 dataset is for the tourist domain. It consists of a total of 2265 dialogues. The training data used for the experiment is 1265 dialogues and 250 dialogues for validation. The testing data for the experiment is 750 dialogues.

5.2. Evaluation metrics

The framework is evaluated with three metrics: Inter Cross-Entropy (ICE), accuracy, and F-score. The ICE is cross-entropy between the probability density from the confidence and the optimal density given by delta functions at the correct values [43] In the paper, it was demonstrated ICE gives a consistent performance ranking for both the confidence scores and the overall correctness of the system. A lower ICE metric indicates good dialogue confidence performance. Accuracy and F-score are in range 0 to 1 and higher values indicate good performance. ICE metric is highly correlated with the performance of SDS. Accuracy and F-score will reflect the performance of the SLU model. Thus for the evaluation of model in the experiment is done with ICE, f-score, and accuracy.

5.3 Baseline Method

The framework used for experimentation is an extension of the semantic decoder [42] The decoder is extended by incorporating CNN for feature extraction. The extended framework experimented with incorporating eleven different classifiers from machine learning. The framework was evaluated on DSTC2 and DSTC3 datasets. The framework was also investigated for robustness for less training data.

5.4 Experimental setting 1

The first set of experimentation was carried out on the DSTC2 dataset. The experiment was conducted on a deep semantic decoder by incorporating 11 different classifiers. For each classifier, the robustness was evaluated by training the model with different quantity of training data and then testing using the entire set. The different quantity of training data was 1765 dialogues (100% training data), 1323 dialogues (75% data), 882 dialogues (50% training data), and 441 dialogues (25% training data). Testing of SLU for all different quantity of training data was done on 1117 testing dialogues. The classifiers investigated on SLU are SVM, SGD, AdaBoost, DT, RVC, MLP, RF, *k*-neighbourhood, ARD, Gaussian naive Bayes, and VRVM.

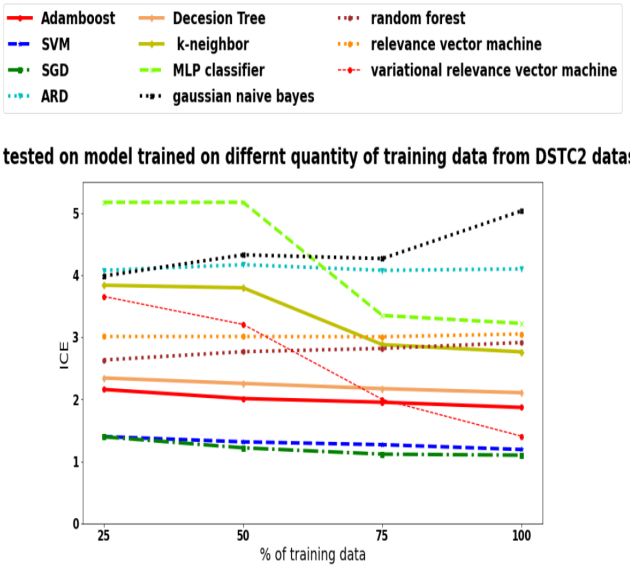


Figure 3: ICE for a model trained on 25% training data, 50% training data, 75% training data, and 100% DSTC 2 training data.

Figure 3 shows the experimental results for ICE measured on the dSTC2 dataset for different quantity of training data. Results show ICE is directly proportional to training data. Almost all classifier ICE was high when training data was less. This indicates the overall quality of distribution was least for 20% of training data and most when 100% training data. Their set of classifiers displaying the best performance are SVM, SGD, and VRVM at 100% data. SVM and SGD displayed the least variation in ICE at the different quantity of training data. The variation in VRVM and MLP was most for different quantity of training data. The set of classifiers displaying the worst performance are ARD, RVC, and MLP. The classifier which had an overall good performance for all quantity of training data and full data is SGD.

The testing accuracy results of the experiment on the DSTC2 dataset are shown in Figure 4. The measurement accuracy is for correct classification of user utterance to semantic form consisting of dialogue act and slot value pair. The classifiers whose topmost semantic hypothesis was correct are SVM, VRVM, decision tree, and SGD. These classifiers were also able to predict the semantic hypothesis for testing data even if the training data was less.

VRVM was not able to maintain its performance for limited training data. Thus SVM, decision tree, and SGD were able to maintain its performance even if training data was scarce. The classifiers displaying the least performance are ARD, RVC, and random forest. The VRVM's testing accuracy measured for 20% training data is 0.827 and for training 100% data it is 0.939. Thus, it is not able to maintain its performance under a scarcity of data. Similarly, MLP and k– neighborhood classifiers were also not able to maintain its performance under data scarceness.

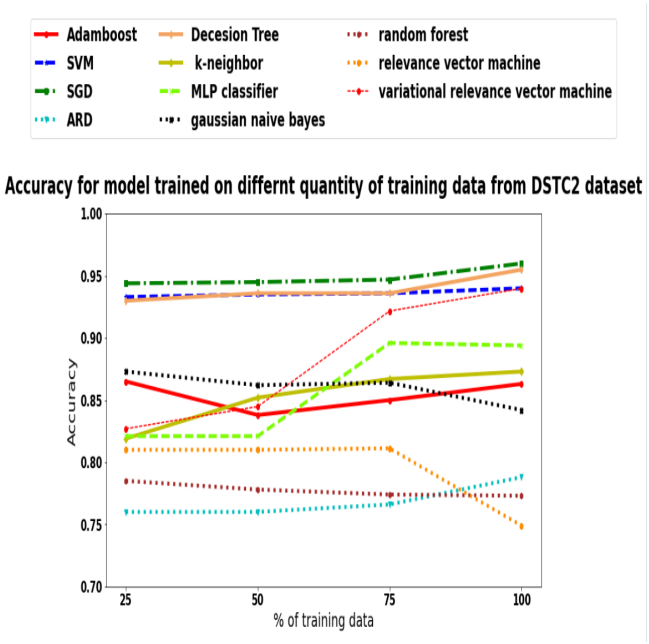


Figure 4: Testing data accuracy for a model trained on 25% training data, 50% training data, 75% training data, and 100% DSTC 2 training data.

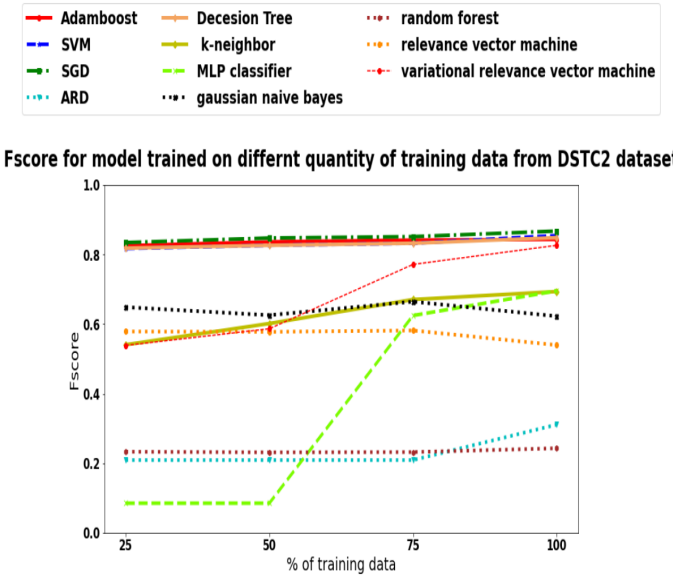


Figure 5: Testing data F-score for a model trained on 25% training data, 50% training data, 75% training data, and 100% DSTC 2 training data.

F-score for testing data DSTC2 is presented in Figure.5. A high value of the F-score will indicate a low false-negative and positive value of the semantic hypothesis. A high F-score was displayed by SLU incorporating Decision Tree, SVM, VRVM, and SGD. SLU's robustness for limited data was increased with these classifiers. VRVM again failed to show robustness for limited data. Low F-score was displayed by random forest and RVC. SLU incorporating MLP also had a lot of variation in F-score for different quantity of training data.

5.5. Experimental setting 2

The second set of experiments was carried out on DSTC3. The different quantity of training data was 1265 dialogues (100% training data), 925 dialogues (75% training data), 600 dialogues (50% training data), and 300 dialogues (25% training data). The testing data consist of 750 dialogues. The eleven classifiers included in an experiment setting 1 is also included in experiment setting 2 as well.

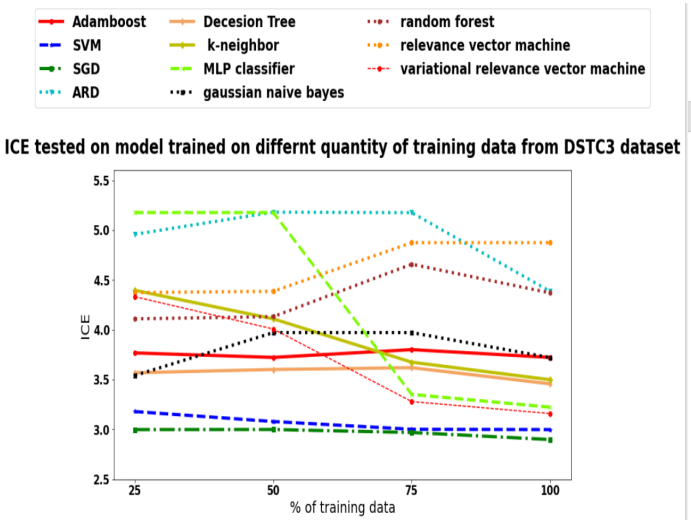


Figure 6: ICE for a model trained on 25% training data, 50% training data, 75% training data, and 100% DSTC 3 training data.

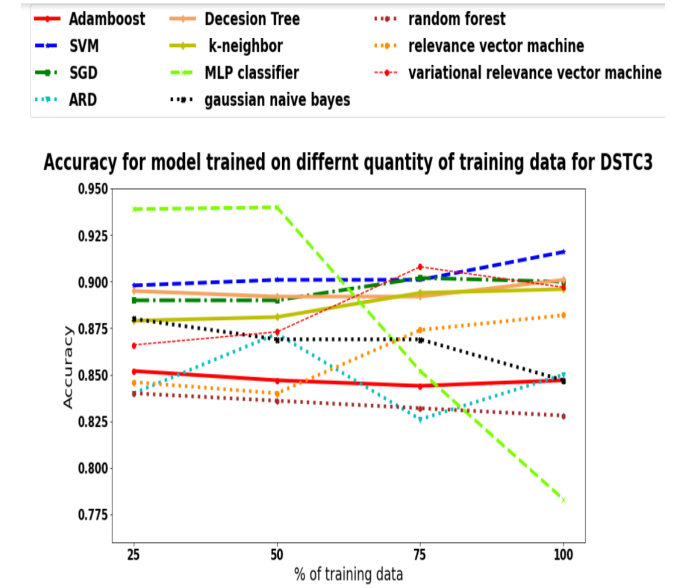


Figure 7: Testing data accuracy for a model trained on 25% training data, 50% training data, 75% training data, and 100% DSTC 3 training data

Figure 6 shows the ICE readings for the different quantity of training data for DSTC3 datasets. The results are similar as found for the DSTC2 experiment. The SLU incorporating the set of

classifiers displaying good results are SVM, SGD, and VRVM. Variation in ICE for different quantity of training data was minimum for SVM and SGD. Variation in ICE was more in VRVM. The SLU incorporating a set of classifiers displaying the worst results are ARD, RVC, and RF. The maximum variation in ICE was displayed by MLP, RVC, and K-neighbourhood.

Testing accuracy for the SLU model for different quantity of training data for the DSTC3 dataset is shown in Figure 7. Again, the results are in line with DSTC2 results. The classifiers which were able to classify user utterance to semantic form most accurately were SVM, SGD, and decision tree. The classifiers performing well under data scarcity are SVM, SGD, and decision tree. The worst performance in terms of data scarcity was showed by the MLP classifier. In terms of accuracy worst performance was shown by ARD, RF, and RVC.

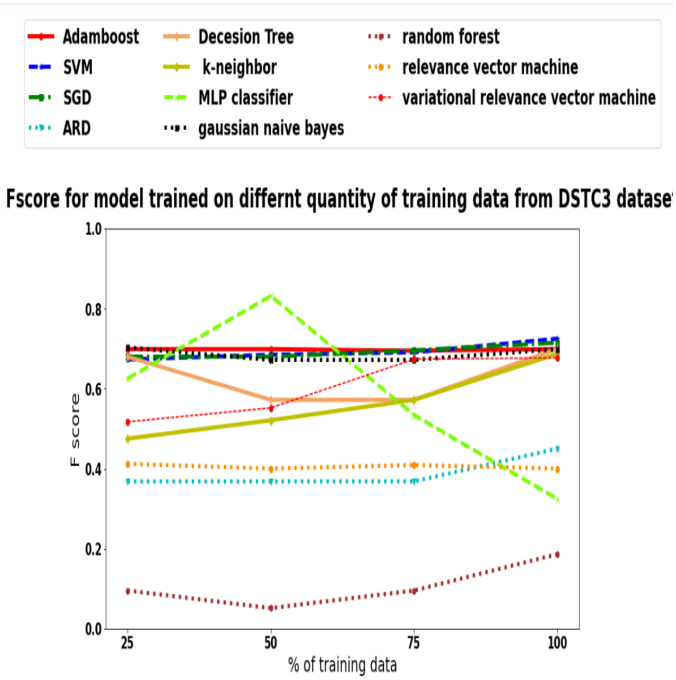


Figure 8: Testing data F-score for a model trained on 25% training data, 50% training data, 75% training data, and 100% DSTC 3 training data.

The results for SLU model performance on the DSTC3 dataset are shown in Figure 8. The results are again similar to the DSTC2 dataset. F-score was high for SLU when incorporated SGD, SVM, VRVM, and decision tree. Low F-score was displayed by SLU when incorporated RVC, RF, and ARD.

Variation in ICE for SGD and SVM when SLU is trained for least and highest train data is 0.296 and 0.209 for the DSTC2 dataset and 0.190 and 0.190 for the DSTC3 dataset, respectively. The least variation in ICE was found for SVM and SGD. Thus, the set of classifiers showing robustness under limited training data for ICE metrics are SVM and SGD. Variation in ICE for MLP and Gaussian naive Bayes when SLU is trained for least and highest train data is 1.95 and 1.45 for DSTC2 and 2.251 and 0.18 for DSTC3 dataset, respectively. The highest variation in ICE was found for MLP and Gaussian naive Bayes. Thus set of classifiers showing the worst result for ICE are MLP and Gaussian naive Bayes.

Table 1: Results on DSTC2 dataset

S.No	Classifier	ICE	Accuracy	F-score
1	Adamboost	1.861	0.863	0.848
2	SVM	1.191	0.940	0.855
3	SGD	1.100	0.960	0.868
4	ARD	4.02	0.788	0.311
5	Decision Tree	2.107	0.955	0.849
6	k-neighbor	2.765	0.873	0.694
7	MLP classifier	3.224	0.894	0.696
8	gaussian naive Bayes	5.03	0.842	0.623
9	random-forest	2.916	0.773	0.243
10	RVC	3.053	0.799	0.540
11	VRVM	1.405	0.939	0.827

Table 1 shows the results for the DSTC2 dataset. Table 2 shows results for the DSTC3 dataset for the tourist domain. Comparison of accuracy, F-score and ICE from table 1 and table 2 can be studied.

Table 2: Results on DSTC3 dataset

S.No	Classifier	ICE	Accuracy	F-score
1	Adamboost	3.723	0.847	0.699
2	SVM	3.180	0.916	0.725
3	SGD	2.999	0.900	0.715
4	ARD	4.380	0.850	0.450
5	Decision Tree	3.458	0.901	0.700
6	k-neighbor	3.499	0.896	0.687
7	MLP classifier	3.780	0.783	0.324
8	gaussian naive Bayes	3.721	0.847	0.699
9	random forest	4.179	0.828	0.186
10	RVC	4.375	0.884	0.400
11	VRVM	0.311	0.897	0.677

The results from table 1 and table 2 are analyzed to find the classifier performing best and worst. The lowest ICE results are obtained by SLU incorporating SGD and SVM when evaluated on both DSCT2 and DSTC3 datasets. The highest accuracy is obtained by SLU incorporating SVM and SGD for DSTC2 and DSTC3. The highest F-score is obtained by SLU incorporating SVM and SGD for DSTC2 and DSTC3.

6. Conclusion

The work extends the classification algorithm in machine learning to the semantic decoder in SLU. Experimental results on both the DSTC2 database for the hotel domain and the DSTC3 database for the tourist domain had similar findings. The SLU's robustness for data scarcity was measured by three evaluation

metrics: accuracy, ICE, and F-score. The SLU model was trained at 25% data (least training data), 50% data, 75% data, and 100% data (highest training data). The work indicates three important findings from the results obtained from experimentation. The first finding is that both the databases are displaying variation in SLU performance for different classifiers indicating SLU performance is influenced by the selection of proper classifiers. The second finding is that there is variation in ICE for least and maximum training data for all the classifiers indicating ICE is highly influenced by the quantity of training data used to train the SLU model. The third finding is the amount of variation in ICE displayed by each classifier is different indicating the robustness of SLU to work under data scarceness is influenced by the right choice of the classifier.

Variation in ICE, accuracy and F-score for SLU incorporating SVM is 0.209, 0.007, 0.035 for DSTC2 database and 0.190, 0.018, 0.052 for DSTC3 database, respectively. Variation in ICE, accuracy, and F-score for SLU incorporating SGD is 0.296, 0.016,0.033 for DSTC2 database and 0.190, 0.010, 0.035 for DSTC3 database, respectively. Thus, the overall performance of SLU, maintaining good performance for all the three evaluation metrics, is by incorporating SVM and SGD. The overall low performance is displayed by SLU incorporating ARD, RVC, MLP, Gaussian naive Bayes, and RF.

References

[1] S. Jagdale, M. Shah, "Evaluation of Stochastic Gradient Descent and Spoken Language Understanding," 2019 International Conference on Nascent Technologies in Engineering (ICNTE), 1–4, 2019, DOI:10.1109/ICNTE44896.2019.8946073

[2] A. Torres-Rodríguez, D. Morillón-Gálvez, D. Aldama-Ávalos, V.H. A. Basavanhally, S. Doyle, A. Madabhushi, "Predicting classifier performance with a small training set: Applications to computer-aided diagnosis and Rutgers, The State University of New Jersey Department of Biomedical Engineering," Biomedical Engineering, **1**, 229–232, 2010, DOI:10.1109/ISBI.2010.5490373

[3] A. Basavanhally, S. Viswanath, A. Madabhushi, "Predicting classifier performance with limited training data: Applications to computer-aided diagnosis in breast and prostate cancer," PLoS ONE, **10**(5), 1–18, 2015, DOI:10.1371/journal.pone.0117900.

[4] V.K. Vijayan, K.R. Bindu, L. Parameswaran, "A comprehensive study of text classification algorithms," 2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017, 2017-Janua, 1109–1113, 2017, DOI:10.1109/ICACCI.2017.8125990

[5] G. Forman, I. Cohen, "Learning from little: Comparison of classifiers given little training," Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 3202, 161–172, 2004, DOI:10.1007/978-3-540-30116-5_17.

[6] A.S. Altheneyan, M.E.B. Menai, "Naïve Bayes classifiers for authorship attribution of Arabic texts," Journal of King Saud University - Computer and Information Sciences, **26**(4), 473–484, 2014, DOI:10.1016/j.jksuci.2014.06.006

[7] S. Pournouri, S. Zargari, B. Akhgar, "An Investigation of Using Classification Techniques in Prediction of Type of Targets in Cyber Attacks," Proceedings of 12th International Conference on Global Security, Safety and Sustainability, ICGS3 2019, 202–212, 2019, DOI:10.1109/ICGS3.2019.8688266.

[8] N. Wang, X. Zhao, Y. Jiang, Y. Gao, "Iterative metric learning for imbalanced data classification," IJCAI International Joint Conference on Artificial Intelligence, 2018-July, 2805–2811, 2018, DOI:10.24963/ijcai.2018/389

[9] G.M. Foody, M.B. McCulloch, W.B. Yates, "The effect of training set size and composition on artificial neural network classification," International Journal of Remote Sensing, **16**(9), 1707–1723, 1995, DOI: 10.1080/01431169508954507.

[10] S.S. Yadav, S.M. Jadhav, "Deep convolutional neural network-based medical image classification for disease diagnosis," Journal of Big Data, 6(1), 2019, DOI: 10.1186/s40537-019-0276-2.

[11] H. Sharma, S. Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining," International Journal of Science and Research (IJSR), **5**(4), 2094–2097, 2016, DOI:10.21275/v5i4.nov162954

[12] N. Patel, S. Upadhyay, "Study of Various Decision Tree Pruning Methods with their Empirical Comparison in WEKA," International Journal of Computer Applications, **60**(12), 20–25, 2012, DOI: 10.5120/9744-4304.

[13] J.R. Quinlan, "Induction of decision trees," Machine Learning, **1**(1), 81–106, 1986, DOI: 10.1007/bf00116251.

[14] P.L. Tu, J.Y. Chung, "A new decision-tree classification algorithm for machine learning," Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, 1992-Novem, 370–377, 1992, DOI:10.1109/TAI.1992.246431

[15] H. Sharma, S. Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining," International Journal of Science and Research (IJSR), **5**(4), 2094–2097, 2016, DOI:10.21275/v5i4.nov162954.

[16] H. Sharma, S. Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining," International Journal of Science and Research (IJSR), 5(4), 2094–2097, 2016, DOI:10.21275/v5i4.nov162954

[17] S. Basu, N. Das, R. Sarkar, M. Kundu, M. Nasipuri, D.K. Basu, "An MLP based approach for recognition of handwritten 'Bangla' numerals," Proceedings of the 2nd Indian International Conference on Artificial Intelligence, IICAI 2005, (March), 407–417, 2005.

[18] M. Abdar, N.Y. Yen, J.C.S. Hung, "Improving the Diagnosis of Liver Disease Using Multilayer Perceptron Neural Network and Boosted Decision Trees," Journal of Medical and Biological Engineering, **38**(6), 953–965, 2018, DOI: 10.1007/s40846-017-0360-z.

[19] T. Windeatt, "Ensemble MLP classifier design," Studies in Computational Intelligence, **137**(5), 133–147, 2008, DOI:10.1007/978-3-540-79474-5_6.

[20] B.M. Gayathri, C.P. Sumathi, "An Automated Technique using Gaussian Naive Bayes Classifier to Classify Breast Cancer," International Journal of Computer Applications, **148**(6), 16–21, 2016

[21] R. Cordoba, L.F. D'Haro, F. Fernandez-Martinez, J.M. Montero, R. Barra, "Language identification using several sources of information with a multiple-Gaussian classifier," International Speech Communication Association - 8th Annual Conference of the International Speech Communication Association, Interspeech 2007, **3**(May 2014), 1913–1916, 2007

[22] P. Guo, Y. Jia, M.R. Lyu, "A study of regularized Gaussian classifier in high-dimension small sample set case based on MDL principle with application to spectrum recognition," Pattern Recognition, **41**(9), 2842–2854, 2008, DOI:10.1016/j.patcog.2008.02.004.

[23] M. Morchid, R. Dufour, P.M. Bousquet, M. Bouallegue, G. Linares, R. De Mori, "Improving dialogue classification using a topic space representation and a Gaussian classifier based on the decision rule," ICASSP, IEEE International Conference on Acoustics, Speech, and Signal Processing - Proceedings, 126–130, 2014, DOI:10.1109/ICASSP.2014.6853571.

[24] H. El Sorady, A. Shoukry, S. Bassiouny, "Improving the parametric Gaussian classifier using neural networks," National Radio Science Conference, NRSC, Proceedings, 1996, DOI:10.1109/nrsc.1996.551118.

[25] C. Lee, E. Choi, B. Baek, C. Yoon, "Analyzing the Gaussian ML classifier for limited training samples," International Geoscience and Remote Sensing Symposium (IGARSS), **5**(C), 3229–3232, 2004, DOI:10.1109/igarss.2004.1370389

[26] Y.G. Zhao, F. Zheng, Z. Song, "Hand Detection Using Cascade of Softmax Classifiers," Advances in Multimedia, 2018, 2018, DOI:10.1155/2018/9204854

[27] D. Song, C. Lu, J. Ma, "Gearbox fault diagnosis based on VMD-MSE and AdaBoost classifier," Vibroengineering Procedia, **14**, 120–125, 2017, DOI:10.21595/vp.2017.19242.

[28] M.L.R. Chandra, B.V. Kumar, B. Sureshbabu, "IoT enabled home with smart security," 2017 International Conference on Energy, Communication, Data Analytics, and Soft Computing, ICECDS 2017, 1193–1197, 2018, DOI:10.1109/ICECDS.2017.8389630.

[29] F. Wang, D. Jiang, H. Wen, H. Song, "AdaBoost-based security level classification of mobile intelligent terminals," Journal of Supercomputing, 75(11), 7460–7478, 2019, DOI:10.1007/s11227-019-02954-y.

[30] S. Wu, H. Nagahashi, "Analysis of generalization ability for different AdaBoost variants based on classification and regression trees," Journal of Electrical and Computer Engineering, 2015, 2015, DOI:10.1155/2015/835357.

[31] C.H. Zhao, Y. Zhang, Y. Wang, "RVM classification of hyperspectral image based on wavelet kernel function," 2011 International Conference on Multimedia Technology, ICMT-2011, 3647–3650, 2011, DOI:10.1109/ICMT.2011.6002159.

[32] L. Wei, Y. Yang, R.M. Nishikawa, M.N. Wernick, A. Edwards, "Relevance vector machine for automatic detection of clustered micro calcifications," IEEE Transactions on Medical Imaging, **24**(10), 1278–1285, 2005, DOI:10.1109/TMI.2005.855435.

[33] K. Pal, B. V. Patel, Emotion Classification with Reduced Feature Set SGD Classifier, Random Forest, 2020, DOI: 10.1007/978-981-15-6648-6.

[34] F. Kabir, S. Siddique, M.R.A. Kotwal, M.N. Huda, "Bangla text document categorization using Stochastic Gradient Descent (SGD) classifier," Proceedings - 2015 International Conference on Cognitive Computing and Information Processing, CCIP 2015, 0–3, 2015, DOI:10.1109/CCIP.2015.7100687.

[35] J. Ding, J. Wang, X. Kang, X.H. Hu, "Building an SVM Classifier for Automated Selection of Big Data," Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017, 15–22, 2017, DOI:10.1109/BigDataCongress.2017.12.

[36] N.V. Kumar, P.V. Kumar, K. Pramodh, Y. Karuna, "Classification of Skin diseases using Image processing and SVM," Proceedings - International Conference on Vision Towards Emerging Trends in Communication and Networking, ViTECoN 2019, 1–5, 2019, DOI:10.1109/ViTECoN.2019.8899449.

[37] T.D. Nguyen, T.D. Do, M.H. Le, N.T. Le, W. Benjapolakul, "Appliance Classification Method Based on K-Nearest Neighbors for Home Energy Management System," 2019 1st International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics, ICA-SYMP 2019, 53–56, 2019, DOI:10.1109/ICA-SYMP.2019.8646278.

[38] S. Taneja, C. Gupta, K. Goyal, D. Gureja, "An enhanced K-nearest neighbor algorithm using information gain and clustering," International Conference on Advanced Computing and Communication Technologies, ACCT, 325–329, 2014, DOI:10.1109/ACCT.2014.22.

[39] M.S. Alam, S.T. Vuong, "Random forest classification for detecting Android malware," Proceedings - 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, GreenCom-IThings-CPSCom 2013, 663–669, 2013, DOI:10.1109/GreenCom-iThings-CPSCom.2013.122.

[40] S. Hwang, M.K. Jeong, "Robust relevance vector machine for classification with variational inference," Annals of Operations Research, **263**(1–2), 21–43, 2018, DOI: 10.1007/s10479-015-1890-9.

[41] D. Kropotov, D. Vetrov, L. Wolf, T. Hassner, "Variational relevance vector machine for tabular data," Journal of Machine Learning Research, **13**(2001), 79–94, 2010.

[42] M. Henderson, M. Gasic, B. Thomson, P. Tsiakoulis, K. Yu, S. Young, "Discriminative spoken language understanding using word confusion networks.," SLT, 176–181, 2012, DOI:10.1109/SLT.2012.6424218.

[43] B. Thomson, "Statistical methods for spoken dialogue management," Ph.D. Cambridge University, (December), t, 2009, DOI: 10.1007/978-1-4471-4923-1_1.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 472-480 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Reliability Improvement of Radial Distribution System by Reconfiguration

Srividhya. P[1], Mounika. K[1], Kirithikaa. S[1], Narayanan. K[1,*] Gulshan Sharma[2], Girish Ganesan. R[3], Tomonobu Senjyu[3]

[1]*EEE Department, SASTRA Deemed University, Thanjavur, 613401, India*

[2]*Department of Electrical Power Engineering, Durban University of Technology, Durban, 4001, South Africa*

[3]*Power and Energy System Control Laboratory, University of the Ryukyus, Okinawa, 903-0213, Japan*

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 20 October, 2020
Online: 20 November, 2020

Keywords:
SAIDI
SAIFI
CAIDI
CAIFI
Reconfiguration

ABSTRACT

In a radial distribution system, component failure leads to service interruption and subsequent disconnection of a few load points. An abnormal variation in the voltage levels and high power loss gives rise to reliability issues. In this work, feeder reconfiguration is performed to improve reliability, minimize voltage deviation and power loss. A Binary Particle Swarm Optimization based search function handles the reconfiguration. The goal of this search function is to identify the optimal tie switches which minimize the multi-objective function. The search algorithm based on Binary Particle Swarm Optimization is utilized to handle the reconfiguration which identifies the optimal tie switches for minimizing the multi-objective function. The effect of the various algorith parameters on the convergence rate is studied. The proposed algorithm has been tested on Standard IEEE 33 and 69 test systems and the obtained results show a boost in the system reliability.

1 Introduction

Electric distribution system (EDS) are exponentially evolving due to the increasing demand and the changing sensitivity of customers loads. An interruption in electric power supply to the customers is caused by many events such as equipment failure, faults at substations, trees and much more. These interruptions cause a severe impact in the load equipment apart from the obvious financial losses and inconvenience. It is proclaimed that the failures in the distribution system causes nearly 80% of the interruptions. The customers of the electric utility desires an interruption free reliable power supply at the end of the day, emphasizing the need for calculating the reliability of EDS.

Reconfiguration is opening and closing of sectionalizing and tie switches. The former is normally closed whereas the latter is normally open switches. The alteration of the status of the switches under normal conditions or during contingencies results in change in topology of the feeders. It reduces the cost of implementation of various protective devices while satisfying the constraints such as maintaining radial condition of the system and feeding all the buses. It has become a practice nowadays to implement distributed generators along with reconfiguration to obtain better results [1]-[3]. But implementing Distributed Generation (DG) increases the overall cost of the EDS. Implementing reconfiguration will ensure that the customer does not face outages by rerouting the required power thereby enhancing reliability of the system. Voltage fluctuation of can be endured only for a brief stretch. If not, the impairment instigated in the equipment connected to the EDS would be high [4]. So reconfiguration is used to maintain the same within limits. Reconfiguration contributes to proper distribution of load and hence lower magnitude of current in the branches thereby reducing the power loss. This increases the loadability of the feeder thus contributing to the enhancement of reliability of the system.

There are several documented ways to solve for the optimal system reconfiguration, namely, heuristic, meta-heuristic and deterministic methods. Deterministic models hold no uncertainity and the output only depends on the intial state of the system and the value of the system parameters. Heuristic techniques are often greedy algorithms pursuing a local optimum instead of a global optimum even though they are problem dependent. Meta-heuristic

*Corresponding Author: Narayanan. K, EEE Department, SASTRA Deemed University, Thanjavur, Tamil Nadu, India. Mob: +91-9790258910 Email: narayanan.mnit@gmail.com

approaches are independent of the problem and not greedy, allowing them to perform and in-depth search of the workspace to get an optimum solution. In most cases, the parameters of interest are power loss, voltage profile and reliability. From the extensive literature cited in this work it is seen that employing meta-heuristic methods are techniques has been a recent trend in the research field.

In [5]-[14] an objective function has been developed consisting of both reliability indices and loss in models with and without a DG while [15]-[18] involve loss reduction and voltage profile improvement. Since 1980, Genetic algorithm(GA) which is classified as an evolutionary algorithms is widely used as an optimization tool [5], [8], [19] and [20]. But it is a slow algorithm [19] as it requires decent sized population and a number of generations to see good results whereas Particle Swarm Optimzation (PSO) entails small quantity of ciphering time not counting the better accuracy in results in contrast to GA. In [8] an Enhanced GA is presented that can resolve the reconfiguration of the feeder proficiently and sturdily with fewer iteration and time. A novel Micro genetic algorithm with reduced number of efficient solution which works well for multi-objective functions is presented in [6].

The technique shown in [7] adopts an Analytic Hierarchy Process for decision making with multiple criteria analysis. In the work presented in [9], Dynamic programming is used to find minimal cutsets and a search based technique for reconfiguration to attain the ideal set of switches has been articulated for RDS using DG. In [11] reconfiguration is solved using a novel Imperialist Competitive Algorithm which intensifies the exploration ability.In [14] the randomness in the reliability data is taken into account and has introduced a robust approach. In [3] a CRO algorithm to deal with complex network reconfiguration with DG present is taken. Different versions of Particle Swarm Optimization (PSO) is used in [10], [15], [17], [19], [21]–[24] where [15] and [23] takes another step by hybridizing PSO with Ant Colony Optimization and Evolutionary Programming respectively, for reconfiguration for various objectives. This algorithm provides better results faster in comparison to the other algorithms.

The contribution of this paper is the enhancement of reliability indices by performing an optimal reconfiguration of the distribution system using a modified version of PSO. The objective function involves minimization of Reliability index SAIDI, Voltage Deviation and Power loss. Four other reliability indices are calculated as well. The influence of certain parameters in PSO are determined using a MATLAB simulation environment and the results prove the validity of the proposed concept.

2 Problem Formulation

Reliability is the ability of a distribution system to provide uninterrupted power supply to customer [25]. Reliability indices are the benchmark parameters for evaluating reliability of the distribution system and performance of the utility. There are five standard reliability performance indices: System Average Interruption Frequency Index (SAIFI), System Average Interruption Duration Index (SAIDI), Customer Average Interruption Duration Index (CAIDI), Average System Availability Index (ASAI) and Average System Unavailability Index (ASUI) which are briefed below.

2.1 Reliability Indices

SAIFI is the average frequency of interrupt occurences that a customer experiences,

$$SAIFI = \frac{\sum_i \lambda_i \times N_i}{N_i} f/yr \tag{1}$$

where,

λ_i - failure rate(failure/year) of ith load point

N_i - Number of customers at point i.

SAIDI is the average duration of interrupt a customer experiences,

$$SAIDI = \frac{\sum_i U_i \times N_i}{N_i} h/yr \tag{2}$$

where,

U_i - product of failure rate and repair rate of ith load point.

Failure rate (λ_i) and Repair rate (r_i) are called load point indices and are calculated using Cutset Approach as the systen is a series-parallel system.

CAIDI is the average interruption time a consumer is affected by interruption per year. It is related to SAIFI and SAIDI indices.

$$CAIDI = \frac{SAIDI}{SAIFI} \tag{3}$$

ASAI represents the percentage of time during a year that the average customer have power.

$$ASAI = \frac{N \times 8760 - \sum_i U_i \times N_i}{N \times 8760} \tag{4}$$

It is the average system unavailability index is the contrary of ASAI and is given by,

$$ASUI = 1 - ASAI \tag{5}$$

2.2 Voltage Deviation and Power Loss

A stable voltage profile alongside low power loss is important as it directly benefits the electric utility operation and subsequently increases customer satisfaction. For this reason, it is embedded into the objective function. Voltage deviation and power loss can be calculated as follows,

$$V\_DEV = \sum_{i=2}^{N} (1 - |V_i|)^2 \tag{6}$$

$$P\_LOSS_i = \frac{R_i(P_i^2 + Q_i^2)}{|V_i|^2} \tag{7}$$

$$P\_LOSS = \sum_{i=1}^{N} k_i P\_LOSS_i \tag{8}$$

where,

P_LOSS_i is the power loss of branch i, P_LOSS is the total power loss, P_i and Q_i are the real power and reactive power of branch i, R_i and is the branch resistance, V_i is the voltage at bus i, N is the total number of branches and k_i is the status of branch i (1=close, 0=open)

2.3 Constraints

2.3.1 Failure and Repair rate constraints

$$\lambda_{i,min} \leq \lambda_i \leq \lambda_{i,max} \tag{9}$$

$$r_{i,min} \leq r_i \leq r_{i,max} \tag{10}$$

2.3.2 Real and Reactive power constraints

$$P_{G,min} \leq P_G \leq P_{G,max} \tag{11}$$

$$Q_{G,min} \leq Q_G \leq Q_{G,max} \tag{12}$$

2.3.3 Voltage constraint

$$V_{i,min} \leq V_i \leq V_{i,max} \tag{13}$$

2.3.4 Radial topology constraint

$$\varphi(L) = 0 \tag{14}$$

$\lambda_{i,min}$, $r_{i,min}$, $\lambda_{i,max}$ and $r_{i,max}$ are the limits of the values of failure and repair rate. $P_{G,min}$ and $P_{G,max}$ are the vectors of lowest and highest available real power generation whereas $Q_{G,min}$ and $Q_{G,max}$ are the vectors of lowest and highest available reactive power generation. $V_{i,min}$ and $V_{i,max}$ are the limits of the branch voltages and $\varphi(L)$ is the number of loops in the system.

2.4 Objective Function

$$Minf = w_1 \times \frac{SAIDI}{SAIDI_T} + w_2 \times \frac{V\_DEV}{V\_DEV_T} + w_3 \times \frac{P\_LOSS}{P\_LOSS_T} \tag{15}$$

Since reliability improvement is the major objective of the work, w_1 is given higher value such that,

$$w_1 + w_2 + w_3 = 1 \tag{16}$$

As the objectives have different dimension they're conveter to per-unit values. $SAIDI_T$,V_DEV_T and P_LOSS_T are the initial case values of SAIDI, Voltage deviation and Power Loss respectively. Thus multiple objectives are unified into a single objective resulting in reduced optimization load. The system condition before reconfiguration is taken as the base.

3 Proposed Method

This section presents the basics of PSO and the formulation an intelligent search method depending on PSO. All the feasible configurations of the distribution system that satisfy the objective function is presented.

3.1 Particle Swarm Optimization

PSO is a meta heuristic population based technique with its genesis from the observation of social behaviour associated with birds flocking. It can be applied to both continuous and discrete systems. In the presented problem, the switches are chosen from a discrete set for different bus systems and therefore discrete PSO is applied in this case.

The algorithm develops a population of particles each of which are a viable solution for the problem. Initially each particle is assigned a random position, velocity and objective function, X_i, V_i and F_i respectively. The evaluation is carried out using positional coordinates and population size as the input. Each particle has a follow up of its position and the best value acquired so far and is stored in P_{best}. . In the meantime the overall best value produced so far by any particle is stored as G_{best}.

$$w = w_{max} - \frac{w_{max} - w_{min}}{iter_{max}} \times iter \tag{17}$$

The interia of the particles has an impact on the future velocities of the particles. Experimentally it is observed that decreasing the value of inertia over the iterations causes exploration and exploitation to occur linearly. Number of particles in the swarm is represented by n. Smaller value takes large number of iterations to converge whereas high value cause a rise in computational complexity and time consumed. So a suitable value is chosen.

$$v_{ij}^{t+1} = w \times v_{ij}^t + c_1 r_{1j}^t [P_{best} - x_{ij}^t] + c_2 r_{2j}^t [G_{best} - x_{ij}^t] \tag{18}$$

where,

v_{ij}^t - particle velocity

x_{ij}^t - particle position

c_1 - cognitive parameter

c_2 - social parameter c_1 and c_2 represents the weights of the stochastic acceleration term that pulls the particles towards P_{best} and G_{best} respectively.

r_1 and r_2 are random values in the range [0,1].

The particles coordinates are updated using the sigmoidal equation,

$$s_{ij}^t = \frac{1}{1 + e^{-v_{ij}^{t+1}}} \tag{19}$$

The *ciel* of the function converts the continuous to discrete values. It is further modified according to the problem at hand to choose the values from a particular set of values. In our work, the set of values corresponds to the switches that are to be opened for reconfiguration from the array of switches from each loop of the bus system if the tie lines are closed.

3.2 Algorithm

1. Add Load bus data, line data maximum iterations and switch value matrix. Initialize population, position, random velocities, P_{best}, G_{best}.
2. Calculate inertia, update velocity and particles coordinates.
3. Check radiality constraint and calculate fitness. Power loss and voltage deviation are calculated from load flow solution. Reliability is calculated using the concept of cutsets.

Backward Forward Sweep Algorithm is used as the load flow solution.

4. Update P_{best} if the fitness function is lesser than the previous value.
5. Update G_{best} if the fitness function is lesser than the previous value.
6. If there is no criteria match, update particle velocity using the previous velocity, distance to P_{best} and distance to G_{best} in the equation
7. Update particle position from provided switches.
8. Terminate if the maximum iterations is reached. Go to step 3. The flowchat of the proposed algorithm is shown in Figure 1.

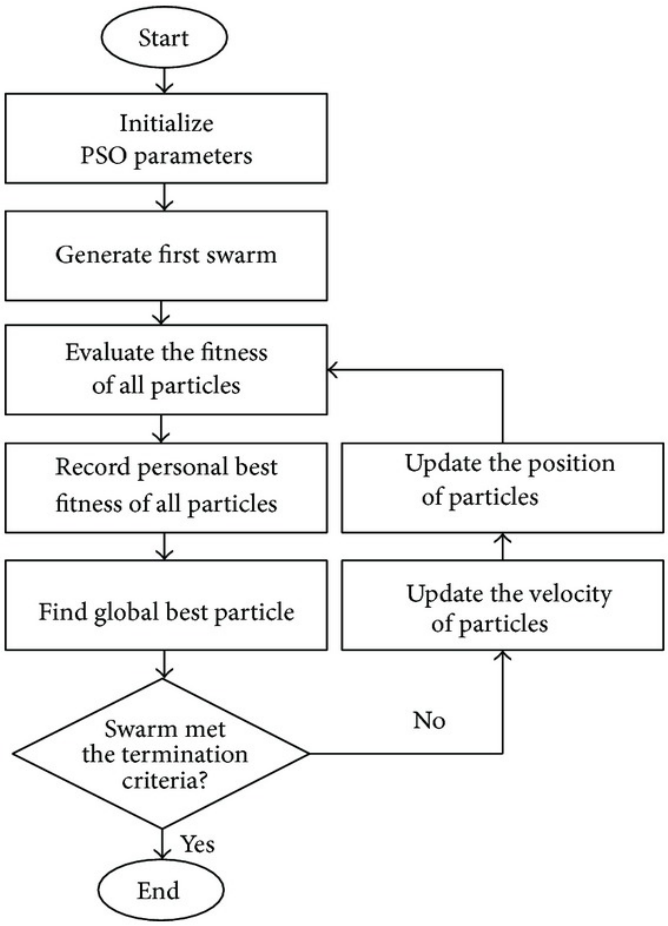


Figure 1: Flowchart of Proposed PSO

4 Results and Discussion

The IEEE 33 and 69 bus system data is presented in [10]. The failure rate and repair rate for the same is presented in [20]. The proposed method is implemented on the test cases and and the effectiveness is tested for the given objective function. Each load point is assumed to have one customer. Four different cases is made out of the objective function presented in Eq.15 with different weightage to each objective but with heighest weightage to SAIDI.

1. Case 1: $w_1 = 1.0$, $w_2 = 0$ and $w_3 = 0$
2. Case 2: $w_1 = 0.75$, $w_2 = 0$ and $w_3 = 0.25$
3. Case 3: $w_1 = 0.75$, $w_2 = 0.25$ and $w_3 = 0$
4. Case 4: $w_1 = 0.5$, $w_2 = 0.25$ and $w_3 = 0.25$

4.1 33 Bus System

It has a voltage rating of 12.7kV with a constant power load of (3.715 +j2.3) MVA. It is generally used to evaluvate the distribution reconfiguration problem. Figure 2 shows the single line diagram for a 33 bus system. In the base configuration, the solid lines represent the NC switches (1 through 32) and the dotted lines represed the NO switches (33 through 37).

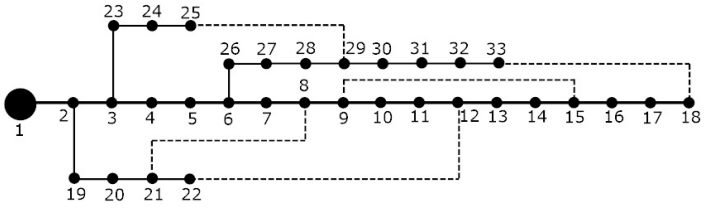


Figure 2: 33 bus system before reconfiguration

4.1.1 Case 1

The main objective of this case is to minimise the reliability index, SAIDI, which is used to find how reliable the system is. The algorithm is trained using this test case in order to find optimal solutions in the upcoming cases for the multi-objective function. For comparison purposes, the values of the indices of the base case is calculated.

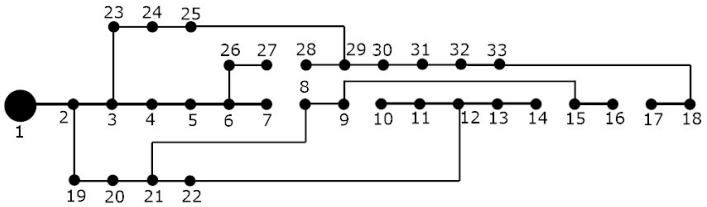


Figure 3: 33 bus system after reconfiguration (Case 1)

Figure 3 shows the bus system post reconfiguration. In this case a drop of 0.8769 p.u for duration index which is a 40% reduction from the original value making the system highly reliable. The Table 1 shows the best possible values, w_{min}, w_{max} and population size. A value of population swarm between 40 and 50 reduces the convergence time. The fitness function converges in the 6_{th} iteration when population is randomized and in the 8_{th} iteration when inertia is randomized.

Table 1: PSO parameters - 33 bus system

Parameter	Value
w_{min}	0.2
w_{max}	0.9
Population	48

The Voltage deviation and power loss have shown a reduction of 24% and 55% for the reconfigured system even without giving weightage to the same as shown in Table 2. Power loss and voltage deviation has shown drastic reduction in its values in spite of not giving weightage to them in this case thus concluding on the fact that all the three objectives are interrelated.

4.1.2 Case 2

The values of the parameters, inertia weight (minimum and maximum) and population of the swarm, are fixed for cases 2, 3 and 4. From Table 2 it is clear that with a 0.5% sacrifice of SAIDI, the voltage deviation and power loss values have been lowered by additional 4% and 2% respectively in comparison with the results of case 1. As shown in Figure 4 the system is radial post reconfiguration.

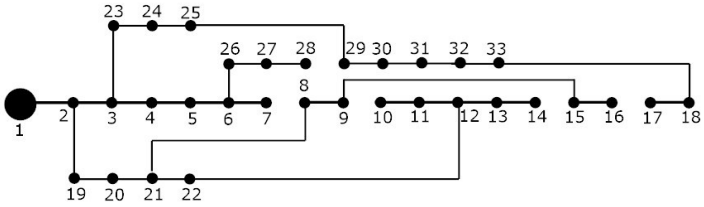


Figure 4: 33 bus system after reconfiguration (Case 2)

4.1.3 Case 3

It was found that the results of Case 3 is identical and hence the single line diagram is not repeated. The loss is reduced by an additional 4% when compared to case 1 while giving up only a 0.5% in SAIDI value with respect to case 1.

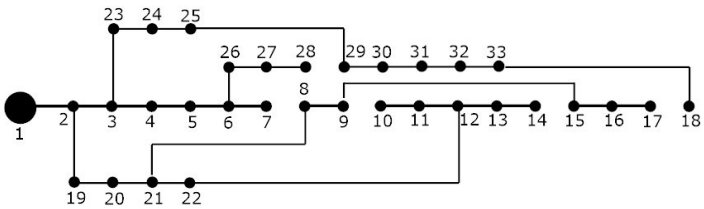


Figure 5: 33 bus system after reconfiguration (Case 4)

4.1.4 Case 4

It can be seen that a further reduction of 4kW power loss in comparison to the previous two cases. Voltage deviation has also reduced by an extra 2% when compared to the past two cases along with the betterment of minimum voltage in the system. Hence it is proved that the system is highly reliable with power loss and voltage deviation improved drastically. Figure 5 displays the bus system configuration after reconfiguration. Due to space issues, Voltage profile and Convergence graph of the final case is alone shown in Figure 6 and 7 to picturize the improvement in the voltage profile and the efficacy of the proposed method.

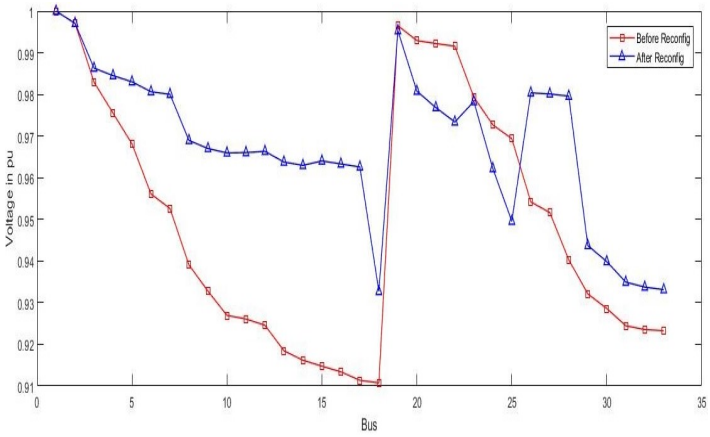


Figure 6: Voltage Profile pf 33 Bus System (Case 4)

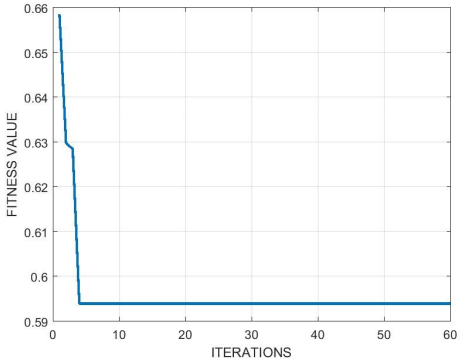


Figure 7: Convergence graph of 33 Bus System (Case 4)

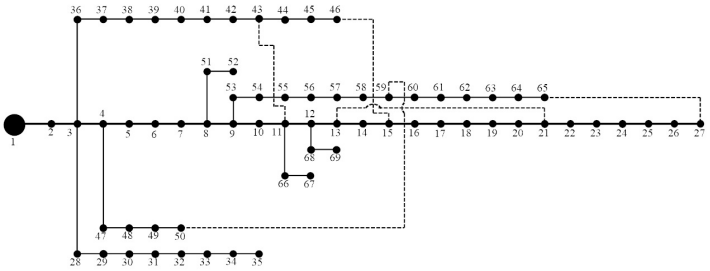


Figure 8: 69 bus system before reconfiguration

4.2 69 Bus System

The proposed topology is tested on a IEEE Standard 69 Bus system. This system has a rated voltage of 12.7kV.NO switches are 69 through 73 as in Figure 8. The reliability indices before reconfiguration are recorded by applying the algorithm to the base case of the 69 bus system. Other than the zero injection buses, each load

Table 2: Consolidated Results of 33 Bus System

INDICES	BASE CASE	CASE 1	CASE 2	CASE 3	CASE 4
TIE SWITCHES	33,34,35,36,37	7,9,14,16,27	7,9,14,16,28	7,9,14,16,28	7,9,14,17,28
SAIFI	2.5741	2.2675	2.258438	2.258438	2.308125
SAIDI	2.2291	1.3522	1.362656	1.362656	1.405469
CAIDI	0.86597	0.59633	0.603362	0.603362	0.608922
ASAI	0.99975	0.99985	0.999844	0.999844	0.99984
ASUI	0.00025446	0.00015436	0.000156	0.000156	0.00016
V_DEV (%)	2.8945	1.3137	1.20724	1.20724	1.137742
P_LOSS (kW)	202.7142	154.2636	150.2211	150.2211	146.3066
V_MIN (p.u)	0.9131	0.9276	0.9291	0.9291	0.9327

point is assumed to have one customer. The Inertia and Population are randomized separately and run hundreds of iterations to arrive at the best values. The values obtained as shown in Table 3 further implicate that w_{min}, w_{max} should be chosen with a difference of 0.5 for accurate results and reaffirms that population value within [40,50] holds the best results.

4.2.1 Case 1

Apart from SAIFI and SAIDI other indices such as CAIDI, ASAI and ASUI are found for base case in order to compare the results of all the cases. The switches that are to be opened post reconfiguration are {14, 18, 56, 64, 69}. When the switches are compared with the failure and repair rate, it is found that the branches with maximum failure rate are opened so that the values of the indices drop according to the Eq. 1 and Eq. 2. Table 4 shows that there is a major cutback of 0.7431 p.u which is 28.8% for SAIFI and 0.3804 which is 22.28% for SAIDI. Values of all the indices are shown as well. The power loss and voltage deviation is determined for the reconfigured system. Loss before and after reconfiguration presents a substantial drop of 47.11%. Voltage deviation has reduced by 71.2%. Figure 9 represents the 69 bus system after reconfiguration maintaining radial topology and its load evened out.

4.2.2 Case 2

In this case duration and voltage deviation are optimized. SAIDI plummeted by a 21% while voltage deviation which was given a weightage of 0.25 in the objective function dropped by 2.49 p.u which is nearly 74% with respect to base improving the voltage profile substantially and thereby improving overall reliability of the power system. Power loss for the above configuration was 119.6465kW which an improvement from the previous case highlighting the impact of better voltage profile on power loss. Figure 10 represents the 69 bus system after reconfiguration maintaining radial topology.

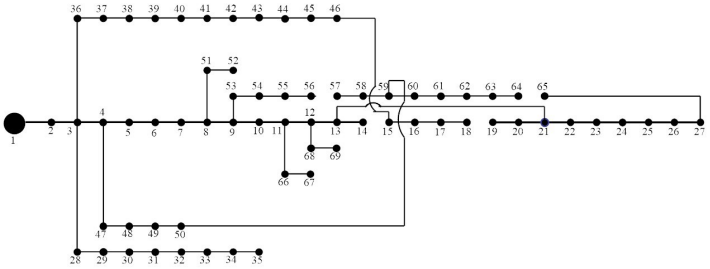


Figure 9: 69 bus system after reconfiguration (Case 1)

Table 3: PSO parameters - 69 bus system

PARAMETER	VALUE
w_{min}	0.2
w_{max}	0.7
Population	46

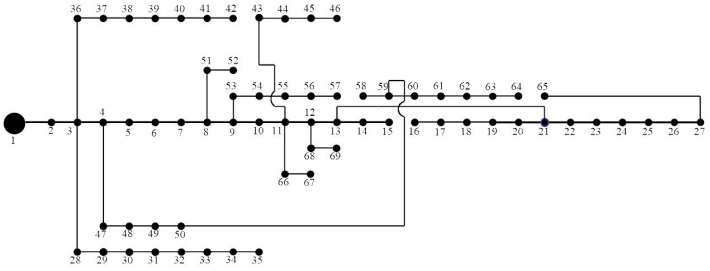


Figure 10: 69 bus system after reconfiguration (Case 2)

4.2.3 Case 3

This case involves optimizing SAIDI and Power loss with a weightage of 0.75 and 0.25 respectively. Table 4 shows the result after

reconfiguration and it is found that loss has reduced to 106.7573kW even though SAIDI raised by a value of 0.04. The voltage deviation calculated after reconfiguration was 0.8958 which is just a 0.02 p.u rise while the loss has reduced considerably. The minimum value of voltage has improved due to minimized power loss. This helps the sensitive equipment from falling a prey for low voltage conditions and hence improves the reliability of the system. Figure 11 represents the 69 bus system after reconfiguration maintaining radial topology.

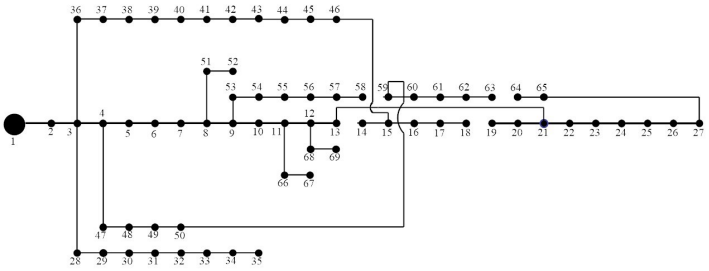


Figure 11: 69 bus system after reconfiguration (Case 3)

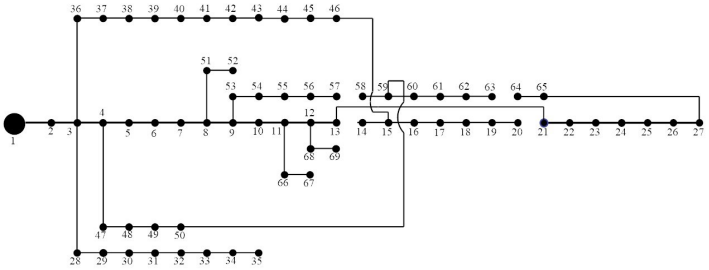


Figure 12: 69 bus system after reconfiguration (Case 4)

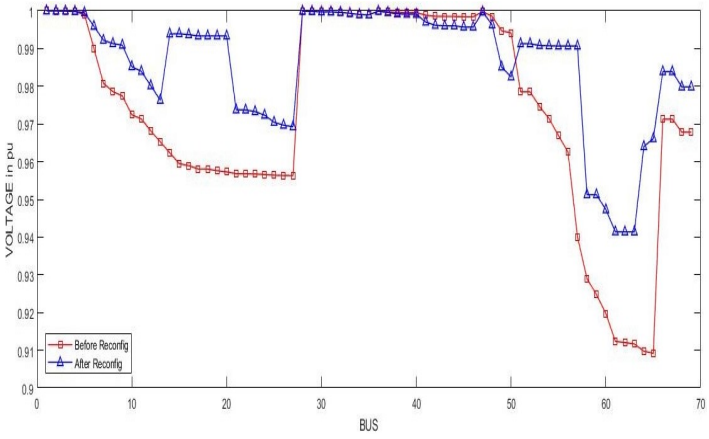


Figure 13: Voltage Profile of 69 Bus System (Case 4)

4.2.4 Case 4

This case includes all the three objectives with equal weightage to voltage deviation and power loss of 0.25 and reliability index SAIDI holding higher importance with weightage of 0.5. It is clearly seen from Table 4 that all three objective functions have their values reduced from case 3 giving us the best possible combination of switches under the above condition. The minimum voltage stays the same as previous case but overall voltage profile has improved. In comparison to the base case there is a drop of 20.4%, 73.45% and 52.57% in SAIDI, Voltage deviation and Power loss respectively as shown in Table 4. Figure 12 sketches the 69 Bus System with loads distributed and following the radial topology constraint. Figure 13 shows the pictorial representation of the improvement in voltage profile while Figure 14 supports the claim of high convergence rate in the proposed method.

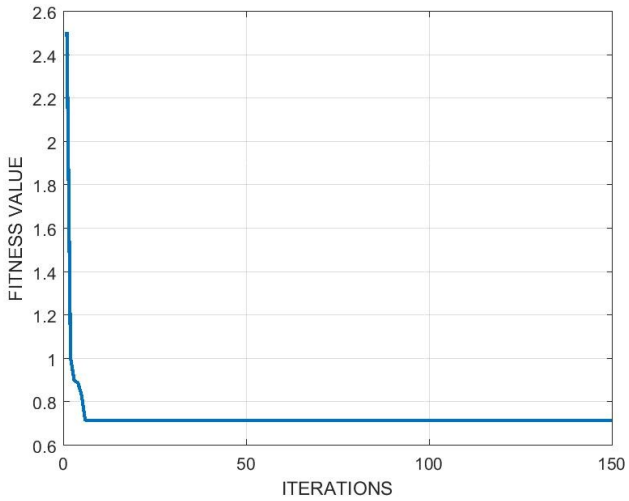


Figure 14: Convergence graph of 69 Bus System (Case 4)

5 Conclusion

This work presents a method for reliability improvement of Radial Distribution System through feeder reconfiguration. It is constructed based on Particle Swarm Optimization. The radial topology is conserved using the graph theory function in MATLAB. The load point indices such as failure rate and repair rate are calculated using Cut-set concept. The reliability indices are calculated using theoretical formulae. Power loss and Voltage deviation is computed by utilizing Backward-Forward Sweep algorithm.

The Inertia weight and Population values in the proposed algorithm clearly influences the accuracy of the results and the time to taken to obtain the same as it regulates the exploration and exploitation of the search space.

In single objective function, the switch selection for reconfiguration solely depends on the values of the load point indices (λ_i & U_i) and hence switches with highest values of the indices are opened. This results in cutting down the number of load points in the longest branch and distributing them to other branches.

In multi objective function, the switch combination chosen not only depends on the failure rate and repair rate of the load points but also on the voltage and power loss. This results in minor increase in the SAIDI and other reliability indices but it improves the voltage profile aiding the loads connected the system. It takes some demand of the utility system due to reduced power losses.

Table 4: Consolidated Results of 69 Bus System

INDICES	BASE CASE	CASE 1	CASE 2	CASE 3	CASE 4
TIE SWITCHES	69,70,71,72,73	14,18,56,64,69	15,42,57,64,71	13,18,58,63,69	13,20,57,63,69
SAIFI	2.5778	1.834623	1.85446	1.912598	1.8446
SAIDI	1.7071	1.326789	1.34073	1.381665	1.358867
CAIDI	0.6622	0.723194	0.722973	0.722402	0.736673
ASAI	0.999746	0.999849	0.999847	0.999842	0.999845
ASUI	0.000254	0.000151	0.000153	0.000158	0.000155
V_DEV (%)	3.3742	0.9717	0.8777	0.8958	0.895569
P_LOSS (kW)	225	119.6465	119.6465	106.7573	106.7085
V_MIN (p.u)	0.9092	0.9299	0.9299	0.9414	0.9414

References

[1] K. Jasthi, D. Das, "Simultaneous distribution system reconfiguration and DG sizing algorithm without load flow solution," IET Generation, Transmission Distribution, **12**(6), 1303–1313, 2018, doi:10.1049/iet-gtd.2017.0338.

[2] J. Wen, Y. Tan, L. Jiang, K. Lei, "Dynamic reconfiguration of distribution networks considering the real-time topology variation," IET Generation, Transmission Distribution, **12**(7), 1509–1517, 2018, doi:10.1049/iet-gtd.2017.1304.

[3] P. Sharma, N. Gupta, K. R. Niazi, A. Swarnkar, "Investigation of network reconfiguration on the reliability and performance of distribution systems using CRO," in 2017 6th International Conference on Computer Applications In Electrical Engineering-Recent Advances (CERA), 80–85, 2017, doi: 10.1109/CERA.2017.8343305.

[4] K. Muttaqi, A. Dt Le, G. Ledwich, M. Negnevitsky, "Minimising Voltage Deviation in Distribution Feeders by Optimising Size and Location of Distributed Generation," Australian Journal of Electrical and Electronics Engineering, **3**, 2007, doi:10.1080/1448837X.2007.11464155.

[5] M. M. Legha, M. Mohammadian, M. Darini, "Loss Reduction and Reliability Improvement with Optimization of DGs Placement Using Genetic Algorithm - A Case Study on the Electrical Distribution Network of North Kerman Area," International Journal of Scientific & Engineering Research, **5**(2), 2014.

[6] J. E. Mendoza, M. E. Lopez, C. A. C. Coello, E. A. Lopez, "Microgenetic multiobjective reconfiguration algorithm considering power losses and reliability indices for medium voltage distribution network," IET Generation, Transmission Distribution, **3**(9), 825–840, 2009, doi:10.1049/iet-gtd.2009.0009.

[7] A. P. C. Mello, D. P. Bernardon, L. L. Pfitscher, L. N. Canha, V. J. Garcia, M. Sperandio, "Reconfiguration of distribution network considering distributed generation and multivariables criteria," in 2013 48th International Universities' Power Engineering Conference (UPEC), 1–6, 2013, doi: 10.1109/UPEC.2013.6714870.

[8] D.-L. Duan, X.-D. Ling, X.-Y. Wu, B. Zhong, "Reconfiguration of distribution network for loss reduction and reliability improvement based on an enhanced genetic algorithm," International Journal of Electrical Power & Energy Systems, **64**, 88 – 95, 2015, doi:https://doi.org/10.1016/j.ijepes.2014.07.036.

[9] P. Pavani, S. N. Singh, "Reconfiguration of radial distribution networks with distributed generation for reliability improvement and loss minimization," in 2013 IEEE Power Energy Society General Meeting, 1–5, 2013, doi: 10.1109/PESMG.2013.6672876.

[10] S. Elsaiah, M. Benidris, Y. Tian, J. Mitra, "A Comprehensive Analysis of Reliability-oriented Optimal Distribution System Reconfiguration," in 2018 IEEE Industry Applications Society Annual Meeting (IAS), 1–9, 2018, doi: 10.1109/IAS.2018.8544624.

[11] M. Esmaili, M. Sedighizadeh, M. Mahmoodi, M., "Reconfiguration of distribution systems to improve reliability and reduce power losses using Imperialist Competitive Algorithm," Iranian Journal of Electrical and Electronic Engineering, **13**, 2017, doi:10.22068/IJEEE.13.3.6.

[12] A. Kavousi-Fard, M.-R. Akbari-Zadeh, "Reliability enhancement using optimal feeder reconfiguration," Neurocomputing, **106**, 1–11, 2013, doi:10.1016/j.neucom.2012.08.033.

[13] N. G. Paterakis, A. Mazza, S. F. Santos, O. Erdinç, G. Chicco, A. G. Bakirtzis, J. P. S. Catalão, "Multi-Objective Reconfiguration of Radial Distribution Systems Using Reliability Indices," IEEE Transactions on Power Systems, **31**(2), 1048–1062, 2016, doi:10.1109/TPWRS.2015.2425801.

[14] J. C. López, M. Lavorato, J. F. Franco, M. J. Rider, "Robust optimisation applied to the reconfiguration of distribution systems with reliability constraints," IET Generation, Transmission Distribution, **10**(4), 917–927, 2016, doi:10.1049/iet-gtd.2015.0558.

[15] M. Amin Heidari, "Optimal network reconfiguration in distribution system for loss reduction and voltage-profile improvement using hybrid algorithm of PSO and ACO," CIRED - Open Access Proceedings Journal, **2017**(1), 2458–2461, 2017, doi:10.1049/oap-cired.2017.1007.

[16] I. A. Quadri, S. Bhowmick, D. Joshi, "Multi-objective approach to maximise loadability of distribution networks by simultaneous reconfiguration and allocation of distributed energy resources," IET Generation, Transmission Distribution, **12**(21), 5700–5712, 2018, doi:10.1049/iet-gtd.2018.5618.

[17] S. R. Tuladhar, J. G. Singh, W. Ongsakul, "Multi-objective approach for distribution network reconfiguration with optimal DG power factor using NSPSO," IET Generation, Transmission Distribution, **10**(12), 2842–2851, 2016, doi: 10.1049/iet-gtd.2015.0587.

[18] Z. Gong, Q. Chen, K. Sun, "Novel methodology solving distribution network reconfiguration with DG placement," The Journal of Engineering, **2019**(16), 1668–1674, 2019, doi:10.1049/joe.2018.8521.

[19] B. Vasudevan, K. Sinha, A., "Reliability improvement of reconfigurable distribution system using GA and PSO," Electrical Engineering, **100**, 1–13, 2017, doi:10.1007/s00202-017-0580-9.

[20] N. Gupta, A. Swarnkar, K. R. Niazi, "Distribution network reconfiguration for power quality and reliability improvement using Genetic Algorithms," International Journal of Electrical Power & Energy Systems, **54**, 664 – 671, 2014, doi:https://doi.org/10.1016/j.ijepes.2013.08.016.

[21] S. Elsaiah, M. Benidris, J. Mitra, "Reliability improvement of power distribution system through feeder reconfiguration," in 2014 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 1–6, 2014, doi:10.1109/PMAPS.2014.6960676.

[22] K. B. Kela, L. D. Arya, "Reliability Optimization of Radial Distribution Systems Employing Differential Evolution and Bare Bones Particle Swarm Optimization," Journal of The Institution of Engineers (India): Series B, **95**, 231–239, 2014, doi:10.1007/s40031-014-0094-z.

[23] M. F. Sulaima, N. H. Shamsudin, H. I. Jaafar, W. M. Dahalan, H. Mokhlis, "A DNR and DG Sizing Simultaneously by Using EPSO," in 2014 5th International Conference on Intelligent Systems, Modelling and Simulation, 405–410, 2014, doi:10.1109/ISMS.2014.75.

[24] I. I. Atteya, H. Ashour, N. Fahmi, D. Strickland, "Radial distribution network reconfiguration for power losses reduction using a modified particle swarm optimisation," CIRED - Open Access Proceedings Journal, **2017**(1), 2505–2508, 2017, doi:10.1049/oap-cired.2017.1286.

[25] P. C. Sekhar, R. A. Deshpande, V. Sankar, "Evaluation and improvement of reliability indices of electrical power distribution system," in 2016 National Power Systems Conference (NPSC), 1–6, 2016, doi:10.1109/NPSC.2016.7858838.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 481-488 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

American Sign Language Recognition Based on MobileNetV2

Kin Yun Lum[*,1], Yeh Huann Goh[1], Yi Bin Lee[1]

[1]*Department of Mechanical Engineering, Faculty of Engineering and Technology, Tunku Abdul Rahman University College, Kuala Lumpur, 43200, Malaysia*

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 20 October, 2020
Online: 20 November, 2020

Keywords:
Sign Language
Convolutional Neural Network
Depthwise Separable Convolution
MobileNet
Transfer Learning

ABSTRACT

Sign language is a form of communication language designed to link a deaf-mute person to the world. To express an idea it requires the use of hand gestures and body movement. However, the bulk of the general population remain uneducated to understand the sign language. Therefore, a translator is required to facilitate the communication. This paper wishes to extend the previously proposed Convolutional Neural Network (CNN) model for predicting American Sign Language with a MobileNetV2-based transfer learning model. The latter model effectively generalized on a dataset which is around 18 times larger with 5 additional groups of hand signs. Over 98% of the recognition accuracy had been reported. Because of its relatively fewer parameters and less intensive computational operations compared to other deep learning architectures, the model was also ideal to be implemented on mobile devices. The model will serve as the key to deploying a sign language translator software on smartphone to enhance communication efficiency between the deaf-mute person and the general public.

1 Introduction

This paper is an extension of work originally presented in conference IEEE International Conference on Signal and Image Processing Applications (ICSIPA) 2019 [1]. The work was enhanced in the following aspects:

1. Dataset: Previous study used only 4800 images with 24 groups of self-collected images except for letters J and Z, since both signs require movement. Furthermore, all images were taken from a fixed distance, which is not generally the case when it comes to actual application. The sign performer can appear at different distances from the smartphone camera. In this paper Kaggle's dataset was used. Each sign has variations in capturing distance and brightness. There are a total of 87,000 images with 29 classes. In addition to the 26 classes belonging to letter A-Z, there are an additional 3 classes for SPACE, DELETE and NOTHING. As per the description of the dataset, these 3 classes are found to be useful in real time application and classification.
2. Method: Previous proposed shallow network failed to generalize well with this new ~18x larger dataset. Instead, transfer learning using MobileNetV2 was implemented. Unlike other deep learning models, the MobileNetV2 gives high prediction accuracy without penalizing too much on computational cost and memory. This met our original intention to have this model deploy as a mobile application in the future.

Spoken language binds a significant section of the population. The spoken language would not, however, support the deaf and mute population. Statistics from the World Health Organization (WHO) shows there are at least 5% or 466 million of people suffer from hearing impairment [1]. An individual with hearing impairment may have a distorted speech or may not speak at all. This creates a barrier to contact between them and the society.

Fortunately, sign language comes to help for this special group of people. It is a kind of visual communication language that uses a combination of hand gesture, facial expression and body posture. It helps the hearing-impaired communities to express their feelings and addresses the general communication issues within their communities.

The sign language is completely different from spoken language. It has its own syntax and the way it expresses itself. For the general population this may be difficult to understand and practice [2].

[*]Corresponding Author: Kin Yun Lum, Department of Mechanical Engineering, Faculty of Engineering and Technology, Tunku Abdul Rahman University College, Kuala Lumpur, Malaysia, lumky@tarc.edu.my

Moreover, most general public are illiterate to sign language. Unless they are surrounded by a deaf-mute person, most appear to be uninterested and disregard the meaning of sign language. This presents a real barrier of contact between the deaf-mute community and the rest of society, as a issue which has yet to be completely resolved.

Innovations of technology in the area of sign language recognition try to break down this communication barrier. To date, it addresses the issues mainly with two different approaches: 1) Sensor based system 2) Vision based system [3].

The sensor based system is a wearable device which utilizes various sensors to capture motion, spatial positions and velocity of the hand. For example, cyber glove is a wearable device that employs flex sensor to detect fingers movements and Inertial Measurement Unit (IMU) sensor to track the hand motions [4]. There are also invasive and non-invasive electromyography (EMG) wearable sensors system to measure human muscle's electrical pulses and collect the bio-signal to detect fingers movements [5, 6]. These wearable system can be costly, cumbersome, inconvenient, heavy and uncomfortable for daily use.

Vision based system has recently become increasingly popular because there is no need to attach sensors physically to the human. There are attempts to use Leap Motion Controller [2] and Microsoft Kinect which have depth sensor [4, 7, 8] to construct the sign language recognition system. However, such technology is not widely accessible and it can be expensive to the users. Furthermore, the systems were implemented using desktop or laptop computer, which is impractical because of its weight and size.

Since hand sign gestures are perceived through vision, computer vision can be a subject of profound interest. One of the solutions is to build a vision based sign language recognition system with the commonly available smartphone camera. Vision-based systems are historically prone to reliability issues. Background noise, colours, and lighting are greatly varied under actual environment. This would result in a lower rate of detection. The latest advance in computer vision and machine learning technologies has shown considerable progress in the classification of images [9, 10]. In the competition for image classification, the algorithm called Convolutional Neural Networks (CNNs) and its derivative networks such as AlexNet, GoogleNet and ResNet achieved more than 90% accuracy [11]. Conventional machine learning relies on manual features extraction. Instead of creating complex handcrafted features, CNNs will automate the feature construction process. That would significantly reduce human error and increase the accuracy of detection.

There are several studies using CNN for the identification of American Sign Language (ASL) but the findings are not satisfactory [12]. Furthermore, while some deep networks may offer high classification accuracy, a large number of parameters and heavy mathematical operations limit their application in smartphones.

Our aim is to develop a computer vision system that can use mobile camera to recognize sign language, so that it can be used everywhere and anywhere. Previous attempt to create a CNN network to recognize 24 classes of sign language has demonstrated a detection accuracy of 95% experimentally. This paper extends the work to draw on a much larger dataset in order to enable greater generalization. The larger dataset contains 5 more classes than the previously used dataset. It included all 26 alphabets and another 3 classes for the sign of SPACE, DELETE and NOTHING to make the whole sign language model more complete. In addition, this dataset has also images which were captured at different distance from the camera, which may reflect a real-life application more closely. Transfer learning using MobileNetV2 was implemented to further improve the model accuracy to more than 98%.

The remainder of this paper is structured as follows: the relevant works will be discussed in Section 2. Followed by the section 3 concept and theory that will address the different CNN architectures and put emphasis on MobileNet, which is the transfer learning model to be implemented in this paper. Dataset description and preprocessing will be described in the methodology in Section 4. Then we will detail our model architecture and setup of experiments. The section ends by showing some evaluation metrics that we used to measure the performance of our model. Section 5 addresses experimental findings with discussions. Lastly, Section 6 ends with the conclusions of this paper with some suggestions for future study.

2 Related Works

The works related to vision based sign language recognition system can be categorized into two, which are non-CNN and CNN approaches. The non-CNN approaches were mainly the traditional methods which involved manual feature extraction and classification. The involved classification algorithms were k-nearest neighbors (KNN) and Neural Network [13, 14]. These approaches normally had lower reported prediction accuracy. On the other hands, the CNN approaches that had better reported results were either proposing their custom CNN models [7, 8, 15] or implementing transfer learning [11, 12, 16, 17]. There was also literature which compared both their custom CNN model and transfer learning [18].

For custom CNN models, Vivek Bheda et. al. used a deep convolutional network to classify ASL with alphabets and digits. The proposed network had 3 cascaded convolutional layers and 1 max-pooling layer. There are two hidden layers with dropout before connecting to the output layer. Accuracy of 82.5% had been reported [7].

Ameen. S. et. al. separated the image and depth information from Microsoft Kinect input and used two convolution layers to perform feature extraction on each of them. Both extracted feature maps were combined in the second stage of convolution before passing to the classifier. The reported precision and recall were 82% and 80% respectively [15].

Similar architecture was used by Pigou. L. et. al. to extract hand features and upper body features separately from Microsoft Kinect dataset. Each CNN was three layers in depth. They had reported an accuracy of 91.7% [8].

For transfer learning, Das, A. et. al. used Inception v3 on custom processed static gesture images and obtained an average validation accuracy of 90% [16]. Alashhab. S. et. al. tested on VGG16, VGG19, ResNet, Xception, Inception V3, MobileNet and SqueezeNet. They were dealing with hand gestures detection instead of sign language. There were 5 classes of hand gestures used. The highest accuracy of transfer learning usign MobileNet was 99.45% [11]. Kania, K.et. al. build a network to recognize ASL with reference to the Wide Residual Networks. Data augmentation was used to increase the number of training sets and reported a high-

est accuracy of 93.3% [12]. Garcia and Viesca used GoogLeNet architecture to recognize 24 classes of sign language with only 70% or reported accuracy [17]. As far as our literature, there was only one work to compare custom CNN model and transfer learning done by Bousbai, K. and Merah, M. Their training set had 1815 coloured images with black background collected from five volunteers. Their custom CNN had a reported efficiency of 98.9%, slightly surpassed the transfer learning model with 97.06% accuracy [18].

In general summary, transfer learning can perform better than most of the custom CNN model when come to pure image recognition. Custom model is only needed usually to cater for the extra input like depth information. Transfer learning model specially designed based on pure images dataset like ImageNet might not doing as good as the custom model. Other than this reason, transfer learning is a more appropriate solution to develop a well performing model. Therefore, it is being used in this paper to enhance the results of our previous work.

3 Concept and Theory

3.1 Convolutional Neural Network (CNN) Architecture

Convolutional Neural Network (CNN) has proven to be very effective to handle image classification problems. It is in use by many industry leaders such as Amazon, Google and Facebook [8]. The convolution layers in the CNN perform multiple discrete convolutions on the input with defined number of trainable weights filters. The filters' weights were updated during training. After applying filtering on each channel, it will capture the image features such as lines, edges, colors and other visual elements. The deeper the network goes, more complex features like shapes and patterns can be generated. Combining all these features together with the activation functions can produce feature maps which can then pass to the ordinary neural network classifier to correctly classify the object.

Due to significant progress has been achieved in image classification, there were various CNN architectures introduced such as VGG16, ResNet and Inception Net. These models were trained and verified on large-scale annotated datasets like ImageNet and proven to be effective. Instead of constructing the CNN from scratch, leverage on the transfer learning on these architectures can significantly reduce the time required to produce a working model for a particular problem. Moreover, since these models were trained on ImageNet which consists of 1.2 million categorized natural images of 1000+ classes. Using the trained weights as a starting point will definitely help the model to converge faster as compared to training from scratch. As discussed by Shin, H. C. et. al, even though there is a disparity between the ImageNet and our sign language images, the deep architecture which is trained comprehensively can as well generalize on other specific domain problem with only minor fine-tuning training [19]. Among different CNN architectures, MobileNet was employed in this paper because it is an efficient network architecture designed to run on any computationally limited platform.

3.2 MobileNetV1

MobileNet had been introduced as a small networks and it is optimized for latency to match the restricted resource application such as in mobile devices, robotics and self-driving vehicles. These application platforms are not as powerful as a general purpose computer, but still requires comparable recognition accuracy under a timely computation.

Unlike other network architectures use shrinking or compression to reduce its size and in return to improve on computational costs, MobileNet was completely rebuilt using depthwise separable convolutions. Depthwise separable convolution is a form of factorized convolutions which decomposes a conventional convolution into a depthwise convolution and a pointwise convolutoin. This factorization ends up reducing computation cost and model size drastically. The rationale behind this is to reduce the costly convolution process due to multiple multiplication as explained below [20].

Examine a standard convolution process, let an input image, F of dimension $D_W \times D_H \times M$ where D_W and D_H correspond to the spatial width and height of the input image and M is the number of input channels (depth). There are N numbers of square filters, K with size of $D_K \times D_K \times M$ where $D_K$is the spatial dimension of the kernel. An example of standard convolution with an input matrix with channel, $M = 3$ and N number of 3×3 kernels is illustrated in Figure 1a. All the convolution processes discussed in this section are assumed to have stride of one and padding. This standard convolution between F and N numbers of K will have a computational cost of:

$$J_c = N\left[(D_W \times D_H) \times (D_K \times D_K \times M)\right] \tag{1}$$

The MobileNet addresses this issue by splitting the convolution process into two steps, namely depthwise convolution and pointwise convolution. The whole process is called depthwise separable convolution. In the depthwise convolution process, instead of using kernel with depth M, single layer kernels are applied on each input channel as showed in Figure 1b. This results in a computational cost of:

$$J_d = M\left[(D_W \times D_H) \times (D_K \times D_K)\right] \tag{2}$$

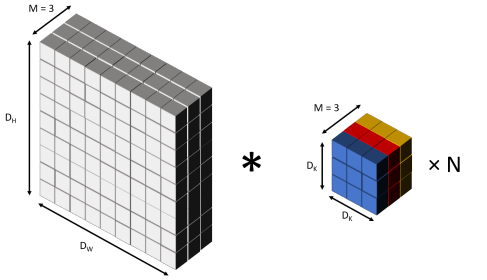
Then, pointwise convolution is performed on the output of the above depthwise convolution feature map. Pointwise convolution is being done by N number of 1×1 kernels with depth of M. This process is also called linear combination since the output matrix is of summation of linear scaling multiples. An pictorial illustration of this process is shown in Figure 1c. The computational cost of the pointwise convolution is:

$$J_p = N\left(M \times D_W \times D_H\right) \tag{3}$$

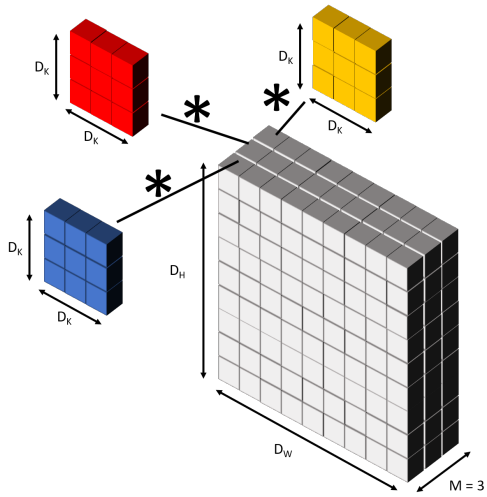
The total computational cost of a complete depthwise separable convolution is $J_{dp} = J_d + J_p$. Comparing to the standard convolution, there is a reduction factor of:

$$J_r = \frac{J_{dp}}{J_c} = \frac{1}{N} + \frac{1}{D_K^2} \tag{4}$$

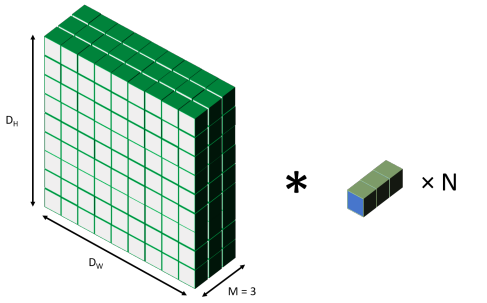
For a depthwise separable convolution with 3×3 kernels used, the computational cost can be 8 to 9 times lesser with only minor trade-off in accuracy [20].



(a) Standard convolution



(b) Depthwise convolution



(c) Pointwise convolution

Figure 1: Comparison between (a) Standard convolution (b) Depthwise convolution and (c) Pointwise convolution

3.3 MobileNetV2

MobileNetV2 is an improved version of MobileNetV1. MobileNetV1 or most deep convolution network uses the Rectified Linear Unit (ReLU) activation function defined by

$$f(x) = \begin{cases} x, & \text{if } x \geqslant 0 \\ 0, & \text{if } x < 0 \end{cases} \tag{5}$$

The ReLU activation function disregards any values smaller than 0. This non-linear transformation is argued to result in the loss of information, especially on input with a lower number of channels. ReLU may have less impact on input with lots of channels because the missing information in particular activation might still be preserved by other channels [10].

To tackle this issue, MobileNetV2 uses a narrow -> wide -> narrow approach in its inverted residual block. The low-dimensional input is first expanded by using a pointwise 1×1 convolution to produce a higher dimensional space which can cater for the information loss caused by the ReLU activation. Then, spatial filtering using depthwise convolution with ReLU activation was implemented in this higher dimensional feature map. Finally, another pointwise convolution was performed to project back to a lower dimensional output feature map. For the last step, linear activation instead of ReLU was used to preserve more information when encoding to a lower dimensional output map. This idea was called linear bottleneck by the authors. The non-linear transformation only happened in the internal expanded higher dimensional space inside the block. When come to lower dimensional output, linear transformation was used [10].

Besides, a skip connection similar to residual block is added between the input and output of the inverted residual block to allow gradient flow during backpropagation. This approach is essential to build a deep network. One last minor adjustment is the use of ReLU6 in this inverted residual block which capped the maximum output to 6 as defined by:

$$f(x) = min(max(0, x), 6) \tag{6}$$

The complete inverted residual block discussed in MobileNetV2 is shown by Figure 2.

4 Methodology

4.1 Dataset

Figure 3 shows the American Sign Language (ASL) for all alphabets. In the previous article, dataset was produced by smartphone camera captures [1]. There were a collection of 4800 images with 24 classes of gestures. The signs for character J and Z were excluded due to dynamic. All the images were taken from at a fixed camera distance with little lighting variation as displayed in Figure 4.

The challenge of developing a usable sign language recognition model lies in the variation in size, position, and shapes of the input image [15]. In real life applications, the sign language performer might stands at different distance relative to the smartphone camera. Provided data augmentation which is widely adopted in training CNN with limited dataset, which introduces more variations with scaling, rotating, flipping, shifting, shearing, etc. However, this image augmentation failed to reproduce the realistic perspective variations as discussed by Wenjin Tao et. al [4].

To extend the work, the dataset from Kaggle named "ASL Alphabet" was used [21] . Each sign has variations in distance capture and illumination. An example of sign B was shown in Figure 5. There are a total of 87,000 images in 29 classes. Besides the 26 classes belong to the letter A - Z, there are three additional classes for SPACE, DELETE and NOTHING. As per dataset description, these 3 classes are found to be helpful in real time application and classification.

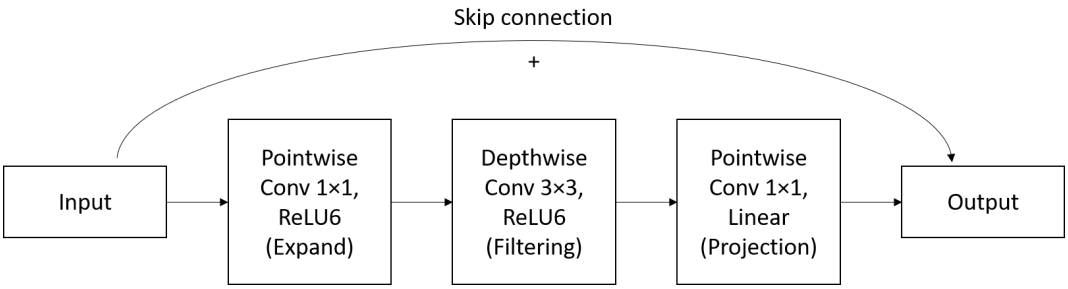


Figure 2: Inverted residual block

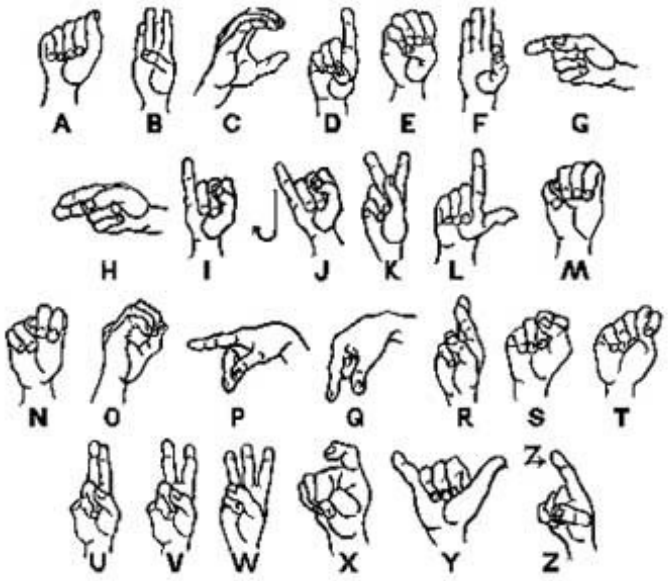


Figure 3: American Sign Language for all alphabets

Figure 4: Example self-captured dataset from previous article. From left to right (top) A, B, C and (bottom) D, E, F [1]

4.2 Model Architecture

The new dataset in this paper is approximately 18 times larger than the previous dataset. The previous proposed CNN network which had only two cascaded convolution layers failed to generalize to this dataset. The model was therefore being reconstructed using transfer learning with MobileNetV2. This CNN architecture was optimized for mobile devices to give high prediction accuracy while keeping the parameters and mathematical operations low. Compared to MobileNetV1, the MobileNetV2 is about 30 - 40% faster when tested on Google Pixel phone due to 2 times fewer mathematical operation and 30% lesser parameters. Not only that, it outperforms MobileNetV1 in terms of recognition accuracy [18].

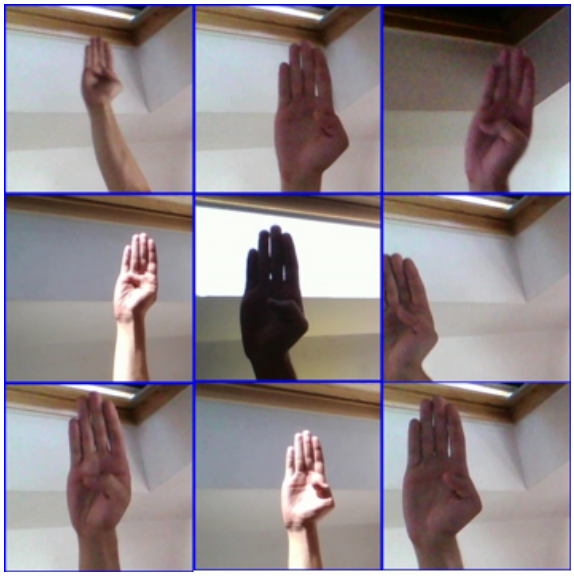


Figure 5: Example sign for letter 'B' under different capturing distance and lighting condition

All 29 classes had been used in this work. The images were resized to 224×224 pixels as suggested in the article by Samer Alashhab. et. al. [11]. Among different input sizes, 224×224 pixels gives best performance by experiment. The inputs were shuffled randomly and normalized. The images were divided into training and validation sets using a ratio of 80:20. The training set has 69,600 images while the validation set has 17,400 images. In view of the larger dataset, no data augmentation was carried out.

The built architecture is shown in Figure 6. The MobileNetV2 was preloaded with weights trained by ImageNet. Average pooling layer was added to the MobileNetV2 output feature map. Continued with a dense neural network layer of 1000 neurons. The activation function for this dense network was standard ReLU. A 50% dropout layer was added after that to prevent overfitting. Final layer was the dense layer that provides classification for 29 classes of images. It had a softmax activation function as defined by:

$$\sigma(z)_j = \frac{e_j^z}{\sum_{k=1}^{K} e_k^z} \tag{7}$$

The model was compiled using categorical crossentropy loss function. The model had about 3.5 millions training parameters. The optimizer used was RMSprop with a learning rate of 1×10^{-4}. Since the sign language images are differ from the ImageNet dataset, all the layers were unfreezed and retrained. Experiments were all conducted on a computer with Intel Core i5 CPU, 12 GB RAM and a NVIDIA GeForce GTX 1060. All of the steps mentioned in this section were performed using Keras deep learning framework with TensorFlow as the backend.

4.3 Performance Metrics

The performance metrics used to evaluate the build model are defined below, where TP stands for True Positive (correct prediction), TN is True Negative (correct rejection), FP is denoted as False Positive (incorrect prediction) and FN is defined as False Negative (incorrect rejection).

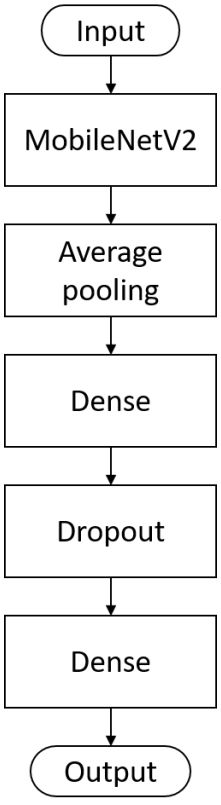


Figure 6: Model architecture

1. Accuracy - the overall correct prediction of the model

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

2. Precision - ratio of correct prediction over all the predicted values of a particular class.

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

3. Recall - ratio of correct prediction over all predicted values that are supposed to be positive.

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

4. F1-Score - weighted average of precision and recall. It tells the balance between precision and recall.

$$F1\ score = 2\left[\frac{Precision \times Recall}{Precision + Recall}\right] \tag{11}$$

5 Results and Discussion

The highest overall accuracy obtained by this model was 98.67%. The confusion matrix is shown in Figure 7. For each of the sign, 28 out of 29 were recorded with accuracy higher than 95%. The highest single class accuracy was 100% and the lowest was 85.67%. Lowest accuracy was recorded for the letter Q. There were only 514 out of 600 images were correctly predicted. The rests were all being predicted as P. This was believed to be the similarity existed in these two signs, where the index finger and thumb are both pointed out.

Looking at the precision, recall, and f1-score for each class as shown in Table 1, most of them obtained a score of 95% and above. The precision of letter P and recall of letter Q was lower due to the above mentioned similarity. Hence it also lowered the f1-score for these two classes. The second lowest recall was the sign for letter X which scored 92.83%. Some of them were being wrongly classified as either S or U due to the similar rounded fist shape.

Table 2 shows the comparison with some previous works. To have a fair comparison, only those works implementing transfer learning were compared. In the comparision, there was only one study done by Alashhab. S. et. al. had slightly better accuracy then this work, but they were not working with ASL and there were only 5 classes of hand signs used. Other than that, this work showed better results than the prior works. For a similar MobileNetV2 transfer learning architecture for ASL, there is also a marginally improvement over the work did by Bousbai, K and Merah, M.

Table 1: Precision, Recall, F1-score for each class

	Precision	Recall	F1-score
A	0.9677	1.0000	0.9836
B	0.9933	0.9917	0.9925
C	1.0000	1.0000	1.0000
D	1.0000	1.0000	1.0000
E	0.9693	1.0000	0.9844
F	0.9950	1.0000	0.9975
G	1.0000	0.9850	0.9924
H	0.9709	1.0000	0.9852
I	1.0000	0.9800	0.9899
J	1.0000	0.9633	0.9813
K	1.0000	1.0000	1.0000
L	1.0000	1.0000	1.0000
M	0.9771	0.9950	0.9860
N	0.9817	0.9850	0.9834
O	1.0000	0.9950	0.9975
P	0.8721	0.9883	0.9266
Q	1.0000	0.8567	0.9228
R	0.9934	1.0000	0.9967
S	0.9769	0.9850	0.9809
T	0.9983	0.9783	0.9882
U	0.9707	0.9933	0.9819
V	1.0000	0.9917	0.9958
W	0.9868	1.0000	0.9934
X	0.9893	0.9283	0.9579
Y	0.9983	1.0000	0.9992
Z	0.9983	0.9983	0.9983
delete	0.9967	1.0000	0.9983
nothing	1.0000	1.0000	1.0000
space	1.0000	1.0000	1.0000

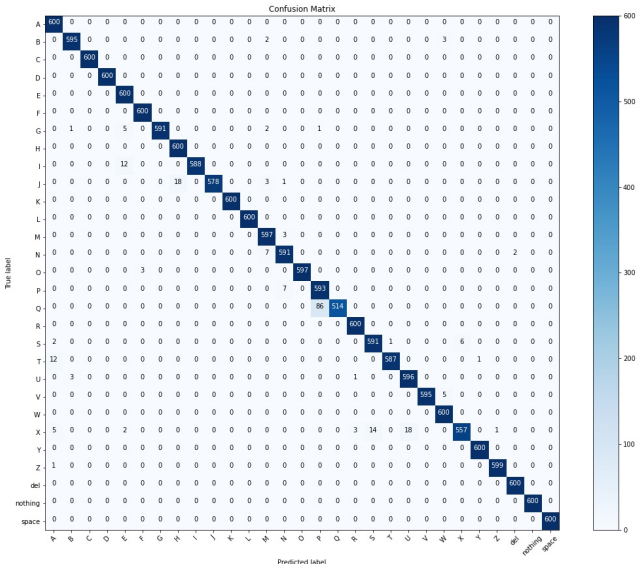


Figure 7: Confusion Matrix

Table 2: Comparison of this work and previous works

Authors	**Description**	**Accuracy**
Das, A. et. al., 2018	Transfer learning using Inception v3 on custom dataset	90.0%
Alashhab, S. et. al., 2018	Transfer learning on multiple architecture like VGGNet, ResNet, etc. on custom 5 classes of hand gestures.	99.45% (MobileNet)
Kania, K.et. al., 2018	Transfer learning using Wide Residual Networks with data augmentation on ASLs	93.3%
Garcia and Viesca, 2016	CNN on 24 ASLs with GoogLeNet tranfer learning	70%
Bousbai, K. and Merah, M., 2019	Compare custom CNN model and transfer learning using MobileNetV2 on ASLs	97.06% (MobileNetV2)
Our previous work, 2019	**Custom CNN on 24 ASLs using phone camera**	**95%**
This work	**Transfer learning using MobileNetV2 on 29 classes of ASLs**	**98.67%**

6 Conclusion and Future Work

This paper presented a model that was able to recognize American Sign Language with 98.67% accuracy. It is an extension of the work originally presented in ICSIPA 2019. Besides the improvement of the model accuracy, the following additional highlights are listed:

1. A much bigger dataset was used with additional 5 classes to make the ASL more complete. That allowed a better generalization of the model. In addition, this distance varying images given in the dataset could also make the model less sensitive to the difference in distance between the ASL users and the camera.

2. Even with the deep transfer learning network used, we still maintained our original intention to make the model suitable for running on mobile devices in order to use the smartphone camera for the convenience of users.

The model was less computational burdening since it had fewer parameters and lesser mathematical computations. With the implementation of the model as an smartphone app, it is believed to be able to improve the quality of communication between the deaf-mute person and the general public.

To offer more usability, this work can be continued with real-time video based sign language recognition. To reduce the effects of ambient noise, there is a need for real-time video processing, such as region of interest segmentation and hand tracking. Apart from that, the image occlusion has yet to be investigated. This can be a challenging issue when part of the performing signs is blocked.

Acknowledgment I wish to acknowledge Tunku Abdul Rahman University College for sponsoring this research.

References

[1] L. Y. Bin, G. Y. Huann, L. K. Yun, "Study of Convolutional Neural Network in Recognizing Static American Sign Language," in 2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 41–45, IEEE, 2019, doi:10.1109/ICSIPA45851.2019.8977767.

[2] T.-W. Chong, B.-G. Lee, "American sign language recognition using leap motion controller with machine learning approach," Sensors, **18**(10), 3554, 2018, doi:10.3390/s18103554.

[3] M. A. Jalal, R. Chen, R. K. Moore, L. Mihaylova, "American sign language posture understanding with deep neural networks," in 2018 21st International Conference on Information Fusion (FUSION), 573–579, IEEE, 2018, doi: 10.23919/ICIF.2018.8455725.

[4] W. Tao, M. C. Leu, Z. Yin, "American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion," Engineering Applications of Artificial Intelligence, **76**, 202–213, 2018, doi:10.1016/j.engappai.2018.09.006.

[5] M. J. Cheok, Z. Omar, M. H. Jaward, "A review of hand gesture and sign language recognition techniques," International Journal of Machine Learning and Cybernetics, **10**(1), 131–153, 2019, doi:10.1007/s13042-017-0705-5.

[6] J. Wu, L. Sun, R. Jafari, "A wearable system for recognizing American sign language in real-time using IMU and surface EMG sensors," IEEE journal of biomedical and health informatics, **20**(5), 1281–1290, 2016, doi: 10.1109/JBHI.2016.2598302.

[7] V. Bheda, D. Radpour, "Using deep convolutional networks for gesture recognition in American sign language," arXiv preprint arXiv:1710.06836, 2017.

[8] L. Pigou, S. Dieleman, P.-J. Kindermans, B. Schrauwen, "Sign language recognition using convolutional neural networks," in European Conference on Computer Vision, 572–578, Springer, 2014, doi:10.1007/978-3-319-16178-5_40.

[9] P. Molchanov, S. Gupta, K. Kim, J. Kautz, "Hand gesture recognition with 3D convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 1–7, 2015, doi: 10.1109/CVPRW.2015.7301342.

[10] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.

[11] S. Alashhab, A.-J. Gallego, M. Á. Lozano, "Hand gesture detection with convolutional neural networks," in International Symposium on Distributed Computing and Artificial Intelligence, 45–52, Springer, 2018, doi:10.1007/978-3-319-94649-8_6.

[12] K. Kania, U. Markowska-Kaczmar, "American Sign Language Fingerspelling Recognition Using Wide Residual Networks," in International Conference on Artificial Intelligence and Soft Computing, 97–107, Springer, 2018, doi: 10.1007/978-3-319-91253-0_10.

[13] R. Hartanto, A. Kartikasari, "Android based real-time static Indonesian sign language recognition system prototype," in 2016 8th International Conference on information Technology and Electrical Engineering (ICITEE), 1–6, IEEE, 2016, doi:10.1109/ICITEED.2016.7863311.

[14] H. M. Ewald, I. Patil, S. Ranmuthu, "ASL fingerspelling Interpretation," University of Stanford, Reports, 2016.

[15] S. Ameen, S. Vadera, "A convolutional neural network to classify American Sign Language fingerspelling from depth and colour images," Expert Systems, **34**(3), e12197, 2017, doi:10.1111/exsy.12197.

[16] A. Das, S. Gawde, K. Suratwala, D. Kalbande, "Sign language recognition using deep learning on custom processed static gesture images," in 2018 International Conference on Smart City and Emerging Technology (ICSCET), 1–6, IEEE, 2018, doi:10.1109/ICSCET.2018.8537248.

[17] B. Garcia, S. A. Viesca, "Real-time American sign language recognition with convolutional neural networks," Convolutional Neural Networks for Visual Recognition, **2**, 225–232, 2016.

[18] K. Bousbai, M. Merah, "A Comparative Study of Hand Gestures Recognition Based on MobileNetV2 and ConvNet Models," in 2019 6th International Conference on Image and Signal Processing and their Applications (ISPA), 1–6, IEEE, 2019, doi:10.1109/ISPA48434.2019.8966918.

[19] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," IEEE transactions on medical imaging, **35**(5), 1285–1298, 2016, doi: 10.1109/TMI.2016.2528162.

[20] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[21] Akash, "ASL Alphabet, https://www.kaggle.com/grassknoted/asl-alphabet," 2018 (Accessed July 1, 2020).

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 489-496 (2020)

www.astesj.com

Special Issue on Innovative Research in Applied Science, Engineering and Technology

ASTESJ

ISSN: 2415-6698

An Adaptive Nonlinear Sensorless Controller of Doubly Fed Induction Generator Driven By Wind Turbine

Radouane Ourhdir*, Mohammed Rachidi

Department of Electromechanical Engineering, Ecole Nationale Supérieure des Arts et Métiers, Moulay Ismail university, Meknès, BP 4024, Morocco

ARTICLE INFO

Article history:
Received: 01 July, 2020
Accepted: 07 September, 2020
Online: 20 November, 2020

Keywords:
Doubly fed induction generator (DFIG)
Control Wind Power Generation
Rotor speed observer
Unknown Aerodynamic Torque
Unknown Rotor Winding Resistance
Adaptive control projection operator
Lyapunov Theory

ABSTRACT

In this study, an adaptive nonlinear sensorless controller for doubly fed induction generator (DFIG) driven by a wind turbine is proposed. The aim is to maximize the extracted power by tracking the wind turbine optimal torque-speed characteristic without the need for rotor speed measurement. This controller ensures a satisfactory tracking of both stator flux and rotor speed. It considers the detailed model of DFIG in an arbitrary (d–q) rotating frame without any simplifying assumption. An observer provides the rotor speed estimation incorporated in the control loop. To guarantee the system stability under parametric uncertainties like the aerodynamic torque and the rotor winding resistance, which harms the efficiency and the robustness of the controller, update laws are established to estimate the uncertain parameters. Lyapunov's theory is used to prove the system stability. The proposed adaptive sensorless controller validity is demonstrated by simulation in Matlab/Simulink environment. The robustness of the controller is confirmed by the comparison between the same controller with and without adaptation.

1. Introduction

Recently, using wind energy to produce electricity has been gradually grabbing the attention of many researchers, thanks to its cleanliness and dependability. Because of its smaller power electronic components rating and its wide speed operation range, the DFIG is seen as the most used wind turbine generator [1–3]. The DFIG rotor windings are connected to the electrical grid via rotor and grid side converters (RSC and GSC) and RL filter, whereas the stator windings are immediately related to the electrical grid as illustrated by Figure 1. In the last decade, many researchers giving much focus and interest to the field of sensorless control of DFIG-based wind turbine (DFIG-WT)[4–7]. In this control strategy, the DFIG-WT systems operate without the need to use the speed sensor, which lowers the whole system cost and enhance its overall reliability by minimizing the number of components that are susceptible to failure.

There has been some research about sensorless control using proportional-integral (PI) loops such as [8,9]. The latter mainly relied on assumptions that the flux or stator voltages are constant, and the stator resistance is negligible. However, these assumptions are useless under grid faults or load variations. Furthermore, the small resistance of the stator may cause an incorrect damped dynamics of the stator flux [10].

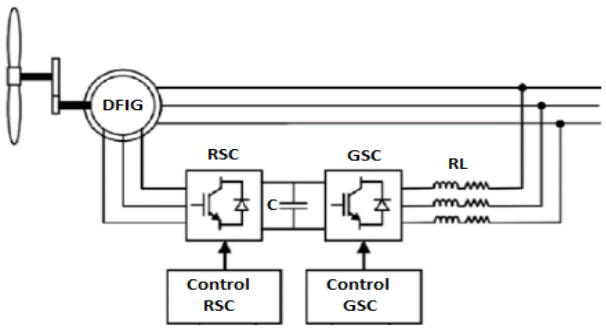


Figure 1: Schematic representation of grid connected DFIG-WT system

For the most control strategies, if the actual values do not correspond to the parameters and variables of DFIG-WT used in the control law, it could result in a failure of the closed-loop control. So, when some parameters and variables are difficult to be measured, some strategies based on estimation theory are important to generate a pertinent control. The frequency, the

*Corresponding Author: Radouane Ourhdir, radouane.ourhdir@edu.umi.ac.ma

temperature and the magnetic saturation may cause the change of DFIG parameters [11]. Therefore, the online parameter estimation or disturbance observer is compulsory to have high precision control.

Much research has addressed the control of DFIGs during parameter variation [12–14]. On the other hand, the modeling and the control of WT is not an easy task, particularly when large scale WTs, in megawatts, are considered [15,16]. Hopefully, to minimize the drive train vibration and to develop the quality of the extracted power, some works use the estimation theory in order to take into consideration the uncertainties arose from the aerodynamic torque, particularly in the stage of the maximum power point tracking (MPPT) [17].

The main focus of this work is to elaborate an adaptive sensorless control for the rotor side converter (RSC), this controller takes into consideration the detailed model of DFIG. Update laws in real-time are established to deal with the parameter uncertainties. To preserve the objectives of the MPPT without the need to use rotor speed sensor , a speed observer is incorporated in this controller. The remaining of this research paper displays as follows: The second section revolves around the model of the DFIG-WT system. The third section puts forward the suggested nonlinear controller: For the start, we present the proposed rotor speed observer. Afterward, we introduce our sensorless controller design, and we point out to the control laws of DFIG, and update laws for the speed observer, as well as the estimated rotor resistance and aerodynamic torque, which are competed using the Lyapunov theory. The last section considers simulation results and discussion. At last, we wrap up by drawing some concluding remarks.

2. DFIG-WT model

2.1. Wind turbine model

Based on wind turbine aerodynamic[18], The power captured from the wind can be expressed as :

$$P_t = \frac{1}{2}\rho\pi R^2 C_p(\lambda,\beta) v^3 \tag{1}$$

where ρ is the air density, R is the blade radius and v is the wind speed. The wind power coefficient $C_p(\lambda,\beta)$ depens on the design the blades; is a function of pitch angle β and tip speed ratio λ (TSR).
where:

$$\lambda = \frac{R\,\omega}{G\,v} \tag{2}$$

In which ω is the mechanical generator speed and G is the gearbox ratio. In the basis of the wind turbine design [18], the expression of power coefficient is:

$$C_p(\lambda,\beta) = c_1\left(\frac{c_2}{\lambda_i} - c_3\beta - c_4\right)\exp\left(-\frac{c_5}{\lambda_i}\right) + c_6\lambda \tag{3}$$

where:

$$\frac{1}{\lambda_i} = \frac{1}{\lambda + 0.08\beta} - \frac{0.035}{\beta^3 + 1}$$

The coefficients c_1-c_6 are listed in Table 1.

Table 1: Values of coefficients c1-c6

c_1	c_2	c_3	c_4	c_5	c_6
0.5176	116	0.4	5	21	0.0068

When the pitch angle β remains unchanged, there is an optimum tip ratio $\lambda = \lambda_{opt}$ that maximizes the power coefficient C_p, In this study, we assume that $\beta = 0$ .So the MPPT is achieved by the control of the generator speed. Conforming to (2) the mechanical speed that optimize the extracted power is:

$$\omega^* = \frac{G\lambda_{opt}}{R_t} v \tag{4}$$

2.2. Induction generator model

Employing the flux-linkage and voltage equations of the induction generator [19], the state variables of DFIG in an arbitrary reference frame rotating at speed ω_0 include d-q rotor currents and stator fluxes i_{dr}, i_{qr}, $\varphi_{ds}, \varphi_{qs}$, and generator speed ω .The considered 5th-order state-space representation of DFIG can be expressed as:

$$\begin{cases} \dfrac{d\varphi_{ds}}{dt} = u_{ds} - \alpha\varphi_{ds} + \omega_0\varphi_{qs} + \alpha M i_{dr} \\ \dfrac{d\varphi_{qs}}{dt} = u_{qs} - \alpha\varphi_{qs} - \omega_0\varphi_{ds} + \alpha M i_{qr} \\ \dfrac{di_{dr}}{dt} = \dfrac{1}{\sigma}u_{dr} - \dfrac{R_r}{\sigma}i_{dr} - \alpha\beta M i_{dr} + (\omega_0 - \omega) i_{qr} \\ \qquad - \beta u_{ds} + \alpha\beta\varphi_{ds} - \beta\omega\varphi_{qs} \\ \dfrac{di_{qr}}{dt} = \dfrac{1}{\sigma}u_{qr} - \dfrac{R_r}{\sigma}i_{qr} - \alpha\beta M i_{qr} - (\omega_0 - \omega) i_{dr} \\ \qquad - \beta u_{qs} + \alpha\beta\varphi_{qs} + \beta\omega\varphi_{ds} \\ \dfrac{d\omega}{dt} = -\dfrac{f}{J}\omega + \dfrac{T_t}{J} - \mu\left(\varphi_{qs} i_{dr} - \varphi_{ds} i_{qr}\right) \end{cases} \tag{5}$$

with:

$$\sigma = L_r\left(1 - \frac{M^2}{L_r L_s}\right), \quad \beta = L_r\frac{M}{\sigma L_s}, \quad \mu = \frac{M}{JL_s}, \quad \alpha = \frac{R_s}{L_s}$$

T_t is the mechanical torque. The rest of the parameters of DFIG-WT system are given in Table 2, Appendix B.

3. Control design

3.1. Speed observer

Achievement of MPPT strategy requires a quick convergence of the mechanical rotor speed to its optimal value. An observer is proposed to estimate the actual rotor speed of DFIG.

$$\dot{\hat{\omega}} = -\frac{f}{J}\hat{\omega} + \frac{\hat{T}_t}{J} - \mu\left(\varphi_{qs} i_{dr} - \varphi_{ds} i_{qr}\right) + \eta \tag{6}$$

Measurement of the rotor currents and the stator fluxes are compulsory for the latter observer; it relies also on the estimated aerodynamic torque $\hat{T}_t$ and the term η which will be elaborated by the means of the Lyapunov approach.

3.2. Reference Signals

Supposing that; the d-q rotating frame is aligned with the stator flux vector as in [20]. In our case, we consider that the stator flux vector is associated with the d-axis of the d-q reference frame giving:

$$\varphi_{qs} = 0;\ \varphi_{ds} = \Psi^* \tag{7}$$

Based on the assumption (7) illustrated by Figure 2, the expressions of the reference signals in the steady-state condition of DFIG can be established:

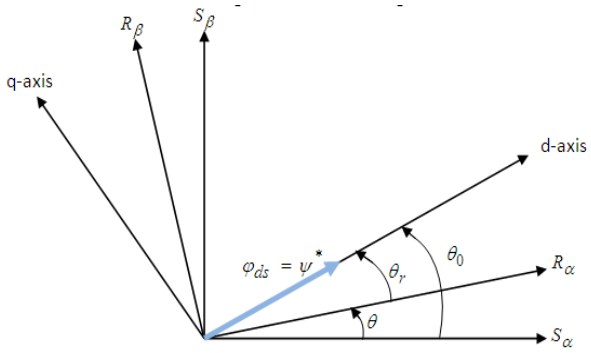


Figure 2: Determination of angles for the DFIG in d-q reference frame

Taking into consideration the latter assumption (7), and according to the first equation of (5), the reference of the direct rotor current can be expressed as follows:

$$i_{dr}^* = \frac{\Psi^*}{M} - \frac{u_{ds}}{\alpha M} + \frac{1}{\alpha M}\frac{d\Psi^*}{dt} \tag{8}$$

From the 5th equation of the system (5), the reference of the quadratic rotor current is established by replacing the aerodynamic torque by its estimate and incorporating a feedback saturation function [21], which relies on the estimated and reference rotor speed:

$$i_{qr}^* = \frac{1}{\mu\Psi^*}\left(-k_\omega sat\left(\hat{\omega} - \omega^*\right) + \frac{f}{J}\omega^* - \frac{\hat{T}_t}{J} + \frac{d\omega^*}{dt}\right) \tag{9}$$

where:

$$sat(y) = \begin{cases} 0.1 & if \quad y \geq 0.1 \\ y & if \quad -0.1 \prec y \prec 0.1 \\ -0.1 & if \quad y \leq -0.1 \end{cases}$$

The saturation function sat(y) has a finite limit at infinity; it is a linear, odd, and differentiable function in a neighborhood of the origin. To satisfy the assumption (7), the d-q reference frame must rotate at a specific speed ω_0 which is expressed using the second equation of (5):

$$\omega_0 = \dot{\theta}_0 = \frac{u_{qs} + \alpha M i_{qr}^*}{\Psi^*} \tag{10}$$

3.3. Control and update laws

The rotor speed sensor has some weak points concerning robustness, maintenance, cost, and cables linking the speed sensor and the control block. Any incorrect value provided by the rotor speed sensor may significantly harm the defined control targets and it can also destroy the system stability. Besides, accurate modeling of the aerodynamic torque is not an easy task, and the rotor winding resistance may vary up during operation due to the heating of the generator. That is to say that it constitutes uncertain parameters for the model which can affect the system stability. Yet, to ensure the system stability and maintain the control objectives in the presence of uncertainties and without needing to measure the rotor speed, the system closed-loop stability is examined to extract the expressions of the additional term η used in the speed observer, and the update laws of the estimated rotor resistance and aerodynamic torque. In this study, we assume that the aerodynamic torque and rotor resistance are unknown and vary slowly with time and should be estimated. To this purpose, the error variables can be introduced as follows:

$$\begin{cases} \tilde{\omega} = \omega - \omega^* \\ \tilde{\varphi}_{ds} = \varphi_{ds} - \Psi^* \\ \tilde{\varphi}_{qs} = \varphi_{qs} \\ \tilde{i}_{dr} = i_{dr} - i_{dr}^* \\ \tilde{i}_{qr} = i_{qr} - i_{qr}^* \\ e_\omega = \hat{\omega} - \omega \\ \tilde{R}_r = R_r - \hat{R}_r \\ \tilde{T}_t = T_t - \hat{T} \end{cases} \tag{11}$$

Using equations (5), (6), (7), (8), (9), (10) and (11) the system of the error dynamics can be expressed as follows:

$$\begin{cases} \dot{\tilde{\varphi}}_{ds} = -\alpha\tilde{\varphi}_{ds} + \omega_0\tilde{\varphi}_{qs} + \alpha M\tilde{i}_{dr} \\ \dot{\tilde{\varphi}}_{qs} = -\alpha\tilde{\varphi}_{qs} - \omega_0\tilde{\varphi}_{ds} + \alpha M\tilde{i}_{qr} \\ \dot{\tilde{i}}_{dr} = \frac{u_{dr}}{\sigma} - \frac{R_r}{\sigma}i_{dr} - \alpha\beta M i_{dr} + (\omega_0 - \omega)i_{qr} \\ \qquad - \beta u_{ds} + \alpha\beta\varphi_{ds} - \beta\omega\varphi_{qs} - \frac{di_{ds}^*}{dt} \\ \dot{\tilde{i}}_{qr} = \frac{u_{qr}}{\sigma} - \frac{R_r}{\sigma}i_{qr} - \alpha\beta M i_{qr} - (\omega_0 - \omega)i_{dr} \\ \qquad - \beta u_{qs} + \alpha\beta\varphi_{qs} + \beta\omega\varphi_{ds} - \frac{di_{qs}^*}{dt} \\ \dot{\tilde{\omega}} = -\frac{f}{J}\tilde{\omega} + \frac{\tilde{T}_t}{J} - \mu\tilde{\varphi}_{qs}i_{dr} + \mu\tilde{\varphi}_{ds}i_{qr} + \mu\Psi^*\tilde{i}_{qr} \\ \dot{e}_\omega = -\frac{f}{J}e_\omega - \frac{\tilde{T}_t}{J} + \eta \end{cases} \tag{12}$$

In order to have a compact error dynamic system, the error variables e_d and e_q are considered instead of $\tilde{i}_{dr}$ and $\tilde{i}_{qr}$:

$$\begin{cases} e_d = \tilde{\varphi}_{ds} + \frac{\tilde{i}_{dr}}{\beta} \\ e_q = \tilde{\varphi}_{qs} + \frac{\tilde{i}_{qr}}{\beta} \end{cases} \tag{13}$$

The error dynamics in (12) becomes:

$$\begin{cases} \dot{\tilde{\varphi}}_{ds} = -(\alpha + \alpha\beta M)\tilde{\varphi}_{ds} + \omega_0\tilde{\varphi}_{qs} + \alpha\beta M e_d \\ \dot{\tilde{\varphi}}_{ds} = -(\alpha + \alpha\beta M)\tilde{\varphi}_{qs} - \omega_0\tilde{\varphi}_{ds} + \alpha\beta M e_q \\ \dot{e}_d = \dfrac{u_{rd}}{\beta\sigma} - \dfrac{\tilde{R}_r}{\beta\sigma} i_{dr} + \left(\varphi_{qs} + \dfrac{i_{qr}}{\beta}\right)e_\omega + \eta_{0d} \\ \dot{e}_q = \dfrac{u_{rq}}{\beta\sigma} - \dfrac{\tilde{R}_r}{\beta\sigma} i_{qr} - \left(\varphi_{ds} + \dfrac{i_{dr}}{\beta}\right)e_\omega + \eta_{0q} \\ \dot{e}_\omega = \dfrac{f}{J} e_\omega - \dfrac{\tilde{T}_t}{J} + v \end{cases} \quad (14)$$

where:

$$\eta_{0d} = -\frac{\hat{R}_r}{\beta\sigma} i_{dr} - \alpha M i_{dr}^* + (\omega_0 - \hat{\omega})\frac{i_{qr}}{\beta} - (\hat{\omega} - \omega_0)\varphi_{qs} + \alpha\Psi^* - u_{ds} - \frac{1}{\beta}\left(\frac{\dot{\Psi}^*}{M} - \frac{1}{\alpha M}\frac{du_{sd}}{dt} + \frac{1}{\alpha M}\frac{d^2\Psi^*}{dt^2}\right)$$

$$\eta_{0q} = -\frac{\hat{R}_r}{\beta\sigma} i_{qr} - \alpha M i_{qr}^* - (\omega_0 - \hat{\omega})\frac{i_{dr}}{\beta} + (\hat{\omega} - \omega_0)\varphi_{ds} + \omega_0\Psi^* - u_{qs} - \frac{1}{\beta}\left[-\frac{\dot{\Psi}^*}{\mu\Psi^{*2}}\left(-k_\omega sat(\hat{\omega} - \omega^*) + \frac{f}{J}^* - \frac{\hat{T}_t}{J} + \dot{\omega}^*\right) + \frac{1}{\mu\Psi^*}\left(-k_\omega \frac{d}{dt} sat(\hat{\omega} - \omega^*)(\dot{\hat{\omega}} - \dot{\omega}^*) + \frac{f}{J}\dot{\omega}^* - \frac{\dot{\hat{T}}_t}{J} + \ddot{\omega}^*\right)\right]$$

The expression of the stator direct voltage derivative $\frac{du_{sd}}{dt}$ is detailed in Appendix A. we construct a Lyapunov function for the system (14) as:

$$V = \frac{1}{2}\left(\tilde{\varphi}_{ds}^2 + \tilde{\varphi}_{qs}^2 + z_d^2 + z_q^2\right) + \frac{1}{2}\lambda e_\omega^2 + \frac{\gamma}{J} e_\omega \tilde{T}_t + \frac{\gamma^2}{\lambda J^2}\tilde{T}_t^2 + \frac{1}{2}\delta\tilde{R}_r^2 \quad (15)$$

where λ, γ and δ are positive design parameters given in Appendix B. The derivative of V along with the time gives

$$\dot{V} = -\alpha\left[\tilde{\varphi}_{ds}^2 + \tilde{\varphi}_{qs}^2\right] - \alpha M\beta\left[\tilde{\varphi}_{ds}^2 + \tilde{\varphi}_{qs}^2\right] + \alpha M\beta\left[e_d\tilde{\varphi}_{ds} + z_q\tilde{\varphi}_{qs}\right] + \tilde{T}_t\left[v\frac{\gamma}{J} - \frac{2\gamma^2}{\lambda J^2}\dot{\hat{T}}_t\right] + \tilde{R}_r\left[-\frac{e_d i_{dr}}{\sigma\beta} - \frac{z_q i_{qr}}{\sigma\beta} - \delta\dot{\hat{R}}_r\right] + e_\omega\left[e_d\left(\varphi_{qs} + \frac{i_{qr}}{\beta}\right) - e_q\left(\varphi_{ds} + \frac{i_{dr}}{\beta}\right) + \lambda v - \frac{\gamma}{J}\dot{\hat{T}}_t\right] + z_q\left[\eta_{0q} + \frac{u_{qr}}{\sigma\beta}\right] + z_d\left[\eta_{0d} + \frac{u_{dr}}{\sigma\beta}\right] - \frac{\lambda f}{J}e_\omega^2 - \frac{\gamma}{J^2}\tilde{T}_t^2 - \left[\frac{\lambda}{J} + \frac{\gamma f}{J^2}\right]\tilde{T}_t e_\omega \quad (16)$$

It's obvious that:

$$\left(\frac{e_d}{2} - \varphi_{ds}\right)^2 + \left(\frac{e_q}{2} - \varphi_{qs}\right)^2 \geq 0 \quad (17)$$

From (17), we can get the following inequality:

$$\left[-\tilde{\varphi}_{ds}^2 - \tilde{\varphi}_{qs}^2 + e_d\tilde{\varphi}_{ds} + e_q\tilde{\varphi}_{qs}\right] \leq \frac{e_d^2 + e_q^2}{4} \quad (18)$$

The control laws are derived as:

$$\begin{cases} u_{dr} = -\sigma\beta\left[\dfrac{\alpha M\beta}{4} e_d + \eta_{0d} + k e_d\right] \\ u_{qr} = -\sigma\beta\left[\dfrac{\alpha M\beta}{4} e_q + \eta_{0q} + k e_q\right] \end{cases} \quad (19)$$

We define the dynamics of the rotor resistance as

$$\dot{\hat{R}}_r = \Pr oj\left[-\frac{1}{\delta}\left[\frac{e_d i_{dr} + e_q i_{qr}}{\sigma\beta}\right], \hat{R}_r\right] \quad (20)$$

where $\Pr oj[g, \hat{R}_r]$ is the projection algorithm[22], which is specified in our case by

$$\dot{\hat{R}}_r = \begin{cases} g & if\ \ R_{rm} \leq \hat{R}_r \leq R_{rM} \\ g & if\ \hat{R}_r < R_{rm}\ \ and\ g \geq 0 \\ g & if\ \hat{R}_r > R_{rM}\ \ and\ g \leq 0 \\ g\left[1 - \dfrac{R_{rm}^2 - \hat{R}_r^2}{R_{rm}^2 - (R_{rm} - \varepsilon_1)^2}\right] & if\ \ \hat{R}_r < R_{rm}\ and\ \ g < 0 \\ g\left[1 - \dfrac{\hat{R}_r^2 - R_{rM}^2}{(R_{rM} + \varepsilon_1)^2 - R_{rM}^2}\right] & if\ \ \hat{R}_r > R_{rM}\ and\ \ g > 0 \end{cases} \quad (21)$$

with R_{rM} and R_{rm} are the upper and the lower bound values of R_r, respectively, and ε_1 is an arbitrary positive constant such that $R_{rm} - \varepsilon_1 \geq 0$.

We define the update laws of the additional term and the aerodynamic torque as

$$\eta = \begin{cases} \varsigma = \dfrac{2}{\lambda}\left[e_q\left(\varphi_{ds} + \dfrac{i_{dr}}{\beta}\right) - e_d\left(\varphi_{qs} + \dfrac{i_{qr}}{\beta}\right)\right] & if\ \hat{T}_t \leq T_{tM} \\ \varsigma = \dfrac{2}{\lambda}\left[e_q\left(\varphi_{ds} + \dfrac{i_{dr}}{\beta}\right) - e_d\left(\varphi_{qs} + \dfrac{i_{qr}}{\beta}\right)\right] & if\ \hat{T}_t > T_{tM} \\ & and\ \dfrac{dP(\hat{T}_t)}{dt}\varsigma \leq 0 \\ \dfrac{\varsigma}{1 + P(\hat{T}_t)} & Otherwise \end{cases} \quad (22)$$

$$\dot{\hat{T}}_t = \begin{cases} \chi = \dfrac{J}{\gamma}\left[e_q\left(\varphi_{ds} + \dfrac{i_{dr}}{\beta}\right) - e_d\left(\varphi_{qs} + \dfrac{i_{qr}}{\beta}\right)\right] & if\ \hat{T}_t \leq T_{tM} \\ \chi = \dfrac{J}{\gamma}\left[e_q\left(\varphi_{ds} + \dfrac{i_{dr}}{\beta}\right) - e_d\left(\varphi_{qs} + \dfrac{i_{qr}}{\beta}\right)\right] & if\ \hat{T}_t > T_{tM} \\ & and\ \dfrac{dP(\hat{T}_t)}{dt}\chi \leq 0 \\ \dfrac{1 + P(\hat{T}_t)}{1 + P(\hat{T}_t)}\chi & Otherwise \end{cases} \quad (23)$$

where:

$$P(\hat{T}_t) = \frac{\hat{T}_t^2 - T_{tM}^2}{2\varepsilon_2\hat{T}_{tM} + \varepsilon_2^2} \quad (24)$$

with T_{tM} is the upper bound value of T_t and ε_2 is an arbitrary positive constant. Substituting (18), (19), (21), (22) and (23) in (16), the time derivative of (15) becomes:

$$\dot{V} \le -\alpha\left(\tilde{\varphi}_{ds}^2 + \tilde{\varphi}_{qs}^2\right) - \underbrace{\left[k - \frac{\alpha\beta M}{4}\right]}_{C}\left(e_d^2 + e_q^2\right) - \lambda\frac{f}{J}e_\omega^2 - \left(\frac{\lambda}{J} + \gamma\frac{f}{J^2}\right)\tilde{T}_t e_\omega - \frac{\gamma}{J^2}\tilde{T}_t^2 \quad (25)$$

The time derivative of (15) can be written in the following way:

$$\dot{V} \le -\min\{\alpha, C\}\left(\tilde{\varphi}_{ds}^2 + \tilde{\varphi}_{qs}^2 + e_d^2 + e_q^2\right) - \lambda\frac{f}{J}e_\omega^2 - \left(\frac{\lambda}{J} + \gamma\frac{f}{J^2}\right)\tilde{T}_t e_\omega - \frac{\gamma}{J^2}\tilde{T}_t^2 \quad (26)$$

Since $\tilde{T}$ and $\tilde{R}_r$ has been designed using a projection algorithm (bounded), from (15) and (26) we can establish that the functions $\tilde{\varphi}_{ds}$, $\tilde{\varphi}_{qs}$, e_d, e_q and e_ω are bounded on $[0, \infty)$. Consequently, from (14) it's clear that $\dot{\tilde{\varphi}}_{ds}$, $\dot{\tilde{\varphi}}_{qs}$, $\dot{e}_d$, $\dot{e}_q$ and $\dot{e}_\omega$ are bounded on $[0, \infty)$. So the functions $\tilde{\varphi}_{ds}$, $\tilde{\varphi}_{qs}$, e_d, e_q and e_ω are uniformly continuous on $[0, \infty)$, as a result, from(13) it can be seen that $\dot{\tilde{i}}_{dr}$, $\dot{\tilde{i}}_{qr}$, $\dot{\tilde{i}}_{dr}$ and $\dot{\tilde{i}}_{qr}$ are bounded, and the functions $\tilde{i}_{dr}$, $\tilde{i}_{qr}$ are also uniformly continuous on $[0, \infty)$.

The closed loop of the system (14) is determined based on (19), (21), (22) and (23) as:

$$\begin{cases} \dot{\tilde{\varphi}}_{ds} = -(\alpha + \alpha\beta M)\tilde{\varphi}_{ds} + \omega_0\tilde{\varphi}_{qs} + \alpha\beta M e_d \\ \dot{\tilde{\varphi}}_{qs} = -(\alpha + \alpha\beta M)\tilde{\varphi}_{qs} - \omega_0\tilde{\varphi}_{ds} + \alpha\beta M e_q \\ \dot{e}_d = -\left(k + \frac{\alpha\beta M}{4}\right)e_d - \frac{\tilde{R}_r}{\sigma}e_d + \frac{\tilde{R}_r}{\sigma}\tilde{\varphi}_{ds} \\ \qquad + e_q e_\omega + \frac{i_{qr}^*}{\beta}e_\omega - \frac{i_{dr}^*}{\sigma\beta}\tilde{R}_r \\ \dot{e}_q = -\left(k + \frac{\alpha\beta M}{4}\right)e_q - \frac{\tilde{R}_r}{\sigma}e_q + \frac{\tilde{R}_r}{\sigma}\tilde{\varphi}_{qs} - e_d e_\omega \\ \qquad -\left(\Psi^* + \frac{i_{dr}^*}{\beta}\right)e_\omega - \frac{i_{qr}^*}{\sigma\beta}\tilde{R}_r \end{cases} \quad (27)$$

We can put the last five equations of (27) in the following form:

$$\begin{aligned} \dot{x}(t) &= [A(t) + B(t)]x(t) + C(t)z(t) \\ \dot{z}(t) &= Dz(t) + Ex(t) + Fw(t) \\ \dot{w}(t) &= Gx(t) \end{aligned} \quad (28)$$

where:

$$x(t) = \left[\tilde{\varphi}_{ds}, \tilde{\varphi}_{qs}, e_d, e_q\right]^T \; ; \; z(t) = \left[e_\omega, \tilde{R}_r\right]^T \; ; \; w(t) = \tilde{T}_t$$

and

$$A(t) = \begin{pmatrix} -(\alpha + \alpha\beta M) & \omega_0 & \alpha\beta M & 0 \\ -\omega_0 & -(\alpha + \alpha\beta M) & 0 & \alpha\beta M \\ 0 & 0 & -\left(k + \frac{\alpha\beta M}{4}\right) & 0 \\ 0 & 0 & 0 & -\left(k + \frac{\alpha\beta M}{4}\right) \end{pmatrix}$$

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{\tilde{R}_r}{\sigma} & 0 & -\frac{\tilde{R}_r}{\sigma} & e_\omega \\ 0 & \frac{\tilde{R}_r}{\sigma} & -e_\omega & -\frac{\tilde{R}_r}{\sigma} \end{pmatrix} \quad C = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \frac{i_{qr}^*}{\beta} & -\frac{i_{dr}^*}{\sigma\beta} \\ -\left(\Psi^* + \frac{i_{dr}^*}{\beta}\right) & -\frac{i_{qr}^*}{\sigma\beta} \end{pmatrix}$$

$$E = \begin{pmatrix} 0 & 0 & \frac{2}{\lambda}f_1\left(\varphi_{ds} + \frac{i_{dr}}{\beta}\right) & -\frac{2}{\lambda}f_1\left(\varphi_{ds} + \frac{i_{dr}}{\beta}\right) \\ 0 & 0 & -f_2\frac{i_{dr}}{\delta\sigma\beta} & -f_2\frac{i_{qr}}{\delta\sigma\beta} \end{pmatrix}$$

$$F = \begin{pmatrix} -\frac{1}{J} \\ 0 \end{pmatrix} \; ; \; D = \begin{pmatrix} -\frac{f}{J} & 0 \end{pmatrix}$$

$$G = \begin{pmatrix} 0 & 0 & f_3\frac{J}{\gamma}\left(\varphi_{qs} + \frac{i_{qr}}{\beta}\right) & -f_3\frac{J}{\gamma}\left(\varphi_{ds} + \frac{i_{dr}}{\beta}\right) \end{pmatrix}$$

with $f_i, i = 1,2,3,$ are bounded functions such that $0 \le f_i \le 1$

Since the functions $\tilde{\varphi}_{ds}, \tilde{\varphi}_{qs}, e_d, e_q, e_\omega$ are bounded, so for all $t \ge t_0$ $A(t), B(t), C(t), D(t), E(t), F(t)$ and $G(t)$ are bounded, with $\|A(t)\| \le A_M$, $\|B(t)\| \le B_M$, $\|C(t)\| \le C_M$, $\|D(t)\| \le D_M$, $\|E(t)\| \le E_M$, $\|F(t)\| \le F_M$, $\|G(t)\| \le G_M$

From (15), the function $V(t)$ can be put into the following inequation:

$$\alpha_1\left(\|x\|^2 + \|z\|^2 + \|w\|^2\right) \le V \le \alpha_2\left(\|x\|^2 + \|z\|^2 + \|w\|^2\right) \quad (29)$$

The time derivative of $V(t)$ (26) can be written in the following way:

$$\dot{V} \le -a_3\|x\|^2 - a_4\|w\|^2 + a_5\|w\|\|z\| \quad (30)$$

with:

$$a_3 = \min\{\alpha, k\}, \; a_4 = \frac{\gamma}{J^2}, \; a_5 = -\left(\frac{\lambda}{J} + \frac{\gamma f}{J^2}\right)$$

According to the proof of lermma 2.1 in [23], the origin $x = 0, z = 0; w = 0$ of (28) is locally exponentially stable.

4. Simulation Results

Simulation of the detailed model of 3MW DFIG-WT using the proposed adaptive nonlinear sensorless controller is performed in MATLAB/Simulink environment, the parameters and characteristics of the latter DFIG-WT along with the design parameters are given in Table 2, and Table 3, Appendix B.

The estimation and tracking performance of the adaptive sensorless controller is shown using a variable wind speed of $(8.5\ m/s)$ mean value as depicted in Figure 3. To show the robustness of the proposed adaptive sensorless controller against the aerodynamic torque and the rotor winding resistance variation the tracking performance for both the sensorless controller with [24] and without adaptation at constant wind speed $(9\ m/s)$ has been compared in Figure 4. The simulation is performed with the variation of the mechanical torque and the rotor winding resistance described by:

$$T_t = \begin{cases} T_t & for\ t \le 20s \\ 0.5T_t & for\ t > 20s \end{cases} \quad (31)$$

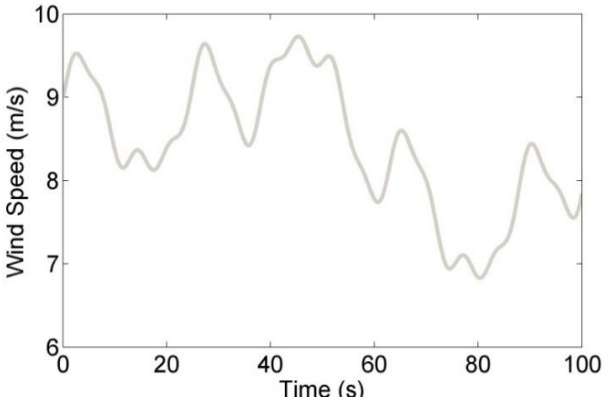
$$R_r = \begin{cases} R_r & for\ t \le 10s \\ 1.5R_r & for\ t > 10s \end{cases} \quad (32)$$

From Figure 3 (a) and 3 (b) the actual rotor speed varies in accordance with the wind speed, which confirms the effectiveness of the MPPT. Figure 3 (c) shows a good tracking performance of the stator quadratic current. According to Figure 3(d) it can be noticed that the estimated torque recovers the applied unknown torque. Similarly, from Figure 3(e) it can be observed that the estimated rotor resistance quickly converges to its true profiles. In Figure 4 the adaptive nonlinear sensorless controller performance is shown by the comparison of the proposed adaptive sensorless and the sensorless controller proposed in [24]. From Figures 4 (a) and 4 (b) it is noted that the stator flux and the rotor speed tracking errors sustain small perturbations caused by the unknown rotor resistance and the unknown aerodynamic torque, but they converge to zero quickly as soon as the estimates of R_r , and T_t converges to their actual values. Whereas in the case of control without adaptation Figures 4 (c) and 4 (d) the rotor speed and the stator flux don't track anymore their references after variation of the rotor resistance and the aerodynamic torque.

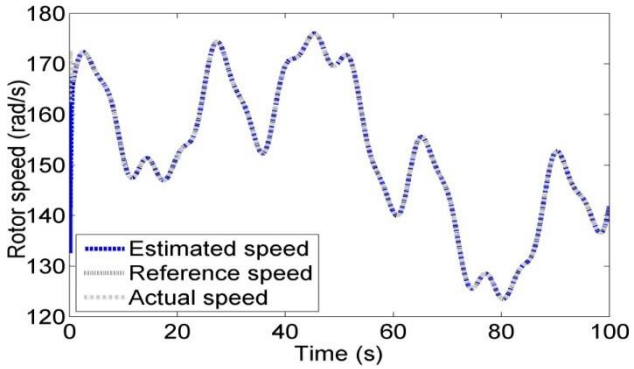
When the actual values do not correspond to their references, the closed loop system is exposed to failure, which destroys the control objectives. These results confirm the robustness of our controller regarding to the rotor resistance variation along with the aerodynamic torque.

5. Conclusion

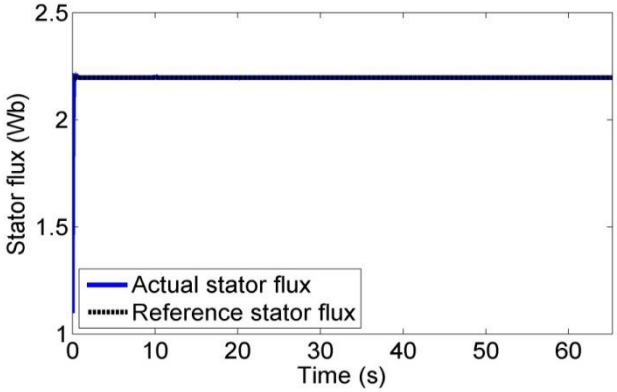
This paper presents an adaptive nonlinear sensorless controller for DFIG-WT to maximize the wind energy extracted under parametric uncertainties and unknown rotor speed. The simulation performed in Matlab/Simulink environment illustrates the good results provided by the controller regarding the good tracking capability of the rotor speed and stator flux, in addition to correct estimation of the unknown rotor speed, aerodynamic torque, and the rotor resistance. Also, the comparison result between the same controller with and without adaptation shows the effectiveness of this work.



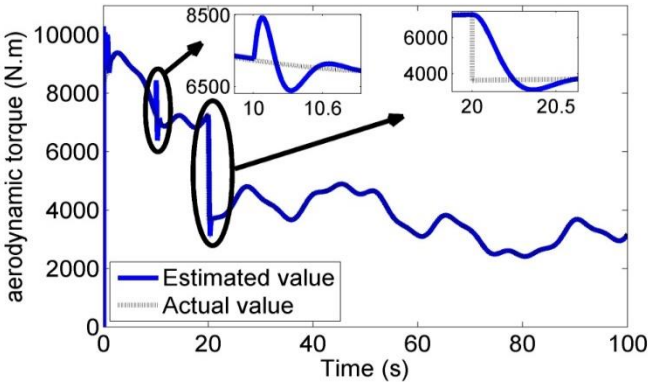
(a) Time varying wind speed



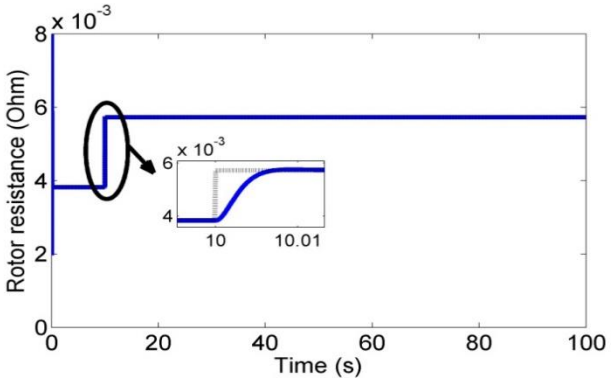
(b) Actual, Reference and estimated rotor



(c) The Actual and reference stator flux

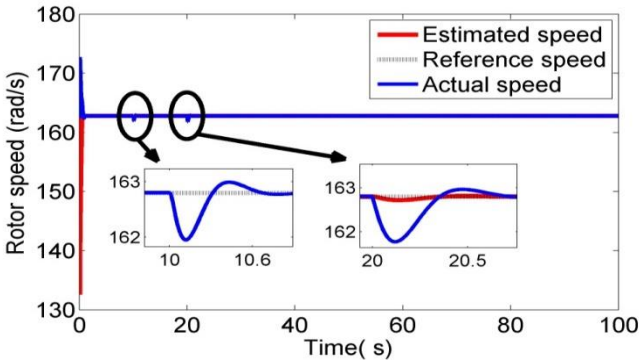


(d) The aerodynamic torque and its estimate

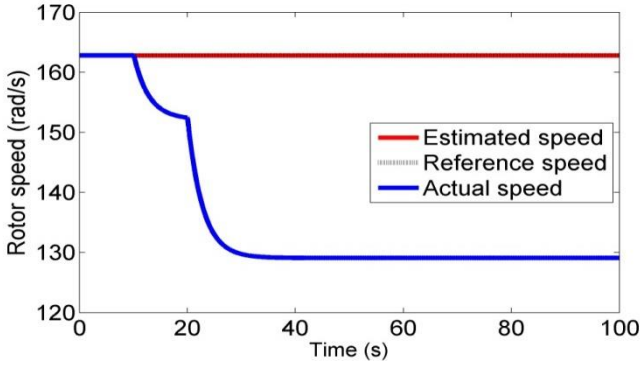


(e) The rotor resistance and its estimate

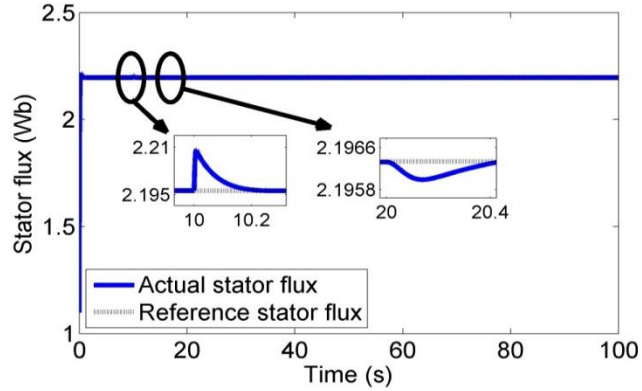
Figure 3: Comparison of tracking performance s under unknown parameters Changes occur at t=10s $\Delta R_r = 50\%$, t=10s $\Delta T_t = -50\%$ and at t=20s



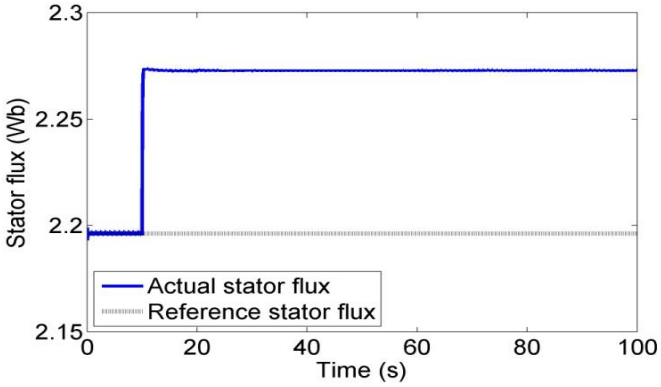
(a) Actual, Reference and estimated rotor speed with adaptation



(b) Actual, Reference and estimated rotor speed without adaptation



(c) The Actual and reference stator flux with adaptation



(d) The Actual and reference stator flux without adaptation

Figure 4: Comparison of tracking performances under unknown parameters
Changes occur at t=10s $\Delta R_r = 50\%$, t=20s $\Delta T_l = -50\%$

References

[1] G.M. As, "Doubly fed induction generator uising back-to-back PWM converters and its application to variable- speed wind-energy generation," 1996.

[2] S. Müller, M. Deicke, R.W. De Doncker, "Doubly fed induction generator systems for wind turbines," IEEE Industry Applications Magazine, **8**(3), 26–33, 2002, doi:10.1109/2943.999610.

[3] G. Boroumandjazi, B. Rismanchi, R. Saidur, "Technical characteristic analysis of wind energy conversion systems for sustainable development," ENERGY CONVERSION AND MANAGEMENT, **69**, 87–94, 2013, doi:10.1016/j.enconman.2013.01.030.

[4] A.B. Ataji, Y. Miura, T. Ise, H. Tanaka, "A Rotor-Current-Based Slip Angle Estimator for Grid-Connected Doubly Fed Induction Generator Requiring the Stator Inductance only," IEEE Transactions on Power Electronics, **32**(6), 4827–4838, 2017, doi:10.1109/TPEL.2016.2598441.

[5] L.Y. Lu, N.F. Avila, C.C. Chu, T.W. Yeh, "Model Reference Adaptive Back-Electromotive-Force Estimators for Sensorless Control of Grid-Connected DFIGs," IEEE Transactions on Industry Applications, **54**(2), 1701–1711, 2018, doi:10.1109/TIA.2017.2765300.

[6] M.W. Kalaga Mbukani, N. Gule, "Comparison of high-order and second-order sliding mode observer based estimators for speed sensorless control of rotor-tied DFIG systems," IET Power Electronics, **12**(12), 1–11, 2019, doi:10.1049/iet-pel.2018.6225.

[7] R.M. Prasad, M.A. Mulla, "Position-sensorless direct torque control of grid-connected DFIG with reduced current sensors," Sadhana - Academy Proceedings in Engineering Sciences, **45**(1), 2020, doi:10.1007/s12046-020-1275-x.

[8] H. Becheri, I.K. Bousserhane, A. Harrouz, I. Colak, K. Kayisli, "Sensorless Control of Wind Turbine Conversion Equipped with a DFIG Using MPPT Strategy," Proceedings - 2018 IEEE 18th International Conference on Power Electronics and Motion Control, PEMC 2018, (3), 605–610, 2018, doi:10.1109/EPEPEMC.2018.8522002.

[9] E.H. Abdou, A.R. Youssef, S. Kamel, M.M. Aly, "Sensorless Wind Speed Control of 1.5 MW DFIG Wind Turbines for MPPT," 2018 20th International Middle East Power Systems Conference, MEPCON 2018 - Proceedings, 700–704, 2019, doi:10.1109/MEPCON.2018.8635209.

[10] B. Yang, L. Jiang, L. Wang, W. Yao, Q.H. Wu, "Nonlinear maximum power point tracking control and modal analysis of DFIG based wind turbine," International Journal of Electrical Power & Energy Systems, **74**, 429–436, 2016, doi:10.1016/j.ijepes.2015.07.036.

[11] G.E.-P. and R.E.-P. López-García , F. Beltran-Carbajal, "A Review of RFO Induction Motor Parameter Estimation Techniques," **18**(2), 271–283, 2003.

[12] J.M. Mauricio, A.E. Leon, A. Gomez-Exposito, J.A. Solsona, "An Adaptive Nonlinear Controller for DFIM-Based Wind Energy Conversion Systems," IEEE Transactions on Energy Conversion, **23**(4), 1025–1035, 2008.

[13] A. Susperregui, I. Lizarraga, J. Jugo, G. Tapia, "Automated control of doubly fed induction generator integrating sensorless parameter estimation and grid synchronisation," IET Renewable Power Generation, **8**(1), 76–89, 2014, doi:10.1049/iet-rpg.2013.0045.

[14] R. Ourhdir, M. Rachidi, "Adaptive input-output feedback linearization control of doubly-fed induction machine in wind power generation," International Review of Automatic Control, **12**(1), 11–20, 2019, doi:10.15866/ireaco.v12i1.15619.

[15] J.M. Pinar Pérez, F.P. García Márquez, A. Tobias, M. Papaelias, "Wind turbine reliability analysis," Renewable and Sustainable Energy Reviews, **23**(March), 463–472, 2013, doi:10.1016/j.rser.2013.03.018.

[16] X. Jin, W. Ju, Z. Zhang, L. Guo, X. Yang, "System safety analysis of large wind turbines," Renewable and Sustainable Energy Reviews, **56**, 1293–1307, 2016, doi:10.1016/j.rser.2015.12.016.

[17] F. Liu, D. Li, P. Li, Y. Song, "Quadratic Optimal Tracking Control of Wind Turbine Systems," IFAC-PapersOnLine, **48**(28), 502–507, 2015, doi:10.1016/j.ifacol.2015.12.178.

[18] S. Heier, GRID INTEGRATION OF WIND ENERGY GRID INTEGRATION OF WIND ENERGY ONSHORE AND OFFSHORE.

[19] I. Press, L. Shafer, G.W. Arnold, D. Jacobson, ANALYSIS OF ELECTRIC MACHINERY AND DRIVE SYSTEMS.

[20] G. Abad, J. López, M.A. Rodríguez, L. Marroyo, G. Iwanski, Doubly Fed Induction Machine: Modeling and Control for Wind Energy Generation, 2011, doi:10.1002/9781118104965.

[21] R. Marino, P. Tomei, C.M. Verrelli, Induction Motor Control Design, 2010, doi:10.1007/978-1-84996-284-1.

[22] L. Praly, "Adaptive Nonlinear Regulation: Estimation from the Lyapunov Equation," IEEE Transactions on Automatic Control, **37**(6), 729–740, 1992,

doi:10.1109/9.256328.
[23] R. Marino, G.L. Santosuosso, P. Tomei, "Robust adaptive observers for nonlinear systems with bounded disturbances," IEEE Transactions on Automatic Control, **46**(6), 967–972, 2001, doi:10.1109/9.928609.
[24] R. Ourhdir, M. Rachidi, "Nonlinear sensorless control of doubly fed induction generator in wind power generation," 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET 2020, **2**(1), 2020, doi:10.1109/IRASET48871.2020.9091997.

Appendix A. Frame transformation and u_{ds} derivative

The direct and quadratic components of the stator voltage in the (d,q) reference frame can be expressed as follows:

$$\begin{pmatrix} u_{ds} \\ u_{qs} \end{pmatrix} = \begin{pmatrix} \cos\theta_0(t) & \sin\theta_0(t) \\ -\sin\theta_0(t) & \cos\theta_0(t) \end{pmatrix} \begin{pmatrix} u_{\alpha s} \\ u_{\beta s} \end{pmatrix} \quad (A.1)$$

where (α, β) stand for the fixed reference frame and θ_0 the electrical angle of the (d, q) rotating reference frame to the (α, β) one, see Figure 2. The components of the stator voltage in (α, β) are:

$$\begin{pmatrix} u_{\alpha s} \\ u_{\beta s} \end{pmatrix} = \underbrace{\sqrt{\frac{2}{3}} \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \sqrt{\frac{3}{2}} & -\sqrt{\frac{3}{2}} \end{pmatrix}}_{T} \begin{pmatrix} u_{as} \\ u_{bs} \\ u_{cs} \end{pmatrix} \quad (A.2)$$

In which T is the Concordia matrix transformation, and u_{as}, u_{bs} and u_{cs} are the 3-phases stator voltage giving by :

$$\begin{pmatrix} u_{as} \\ u_{bs} \\ u_{cs} \end{pmatrix} = V_m \begin{bmatrix} \sin(\omega_s t) \\ \sin\left(\omega_s t - \frac{2\pi}{3}\right) \\ \sin\left(\omega_s t + \frac{2\pi}{3}\right) \end{bmatrix} \quad (A.3)$$

Using the conventional derivative proprieties for scalar, vector and matrix, we can obtain the direct stator voltage component derivative $\frac{du_{ds}}{dt}$ used in (14). From (A.1) the expression of the direct stator voltage is:

$$u_{ds} = \begin{pmatrix} \cos\theta_0(t) & \sin\theta_0(t) \end{pmatrix} \begin{pmatrix} u_{\alpha s} \\ u_{\beta s} \end{pmatrix}$$

From (A.2), (A.3) and (10), the derivative of u_{ds} is given by:

$$\begin{aligned} \frac{du_{sd}}{dt} = {} & \begin{bmatrix} u_{as} & u_{bs} & u_{cs} \end{bmatrix} T^T \frac{u_{qs} + \alpha M i_{qr}^*}{\Psi^*} \begin{bmatrix} -\sin\theta_0(t) \\ \cos\theta_0(t) \end{bmatrix} \\ & + \begin{bmatrix} \cos\theta_0(t) & \sin\theta_0(t) \end{bmatrix} T \omega_s V_m \begin{bmatrix} \sin\left(\omega_s t + \frac{\pi}{2}\right) \\ \sin\left(\omega_s t - \frac{\pi}{6}\right) \\ \sin\left(\omega_s t - \frac{5\pi}{6}\right) \end{bmatrix} \end{aligned} \quad (A.4)$$

$$\cos\theta_0(t) = \frac{\varphi_{\alpha s}}{\sqrt{\varphi_{\alpha s}^2 + \varphi_{\beta s}^2}} \text{ and } \sin\theta_0(t) = \frac{\varphi_{\beta s}}{\sqrt{\varphi_{\alpha s}^2 + \varphi_{\beta s}^2}}$$

Appendix B. characteristics/parameters of 3 MW DFIG-WT and the adaptive sensorless controller

Table 2: Characteristics and parameters of DFIG-WT

Parameter	Definition	Value
	Induction generator	
P_{rat}	Rated power	$3\,MW$
U	Rated stator voltage	$690\ V$
ω_s	Stator angular frequency	314 rad/s
p	Number of pole pairs	2
R_r	Rotor winding resistance	0.00382 Ω
R_s	Stator winding resistance	0.00297 Ω
L_r	Rotor inductance	0.0122 H
L_s	Stator inductance	0.0122 H
M	Mutual inductance	0.01212 H
	Wind Turbine	
R	Blade Radius	$45\ m$
G	Gearbox ratio	100
$C_{p\max}$	Maximal Power coefficient	0.48
λ_{opt}	Optimal Tip Speed Ratio(TSR)	8.14
	Generator and Turbine	
J	Moment of inertia	$254\ kg.m^2$
f	Dumping coefficient	0.24
f_r	grid frequency	50 Hz
U	grid voltage	$690\ V$

Table 3: parameters of the adaptive sensorless controller

Parameter	Value
k	6000
k_ω	100
δ	70000
γ	0.0098
T_{tM}	10000
R_{rm}	$0.5R_r$
R_{rM}	$2R_r$
ε_1	$0.001R_r$
ε_2	5

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 497-506 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

An Economic Theory Perspective for the Fight Against Poverty in the Peruvian Andes

Robert Antonio Romero-Flores*

National University of the Altiplano, Department of Systems Engineering, Puno, 21001, Peru

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 05 November, 2020
Online: 20 November, 2020

Keywords:
Complex systems
Mathematical modelling
Economic theory of water
Systems simulation

ABSTRACT

The fight against poverty in the Peruvian Andes is a complex task in which various professionals, such as engineers, economists, anthropologists, among others, participate. The uncertainty of the decisions taken today, no matter how appropriate they may seem, such as million-dollar investments in irrigation infrastructure, can result in over-production and, therefore, in economic recessions. For this reason, a new mathematical simulation model is proposed using system dynamics to predict recession phenomena that can occur in months or after a few years of auspicious economic growth, and that can cause sales prices to be below production costs. The author has developed the conceptualization of the production system of irrigation improvement projects in several years of multidisciplinary work in the Cusco region of Peru. The primary objective of irrigation projects is to improve the socio-economic conditions of the farmer. Techniques as the fulfillment of goals have been used to quantify qualitative dimensions such as strengthening organizations and trainings that are key to guaranteeing irrigation improvement projects' sustainability in the long term. Therefore, it has been possible to identify the variables and relationships of this type of socio-economic system. To validate the model, we verified that the simulated data are consistent with the historical data collected. Likewise, if the values of the various proposed models' variables are adequately modified, these can be applied to other types of production systems under different market conditions. The dimensions addressed, such as supply, demand, sale price, land, production volume, public budget, etc., enhance the research's importance, making the simulation model formally expressed also acquire nuances from economic theory for the fight against poverty-based on water. One of the study's conclusions is to understand the production systems, it is necessary to see them in the context of their regional economy's behavior.

1. Introduction

One of the most important policies in the fight against poverty of the Peruvian government is implementing irrigation improvement projects. In this paper, we will address issues in greater depth about the experience of Plan Meriss Inka (PEPMI) to deal with the social component. We will highlight the characteristics that make this research an economic theory of water perspective that could not be exhibited in work originally submitted in the 2019 IEEE World Engineering Education Conference (EDUNINE) [1].

The primary nature of the irrigation improvement projects implemented by PEPMI is multidisciplinary, in which professionals of different natures such as engineers, economists, biologists, anthropologists, among others. One of the characteristics that stand out from this type of project is the high social component [2]. Since the first projects' implementation, it was observed that, despite providing rural communities with modern irrigation infrastructure. Farmers were reluctant to use it. A situation that has required social partners to decipher this peculiar impasse. Anthropologists' role in these cases has been essential to understanding the farmer's idiosyncrasy in terms of the intervention of new irrigation projects in their communities it refers.

From work carried out, it was concluded that the resistance to using the new irrigation infrastructure was because the farmer only uses the irrigation infrastructure that they have built and, therefore, they consider it as their own. So, not having participated in constructing this new irrigation infrastructure donated by the government, they did not consider it their own. To overcome this

*Corresponding Author: R.A. Romero-Flores, Urb. Chanu Chanu III S10, +51951885805, rromero@unap.edu.pe

problem, Community participation is incorporated as part of the project's financing, which consists mainly of the farmers' workforce's valuation in the construction of the irrigation infrastructure [3]. Similarly, there are cases of projects in which it has been possible to significantly improve water resources available without the need to build new irrigation infrastructure; it was only required to improve water management and irrigation techniques to increase agricultural productivity. This experience leaves an important lesson for professionals who are not from the social area, which is that, to achieve the objectives of the projects, they should improve their communication skills and understanding of the farmer idiosyncrasies in the area [3]. The social component is just one of the many problems. It gives an idea of the complexity of understanding and managing irrigation improvement projects in the fight against poverty in the Peruvian Andes. The ability and experience to face these peculiarities have made the PEPMI to be considered as a model project in South American by the German donors German Technical Cooperation (GTZ) and the State Development Bank of the Federal Republic of Germany (KFW).

In this understanding, the primary objective to be achieved with social intervention projects is their sustainability in strengthening the organization of farmers and profitability. In other words, farmers must manage their own resources without the accompaniment of PEPMI staff and only through its own organization and self-financed by the profit that results from the sale of their products.

In the Peruvian Andes, there is the occurrence of weather phenomena such as droughts and frosts that cause the loss of crops. The main effect of having adequate irrigation infrastructure is to ensure cultivation; firstly, supplying water in case of droughts and making it possible to recover plants from frosts. So that at least one harvest per year can be ensured. This is also known as crop safety. It is possible to increase the number of agricultural seasons per year by up to two in the best cases. Then, water administration is a strategy ancestrally used since the Incas to face unpredictable climatic changes and to guarantee food for the population [4].

To overcome these problems, millions are invested in infrastructure and training. But in the projects implemented in the Chumbivilcas-Espinar provinces, it has been shown that this does not always happen. Due to the overproduction of dairy products, there was a regional recession. A phenomenon in which supply exceeds demand and produces a drop in prices. In this case, prices fell even below production costs. So, the right short-term decisions can lead to long-term problems [5].

In this research, we are concerned with developing a production model to understand the factors that participate in it. The model simulation process has allowed the model to be validated. However, to know if the production process is profitable. It is required to evaluate it in the environment of its regional economy.

Therefore, it is also necessary to conceptualize and formulate economic factors. This gives greater importance to our research as it analyzes the Andes' economy from a perspective of fighting poverty. The methodology used to achieve our objectives is mainly system dynamics, as it includes the methods of thinking of systems and servomechanisms. The same ones that are suitable for complex systems.

As a result of the research, we have obtained the formulation of the different production sub-systems such as training, investment, available water, infrastructure, and others. And we have also managed to formulate the subsystems of the regional economy such as supply, demand, sale price, competition, etc. This makes it possible to project over time the impact of the irrigation improvement projects' anti-poverty policies. It also allows us to carry out a sensitivity analysis to know if the decisions we make in implementing the project are the correct ones, mainly to avoid regional recessions or falls in products sales prices [1].

2. Related Works

2.1. System dynamics

System dynamics is a popular simulation methodology developed by J. Forrester at the Massachusetts Institute of Technology (MIT). This simulation methodology is based on the theory of servomechanisms and feedback. It is also related to areas such as general systems theory and cybernetics. It also uses methods to study complex systems that act as an interlocutor between engineering methods and the methods for social systems [6]. One of the most important applications of System Dynamics is the world model published in 1970. A simulation of the behavior of society is shown under current "unplanned" growth conditions. The results of the simulation of the model show an over-exploitation of resources and a drop in population. The more population, the more waste is generated, and the more waste, the more diseases are generated is one of the conclusions of the model [7].

In the work of Fifth Discipline, considered to impact the business world greatly, revalues Forrester's systems dynamics. The author in [5] considers important for the understanding of modern organizations the study of dynamic systems, identifying the system's environment, and feedback mechanisms with it. He does not consider essential to arrive at the formulation of mathematical models. As a conclusion to his studies, he proposes the following disciplines: systems thinking, personal mastery, mental models, the construction of a shared vision, and team learning.

For irrigation research, dynamics can also be applied to the study of water management jobs, as demonstrated in the proposal for an optimization model for irrigation management in Australia [8]. In [9], the authors also developed a simulation model for China's water transport problem. In both works, positive results have been obtained using system dynamics.

2.2. Systems thinking

Systems thinking is the fifth discipline; the importance of systems thinking lies in being the body that unites the other disciplines of intelligent organizations. This job is incisive in pointing out that one of the factors for the failure of projects is the lack of systems thinking in their implementation [5]. Systems thinking is based mainly on the method of extension or synthesis. This becomes a premise for troubleshooting, where it is recommended to first view the system as part of a bigger system [10]. In this regard, we must comment that the classical scientific method, as we know, uses the analytical or reductionist method,

which, contrary to the synthetic method, what it does is studying the system in its components. The author in [11] categorically considers as a limitation of the scientific method. However, one of the most renowned authors on epistemology comments that systems thinking is an unfounded fashion issue and cannot generate scientific knowledge [12]. The author in [10] refuted this position, who maintains that the analytical and synthetic methods do not replace each other. On the contrary, they complement each other. In the present work, we show that it is possible to generate knowledge in the scientific method's formal language, between the system and its environment (exogenous variables) and the feedback mechanisms for adjusting the endogenous variables.

2.3. Fight against poverty

Thus, water is the engine for the economy of the Andes, although its cost is relatively low given its usefulness. To highlight the importance of water, we will mention the water-diamond paradox that tries to explain the low cost of water concerning diamonds that cannot generate life. This explains that, although the marginal utility of diamonds is much greater than the marginal utility of water, water's total utility is always greater than the total utility of diamonds [13].

To eradicate poverty, the author in [14] comments that the financial system has been saved on a global scale; for example, in Mexico, the banking system has been saved. Consequently, the cost of helping the poor is much lower and, in financial terms, more profitable if they are to be part of the aggregate demand. And, despite extreme poverty, we seem to have plenty of resources. At the same time, another study concludes that almost a billion people go to sleep hungry every day. In comparison, another billion people suffer from obesity, and 30% of food production is wasted, even from its mismanagement in harvest, sales, and post-consumption [15].

3. Methodology

The experience that served as inspiration for this work was that of the irrigation improvement projects in the Chumbivilcas-Espinar provinces that, due to their height (over 3820 m.a.s.l.) and geography, these are areas where low temperatures predominate (even below zero), all this leads to the existence of extreme poverty in the area [16]. Conditions that, in the first instance, made interventions in the so-called “high” provinces impractical. However, initially, they gave unexpected positive results thanks to the experience of the PEPMI and the adaptation of strategies that consisted, mainly, in cultivating imported pastures that resisted low temperatures and that were the main engine of the livestock industry and its derivatives in the area. Which brought the economic growth of the area. Despite this, after a few years, there was overproduction that manifested itself in the drastic decrease in sales prices. Therefore, it is necessary to manage uncertainty, the consequence of the decisions we make, and that, at first, seem the correct ones in the medium and long term, can become counterproductive. The explanation for this phenomenon was that an adequate market study was not carried out. Thus, to understand the interaction of the multiple factors, it was necessary to understand the behavior of the market and the mechanisms by which the sale price was mainly obtained. In an economic theory, appears the next components: consumers, producers, the market for goods and factors, and government intervention [13]. To formulate the simulation model of the production system, all these dimensions have been taken into account. Therefore, in this document, we will extend the original work to a perspective of economic theory whose main engine is water. In Figures 1 and 2, we can see the Sutunta lagoon and the dam built to store 4 million cubic meters to irrigate 6000 hectares.

Figure 1: Location map of the Sutunta lagoon in the province of Espinar-Cusco over 4000 m.a.s.l.

Figure 2: The dam built in Sutunta lagoon

In the conceptualization and formulation of the new production model, the systems dynamics methodology was used. The same one is based on the general theory of systems proposed in [11], the author that deals mainly to provide a general framework for science. To do this, it recognizes the concept of equilibrium or homeostatic point. The same that we must reach in the interaction of supply and demand in the market model also proposed [1]. Nash’s equilibrium abstracts the theory of non-cooperative games that involve sellers' participation, sets of strategies, and profits [17].

In the conceptualization phase, approximately 285 variables have been identified. The same ones have been organized in the dimensions of investment, training, organization, water for irrigation, irrigation infrastructure, production, market, environment, public budget, and water supply. As mentioned

above, the organizational dimension is primarily qualitative, and techniques have been used to allow it to be considered and provide qualitative information. This dimension has been adequate to be considered in the proposed production model. Another characteristic is that the production model allows obtaining a production volume that is considered supply and considering the existence of competition. Therefore, it is not an empirical model for educational purposes [6]; otherwise, it can provide information in a real environment. The number of variables and the peculiarities described makes it different from the authors' production models in [18], who propose a model of positive mathematical programming for models of regional production in agricultural-environmental programs and the classic spider web model [13].

The simulation of the production model has been carried out in the VENSIM program; the simulation results have served to validate the model. For validation, the Anderson-Darling normality test was used, whose results have been superior to $\alpha = 0.05$. As a result of the optimization model, non-linear programming with restrictions has been used to develop an optimization model that allows us to find the global optimum (profit maximization, cost minimization, and equilibrium point). Likewise, an analysis was carried out on the optimization model using nonlinear programming and genetic algorithms to determine the selection operators' efficiency in restricted non-linear problems.

4. Discussion

4.1. Conceptualization of the elements of the economic system for the fight against poverty based on water

In South America, the Inca culture developed. One of the most important cultures in the world whose cultural richness was based on values such as community work, also known as the "ayni" and the "minka". The Incas built large hydraulic works with the primary purpose of supplying food to the entire population. The importance of institutions like the PEPMI lies in need to revalue the cultural richness that, in some way, was being forgotten. In the field of intervention, they can be observed as problematic situations: inadequate irrigation infrastructure (inland canals that filter water), little farmer organization, and little knowledge of irrigation techniques [4]. This situation can even be complex if the following characteristics exist: conflicts, undefined access rights to water, water scarcity, and geological problems [3].

To solve these problematic situations, PEPMI implements irrigation improvement projects whose objective is to provide the population in the project area with the possibility of improving production and, therefore, improve their quality of life through access to more and better education, health, etc. services.

4.1.1. Production systems modeling in irrigation improvement projects

Consistent with the fifth discipline [5], a work in which the author concludes that one of the causes of project failure is the lack of a systems approach. If we review the experience of the projects implemented in Chumbivilcas-Espinar, in which there were problems of regional recession, after a few years of economic growth, it can be seen that, as part of the bigger system, the "market" is an exogenous component that is always present in the environment of the system. So, it is necessary to carry out an adequate market study and, for simulation purposes, to determine its feedback mechanisms. In figure 1, it can be seen through a block diagram, the main identified subsystems. Given the complexity of these projects' implementation, approximately 285 variables and 90 feedback loops have been observed and properly grouped into subsystems, as shown in Figure 3 [1].

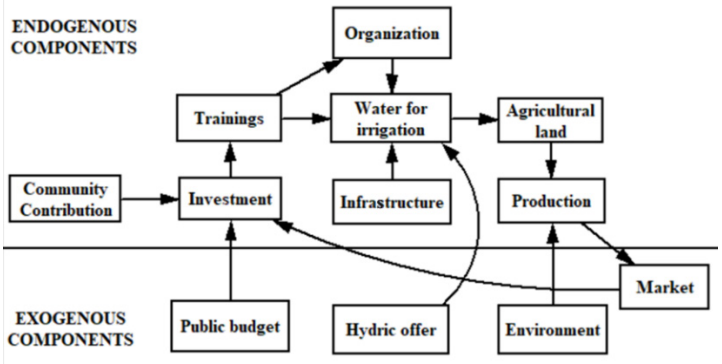


Figure 3: Block diagram of the proposed model for the agricultural production system

Figure 1 also identifies the necessary elements that participate in an economic theory (producers, consumers, goods markets, state intervention, etc.), having, in this case, water as the engine for the fight against poverty [13] or as "Theory of the economic and social value of water" states in a perfect market where there are several buyers and sellers, each commodity in the economy will be given its value [19]. As previously mentioned, the sustainability of projects is based on the strengthening of organizations and profitability. It is, then, necessary to know the profitability to identify the elements that participate under the economic theory defended by various authors [13].

4.2. Formulation and validation of the simulation model

The first objective of the work is to know the future behavior of the production system of the irrigation improvement projects implemented by the PEPMI. So, the planners will have a tool that allows them to know if they are making the right decisions or if the millionaire investments in infrastructure can cause short-medium or long-term adverse effects such as a recession. These phenomena have occurred in the provinces of Chumbivilcas-Espinar and the northern mountain range of Lima [20]. The causal diagrams made in the conceptualization phase have to be converted to mathematical logic models based on Forrester diagrams. For which the Vensim Personal Learning Edition (PLE) software has been used.

4.2.1. Subsystems formulation

Each of the subsystems shown in figure 1 are explained below:

A. Water supply

Determining the future supply of water is one of the main tasks to know. This to guarantee the water supply at the head of the plot. To achieve this, the historical information of the Pampaconga project has been considered. After analyzing the information, it has

been considered to perform a probability function for each month, using discrete probability distribution functions based on histograms. Requirements for the Monte Carlo simulation method [21]. Due to the limitations of Vensim Personal Learning Edition (PLE), the treatment of each probability function has been carried out in Excel and the results added in lookup functions in the implemented model.

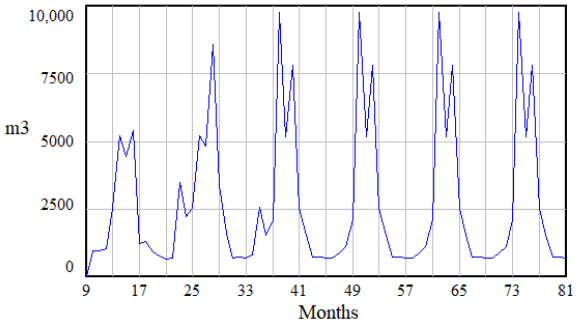


Figure 4: Projection of water availability due to rainfall

As shown in Figure 4, the hydrological year in Peru begins in September (month 9), and the high peaks represent the El Niño phenomenon. With climate change, this phenomenon will be increasingly present in the Peruvian Andes. When this occurs within the project's scope, the loss of production, and even irrigation infrastructure and all kinds of infrastructure is almost inevitable [22].

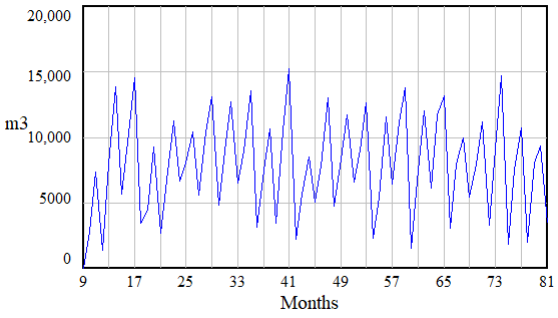


Figure 5: Projection of total water availability

The total water availability is made up of the rains and the flow of the rivers. These are exogenous variables that cannot be influenced. The projection of water availability is shown in Figure 5.

B. Investment

This subsystem is born from the public budget and external debt. For external indebtedness, there is German technical cooperation. According to the experiences of the first works implemented, the farmers showed resistance to change, to use and manage the modern irrigation infrastructure implemented by PEPMI. In addition, there is also the possibility of conflicts between communities over access to water and/or land. This demonstrates the high social component of the irrigation projects implemented [3]. Sometimes the engineer has to act as a sociologist to meet its goals [12]. To overcome these problems, the multidisciplinary team includes anthropologists, professionals who serve as important interlocutors between farmers and PEPMI.

In conclusion, the resistance to the change presented to use the new and free irrigation infrastructure was due; because farmers only use the infrastructure in whose construction they have participated. Therefore, project financing includes the community contribution, which is approximately 15 to 20% of the total project budget. This amount guarantees the participation of community members in the works to be built [3].

Another characteristic to take into account is that the Peruvian government considers irrigation improvement projects as social programs. And, given the appearance of economic problems, as the global pandemic caused by Covid-19, these programs are the first to reduce the public budget.

In the conceptualization of the production model in its interaction with the market, the investment subsystem receives feedback from the market, specifically, of the profit obtained from the sale of the products and the availability of financing that predisposes the advance of trainings and irrigation subsystems infrastructure.

C. Trainings

Training is an essential component to guarantee the sustainability of the project in its component of strengthening organizations and in terms of increasing productivity in the components of water management and the application of irrigation and cultivation techniques.

The PEPMI, being a state project, depends on the monthly budget. According to the monthly budget, it has the necessary factors to carry out the training: operational equipment, material, and workforce.

Figure 6 shows the Forrester diagram that explains the availability of resources to conduct the trainings.

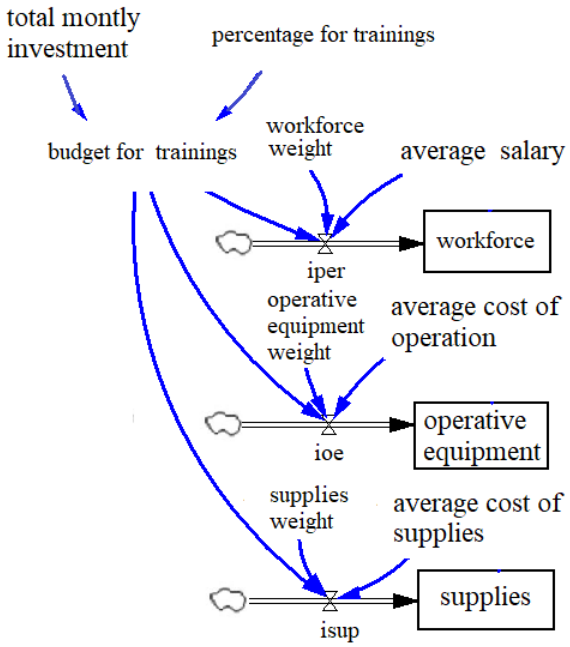


Figure 6: Forrester diagram: Availability of resources for training

The equations that govern the Forrester diagram in Figure 6 are as follows:

$$IPER = (BT*WW)/AS \quad (1)$$

where:

IPER: Personal ratio.

BT: Budget for trainings.

WW: Workforce weight = 0.5.

AS: Average salary = S/ 2000.00.

Percentage for trainings = 0.15.

$$IOE = (BT*OEW)/ACO \quad (2)$$

where:

IOE: Operative equipment ratio.

BT: Budget for trainings.

OEW: Operative equipment weight = 0.3.

ACO: Average cost of operation = S/ 1000.00.

$$ISUP = (BT*SW)/ACS \quad (3)$$

where:

ISUP: Supplies ratio.

BT: Budget for trainings.

SW: Supplies weight =0.2.

ACS: Average cost of supplies = S/ 300.00.

The trainings are carried out according to the availability of resources. The progress ratio is calculated according to the input targets programmed versus those available. And the total training per month predisposes the increase of water management, efficiency in the application of irrigation techniques, efficiency in the application of cultivation techniques, and strengthening of organizations that are governed by the following equations:

$$WM = 0.065*TPM \quad (4)$$

$$EAIT = 0.06*TPM \quad (5)$$

$$EACT = 0.07*TPM \quad (6)$$

$$SO = 0.065*TPM \quad (7)$$

where:

TPM: Trainings per month (the objective is 4 per month).

WM: Water management.

EAIT: Efficiency in the application of irrigation techniques.

EACT: Efficiency in the application of cultivation techniques.

SO: Strengthening of organizations.

As can be seen in Figure 7, in month 28, approximately the objective is achieved. This indicates that the trainings have made the farmer achieve 100% adequate water management. Similar behavior is observed in the variable's efficiency in applying irrigation techniques, efficiency in applying cultivation techniques, and strengthening organizations.

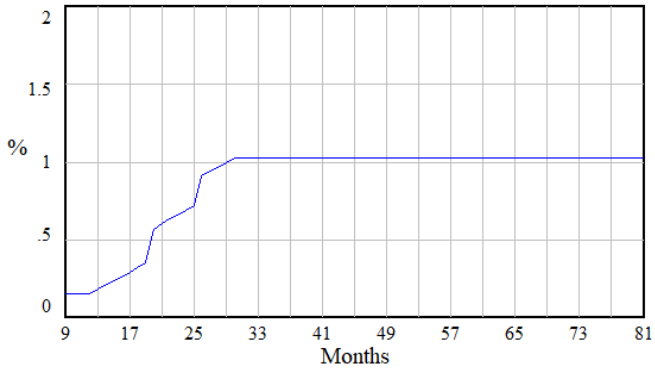


Figure 7: Advancement in water management

D. Irrigation infrastructure

Irrigation infrastructure is the component with the highest cost, depending on the size of the project. Its importance lies in the possibility of transporting water for kilometers without loss due to seepage and optimizing its storage and availability through dams and reservoirs. Advantages that the channels made on land do not have due to the high filtration rate. Irrigation infrastructure is also important to face adverse weather phenomena such as droughts, as it allows managing water scarcity, guaranteeing harvest (crop safety) and counteract adverse weather phenomena such as frost and hailstorms, since having a permanent water supply crops can recover, and, in areas where optimal conditions exist, the number of crops per year can be increased. What is also known as an increase in land use, which is the first step to fight poverty.

Then, the countries need to provide adequate irrigation infrastructure to manage the scarce water resource. This is because the imminent effects of climate change will affect primary sectors such as fisheries and agriculture, such as by 6% to the Peruvian GDP for the year 2030. One of the author's conclusions in [23] is that the emission of polluting gases and, therefore, the effects of climate change will continue. This situation evidently increases the total utility of water as part of managing scarcity and climate variability. Then, new efficient irrigation techniques such as dripping and sprinkling should also be considered. These aspects are already being considered in the making of water and irrigation laws.

Figure 8 shows the Forrester diagram that explains the availability of resources to carry out the irrigation infrastructure.

The equations that govern the Forrester diagram in Figure 8 are as follows:

$$Percentage\ for\ infrastructure = 0.85$$

$$ITS = (BI*TSW)/TSS \quad (8)$$

where:

ITS: Technical stuff ratio.

BI: Budget for infrastructure.

TSW: Technical stuff weight =0.06.

TSS: Technical stuff salary = S/ 2000.00.

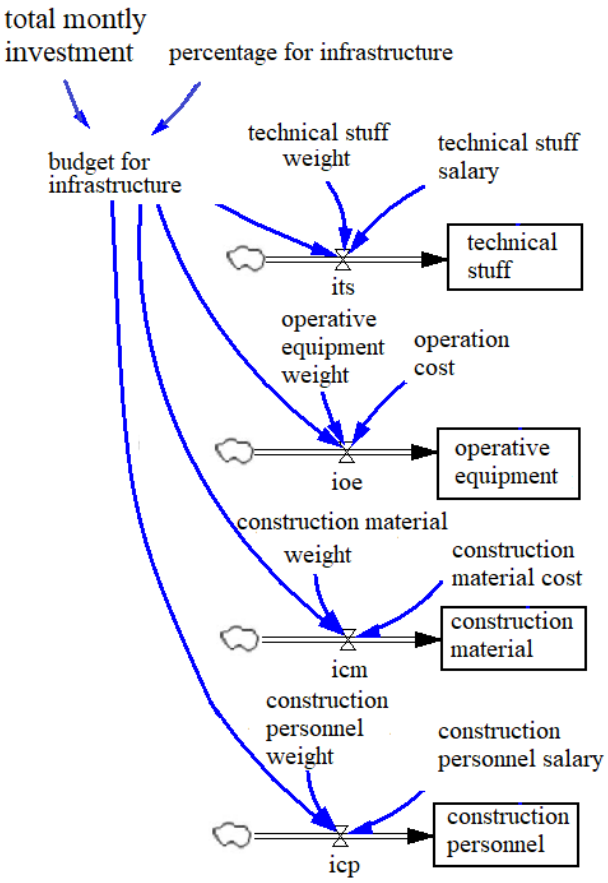


Figure 8: Forrester diagram: Availability of resources for irrigation infrastructure

$$IOE = (BI*OEW)/ACO \quad (9)$$

where:

IOE: Operative equipment ratio.

BI: Budget for infrastructure.

OEW: Operative equipment weight = 0.24.

ACO: Average cost of operation = S/ 4000.00.

$$ICM = (BI*CMW)/CMC \quad (10)$$

where:

ICM: Construction material ratio.

BI: Budget for infrastructure.

CMW: Construction material weight =0.11.

CMC: Construction material cost = S/ 15.00.

$$ICP = (BI*CPW)/CPS \quad (11)$$

where:

ICP: Construction personnel ratio.

BI: Budget for infrastructure.

CPW: Construction personnel weight =0.59.

CPS: Construction personnel salary = S/ 800.00.

Once the necessary resources are in place to carry out the irrigation infrastructure, the progress of the works depends on the availability of resources versus what is programmed and the weather conditions that allow the construction to continue. The construction progress is shown in Figure 9.

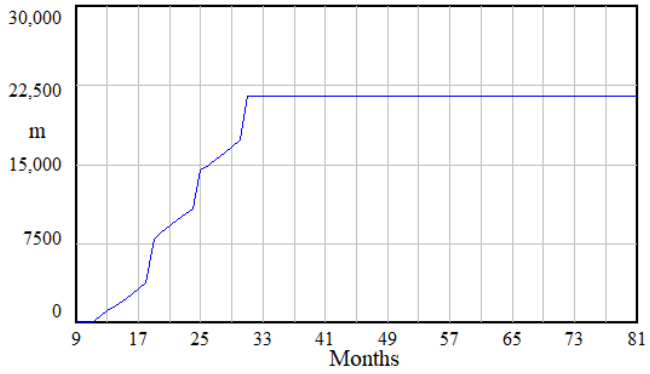


Figure 9: Progress in the construction of irrigation infrastructure

E. Water for irrigation

According to the progress of the irrigation infrastructure and the administration of water, the water for irrigation is determined according to the water demand calculated in the study phase. Figure 10 shows the Forrester diagram that explains the interaction of variables that determine the availability of water for irrigation.

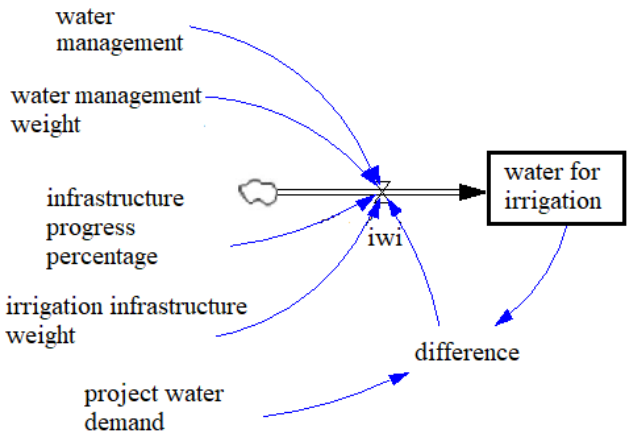


Figure 10: Forrester diagram: Water for irrigation

The equations that govern the Forrester diagram in Figure 10 are as follows:

$$IWI = [(D*WMW)*WM] + [(D*IIW)*IPP] \quad (12)$$

where:

Project water demand: 17,172.00 m3

D=PWD-WI

PWD: Project water demand.

WI: Water for irrigation.

WMW: Water management weight

WM: Water management.

IIW: Irrigation infrastructure weight

IPP: Infrastructure progress percentage

$$Water\ for\ Irrigation\ (t) = \int IWI(t)dt \quad (13)$$

where:

IWI: Water for irrigation rate

Figure 11 shows the increase in irrigation water until the project water demand is met, an important parameter that also serves to design the irrigation infrastructure (channels, reservoirs, canoes, rapids, etc.). The expansion of the agricultural frontier has a similar behavior within the project's scope as planned.

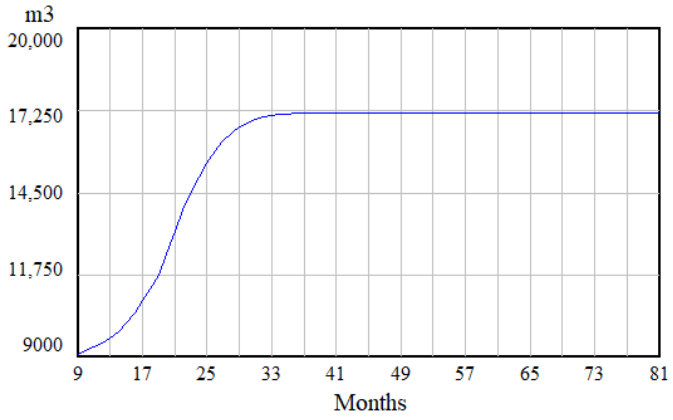


Figure 11: Increase in water for irrigation

F. Production function

The proposed simulation model for production in irrigation improvement projects includes the variables of water, supplies, training, machinery, and the human factor, whose interaction determines the quality and quantity of the cultivated land [1]. This is expressed in equation 14:

$$Adequate\ land\ for\ cultivation = \quad (14)$$

$$\begin{cases} 0 & ; WF \leq 0 \\ [WF*WWF+SF*SFW+TF*TFW+MF*MFW+HRF*HRFW]*PL & ; WF > 0 \end{cases}$$

where:

WF: Water factor.

WWF: Water factor weight (0.6).

SF: Supply factor.

SFW: Supply factor weight (0.1).

TF: Trainings factor.

TFW: Trainings factor weight (0.15)

MF: Machinery factor.

MFW: Machinery factor weight (0.1).

HRF: Human resource factor.

HRFW: human resource weight (0.05).

PL: Production land.

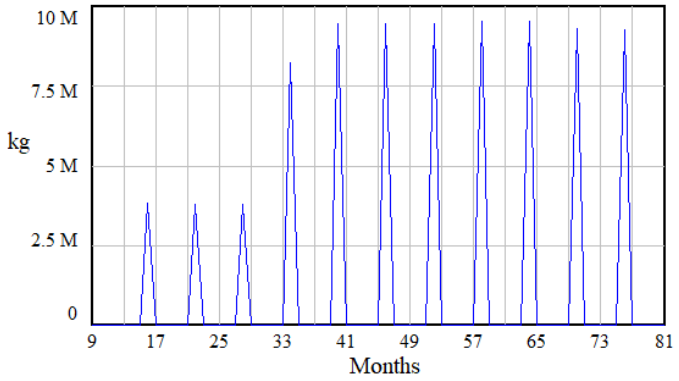


Figure 12: Production volume

G. Market

Once the production model has been obtained to determine the sustainability of the project, we must determine if there is a profit for the farmer due to sales in local markets. At this point, the proposed work requires to know the behavior of the market whose proposed model is of perfect competition [13].

Therefore, the present research becomes a broader study of the regional economy and an economic theory approach to fighting poverty, in this case, based on water as the main engine. Whose total utility in the medium term will allow farmers to access better living conditions such as health, housing, education, etc. Among these factors, we must highlight the greater and better access to education of the Andean population, as it will allow access to better economic income. About education, the winner of the last Nobel Prize in economics, in his work in Kenya,

concludes that better innovations must be made that allow more people to benefit economically from education [24].

H. Price determination

The interaction of supply and demand (which we have considered dividing into current demand and potential demand) and the seasonality of both allows us to know the product's sale price. Compared to the original work [1], in the present, we are going to present a new model described in equation 15.

$$Price = AP + [AP*(1-(SM+SMF))] \quad (15)$$

where:

AP: Average price.

SM: Satisfied market.

SMF: Stocks in the Market Factor.

$$Satisfied\ Market = Production\ volume / Current\ demand \quad (16)$$

Figure 13 shows the sale price behavior, in which it can be seen that the price reaches S/ 1.23 in times of scarcity and S/ 1.00 when there is overproduction. To carry out a sensitivity analysis, a scenario was constructed. The project's production increases by 30%, and due to market capacity (demand) occurs overproduction and, therefore, the sale price drops to S/ 0.66 per kg. These results, in a real environment, could occur in a few months or years. Note that the simulation is for 72 months, and, during the year, there are various seasons. This has been considered from the conceptualization of production and the market model. It represents an advantage and differentiates our proposal in relation to other production and market models such as the spider web.

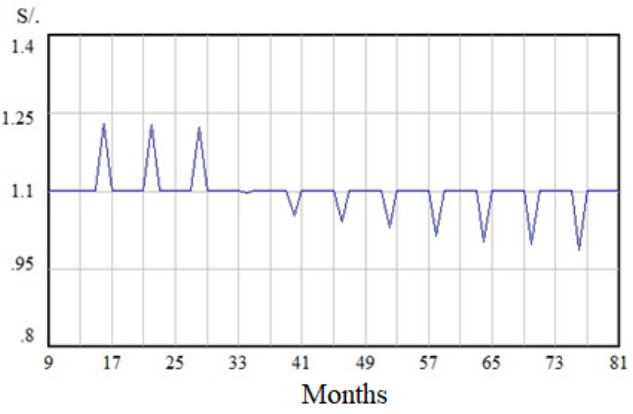


Figure 13: Sale price behavior obtained in simulation

4.3. Production optimization

Another aspect contemplated in economic theory is the maximization of producers' profits, who need to maximize the difference between total income and total cost. The break-even analysis allows us to know the minimum level of sales we must reach to avoid losses. Regarding the minimization of costs, the author in [25] recommends that this is the area we should focus on because it is difficult for producers to influence the sale price. Who governs us and defines the sale price is the market. Therefore, it is essential to reduce costs by implementing modern management techniques within organizations to obtain greater benefits. To define the optimization models, as observed through this work, we have the great restriction of the market (competition, seasonality, weather conditions, etc.). Therefore, it has been determined that the optimization of production is a restricted problem of the non-linear type. The objective function to maximize utility is shown in equation 17.

$$Max\ Z=16\varepsilon X_1 X_2 - 11.14X_1 \quad (17)$$

Subject to:

$X_3 \leq 3600+16X_1$; *Stock on the market*

$X_3-16X_1 \leq 3600$

$X_2 \leq AP+AP*[1-16\ X_1/X_3+X_4]$; *Sale price*

Replacing and solving:

Being: AP = S/ 1.10 (Average price)

$X_2 \leq 2.2-17.6X_1/X_3+1.1X_4$

$X_1 \leq 400$; *Arable area*

$X_4 = 0.1$

Solving, it is obtained that the maximum profit is S/ 5 784,000.00 and the sale price results S/ 1.6 (X_2).

5. Conclusions

The new proposed simulation model allows for understanding the behavior of the production process of irrigation improvement projects. In this paper, the systems dynamics methodology has been used mainly for its formulation. The new model allows the projection of data in the future and, in this way, determine the impact of the decisions that are made and if they do not lead to long-term problems. In the conceptualization phase, it has been possible to identify endogenous and exogenous variables; Among the exogenous variables, the behavior of regional markets stands out, determining the sale price of products. For the validation of the model, the Anderson-Darling normality test has been used for all the identified subsystems (training, strengthening of organizations, irrigation water, farmland, etc.). We obtained from the normality tests values higher than $\alpha = 0.05$.

The production factors that we identify for irrigation improvement projects are water, inputs, training, machinery, labor, and land. The respective weights can be seen in equation 14. The irrigation infrastructure and water management determine the availability of water. All these factors in the model determine the amount of cropland and the production volume to obtain. However, to measure the profitability of production, it is necessary to know its efficiency concerning the regional economy. That's why we formulate models that allow understanding of the market's behavior and sales prices in the region. The proposed models also make it possible to carry out sensitivity analyzes and

detect the appearance of regional recession phenomena. All these models have been expressed in a formal language, a requirement of epistemology for the generation of knowledge. Then our study becomes a perspective of economic theory for the fight against poverty based on water and its correct use.

For the formulation of optimization models, the nonlinear programming techniques based on objective functions and restrictions have been considered. The dimensions evaluated were: profit maximization, cost minimization, and breakeven point. The proposed production and market models have become constraints or objective functions as appropriate in the optimization model. To validate the optimization model, it has been verified that they comply with the Kuhn Tucker conditions. The optimization problem has also been analyzed using genetic algorithms, whose application has made it possible to know that the sexual selection operator is the most efficient for achieving global and local optimums.

The difficulties we had were access to information and the models' calibration, which is a process difficult and slow in which VENSIM has been very helpful. In future works, we will apply the proposed models in the Titicaca lake basin, and we will contrast them with time series.

References

[1] R.A. Romero Flores, "Production Model for Irrigation Improvement Projects Implemented in Cusco - Peru Using System Dynamics," in 2019 IEEE World Conference on Engineering Education (EDUNINE), 1-6, 2019, doi:10.1109/EDUNINE.2019.8875796.

[2] F. Hurtado Huamán, What You Should Remember When Formulating a Rural Development Project, Editorial of the National University of San Antonio de Abad del Cusco, 2014.

[3] Plan Meriss Inka, The Water in Our Hands, Gráficos S.R. Ltda., 1995.

[4] GPER INKA, Training in Water Management and Management, Bizz Editores EIRL, 1995.

[5] P. M. Senge, The Fifth Discipline: The Art & Practice of The Learning Organization, Doubleday, 2006.

[6] J.D. Sterman, Business Dynamics: Systems Thinking and Modelling for a Complex World, McGraw-Hill, 2000.

[7] U. Bardi, The Limits to Growth Revisited, Springer-Verlag, 2011.

[8] A. Elmahdi, H. Malano, T. Etchells, K, Shahbaz, "System Dynamics Optimization Approach to Irrigation Demand Management," in 2005 MODSIM International Congress on Modelling and Simulation, 196-202, 2005, doi:10.13140/2.1.3408.4648.

[9] Z.Zhang, W.X.Lu, Y. Zhao y W. Song, "Development tendency analysis and evaluation of the water ecological carrying capacity in the Siping area of Jilin Province in China using system dynamics and analytic hierarchy process", Ecological Modelling, **275**, 9-21, 2014, doi: 10.1016/j.ecolmodel.2013.11.031.

[10] R. Ackoff, Redesigning the Future, Limusa Noriega, 1994.

[11] L. von Bertalanffy, General Systems Theory, George Braziller Inc., 2015.

[12] M. Bunge, Science, its method and philosophy, Laetoli, 2018.

[13] W.B. Zhang, The General Economy, Springer Cham, 2020.

[14] J. F. Gonzáles Bermúdez, "Income distribution and poverty in Latin America: The cases of Argentina, Brazil, Chile and Mexico. Mexico DF," PhD Thesis, National Autonomous University of Mexico, 2013.

[15] R. Tirado, Greenpeace, Ecological agriculture, Greenpeace Internacional, 2015.

[16] V. Laime Mantilla, Chumpiwillka Nation, Súper Gráfica, 2013.

[17] N.J. Forbes, "Equilibrium Points in N-person Games," Proceedings of the National Academy of Sciences of the United States of America, **36**(1), 48–49, 1950, doi:10.1073/pnas.36.1.48.

[18] O. Rohm, S. Dabbert, "Integrating Agri-Environmental Programs into Regional Production Models: An Extension of Positive Mathematical Programming," American Agricultural Economics Association, **85**(1), 254-265, 2003, www.jstor.org/stable/1244941.

[19] F.H. Arceo, "Theory of the Economic and Social Value of Water Resources", Master´s Thesis, National Autonomous University of Mexico, 2006.

[20] Ministry of Agriculture and Irrigation, Potato: Characteristics of National Production and Marketing in Metropolitan Lima, MINAGRI, 2017.

[21] V.P. Singh, System Modeling and Simulation, new age international (p) limited, 2009.

[22] A Klauer, El Niño La Niña: The ocean-atmospheric phenomenon of the South Pacific, a challenge for science and history, ISBN: 9972–817–09–1, 2000.

[23] P. Vargas, Climate Change and its Effects in Peru, Central Reserve Bank of Peru, 2009.

[24] E. Duflo, P. Dupas, M. Kremer, "School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools," Journal of Public Economics, **123**, 92-110, 2015, doi: 10.1016/j.jpubeco.2014.11.008.

[25] E. Milan, M. Stefanovic, A. Djordjevic, N. Stefanovic, M. Misic, N. Abadic, P. Popovic, "Optimization of Production Process Parameters with a New Model Based on Genetic Algorithms and ABC Classification Methods," Advances in Mechanical Engineering, **8**(8), 1-18, 2016, doi:10.1177/1687814016663477

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 507-513 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Microcontroller Based Data Acquisition and System Identification of a DC Servo Motor Using ARX, ARMAX, OE, and BJ Models

Mokhlis Salah-eddine*, Said Sadki, Bahloul Bensassi

GITIL Laboratory, Department of physics, Faculy of sciences Ain Chock Hassan II University, Casablanca, 20070, Morocco

ARTICLE INFO

Article history:
Received: 27 August, 2020
Accepted: 05 November, 2020
Online: 20 November, 2020

Keywords:
System identification
polynomial models
PEM
Data acquisition card

ABSTRACT

This paper is carrying out a practical identification of a DC servomotor from input and output data. The system's step response was obtained from an acquisition card that we have developed using a microcontroller. The data is collected and downloaded to the PC Thought RS232 Protocol to process the identification process using polynomial models (ARX, ARMAX, OE, and BJ). Afterward, a comparison between the resulting models has been made using the validation criterions R^2 and FPE. The results show that the ARMAX, OE, and BJ models have reproduced a better fit compared to the ARX model.

1. Introduction

Servomotors have been developed at the origin of different uses. For example, radio-controlled aircraft, biped robots, and robotic arms. And nowadays, they are widely used in the robotics field, due to their cost, compactness, ease of use, and their high torque-to-weight ratio, which made them a good choice to achieve a compact and less expensive mechatronic system [1]. However, DCSM does not provide a good position control when it is in contact with a load and sometimes it provides a chattering motion. Thus, a mathematical model needs to be built to analyze and develop its performances. To do so, we tried to identify the mathematical equation of the system based on polynomial models.

A mathematical model is a multitude of equations that describe the relationship between the variables of a process or a phenomenon. It can take many forms depending on the studied system, and we can distinguish between two major classes.

1.1. Theoretical models

Theoretical models, Also known as whit-box, are models, which mostly used for simulations and theoretical studies. They are based on mathematical and physical laws, which made them often complex and hard to define [2]; however, it can give a complete description of the system.

1.2. Experimental models

Experimental models, also known as grey-box or black-box models are models used to define a mathematical model from the experiment's data. As mentioned in [3], in system identification, we could distinguish between two types.

- non-parametrical model is usually presented with numerical values, a graph, or a table, and it has an infinite number of parameters and an undefined structure.
- parametrical models, which are defined models structure who have a finite number of parameters.

According to [4], identifying a system or a process is about estimating a mathematical model that reproduces a static and dynamic response to a test input as close as possible to the real system. To do so, and to get a valid model, we need to follow a certain procedure [5], described in fig.1. It starts with the experiment preparation by choosing the sampling frequency and the input signal as a first step. Then choosing the model structure type, order, and delay as a second step. After that, selects the criterion that minimizes the error between the estimated model and the experimental data as the third step. Finally, the validation step that consists of applying several tests to prove the usability of the model. The resulting models' main objectives can be to understand the behavior of the studied system [6], or building the system control loop [7].

In this paper, Sections 2 and 3 cover a brief introduction of the linear polynomial models and parameter estimation process. Section 4 and 5 highlight the necessary steps for the implementation of the application. In sections 6 and 7, the results obtained from the experiment are presented and discussed.

*Corresponding Author: Mokhlis Salah-eddine, Email: mokhlis.id@gmail.com

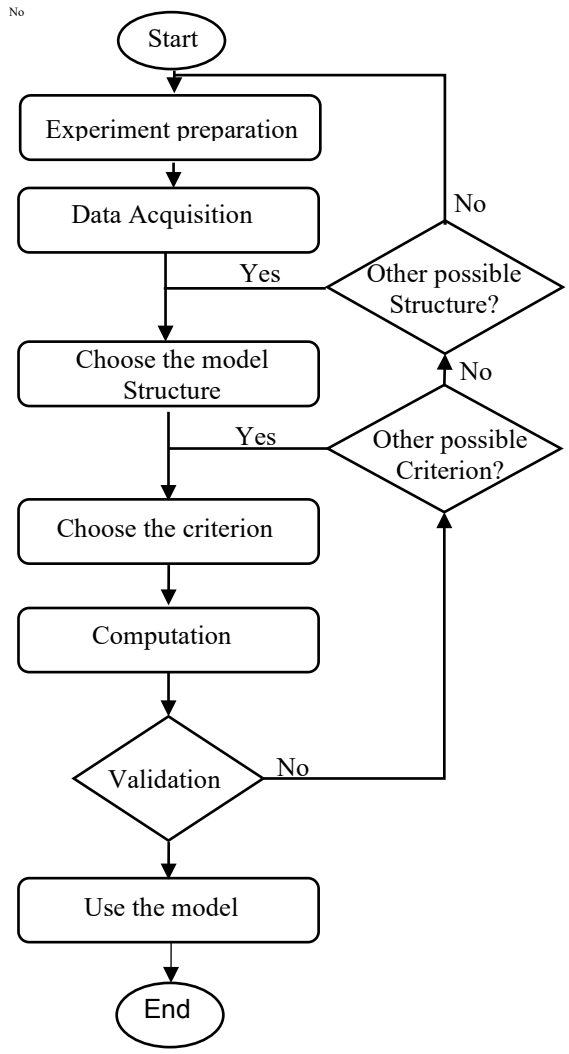


Figure 1: Experimental approach for defining a parametric model of a system

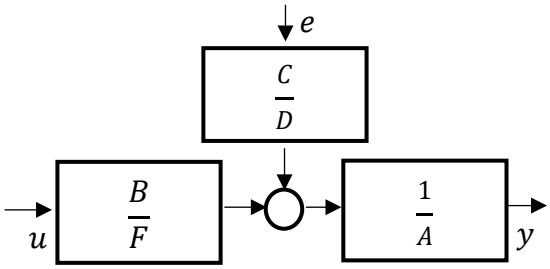


Figure 2: Generalized transfer function Structure

2. Parametrical Models

2.1. Generalities

A discrete linear system can be presented as a general-linear polynomial model, which is considered as a transfer function whose parameters need to be estimated. This estimation is done by recursive and none recursive algorithms like maximum likelihood, instrumental variable or less square. This presentation, also gives flexibility for modelling both stochastic and system dynamics, and it is widely used in many real-world applications from different disciplines, e.g., [8–10].

Polynomial models are described in [11]. They can be defined by the generalized transfer function expressed by (1) and the signal flow represented in figure 2 [5]:

$$A(q)y(k) = \frac{B(z)}{F(z)}z^{-nk}u(k) + \frac{C(z)}{D(z)}e(k) \quad (1)$$

where:

$e(k)$: *is the system disturbance*

nk: *is the system delay*

$$A(z) = 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a} \quad (2)$$

$$B(z) = b_1 z^{-1} + \cdots + a_{n_b} z^{-n_b} \quad (3)$$

$$C(z) = 1 + c_1 z^{-1} + \cdots + a_{n_c} z^{-n_c} \quad (4)$$

$$D(z) = 1 + d_1 z^{-1} + \cdots + d_{n_d} z^{-n_d} \quad (5)$$

$$F(z) = 1 + f_1 z^{-1} + \cdots + f_{n_f} z^{-n_f} \quad (6)$$

and:

n_a : The order of $A(z)$

n_b : The order of $B(z)$

n_c : The order of $C(z)$

n_d : The order of $D(z)$

n_f : The order of $F(z)$

By setting one or more n_a, n_c, n_d and n_f to 0 can create simpler models such as:

2.1.1. ARX model

ARX structure presented in equation (7) [5] is the result of setting n_c, n_d *and* n_f to 0 the signal flow is presented in fig.3. In this structure the noise model $\frac{1}{A(z)}$ is coupled with the systems dynamics, it is used to obtain simple model with good signal to noise ratio.

$$Y_{arx}(k) = \frac{B(z)}{A(z)}z^{-nk}u(k) + \frac{1}{A(z)}e(k) \quad (7)$$

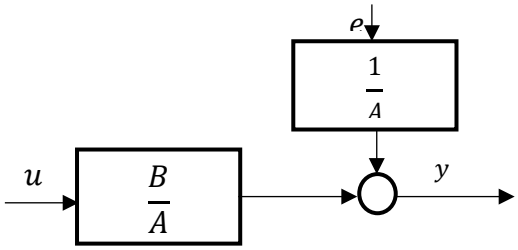


Figure 3: The ARX Structure

2.1.2. ARMAX model

The ARMAX structure presented in the equation (8) [5] is the result of setting n_d and n_f to 0 in the generalized transfer function (1), the signal flow is presented in fig.4. this structure provides more flexibility to model the noise by adding the $C(z)$ parameters

which is considered as a moving average of white noise. It is used when the noise enters in the input.

$$Y_{armax}(k) = \frac{B(z)}{A(z)} z^{-nk} u(k) + \frac{C(z)}{A(z)} e(k) \quad (8)$$

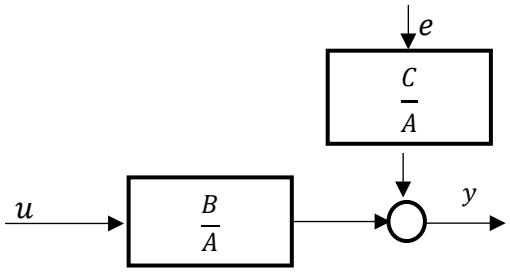


Figure 4: The ARX Structure

2.1.3. Output Error Model

The OE presented in the equation (9) [5] is the result of setting n_a, n_c and n_d to 0 in the generalized transfer function (1), the signal flow is presented in fig.5. In this structure, the noise model is considered as 1. It is used to parameterize the dynamic of the system only.

$$Y_{OE}(k) = \frac{B(z)}{F(z)} z^{-nk} u(k) + e(k) \quad (9)$$

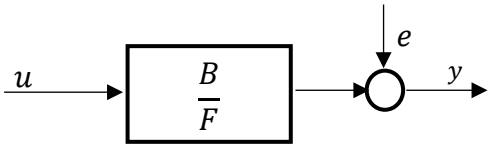


Figure 5: The OE Structure

2.1.4. Box- Jenkins Model

The BJ structure presented in the equation (10) [5] is the result of setting n_d to 0 the signal flow is presented fig.6. This structure provides more flexibility because the noise model is independent of the dynamics model. It is used when the noise enters as a measurement disturbance.

$$Y_{BJ}(k) = \frac{B(z)}{F(z)} z^{-nk} u(k) + \frac{C(z)}{D(z)} e(k) \quad (10)$$

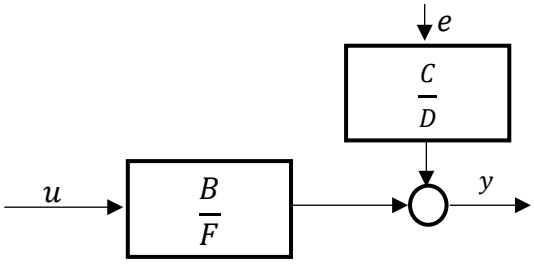


Figure 6: The BJ Structure

3. Parameters Estimation

Parameter estimation is a mathematical process that uses experimental data to estimate the system parameters, in this paper; this process is done by the prediction error method described in [12] and fig 7. this method aims to minimize the error expressed in Eq (13) between the estimated and the real system [13].

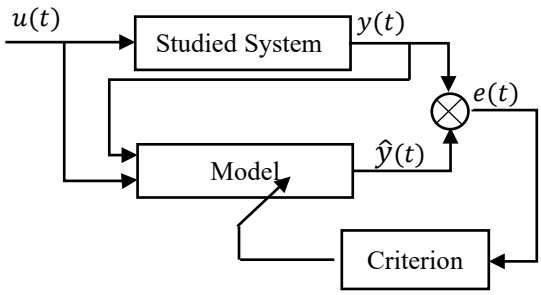


Figure 7: Prediction Error Method

According to [14], to use the Prediction Error Method, we need to choose the model structure, the one-step-ahead predictor $\hat{y}(k|k-1,\theta)$ described with the equation (12) and the criterion function $J(\theta, Z^N)$ given by Eq (11):

$$J(\theta, Z^N) = f\left(R_N(\theta, Z^N)\right) \quad (11)$$

where:

$$R_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N} e(k,\theta) e^T(k,\theta)$$

$$\hat{y}(k|k-1,\theta) = H^{-1}(q)G(q)u(k) + (1 - H^{-1}(q))y(k) \quad (12)$$

And the prediction error is given by:

$$e(k|k-1) = y(k) - \hat{y}(k|k-1,\theta)$$
$$= H^{-1}(q)(y(k) - G(q)u(k)) \quad (13)$$

4. Model Validation

Model validation is the final step, which goal is to compare and validate the usability of the estimated model by testing several criterions like those in this paper:

1.3. Akaike Final Prediction Error

This criterion described in the equation (14) [5] was presented by Akaike (1967) as a final prediction error (FPE), based on Akaike's theory, the best model has the smallest FPE.

$$\mathrm{FPE} = \det\left(\frac{1}{\mathrm{N}}\sum_{1}^{\mathrm{N}} \mathrm{e}(\mathrm{t},\theta_{\mathrm{N}})(\mathrm{e}(\mathrm{t},\theta_{\mathrm{N}}))^{\mathrm{T}}\left(\frac{1+\mathrm{d}/\mathrm{N}}{1-\mathrm{d}/\mathrm{N}}\right)\right) \quad (14)$$

where:

N: The number of values in the estimation data set

$e(t)$: The vector of prediction error

θ_N: The estimated parameters

d: The number of the estimated parameters

4.1 Best Fit

The RMSE criterion measures the similarity between the data and the estimated model and can be calculated using equation (15) [15].

$$\text{Best Fit} = 100\left(1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right) \quad (15)$$

where:

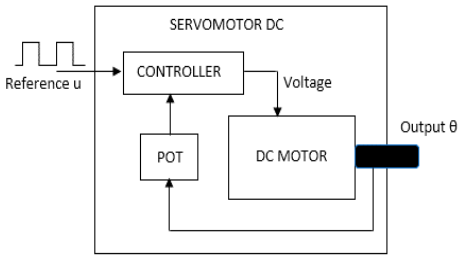
y_i : The measured output

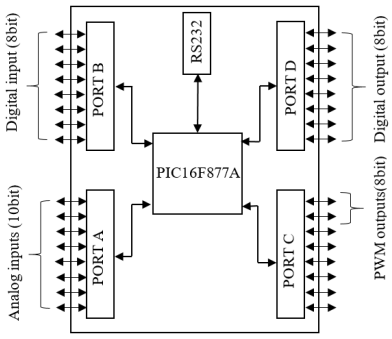
$\hat{y}$: The estimated output

$\bar{y}$: The mean of the output

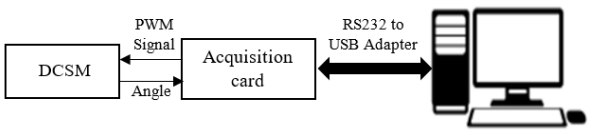
5. Application

A DCSM is composed of a potentiometer and integrated circuit that ensures a dc motor's position control based on a reference signal received from a microcontroller or an external organ (fig.8). However, despite its different applications, there is a major problem in achieving good performance with those servo motors due to the integrated circuit [1]. Therefore, the goal is to estimate all the system's internal and external components' mathematical model, to design a new controller or exploit it in more complex systems. In this application, the reference signal is in the form of a pulse with modulation signal who does not exceed 20ms between two rising edges, and the DCSM supply voltage is 5V.





To extract reliable input and output data, we build the acquisition card with pic16f877a. As shown in fig 9, this microcontroller has eight analogic inputs of 10-bit resolution, one digital input, one digital output of 8-bit resolution, and two PWM output of 8-bit resolution. Moreover, it is linked to the computer using RS232 communication with a speed of 115200bit/s to send and receive data necessary for the experiment. However, today's computer does not support this type of communication. To solve this problem, an RS232 to USB adapter is used, as shown in fig 10.



Afterward, to know the sampling time Ts, which is related to the uncertainty of the internal and external components and the instructions execution time of the microcontroller. The analogic input of the acquisition card is subjected to a periodic signal of known frequency. The samples collected from the acquisition card are represented on a graph. Afterward, the frequency of the test signal is adjusted until a signal similar to that of the input is obtained, which $Ts = 0.0034$ deduced in this application. This will allow having a speed of 300 sps/s.

Next, the analysis and parameter estimation process is applied to the collected data from the acquisition card.

6. Results

Model order determination in parameter estimation is essential because a lower order model cannot adequately describe the system's dynamics, and a higher-order model may have uncertainties. Thus, to extract the best model and to see the effect of changing the order of the parameters, we applied the experimental approach of identifying a parametric model described in fig 1, and we fixed nk to 1 then:

- Over 100 ARX models are calculated where $n_a$and n_b are in the range of 1 to 10.
- Over 125 ARMAX models are calculated where $n_a, n_b \; and \; n_c$ are in the range of 1 to 5.
- Over 25 OE models are calculated where $n_f and \; n_b$are in the range of 1 to 5.
- Over 625 BJ models are calculated where $n_a, n_a, n_a \; and \; n_a$ are in the range of 1 to 5.

Afterward, the Best-fit criterion is used to select the best five structures in each model, and the PFE criterion is used to choose the best structure in each model. Those results are in tables 1 to 4.

7. Discussion

As presented in tables 1 to 4, all the extracted models in each structure have closely a similar FPE. Then the choice of the model is based on the best-fit only. Besides, no significant change occurs in the best-fit criterion by increasing the models' order in each model type. Thus, the selected model will have the lowest order. This will allow a better processing time while doing online parameter estimation in real-time applications.

- The best-estimated ARX model is the ARX331 shown in Figure 11 and Table 1. It has an order of (nb=3 and na = 3), and a fit of 72,44
- The best-estimated ARMAX model is the ARMAX2411 shown in Figure 12 and Table 2. It has an order of (na = 2, nb = 4, nc=1), and a fit of 94,75%.
- The best-estimated OE model is the OE131 shown in Figure 13 and Table 3. It has an order of (nf = 3, nb = 1), and a fit of 95.009.
- The best-estimated BJ model is the BJ12331, shown in Figure 14 and Table 4. It has an order of (nb = 1, nc = 2, nd = 3, nf = 3), and a 95.02% fit.

Based on those results and figure 16. The ARX models provided a fit of 72%, which is a poor fit compared to ARMAX, BJ and OE Models. However, if we want to use a model in real-time applications, the best choice will be OE131 because of the resulting good fit with the measured output and fewer parameters.

Table 1: The estimated ARX models

Model Type	Model Parameters	'Best Fit'	'FPE'
'ARX431'	A(z) = 1 - 1.065 z^-1 + 0.005417 z^-2 + 0.01607 z^-3 + 0.04765 z^-4 B(z) = 3.841e-05 z^-1 + 0.002391 z^-2 + 0.001677 z^-3	72,6272093	0,3764651
'ARX441'	A(z) = 1 - 1.065 z^-1 + 0.005351 z^-2 + 0.01617 z^-3 + 0.04766 z^-4 B(z) = 3.741e-05 z^-1 + 0.002391 z^-2 + 0.002752 z^-3 - 0.001086 z^-4	72,5563214	0,37670021
'ARX421'	A(z) = 1 - 1.065 z^-1 + 0.005566 z^-2 + 0.01616 z^-3 + 0.04759 z^-4 B(z) = 3.772e-05 z^-1 + 0.004052 z^-2	72,5177552	0,37624736
'ARX331'	A(z) = 1 - 1.069 z^-1 + 0.005651 z^-2 + 0.06704 z^-3 B(z) = 0.000315 z^-1 + 0.002366 z^-2 + 0.001609 z^-3	72,4480688	0,37718277
'ARX341'	A(z) = 1 - 1.068 z^-1 + 0.005065 z^-2 + 0.06748 z^-3 B(z) = 0.0003129 z^-1 + 0.002366 z^-2 + 0.002673 z^-3 - 0.001074 z^-4	72,386256	0,37731582

Table 2: The estimated ARMAX models

'Model Type'	Model Parameters	'Best Fit'	'FPE'
'ARMAX3421'	A(z) = 1 - 2.598 z^-1 + 2.202 z^-2 - 0.6041 z^-3 B(z) = 0.001024 z^-1 + 0.002355 z^-2 - 0.008141 z^-3 + 0.004814 z^-4 C(z) = 1 - 1.576 z^-1 + 0.5755 z^-2	94,8596855	0,36096562
'ARMAX2421'	A(z) = 1 - 1.985 z^-1 + 0.9851 z^-2 B(z) = 0.001351 z^-1 + 0.0005912 z^-2 - 0.006229 z^-3 + 0.004419 z^-4 C(z) = 1 - 0.9752 z^-1 - 0.02485 z^-2	94,7854814	0,36124302
'ARMAX4231'	A(z) = 1 - 2.362 z^-1 + 1.557 z^-2 - 0.02015 z^-3 - 0.1743 z^-4 B(z) = -0.0004067 z^-1 + 0.0004638 z^-2 C(z) = 1 - 1.312 z^-1 + 0.1561 z^-2 + 0.156 z^-3	94,7787217	0,36258016
'ARMAX2411'	A(z) = 1 - 1.986 z^-1 + 0.9859 z^-2 B(z) = 0.001361 z^-1 + 0.0006902 z^-2 - 0.006188 z^-3 + 0.004261 z^-4 C(z) = 1 - z^-1	94,7591717	0,3611956
'ARMAX3411'	A(z) = 1 - 2.022 z^-1 + 1.059 z^-2 - 0.03639 z^-3 B(z) = 0.001229 z^-1 + 0.0009276 z^-2 - 0.00656 z^-3 + 0.004529 z^-4 C(z) = 1 - z^-1	94,7457863	0,36131993

Table 3: The estimated OE models

'Model Type'	Model Parameters	'Best Fit'	'FPE'
'OE431'	B(z) = 0.01849 z^-1 - 0.05449 z^-2 + 0.05356 z^-3 - 0.01755 z^-4 F(z) = 1 - 2.947 z^-1 + 2.894 z^-2 - 0.9475 z^-3	95,0341979	6,13354962
'OE131'	B(z) = 7.423e-06 z^-1 F(z) = 1 - 2.933 z^-1 + 2.867 z^-2 - 0.9339 z^-3	95,0098583	6,28005407
'OE341'	B(z) = 0.01923 z^-1 - 0.03923 z^-2 + 0.02007 z^-3 F(z) = 1 - 2.14 z^-1 + 0.8912 z^-2 + 0.6443 z^-3 - 0.3956 z^-4	94,9929819	6,21730334
'OE321'	B(z) = 0.04349 z^-1 - 0.08893 z^-2 + 0.04559 z^-3 F(z) = 1 - 1.984 z^-1 + 0.9845 z^-2	94,9848311	6,21616865
'OE221'	B(z) = -0.001383 z^-1 + 0.001513 z^-2 F(z) = 1 - 1.985 z^-1 + 0.9851 z^-2	94,8945691	6,55918215

Table 4: The estimated BJ models

'Model Type'	Model Parameters	'Best Fit'	'FPE'
'BJ23241'	B(z) = 6.176e-05 z^-1 - 5.46e-05 z^-2 C(z) = 1 - 0.6253 z^-1 + 0.01001 z^-2 + 0.02315 z^-3 D(z) = 1 - 1.629 z^-1 + 0.6427 z^-2 F(z) = 1 - 2.44 z^-1 + 1.393 z^-2 + 0.5362 z^-3 - 0.4889 z^-4	95,0238364	0,360708
'BJ12331'	B(z) = 7.101e-06 z^-1 C(z) = 1 - 1.609 z^-1 + 0.9796 z^-2 D(z) = 1 - 2.46 z^-1 + 2.254 z^-2 - 0.7828 z^-3 F(z) = 1 - 2.936 z^-1 + 2.872 z^-2 - 0.9367 z^-3	95,0160798	0,33251152
'BJ14131'	B(z) = 7.24e-06 z^-1 C(z) = 1 + 0.04146 z^-1 + 0.03341 z^-2 + 0.01694 z^-3 + 0.03648 z^-4 D(z) = 1 - 0.9637 z^-1 F(z) = 1 - 2.934 z^-1 + 2.87 z^-2 - 0.9354 z^-3	95,0135709	0,36037501
'BJ13131'	B(z) = 7.263e-06 z^-1 C(z) = 1 + 0.04056 z^-1 + 0.03335 z^-2 + 0.01701 z^-3 D(z) = 1 - 0.9659 z^-1 F(z) = 1 - 2.934 z^-1 + 2.869 z^-2 - 0.9351 z^-3	95,0131615	0,36038438
'BJ13231'	B(z) = 7.26e-06 z^-1 C(z) = 1 - 0.6241 z^-1 + 0.009762 z^-2 + 0.02454 z^-3 D(z) = 1 - 1.627 z^-1 + 0.6415 z^-2 F(z) = 1 - 2.934 z^-1 + 2.87 z^-2 - 0.9353 z^-3	95,0129791	0,35994338

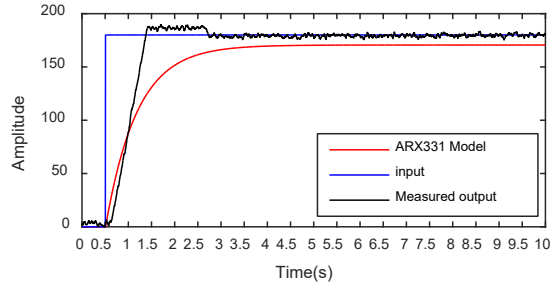


Figure 11: Comparison between the estimated model (ARX331) and measured output

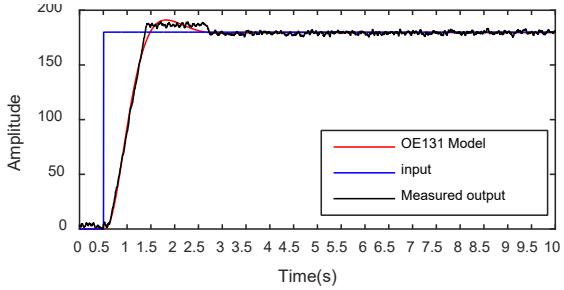


Figure 12: Comparison between the estimated model (OE131) and measured output

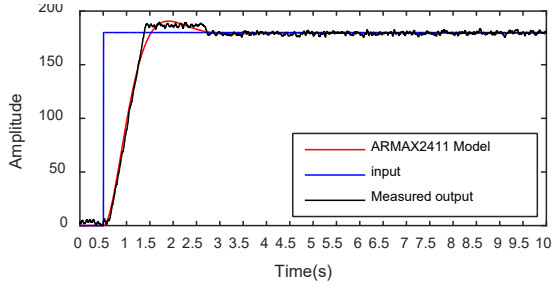


Figure 13: Comparison between the estimated model (ARMAX2411) and measured output

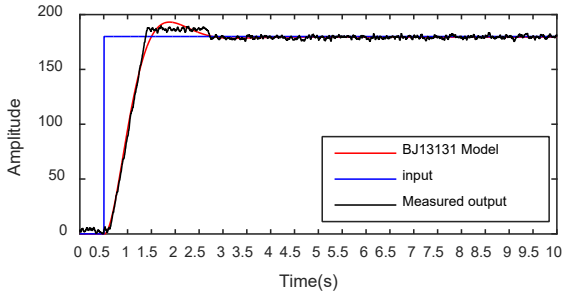


Figure 14: Comparison between the estimated model (BJ1313) and measured output

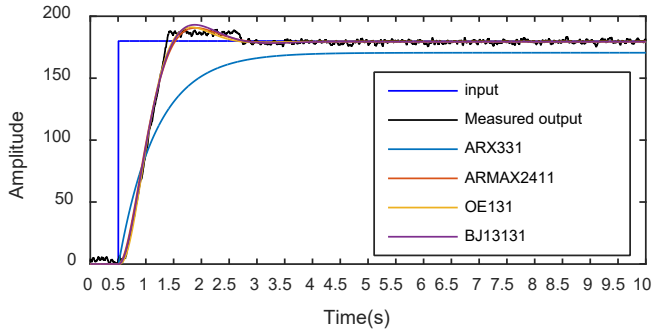


Figure 15: Comparison between the estimated Polynomial models (ARX331, ARMAX2411, OE131, and BJ1313) and measured output

8. Conclusion

In this article, a practical identification of DCSM has been elaborated using polynomial models. To meet application requirements and to collect input-output data, a microcontroller-based acquisition card has been developed. Finally, the comparison between the data and the estimated models using validation criterions shows that the ARMAX, BJ, and OE model has provided a good fit with experimental data and can be used in simulations or a real application.

In future work, the data acquisition system will be based on PLC to be adequate for industrial applications and to provide more sampling frequency and processing time. Moreover, other identification techniques will be applied to more complex non-linear systems to build proper algorithms for the control loop or fault detection.

References

[1] T. Wada, M. Ishikawa, R. Kitayoshi, I. Maruta, T. Sugie, "Practical modeling and system identification of R/C servo motors," in 2009 {IEEE} International Conference on Control Applications, IEEE, 2009, doi:10.1109/cca.2009.5280987.

[2] L. Ljung, "Perspectives on system identification," Annual Reviews in Control, **34**(1), 1–12, 2010, doi:10.1016/j.arcontrol.2009.12.001.

[3] R. Isermann, M. Münchhof, Identification of dynamic systems: an introduction with applications, Springer Science & Business Media, 2010.

[4] M. Villain, Systèmes asservis linéaires, Ellipses, 1996.

[5] L. Lennart, "System identification: theory for the user," PTR Prentice Hall, Upper Saddle River, NJ, 1–14, 1999.

[6] M. Jirgl, L. Obsilova, J. Boril, R. Jalovecky, "Parameter identification for pilot behaviour model using the {MATLAB} system identification toolbox," in 2017 International Conference on Military Technologies ({ICMT}), IEEE, 2017, doi:10.1109/miltechs.2017.7988824.

[7] S. Rai, S.-Y. Yang, T.-C. Tsao, "Wind turbine system identification and individual pitch control," in 2017 American Control Conference ({ACC}), IEEE, 2017, doi:10.23919/acc.2017.7963144.

[8] S. Rachad, B. Nsiri, B. Bensassi, "System Identification of Inventory System Using {ARX} and {ARMAX} Models," International Journal of Control and Automation, **8**(12), 283–294, 2015, doi:10.14257/ijca.2015.8.12.26.

[9] W. Araydah, T.A. Tutunji, I. Al-Naimi, "System identification for a liquid flow process," in 2017 {IEEE} Jordan Conference on Applied Electrical Engineering and Computing Technologies ({AEECT}), IEEE, 2017, doi:10.1109/aeect.2017.8257756.

[10] S.L. Chavan, D.B. Talange, "System identification black box approach for modeling performance of {PEM} fuel cell," Journal of Energy Storage, 18, 327–332, 2018, doi:10.1016/j.est.2018.05.014.

[11] L. Ljung, "Black-box models from input-output measurements," in {IMTC} 2001. Proceedings of the 18th {IEEE} Instrumentation and Measurement Technology Conference. Rediscovering Measurement in the Age of Informatics (Cat. No. 01CH 37188), IEEE, 2001, doi:10.1109/imtc.2001.928802.

[12] L. Ljung, "Prediction error estimation methods," Circuits, Systems, and Signal Processing, **21**(1), 11–21, 2002, doi:10.1007/bf01211648.

[13] M. Brunot, A. Janot, F. Carrillo, M. Gautier, "A Separable Prediction Error Method for Robot Identification," {IFAC}-{PapersOnLine}, **49**(21), 487–492, 2016, doi:10.1016/j.ifacol.2016.10.650.

[14] Y. Zhao, B. Huang, H. Su, J. Chu, "Prediction error method for identification of {LPV} models," Journal of Process Control, **22**(1), 180–193, 2012, doi:10.1016/j.jprocont.2011.09.004.

[15] L. Ljung, System identification toolbox. Getting started guide. The Mathworks Inc, 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 514-519 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Facebook Translation Service (FTS) Usage among Jordanians during COVID-19 Lockdown

Zakaryia Almahasees[1,*], Helene Jaccomard[2]

[1]*Applied Science Private University, Department of English Language and Translation, Faculty of Arts and Science, Amman, 00962, Jordan*

[2]*University of Western Australia, Department of Translation, School of Humanities, Perth, 0061, Australia*

ARTICLE INFO

Article history:
Received: 04 September, 2020
Accepted: 30 October, 2020
Online: 20 November, 2020

Keywords:
Facebook Translation service
Facebook and Covid-19
Facebook Posts
Translation Studies
English and Arabic

ABSTRACT

The world is facing an unprecedented virus outbreak, COVID-19, hitting more than 200 countries. Governments have been striving to prevent the spread of the virus through the lockdown. During the strict lockdown in Jordan, people needed to stay home, and they used available social networks to keep updated on COVID-19, with Facebook, the most popular social media platform. The study aimed to elicit information about assessing the use of FTS as a source of information in general and on COVID-19, in particular, FTS for those interested in English posts. However, it cannot read them and how reliable users think FTS is. The questionnaire was sent through the available networking sites, such as Facebook Messenger. The study found that 94.3% use Facebook daily; 87.1% of the participants activated Facebook Translation Service (FTS). It is found that 62.2% of the participants considered Facebook as a primary source of information regarding COVID-19 and 27.8% as secondary source. In terms of FTS usage, 87.3% used FTS in translating English Facebook posts into Arabic, and 83.8% used FTS in translating English Facebook COVID-19 posts into Arabic during the lockdown. On the other hand, it is found that the majority found FTS committed minor errors in terms of adequacy and fluency. This success is due to the usage of Neural Machine Translation (NMT)approach and bilingual text corpora. The advantage is that FTS is a well-trained database that can provide more accurate translation than other model. In conclusion, disregarding FTS output quality, our research shows that Facebook and FTS became a significant source of information during abrupt crises. Such research would encourage government officials to better use Facebook and FTS as complements to their national health campaigns.

1. Introduction

The translation is the medium of human communications. It bridges the gap between human communities and ensures the best communication among the global communities [1]. Nowadays, the world is facing the most frightening Coronavirus (COVID-19) outbreak in the last century. International Health Organizations, Medical Specialists, health authorities, political leaders, and the public write posts round the clock on social networks, making the influx of information massive. The flow of information is unprecedented, and therefore, there should be a way to facilitate transferring this information into all-natural languages. It is predicted that the demand for translation industry will reach unprecedented number $56 billion in 2021 [2]. This growth in translation demand has been occurring at a rate beyond professional human translators' capacity to keep up with, and the cost associated with human translation is too high for many purposes [3]. For example, It is indicated that they can translate up to 10 billion words daily. Translating the massive flow of information is beyond human translators' capacity, hence the central role of Machine Translation (MT). This tool can cope with large amounts of data at a split-second rate. However, there is no relation between the speed of translation and the quality of the translation [4].

In facing COVID-19, MT works as a tool to challenge and fight the virus. It is stated that "The challenge is how to communicate rapidly changing data across language borders, so that essential information is not lost in translation." In this concern, it has been shown that machine translation can help fight COVID-19. The

*Corresponding Author: Zakaryia Almahasees, Email: z_almhasees@asu.edu.jo

current study aims to assess the popularity and use of Facebook Translation Service (FTS) in translating English Facebook posts in general and COVID-19 in particular among Facebookers in Jordan.

Social Networking is a flagship of human communications nowadays. Social networking refers to online platforms to keep humans connected regardless of their locations and territories. Social networking connects people globally of different groups and ethnicities for social, business and other purposes, such as Facebook, Twitter and the like [5]. More than 3.81 billion users of social networks worldwide and lists the most famous six social networking sites worldwide based on their monthly users: Facebook, YouTube, WhatsApp, Facebook Messenger, Weixin(WeChat), and then Instagram. Facebook tops the other social networking sites with 2.5 billion users monthly [6].

Regarding Jordan, it has been listed Facebook as the most popular, with 88.6% of all social networking users. When it comes to languages [7]. English has been classified as the top language with 25.3% of internet content, while Arabic content represents no more than 4.8% [8], [9]. For English content published on Facebook, FTS allows Arabs to read the English posts in Arabic.

However, the number of Facebook active monthly users in the Middle East reaches 200 million [10]. A study conducted in Jordan to elicit information about the source of information for a sample of 500 medical students in three Jordanian Universities. They found that 83.4% used social media and 83.4% favored online search to access COVID-19 related information. However, the study did not specify the most popular social networks among medical Students [11].

Social media becomes an essential means of communication since COVID-19 updates take place around the clock. Such widespread use, particularly at the time of a health crisis mandating social isolation, raises questions about the content Facebook provides in terms of its effectiveness and limitations to end users since Arab Facebookers are mostly exposed to posts written in English. Some Arab Facebookers can read and understand English. However, there is still a large portion of users who are not sufficiently competent to read posts in English, and instead, use FTS, Facebook's translation service since 2011. Facebook Translate helps users communicate with each other in different languages. Facebookers can easily activate the translation service on their profiles. Users can select their region and language and the language they wish their posts to be translated into in their settings.

2. Facebook Translation Service (FTS)

Historically, Facebook initially used Google Translate service to translate the posts into 50 languages. Facebook launched its first Translation Service in 2011, called In-line Translation Facebook Tool [12], this time powered by Microsoft Bing. For example, if a non-Spanish speaker gets a cross a Spanish Facebook post, he will be able to see the translation of the post in his own language.

Currently, FTS offers translation services for 89 languages, and they will continue to add more languages. FTS adopted a phrase-to- phrase MT approach by translating whole sequences of words, where the lengths may differ. In 2017, Mark Zuckerberg, the founder of Facebook, explained that translation is the best means to connect humans globally. Language understanding brings people close and thus help them to share and get access to information universally [13]. In the same year, the Facebook research team developed a new MT approach using Convolutional Neural Networks (CNNs) instead of Google Translate. The recurrent neural networks can help translate languages more accurately (increase quality on a BLEU scale) and up to nine times faster than the traditional recurrent neural networks (RNNs). Neural MT has been described approach as a tool to provide instant translation accurately across languages faster than previous MT approaches [14]. This is confirmed by [15], [16].They have shown that "CNNs have been very successful in several machine learning fields, such as image processing. However, recurrent neural networks (RNNs) are the incumbent technology for text applications and have been the top choice for language translation because of their high accuracy".

In late 2019, the Facebook research team announced new advances in NLP, boosting Facebook Translation accuracy. They have also introduced a new self-supervised pretraining approach, RoBERTa, that surpassed all existing Natural Language Understanding (NLU) systems on several language comprehension tasks. They have also collaborated with New York University (NYU), DeepMind Technologies, and the University of Washington (UW) to promote their future research [16].

The present study focusses on Facebook translation of English Facebook posts related to COVID-19 released by International Organizations such as WHO, political leaders, medical specialists, and the general public into Arabic. The study provides constructive feedback about whether FTS is considered a reliable source of information or not and the degree of trust among the end-users. It addresses the following questions:

What, according to participants, is the percentage of FTS Activation in Jordan?

What is the rate of using FTS in translating English Facebook posts and Covid-19 Facebook posts into Arabic?

As a source of information, how often Jordanians use FTS?

To what degree FTS output is adequate and fluent in English to Arabic translation?

To what extent, Jordanian rated FTS output?

3. Literature Review

The outbreak of COVID-19 makes social networking sites a source of information for many people around the globe. COVID-19 is an emergent crisis, unprecedented in the size of information posted daily about the pandemic. Facebook usage increases by 40% due to the COVID-19 crisis. Facebook becomes a medium of communication. Machine Translation plays an influential role in facilitating communication among human communities [17], [8]. The study is mainly concerned with the popularity of FTS among Jordanians in translating the English Facebook posts into Arabic. To the best of knowledge of the author, this is the first research to fill the gap and provide the literature with information about the role of FTS in educating Jordanians and fighting COVID-19.

Social networking sites have become a useful tool in confronting the crisis and keeping people close to each other when they are isolated due to lockdown in most of global countries. He

adds that the public health experts have used social networking sites to interact with the public and provide healthcare advice to fight and prevent the spread of the outbreak[[18]]. The acceleration in digital life changes how health information is approached. The health care providers used this opportunity to disseminate their advice and right information for the public [[19]].

On the other hand, Facebook becomes the source of information for 100,000 health care providers regarding the COVID-19 outbreak, "For the past few weeks, after I get home from my shifts in the emergency room, I have been scrolling through Facebook pages on my laptop, getting firsthand stories from doctors in Italy, China, and Iran" [20]. Another study has been conducted a descriptive study on medical students in Jordan to assess knowledge, attitude, perceptions, and precautionary measures toward COVID-19. They have found that 83.4% used social media sites as their preferred source of information regarding COVID-19 [11]. Moreover, another study has been carried out in Jordan to assess university students' knowledge, practice, and attitude regardless of their majors. They have found that university students' information source is social media, internet, and television. No significant difference was noticed between medical and non-medical college students on the sources of their information. On the other hand, social networking sites have become platforms for survival against social isolation during the COVID-19 [21]. In this regard, Facebook becomes the central platform to help the community and those in need during this pandemic [22].

Another use of social networks during the crisis is highlighted by as a platform to preach and provide spiritual support to conduct digital prays and biblical lesson [23]. On the other hand, social media shows how social networks become the platform for sharing controversial messages to Chinese people. Some users blamed China as a source of the disease and therefore Chinese became the target of many hateful and offensive posts on most social networking sites., such as Facebook and the like. Facebook and other social media have been clamped down the spread of fake news concerning COVID-19 [24], [25]. The study highlights the FTS infighting world crisis, COVID-19, in provide translation for the English Facebook posts related to COVID-19 into Arabic.

4. Methodology

Arabic is the official language of 24 Arab countries with a population of 300 million. Jordan is one of Arab countries and the research on Facebook Translation service in Jordan would of great interests and a source of information for all Arabs. Descriptive cross-sectional study design was used in this work. Our sample consisted of 200 Jordanians from five cities of Jordan: Ajlun, Amman, Irbid, Jerash, Kerak. The study population has different educational backgrounds except for English language specialists (Linguistics, literature, and Translation studies) and were asked to fill the questionnaire. The study used an online questionnaire delivered to participants in the period between March 10 to July 29 during the lockdown in the Jordanian government's effort to control the spread of COVID-19. The online questionnaire was created on Google Forms and sent to the participants through Facebook Messenger to ensure the social distancing in line with the Jordanian government's recommendations. Therefore, all participants in this study have Facebook pages. The study's total population was 200 responded to the study, 106 females (52.7%) of the sample, and 95 male participants (47.3%) (see appendix 1) An independent T-test was conducted to identify a relation between rural and urban areas regarding rating the adequacy and fluency of FTS. It is evident from the table 2 that a value of 1.076 at sig = 0.283 indicates that there are no statistically significant differences at 0.05 between rural and urban residents in the evaluation of the fluency of FTS. Moreover, another t-test was conducted to verify the relation between gender and fluency rating. It is evident from the table 3 that a value of t was 1.145 at sig = 0.254, indicating that there were no statistically significant differences at 0.05 between males and females in evaluating the fluency of Facebook Translation Service. Another also t-test conducted on the relation between the education level and the adequacy. The study found that there are statistically significant differences at 0.05 in evaluating the fluency of FTS according to the educational level, and by conducting post-tests to test the analysis of variance, it was found that the source of the differences is the existence of statistically significant differences between holders of a diploma and those who hold a master's degree in favor of a diploma.

4.1. Research Instrument

The researchers designed a dichotomous questionnaire that was created on the Google Forms platform. The questionnaire has yes/no questions to elicit Jordanians' responses concerning their usage of Facebook Translation service in translating English Facebook posts into Arabic and, in particular COVID-19 posts.

The online questionnaire consisted of seven main divisions main sections. The first section collected demographic data on gender, age, education, cities, and living areas. The second section consisted of the following questions:

Have you activated FTS on your Facebook Page?

Have you used FTS in translating English Facebook posts into Arabic?

Have you used FTS in translating English Covid-19 Facebook posts into Arabic?

Have you considered FTS output as a Primary source of your information regarding COVID-19?

Irrespective of the English posts' contents (maybe not reliable, fake, or misleading), do you find the translation by FTS acceptable?

Irrespective of the English posts' contents (maybe not reliable, fake, or misleading), do you think Facebook Translation Service has translated them accurately?

Have you rated FTS?

Two professors reviewed the survey, and proper modifications were completed before delivering to participants. Participation in the study was voluntary, and personal Facebook pages and identifiers were not collected. Data was imported into Excel for management and then SPSS for analysis. Numbers and percentages were presented for all variables.

5. Results

5.1. FTS activation among Jordanians

First, we examined the number of Facebookers who activated FTS on their Facebook pages. As mentioned above, the questionnaire was sent via Facebook Messenger, and therefore all participants have Facebook pages. Our analysis showed that (87.5%) of the participants activated FTS on their Facebook, while only (12.5%) of the participants did not activate FTS. This finding shows that FTS is used mostly by Jordanian. Moreover, this result correlates with [[7]] that Facebook is the most popular social media.

Table 1: Percentage of FTS Activation

Classification	Sum of Frequency	Sum of Percent %
No	25	12.5
Yes	175	87.5
Grand Total	**200**	**100**

5.2. Have you used FTS in rendering English Facebook posts into Arabic?

Next, we wanted to know if Jordanians used FTS in rendering English Facebook posts into Arabic in general and English posts concerning COVID-19 rendering into Arabic. Our analysis showed that (87.6%) of Jordanians used FTS in rendering English posts into Arabic, while 12.4% did not use FTS in translation. These results highlight the role of FTS in facilitating communications between English and Arabic.

Table 2: FTS usage among Jordanians at the time of Covid-19

Have you used Facebook Translation Service in rendering English posts into Arabic?		
Classification	**Sum of Number of respondents**	**Sum of Percent %**
No	25	12.4
Yes	175	87.6
Grand Total	**200**	**100**

5.3. Have you used FTS in translating English Facebook Covid-19 posts into Arabic?

We aimed to elicit information about the percentage of FTS in translating English Covid-19 posts into Arabic. The study's analysis showed that (83.1%) have used FTS in translating English Facebook COVID-19 posts into Arabic. These figures showed that FTS played a significant role in facilitating communications among Jordanians at the time of the COVID-19 outbreak in Jordan.

Table 3: FTS rendering English posts about Corona Virus (COVID-19) into Arabic

Have you used Facebook Translation Service in rendering English posts about Corona Virus (COVID-19) into Arabic?		
Classification	**Sum of Frequency**	**Sum of Percent %**
No	33	16.9
Yes	167	83.1
Grand Total	**200**	**100**

5.4. Have you used FTS daily at the time of COVID-19?

Next, we aimed to identify how often Jordanians used FTS during the time of lockdown in Jordan. Our analysis showed that most of Jordanians used FTS daily with (94. %). This correlates with [[7]] indicated that Facebook usage in Jordan reached 92% over the COVID-19 period, while only (5.97%) did not use it daily.

Table 4: Frequency Facebook and FTS among Jordanians at the time of Covid-19 lockdown

Classification	Sum of Frequency	Sum of Percent %
Daily	188	94
Rarely	1	0.5
Weekly	11	5.5
Grand Total	**200**	**100**

5.5. Have you considered FTS output as a Primary source of your information regarding COVID-19?

First, we examined how the participant deals with FTS as a source of information. The question aimed to identify whether Jordanians considered FTS output as sources of information at the time of lockdown. Our analysis showed that most participants with (62.2%) considered FTS output as the primary source of their information regarding COVID-19. In comparison (27.8%) showed that they considered FTS as a secondary source of their information.

Table 5: Facebook as a Sources of Information Related to COVID-19

Category	Sum of Frequency	Sum of Percent %
Partial	3	1.5
Minor	16	8.5
Primary	125	62.2
Secondary	56	27.8
Grand Total	**200**	**100**

5.6. How do you rate the fluency of FTS output?

The above chart illustrates how the study's participants rated FTS. The analysis showed that the respondents rated FTS output at the category of almost all of the meaning in the translation with 41%, followed by 40.5% who believed that FTS output is complete as fluent. However, 14% of the responses showed that the translation is not fluent, and FTS output contains some of the ST meaning, followed by 4.5% for FTS as disfluent.

Table 6: FTS Adequacy

Adequacy	Sum of Frequency	Sum of Percentage
All the meaning in the source is contained in the translation, no more, no less.	81	40.5
Almost all the meaning in the source is contained in the translation.	82	41
Fragments of the meaning in the source are contained in the translation.	28	14
None of the meaning in the source is contained in the translation.	9	4.5
Grand Total	**200**	**100**

5.7. Have you rated the adequacy of FTS output?

The above chart elucidates how the study respondents rated FTS output fluency (grammar). It showed that the majority of respondents rated FTS output as fluent with minor errors at the percentage of 69.65%, followed by 23% indicated that FTS output rated FTS output as fluent with no errors. On the other hand, 6.5% rated FTS output as disfluent.

Table 7: FTS Fluency

Fluency	Sum of Frequency	Sum of Percentage
Refers to a perfectly flowing text with no errors.	46	23
Refers to a smoothly flowing text even when a number of minor errors are present.	139	69.5
Refers to a text that is poorly written and difficult to understand.	13	6.5
Refers to a very poorly written text that is impossible to understand.	2	1
Grand Total	**200**	**100**

5.8. Have you rated FTS output?

Then, the study aimed to elicit information about the FTS output rating by the participants. The analysis showed that 55.5% of respondents rated FTS sometimes, 42% rated FTS always, while 2.5% were not sure they rated FTS. The study here emphasizes that rating is crucial to improve its service and provide the end-users with information about the end users' trust of FTS output development of FTS service.

Table 8: FTS Rating

Classification	Sum of Number of respondents	Sum of Percent %
Always	84	42
Not sure	5	2.5
Sometimes	111	55.5
Grand Total	**200**	**100**

6. Discussion

The current descriptive study assessed the popularity of FTS in translating English Facebook posts into Arabic and Posts related to COVID-19 in particular. Participants were found to be familiar with FTS usage during the COVID-19 lockdown. The results showed that social networking sites have a significant role in our modern life. Facebook becomes one of the most popular social media websites. Facebook becomes a source of information during the time of COVID-19 lockdown since Facebookers read posts around the clock. The study showed that (94%) of participants used Facebook daily to read, share, and post. This percentage complied with [[7]] that Facebook occupied the first rank in social media stats with (88.69%) over June 2020 in Jordan.

Our assessment of the information sources showed (62.2%) of Jordanians considered FTS as the primary source of their information during the COVID-19 lockdown. The study attunes with what [[11]] and [[21]] said in their research that Facebook becomes the source of information for medical students in Jordan (83.4%). This means that Facebook becomes a popular online platform that supports the efforts to fight COVID-19 at the government side, where the government officials and health care providers to convey their messages to control the outbreak of COVID-19. Even though FTS became a source of information in translating English Facebook posts into Arabic, but the need for a thorough analysis of the output of FTS by a translation scholar is of great importance to highlight the limitations and strengths of FTS.

The study indicated that FTS was used daily, equal to (94%). Moreover, it is found that Jordanians used FTS in translating COVID-19 posts into Arabic, with a percentage of (83.1%). This finding complied with [[11]] finding that (83.4%) medical students in Jordan used Social Media to learn about COVID-19. However, they did not ask the participants about their preferred social media platform. This finding motivates the researcher to conduct further study to verify the capacity of FTS to provide adequate and fluent output to the end-users.

The study found that the participants rated FTS output adequacy with (41%) almost the meaning of the ST in the TL; (40.5%) showed that FTS output is fully adequate to the end-user. Furthermore, it is found that (69.65) of FTS output has minor errors, while (23%) is free of errors. On the other hand, the study highlighted that Jordanians sometimes tended to rate FTS service with (55.5%) for rating FTS and always rating for FTS with (42%). The rating provides a review for the FTS team to improve its services. The rating benefits Facebook in two ways: help the users make decisions to use FTS or not and help Facebook gain the users' credibility.

7. Conclusion

Overall, Facebookers in Jordan tend to use FTS in rendering English posts in general and posts related to COVID-19 into Arabic. This means that Facebookers used Facebook as their primary source of information. The study concludes that in such an unprecedented COVID-19 crisis, the health officials should create Facebook Profiles with a blue tick for medical information during crises. Countries where the outbreak is hitting hard should focus on educating the public. Moreover, medical specialists and translation scholars should form evaluation groups to ensure that FTS can render COVID-19 medical posts into Arabic fluently and adequately. Finally, the study recommends further studies to scrutinize FTS outputs thoroughly to highlight the service's limitations and strengths.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to extend their thanks to all participants in this study. The research has been funded by Applied Science Private University, Jordan.

References

[1] Z. Almahasees, Diachronic Evaluation of Google Translate, Microsoft Translator and Sakhr in English-Arabic Translation — the UWA Profiles and Research Repository, 2020.
[2] Jack, The Top 3 Translation Industry Trends for 2019, 2019.
[3] Z.M. Almahasees, "Assessment of Google and Microsoft Bing Translation of Journalistic Texts," International Journal of Languages, Literature and Linguistics, 2018, doi:10.18178/ijlll.2018.4.3.178.

[4] Omniscen Technologies, How fast is machine translation? : Language Studio Help Desk, 2016.
[5] KH, How Big Data And Machine Translation Combine To Fight COVID-19, 2020.
[6] Kenton, Social Networking Definition, 2020.
[7] Statcounter, Social Media Stats Jordan | StatCounter Global Stats, 2020.
[8] Z.M. Almahasees, "Assessing the Translation of Google and Microsoft Bing in Translating Political Texts from Arabic into English," International Journal of Languages, Literature and Linguistics, 2017, doi:10.18178/ijlll.2017.3.1.100.
[9] Sistanis, Top 10 Languages Used On the Internet for 2020 - Speakt.com, 2020.
[10] Debusmann, Facebook teams up with Middle East firm to combat Covid-19 rumours - Arabianbusiness, 2020.
[11] A.I. Khasawneh, A.A. Humeidan, J.W. Alsulaiman, S. Bloukh, M. Ramadan, T.N. Al-Shatanawi, H.H. Awad, W.Y. Hijazi, K.R. Al-Kammash, N. Obeidat, T. Saleh, K.A. Kheirallah, "Medical Students and COVID-19: Knowledge, Attitudes, and Precautionary Measures. A Descriptive Study From Jordan," Frontiers in Public Health, 2020, doi:10.3389/fpubh.2020.00253.
[12] Singh, How to enable new Facebook In-Line Translation Tool on Facebook Pages | Tech 24 Hours, 2020.
[13] Knowledge Complexity, A recipe for intimacy?, 2017.
[14] J. Gehring, M. Auli, D. Grangier, D. Yarats, Y.N. Dauphin, "Convolutional sequence to sequence learning," in 34th International Conference on Machine Learning, ICML 2017, 2017.
[15] Facebook AI, New research awards in NLP and machine translation, 2019.
[16] Facebook Engineering, A novel approach to neural machine translation - Facebook Engineering, 2017.
[17] Z. Almahasees, "Machine Translation Quality of Khalil Gibran's The Prophet," Arab World English Journal For Translation and Literary Studies, 2017, doi:10.24093/awejtls/vol1no4.12.
[18] T. Loren, Social Media & Covid-19: How to Communicate During a Crisis - Later Blog, 2017.
[19] Garza, How Social Media Is Shaping Our Fears of the Coronavirus | Time, 2020.
[20] Zakrzewski, The Technology 202: Coronavirus could change how social networks approach public health - The Washington Post, 2020.
[21] H. Alzoubi, N. Alnawaiseh, A. Al-Mnayyis, M. Abu-Lubad, A. Aqel, H. Al-Shagahin, "Covid-19 - Knowledge, attitude and practice among medical and non-medical university students in Jordan," Journal of Pure and Applied Microbiology, 2020, doi:10.22207/JPAM.14.1.04.
[22] H. Ouyang, "At the Front Lines of Coronavirus, Turning to Social Media," The New York Times, 2020.
[23] Heilweil, Coronavirus social distancing leads to empty churches and a rise in apps - Vox, 2020.
[24] Wong, Coronavirus sparks a different kind of problem for social networks - CNET, 2020.
[25] Scott, Facebook to tell millions of users they've seen 'fake news' about coronavirus - POLITICO, 2020.

Appendices:

Appendix 1: Frequency and Percent for demographic variable

Variable	Classification		Frequency	Percent %
Gender	Male		95	47.3
	Female		105	68.2
Education	Bachelor		46	22.9
	Diploma		87	43.3
	High School		10	5.0
	Master		35	17.4
	PhD		21	10.4
Cities	Amman	Urban	50	25%
		Rural	0	0
	Ajlun	Urban	19	9.5
		Rural	11	5.5
	Jerash	Urban	26	13
		Rural	9	4
	Irbid	Urban	30	15
		Rural	20	10
	Kerak	Urban	21	10.5
		Rural	15	7.5
Total			200	100

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 520-530 (2020)

www.astesj.com

Second International Virtual Conference on Multidisciplinary Research 2020

ASTESJ

ISSN: 2415-6698

Accelerating Decision-Making in Transport Emergency with Artificial Intelligence

Alexander Raikov*

Institute of Control Sciences, Russian Academy of Sciences, 119526, Russia

ARTICLE INFO

Article history:
Received: 04 March, 2020
Accepted: 10 July, 2020
Online: 20 November, 2020

Keywords :
Artificial intelligence
Big data analyses
Cognitive modeling
Convergent approach
Virtual meetings
Transport emergency

ABSTRACT

The paper addresses speeding up meetings in a networked environment during rescue works in a transport emergency. Several groups of representatives of various services and observers participate in those meetings. The number of wrong decisions tends to increase because remote participants cannot understand each other quickly. First, the meetings must be efficiently held to avoid making wrong decisions, including medical diagnoses for injuries. The ultimate goals are to improve injured' health and life. Artificial intelligence (AI), big data analysis, and deep learning methods suggested in this paper for decision-making support have a cognitive character, i.e., try to take into account the thoughts and emotions of participants. The author's convergent approach ensures the purposefulness and sustainability of decision-making. This approach transforms divergent decision-making processes into convergent. The approach is based on the inverse problem-solving method in topological space, genetic algorithms, control thermodynamic theory, and using the ideas of creating AI models' cognitive semantics with quantum mechanics methods. This approach gives meetings' members the list of decision-making rules with accelerating consensus achievement. The examples of the rules are: the goals have to be arranged as a 3-level tree and ordered by importance; semantic interpretations of computer models' factors and their connections must be separated; rescue resources must be represented in a finite number of separated components, and so on. The approach also exploits traditional technical tools of augmented reality, virtual collaboration, and situational awareness. It has been repeatedly used to build socioeconomic and manufacturing sectoral strategies and is currently being adapted for emergencies.

1. Introduction

Every second has great importance after the moment of the outbreak of transport emergencies. The first hour after the moment is named the "golden hour"—it is a critical time for saving the injured and suffered lives. Different types of transport entities may take part in automobile, railway, aircraft, all-terrain vehicle (ATV), motorcycle, bicycle, etc. The authorities and their departments, rescue teams, police, ambulance staff, representatives of the population, etc. may participate in the processes. They all must understand each other and get agreement on decisions very quickly.

Rescue teams usually have a special pre-designed guide to synchronize their participants' activities as usual. It may be represented by way of detailed prewritten instructions, plans, or models of activity. The members of the rescue team must know it in advance. But every emergency is characterized by its own features and new participants, and the standard guide sometimes can only confuse the situation. The activity plans of all participants must be synchronized in any circumstances and for different environments. The integrity of the team member tactics' implementations has to achieve the proper synergistic effect.

Special decision support systems have been created based on virtual reality tools and artificial intelligence (AI) technologies. A sharp rise of research and the creation of such systems for this activity began after the 9/11 attacks and the 2009 crash of an Air France Airbus A330. The situational awareness (SA) approach [1, 2] was suggested and used in researching these disasters. Profound technical decisions were found, and obvious progress was made [3].

Simultaneously, the crucial importance of accelerating the collective decision-making by getting agreement between different participants in extreme conditions makes the issue of speeding up virtual meetings in emergencies always relevant. Some author's

*Corresponding Author: Alexander Raikov, alexander.n.raikov@gmail.com

papers are devoted to accelerating distributed collective decision-making that can be held in situational centers (SC) or a networked environment [4]. It was already suggested the technologies that help: speeding up a corporate strategic conversation [4], raising the quality of networked experts' procedures [5], improving medicine meeting (conversation) [6], or even accelerating networked democracy processes [7] and using for these methods from quantum theory [8,9].

Brainstorming procedures are divergent as usual; they have goals of creating a lot of new ideas. But these ideas may not be consistent, and as a result, synergy cannot be achieved. The author's convergent approach ensures the purposefulness and sustainability of networked collective meetings [4, 5]. It has been tested in different fields of the social-economy sphere and for decision-making in SC. It is now suggested for networked and collective decision-making support in emergencies when many different participants and physical automated tools (computers, cars, and intellectual robots) take part.

Three meeting modes may be distinguished: monitoring the situation with an operational solution of ongoing issues, planned decision-making with long-term preliminary preparation, and decision-making in an emergency. The latter is the most difficult. It claims a special approach that will ensure the convergence of distributed meetings and conversations to save the lives of as many injured people as possible. The paper has devoted this mode.

The paper includes a description of a transport emergency's characteristics, considers the limitations of the available rescue means of situational awareness. The concepts of denotative (formalized) semantics and cognitive (non-formalized) semantics for AI models are introduced, wherein, shows that modern AI predominantly uses only the former one. Further, the paper examines the possibilities of advanced AI and proposes to consider the possibility of using quantum theory to construct cognitive semantics. The author's convergent approach is then proposed to speed up networked distributed meetings in conditions of transport emergency. Simultaneously, the author's approach to the organization of networked expertise and cognitive modeling is considered in more detail. Then, some practical implementations of this approach are demonstrated. The paper ends with the formulation of discussion questions.

2. Emergency Features

The following features characterize an emergency transport situation (railway, aviation, highway, etc.):

- Presence of injured and/or dead.
- Victims need quick medical attention.
- Extreme lack of time.
- Possibility of destruction or contamination of the area.
- Participation of many departments, organizations, and rescue teams in decision-making.
- Distribution of collective decision-making process.
- Lack of material resources in the proper amount.
- Experienced experts are not available, etc.

In such situations, special tools have been created to speed up and improve the quality of collective decision-making, including virtual collaboration systems and SA systems [3]. They differ significantly from traditional teleconference. They are based on the advanced technology of discourse processing to ensure accelerated mutual understanding during participants' communications. SA is associated with multidimensional visualization, targeted information. It supports group decision-making, perception, prediction of the external environment changes, and mutual understanding between participants.

It is shown that in emergencies, network teams act faster than hierarchical ones [1,5,7]. The critical factors are as follow lack of necessary information, a holistic view of the situation at all levels of management, and the absence of an overall operational picture. Moreover, the growth of data volumes and the number of precedents do not always improve SA and quality decisions. The main sources of the problems that create information gaps are the team members themselves. They are focusing on solving their own problems. Inaccurate and belated information about events is the reason for information gaps too. The aspect of the lack of trust is important. Recently, the concept of SA is increasingly a group character: participants have to get an equally understanding of an operational picture when information is visualized on a map.

The analysis of the issue of accelerating the group decision-making in emergency using SA and virtual collaboration approaches showed the emphasis of the engineers' and researchers' efforts on the following areas:

- The creation of a distributed SC for monitoring and managing the processes of eliminating the consequences.
- Forecasting the danger of traffic accidents for ensuring effective decision-making.
- Integrating methods of information and analytical support crowdsourcing and networked expertise processes.
- Ensuring accelerated convergence of decision-making processes based on cognitive modeling and inverse problem-solving methods, etc.

One of the most efficient approaches to improve group decision-making processes in emergencies is SA. Let us take a look at it in more detailed.

3. Situational Awareness

The SA process embraces individuals, teams, AI systems, machines, and computer tools. It includes several stages [1,3,10], has a gradation in levels [11], and is characterized by different models [12]. The process may have abnormal behavior that is defined as behavior to deviate from the global plan in some circumstances.

It is known that SA systems do not always behave sustainable and purposeful. Some studies try to represent cognitive non-formalized processes by linking participants with AI systems. It allows each actor to get information in special templates from other actors that help to increase the opportunities of all actors [13]. Non-predicted and false responses are still common [3]. It is usually considered that SA is a suitable framework for collective cognitive

modeling that helps to make effective decisions by using explicit and implicit information.

But most papers do not consider the chaotic nature of human cognition, especially in cases of responding to abnormal machine behavior. Participants' thoughts, emotions, free will, experience, senses, etc. cannot be described logically. It cannot be represented in computers' logical knowledge bases. As soon as an attempt to describe an emotion is made, this emotion disappears or loses its nature and creative power. This fact compels to consider simultaneously two types of cognitive processes—systems 1 and 2 as shown in Table 1.

Table 1: Two types of cognitive systems

System number	Operating means	Semantics type	Serves to ensure
1	Non-logically, unconsciously, uncaused, associatively, intuitively	Cognitive	Provides creativity, helps to get insights[8], "Eureka effect" [15]
2	Logically, deliberately, rationally and analytically; neuron deep learning (partially)	Denotative	Logical inference; provides necessary conditions for purposeful and sustainable decision-making [4].

The systems are interacting. Some researchers argue that the system 1 has an advantage over the second one. But it obvious that intuition is not only the ability to make a decision.

Relevant information of the system 2 may be claimed to avoid mistakes in judgment [14]. This system gives the possibility to make the features of human activity and task characteristics distinct. An example of the presentation of system 2 is big data. They are used to build denotative semantics. However, it should be noted that the juxtaposition of each factor or connection of the AIs' models with some relevant subset of big data will require the use of special methods for solving problems on cognitive models, namely, methods for solving large-scale global optimization problems [15]. But due to the very high requirements for the promptness of solving problems in emergencies, most likely, such an approach for such situations is unacceptable. This paper proposes replacing big data subsets with their projections on a deep neural network (See Section 8).

The creative process of getting an insight may be non-logical; the result appears after deliberate and sometimes long-time unsuccessful thinking. It needs to use logical efforts to speed up the process. [16]. It was suggested to put a cognitive continuum of the first above-mentioned process in the unconscious end of some scale, but the second one—in the conscious end [17]. Distinctions between these two processes have a continuum power. Effective decision-making may require flexibility on this continuum. And intuitive decision-making can be partly conscious, and analytical decisions can be partly unconscious.

Situational awareness in emergencies is working with machines and humans. It needs to control machines' abnormalities and supports correct group decision-making during abnormal machine behaviors. The three-level SA model was suggested for these cases [18]. It includes the following components:

- Perception of the situation (level 1).
- Comprehension of the situation's meaning and assessment relevance to certain goals (level 2).
- The projection of its status and dynamics for the near future (level 3).

Many relevant issues in different safety domains (aircraft, air traffic control, power plants, maritime operations, manufacturing systems, collaborative and isolated work, healthcare, etc.) benefit from the application of the three-level SA approach.

As has been shown, SA can cover a wide range of cognitive concepts. It is affected by psychological phenomena such as stress, mental workload, and trust [19]. The three-level SA approach was used in deliberate sense-making, story construction, recognizing main decisions, heuristics, and implicit cues of sociotechnical systems. The three-level SA approach has been used to explain deliberate attempts of detecting, diagnosing, and forecasting and increasing collective understanding of the problem situation.

An important aspect of SA is the sense-making process that explains how humans understand the environment by assigning meaning to experiences [20]. Sense-making is about values, priorities, and preferences. It holds different elements together to guide actions. Sense-making is a complex process and involves previously accumulated information. Different authors have tried to model cognition by identifying cues, diagnosing problems, taking action, questioning, elaborating, comparing, switching, and abandoning the search for a frame [21].

Distributed SA (DSA) is a more difficult construction than SA [22]. Its attention devotes to special issues of covering distributed cognition and focuses on collective SA, or awareness between agents. Humans and machines can represent them. DSA is used in modeling virtual collaborative processes.

DSA differs from SA substantially. For example, for DSA, it does not matter if the human agent does not know anything. Other components of the distributed system can augment the lack of his or her knowledge. On the contrary, in the case of SA, if the human does not know something, it may cause a significant problem or accident. But SA can support a broader spectrum of cognition because of the implicit character of thinking.

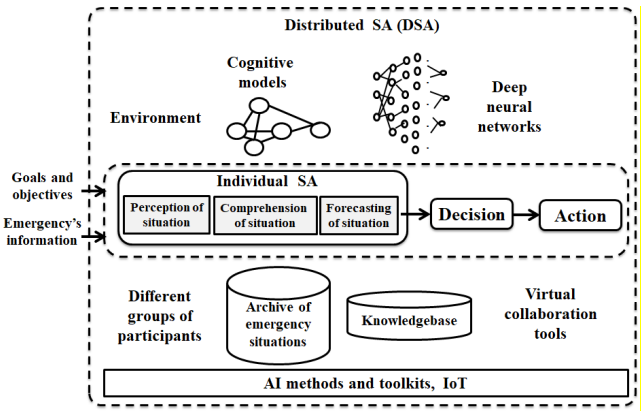


Figure 1: SA and DSA systems' components

It was identified that there are four classes of failure transactions that can be the reason for the accident [3]: absent transactions, inappropriate transactions, incomplete transactions, and misunderstood transactions. Every one of these transactions has implicit and explicit interpretations, intangible and tangible representations. SA and DSA have framework constructions that can be represented in a formalized way by frames, ontology, or proposition logic. The components of SA and DSA systems are illustrated in Figure 1, where the idea of individual construction has been taken from [3].

Thus, SA and DSA have two semantics' interpretations: cognitive (non-formalized process or System 1) and denotative (formalized process or System 2).

4. Cognitive and Denotative Semantics

Emotions, thoughts, insights, intuition, inspiration, etc. come without awareness. They are usually defined as the capability to act or make decisions appropriately without consciously balancing alternatives, without following a certain rule.

Human and AI cognitive abilities are deep discussed in [23]. This paper analyzed 2500 publications. There are about three hundred cognitive architectures that describe human cognitive abilities. This paper's main attention was made to a logical description of cognitive phenomena. It was also noted that the topic of cognitive architectures is rather rare in works devoted to Artificial General Intelligence (AGI).

Therefore, SA or DSA does not cover this communication layer and participants' activities or covers it only in a very restricted means. They do not fully reflect the semantics generated by thoughts and feelings. In the SA or DSA paradigm, only formalized semantics can be represented. Moreover, in an emergency, technical devices and computers reflect only formalized, material aspects of the situation.

A machine has its own personality, which could be defined as only the technical perception of everything the participants know and feel [24]. Humans have the ability to apply their senses to identify and interpret signals to get their own useful descriptions of machine behavior. As a result, tacit knowledge [25] cannot be embraced by computer AI models; the computer cannot understand them; therefore, digital information from sensors may be estimated in the wrong way.

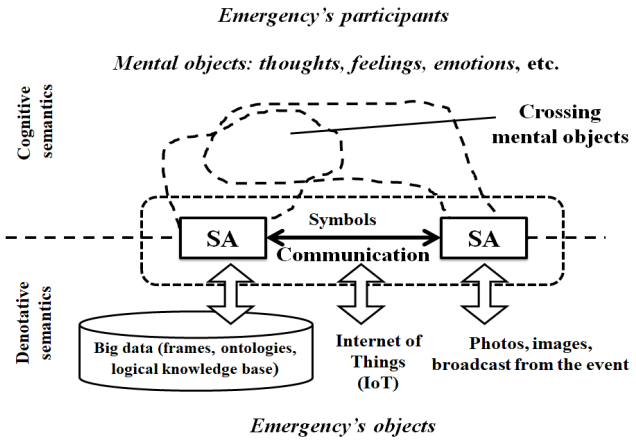


Figure 2: SA and its semantics: cognitive and denotative

Thus, there are two types of the semantics of SA or DSA: denotative (formalized) and cognitive (non-formalized)—see above Table. These semantics are illustrated in Figure 2.

Figure 2 demonstrates that the denotative semantics of SA and DSA has a formalized structure and can be implemented using logical ontologies, symbols, and frames. The instruments of the above-mentioned SC [4] can be useful for centralizing the control of monitoring problems' situations' and ensuring coherence of group decision-making and actions of various teams and organizations. SC helps make group decision-making processes stable (sustainable) and purposeful by involving AI and virtual collaboration technology to give participants the power to achieve synergy in the decision-making processes [26]. SC is using for solving problems that can be characterized by the following specific features:

- Ill-defined and weakly formalized problems.
- Require urgent collective decisions.
- Associated with high risks.
- Poorly predictable and chaotic behavior.
- The process can be interrupted.
- Divergent trait of breakthrough procedures.
- Lack of retrospective experience, etc.

SC usually exploits modern technologies by integrating:

- Strategic analysis methods;
- SA and DSA technologies.
- Advanced tools of visualization.
- Networked convergent conversation technology.
- Cognitive modeling.
- Inverse problem-solving methods.
- Logical analysis, statistical forecasting.
- Big data analysis, knowledge engineering.
- Different AI technologies, etc.

Simultaneously, the cognitive semantics cannot be fully taken into consideration by these technologies because of its non-formalized stuff. They have a more qualitative and non-metric nature than quantitative and metric ones. The cognitive aspects of decision-making and modeling in the SC have to ensure new requirements for cognitive semantic interpretations of SA or DSA. The main idea of making cognitive semantic interpretation by logical computer models consists of:

- Taking into consideration individual and collective unconscious phenomena using indirect and inverse methods of decision support.
- Impacting on the problem situation through a self-developing reflexive-active environment.

Using the ideas of AGI and Strong AI perhaps can help realize this idea. Strong AI and AGI are usually considered different

phenomena [27], but in this paper, we will consider them synonyms without loss of generality. However, considering the practical need to introduce AI tools shortly, the main emphasis will be on using traditional AI, although more developed cognitive semantics are used in its application.

5. Artificial General Intelligence

The traditional AI has been developed on a logical basis. It uses predicate calculus, logical ontologies, frames, and neural networks (NN). It may be noticed that the latter can be represented by logic also. This AI has a strong potential for solving different issues in real practice, such as predicting human behavior and different situations based on experience. This traditional approach can help to recognize the problem situation based on retrospective deep learning. Since traditional AI has a formalized nature, its semantics is denotative.

Currently, the traditional formalized AI gives way to non-formalized and uncaused AI, the semantics of which have cognitive features. The design of AGI can obey reverse-engineering rules. Simultaneously, in the history of AI development, there are roughly four methodologies to achieve kinds of AI: symbolism, connectionism, behaviorism, and statisticalism [27].

AGI is usually considered the machine intelligence that can perform intellectual problems in different fields of knowledge more efficiently than a human being can adapt to the environment and forms its own self-awareness. It is fundamentally non-formalizable human-machine (hybrid) AI. The AGI encompass the subjective, mental, creative, and intellectual activity of human being. The distinctive claims of the AGI are as follows [28]:

- Creation of the necessary conditions (see Section 7) to ensure the convergence of generating ideas and thoughts to ill-defined goals.
- Complex cognitive semantic interpretation of different problem situations' models.
- Taking into consideration the conceptual aspects of the situation.
- Ensure the sustainability and purposefulness of current and future development.
- Computers' models may be changed in a quantized way.
- The behavior of models is synchronized with the behavior of remote objects, the states of which are unknown.
- The space of modeling is conceptual, phenomenological, and uncaused.
- An external impact on the problem situation can be changed unpredictably, etc.

Thus, these claims have to be taken into account during organizing distributed meetings. It requires changing the methodological that was based on traditional AI. The possibilities of AGI's elements have to be taken into account. AGI includes all the advantages of AI and is intended to make a wide range of human cognitive capabilities. But for a while, it is obvious that there is a large gap between AI and AGI. It may be just said that AGI is the inverse of AI [27,28]. In this context, it is necessary to consider such aspects of AGI as:

- Imitation of human consciousness, perhaps, with non-classical methods from such disciplines, as relativistic and quantum physic, thermodynamics, and optics.
- Take into account cosmological, relativistic, atomic, quarks, and microtubule components of the brain.
- Take into account the physical fields (gravitational, electromagnetic, strong, and weak) on the thoughts' processes.
- The group theory and category theory application.
- Inverse problem solving on topological spaces, etc.

The development of AGI requires an interdisciplinary synthesis of different approaches, methods, and technologies. The synergy of these components of AGI has to overcome the non-finite complexity of cognitive semantics. These effects imply the modeling beyond the traditional formalized mechanisms and models of thinking.

E.g., in the large paper [29] devoted to building machines that think like people note the importance of drawing a distinction between AI that purports to emulate or draw inspiration from human cognition and AI that does not. This paper focuses on the former that looks like AGI. The latter undoubtedly is a perfectly reasonable and useful approach to developing AI algorithms for real practice: avoiding cognitive inspiration and claims of neural plausibility. But for emergency requiring rapid collective decisions, this is clearly not enough.

The key aspects of AGI are creating cognitive semantics that can be embraced only by indirect approaches. One of these approaches is quantum semantics.

6. Quantum Semantics

The description of the disaster situation is usually infinite-dimensional. In such cases, the mechanisms of creating quantum semantic interpretations for computer models can find practical application. The quantum principle of complementarity helps to expand and deepen the semantic interpretation of computer models. Applying the special quantum operators to sets of messages created during virtual conversations between participants [8, 9].

The uncaused nature of feelings, emotions, or free will force studying the quantum semantics approach [30,31]. The special quantum operator can help to embrace these unconscious components in decision-making processes [8].

A special place in the AGI approach is occupied by quantum-mechanical effects, such as superposition, complementarity, and non-locality effects [30]. For more complete and detailed coverage of the quantum and relativistic aspects of the cognitive semantics, an attempt can comprehensively study the thought processes using the group theory, which investigate sets with operations on them. The elements of spatial sets will then be various entities: events, ideal and virtual objects, complex numbers, vectors with their scalar products. Each event will be associated with a point in space, considered a mathematical object, different from others. Thus,

both the Fourier transforms, and transformations over the 4-dimensional Minkowski event space elements can be covered. The points of this space can have a non-numerical character, the coordinates of which serve as one of the many ways to specify these points. It helps to create cognitive semantics. The quantum cognitive semantics allow raising the power of AI substantially, e.g.:

- The quantum particle is represented in infinite-dimensional space—it may look like a symbolic representation of the spiritual phenomenon.
- The quantum particle is both a particle and a wave and gets another type in another place—it can reflect that a word's meaning depends on the context.
- Attempts to detect a wave of a quantum particle lead to the disappearance of the interference effect— it looks like the inability to logically describe cognitive semantics.
- The particle's behavior is described with the complex numbers—it can be used for representing the "shadow" components of the cognitive semantics.
- A quantum particle cannot have zero values of its parameters and occurs in a jump-like manner—it looks like the Eureka effect [16].
- The non-locality effect ensures the dependence of the quantum particle's behavior on other particles for which the behavior is not defined—it is associated with the interdependence of natural phenomena, etc.

The quantum semantics features increase the power of AI, making forecasting and solving processes more reliable. The way of realizing the possibilities of quantum semantics consists of emulating quantum operators on a common computer. For example, the paper [8] demonstrates using the Hadamard quantum operator, which interprets the quantum superposition function on a set of words with their frequencies (frequency histograms) in documents. The quantum operator changes the frequency histograms of words from the set of documents, making the coverage of these sets more complete. It also ensures the mapping of different sets of documents and words closed (in mathematical mean) that is the prerequisite for ensuring correctness of the inverse problems solving process (see sections 7, 8, and [4]).

7. The Convergent Approach

The collectives' decision-making processes have to be purposeful and sustainable in emergencies. The participants generate different ideas based on incoming information and make quick assessments of the various solutions' alternatives and suggestions. Creating ideas and making assessments include many interactions between participants using different units of information. The process may be influenced by false or latent information. This influence can greatly reduce the efficiency of group decision-making. This process may be divergent and has to be transformed into convergent (purposeful and sustainable).

Different types of operational objectives during group decision-making are suggested. Let's call them, for example, usual (traditional) and unusual (best) goals. Different values between these extreme values of the types of goals can be placed on the scale.

Usual objectives can be prescribed by automatically extrapolating retrospective statistics by traditional AI methods based on previous experience or retrospective information. The expert system, big data analysis, deep learning, ontology, etc. can be used for this aim. Most of the evaluations of those systems of goal setting are based on pattern recognition methods on retrospective data analysis.

For achieving the unusual, high-quality goals that cannot be achieved with traditional decision-support systems (DSS), the convergent approach was suggested [4, 8]. It has the opposite (inverse) character concerning DSS and AI. To ensure the efficiency of creating group high-level advice (decision-making), it is necessary to create special necessary structural conditions for support purposeful and sustainable development of the decision-making process. The convergent approach is based on the inverse problem-solving method on topological spaces and cognitive modeling. The provisions, algorithms, and programs to accelerate the decision-making process in emergencies with these methods were developed. The difference between the direct and inverse approaches is illustrated in Figure 3.

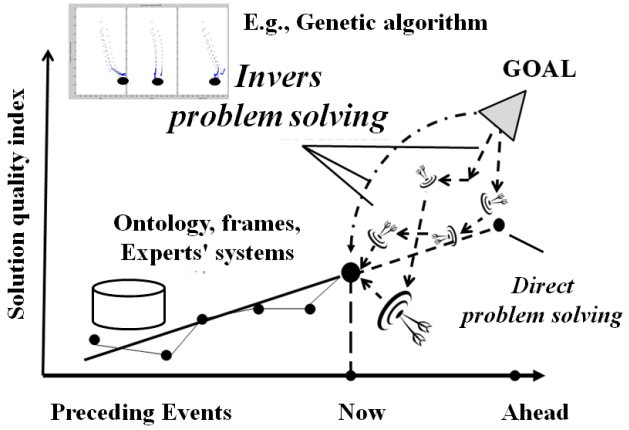


Figure 3: The difference between two approaches of decision-making

If participants of an emergency meeting try to use both these two different approaches (direct and inverse), they apply AI and AGI technologies. For inverse problem-solving, the Tychonoff theorem [32] on topological spaces was applied to ensure the meeting's stable purposefulness.

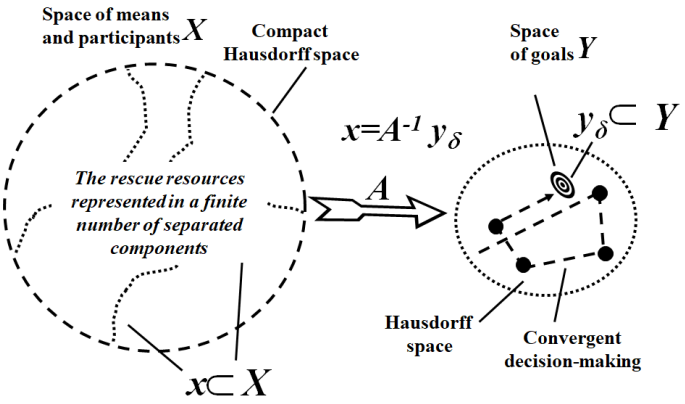
It was confirmed that the semantic interpretation of cognitive models in the decision-making process and the methods of inverse problems solving increase the purposefulness of the actions of the leader of a group of people (various services, rescue teams, spontaneous groups of people, etc.) and, therefore, the speed and quality of decision-making in emergencies.

The convergent approach provides the necessary conditions for the sustainable convergence of collective decision-making processes to ill-defined goals. To ensure these necessary conditions, category theory and the topology theory were used, and the concept of a "convergent monad" (CM) was introduced for which the classical monad was improved with the axioms of CM ℵ [33]:

- D: Top → Top; the number of points in the system of sets Top (topology) is infinite, and the graph of maps is closed (the complement of an open set).
- ℵ is a topological space where its open covers have non-empty finite subcover (*The Compact space*);
- Any two points of ℵ (in the topological sense) have disjointed neighborhoods (*The Hausdorff space*).

These axioms follow from the statement: if X and Y are Hausdorff topological spaces and X is a compact space, and the graph A is closed, these are necessary (but not sufficient) conditions for ensuring convergence of the inverse problem solving [34]. The convergent methodology allows dividing the problem into some parts and then assembling them into a holistic decision.

The following sets and an inverse operator were selected for reflecting the search of the solution in an emergency: the space of goals Y (the exact goal y_0 and the set of inaccurate goals $\{y_\delta\}$); the space of participants and means X for achieving y_0; the operator $A: X \to Y$, and the reverse to it. The domain of definition of the operator A can be narrowed down to some subset of X, and the set of decisions can be narrowed down to a subset of Y (Figure 4).



It is shown that a countable set and an infinite set of real numbers are clearly not suitable for identifying components and describing the space of solutions. This result allows supplementing the structure of the data using the requirements that were formed earlier [4]. For example, for increasing the convergence of a decision-making process, it is required that the number of parameters (factors) that characterize the problem situation should be finite. In this way, these factors and their semantics have to cover the description and representation of the problem and have to be separated, or the border could be created between every two parameters. Considering these requirements, a network brainstorming method based on a genetic algorithm was developed for inverse problem solving [35]. It can be used in emergency conditions for convergent group generation of ideas and formation of an agreed action plan.

Thus, for ensuring the stability and purposefulness of decision-making in emergencies, the components of the goals, rescue means, and ways (algorithm) of the achievement of the goals should be highlighted and structured especially. The emphasis should be focused on the possibility of accelerating the adoption of group decisions in networks and mobile environments and an interdepartmental system of distributed SC. The collective decision-making process involves the networked expertise approach, cognitive modeling, and mathematical apparatus for inverse problems solving [5,7,8].

It was also shown [4] that to ensure the necessary conditions for a stable and purposeful synthesis of effective solutions in a problem situation; it is also advisable to distinguish the components: the whole, order, and chaos. To ensure the desired stability of decision-making processes, these components and their rate of change must be in a certain ratio (Figure 5).

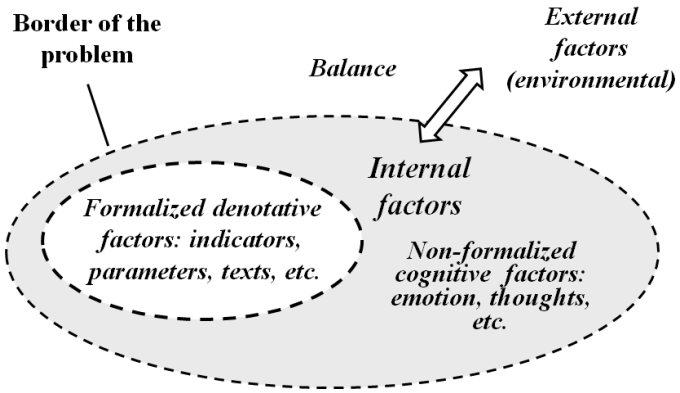


Figure 5: Chaotic component of ensuring the desired stability

The convergent approach offers a list of rules for structuring the information that is created during emergency meetings, as follows:

- The goals have to be constructed as a 3-level tree with main, external and internal goals;
- The goals must be ordered by importance with, for example, the analytic hierarchy method [36];
- Cognitive and denotative semantic interpretations of the model's factors, and their connections must be separated;
- The set of decisions must be represented by a finite and observable set of factors, etc.

Such recommendations help create conditions that ensure the purposefulness and sustainability of the decision-making process in an emergency. As shown by practical testing of the proposed approach's use with the creation of appropriate conditions, the duration of holding collective strategic meetings is reduced by 4-5 times compared to cases when these conditions are not met. However, for emergencies, the acceleration of group decision-making should be significantly greater.

The participants and engaged network experts must have the possibility of putting qualitative information into the process of decision-making in an interactive way by using the cognitive modeling method [4, 5, 35].

8. Networked Expertise and Cognitive Modeling

Different members can participate in disaster meetings to manage an emergency, and some remote experts can be invited [5]. On the base of convergent technology, networked expertise

procedures can be ordered, as shown in Figure 6. This order accelerates the decision-making processes by speeding up getting agreement between the emergency meeting participants regarding goals and ways of action. The SC can centralize monitoring of the situation and ensure the coherence of actions of various teams and organizations.

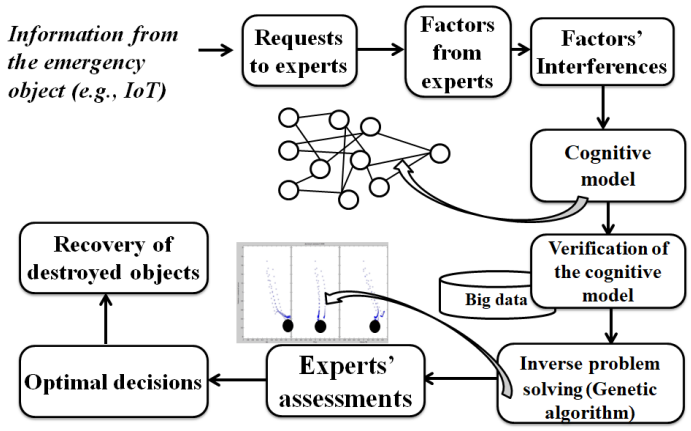


Figure 6: The order of the networked expertise and decision-making

The main steps of the networked expertise are as follows:

- Make the requests to the experts.
- Describe the situation shortly; formulate goals of improving the situation.
- Suggest some conceptual factors that characterize the opportunities to improve the situation.
- Define factors' interferences and create the cognitive model.
- Assess and explore the project of the cognitive model in remote SC.
- Solve the inverse problem with a genetic algorithm for getting the optimal decisions.

As a result of the network experts meeting, the final strategy of activity is formulated. The most difficult part of a group expert procedure is cognitive modeling. For simplifying and speeding up the process of cognitive modeling, an array of reference cognitive models, methods of Big Data analysis, and Deep learning should be used.

In a transport emergency, which claims fast decision-making and effective planning, the traditional manual creation of formalized computer models is impossible because it may take a lot of time. The problems have the characteristics of randomness, chaos, quantization, uncertainty, and non-causality. In these conditions, the participants must apply intuition, experience, and experts' judgments.

For such problems, the approach with the automation of cognitive modeling could be applied [37,38]. The cognitive modeling method is used when non-metric (qualitative) concepts are more important than metric (quantitative) ones. It ensures factors and relations between factors that characterize the problem. It may be represented in a matrix or a directed graph, where nodes represent factors, and arrows represent relations. The process helps to understand the structure of the problem and improves the understanding between participants.

The retrospective experience can help to create the cognitive model automatically. The creation of a cognitive model could be accelerated by using Big Data analysis. It helps to verify early prepared cognitive models and supply ones with denotative semantics in a formalized way. For this, the cognitive model's factors and their connections have to be mapped into the relevant subset of Big Data.

The deep learning of the neural network method was carried out for automating cognitive model creation and verification. The features of this process are as follows:

- Check the adequacy of the cognitive model to the problem situation.
- Ensure the availability of relevant subsets of Big Data.
- Use standard software for training the neural network.

An approach with a replacement of direct access to Big Data by addressing a deep neural network (DNN) was proposed to construct the denotative cognitive model's semantics. This replacement is that access to Big Data in an emergency cannot always be provided. The tool for permanent training of the DNN on retrospective examples for using in a concrete situation is worked out.

The expediency of using DNN was shown on the model with LSTM (Long Short-Term Memory), convolutional, and recurrent [39]. The LSTM model is well suited for classifying, processing, and predicting time series cases where time lags separate events with uncertain duration. The LSTM module can store values for both short and long periods of time. It does not use the activation function inside its recurrent components. The stages of creating a DNN are:

- Downloading the clusters of relevant documents from available data sets.
- Tokenization, lemmatization, and filtering of documents' texts.
- Constructing the LSTM model for training.
- Training the LSTM model.

The process has to consider qualitative aspects of the situation, ensuring the denotative and cognitive semantic interpretation of models. Participants conduct cognitive modeling in the following order:

- Some of the first factors of the emergency are revealed for creating the cognitive model.
- Some of the first mutual influences between factors are evaluated.
- The project of the cognitive model is created automatically with using Big Data analysis or DNN.
- The result of automated cognitive model creation is estimated by a group's leader.
- Goals and control factors in the model are identified.

- The project of the cognitive model is corrected by a group's leader.
- The direct and inverse problems on the cognitive model are solving.
- The decision is suggested.

Figure 7 illustrates the preliminary creation and verification of cognitive models and synthesizes ones during decision-making.

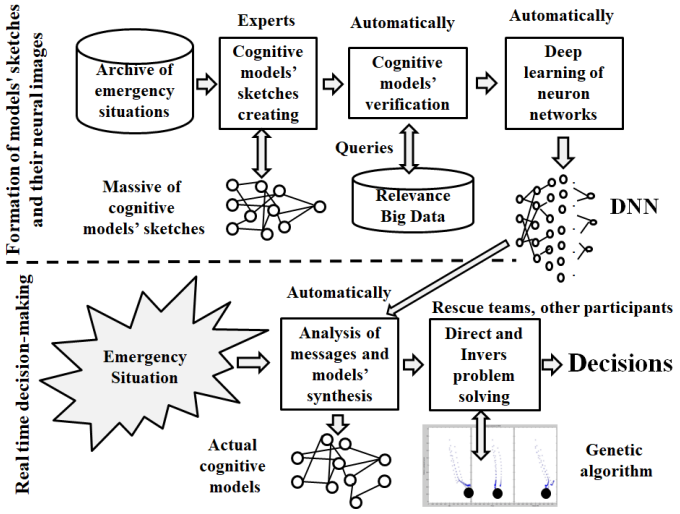


Figure 7: Cognitive model creation and verification

The testing was made on four different sample sizes of documents: 10, 100, 1000, and 4868. As the experiment showed, on the largest of these samples, the accuracy of determining the cognitive model factors reached 93%. This result is so good that it allows raising the question to automate the cognitive model's construction in an emergency.

This group decision-making process in an emergency with the creation of a cognitive model may take only some minutes, and thereby it improves the process of decision-making and, as a result, increases the chances of saving human lives.

9. Convergent Emergency Meeting in Practice

The suggested convergent approach with cognitive modeling was exploited in the distributed SC for speeding up strategic conversations and emergency meetings. The convergent approach helps to reduce the duration of such meetings resolutely. The collective strategic meeting duration was reduced from four days to one day or even four hours.

But convergent distributed meetings in an emergency with the quick agreement of the participants on the emergency place and heads of departments of the authorities and experts still remain the subject for further implementation in practice.

As a result of applying the convergent approach, many drafts of the strategies for the socio-economic development of the cities, regions, and industries were created. During the strategic meeting (conversation), the group of 30–40 participants carries out brainstorming sessions, conducts a strategic analysis, creates a cognitive model, solves a direct and an inverse problem, and works out an optimal strategic solution. The main participants of the conversation and strategic planning are the administration of the controlled object, enterprise management, and remote experts.

Everybody from administration and rescue groups has to know the convergent approach's rules to reduce the duration of the emergency meeting by conducting and structuring the information in a convergent way. Currently, the duration of a strategic conversation takes around 4–6 hours. During this time, every participant makes many decisions regarding the goals of actions. It may coincide or not with the global decision of his group. The behaviors of various participants are different; changing the situation brings chaos to participants' minds.

For example, in [40], it is shown that the strategic meeting for creating the cognitive model of the large megapolis tourism development took 4 hours due to using the convergent methodology. For this, 35 brainstorming were run simultaneously. From the beginning, the goal tree was made. Then the SWOT-analysis (strengths, weaknesses, opportunities, and threats) helps to identify about 70 factors of the situation, which can influence the megapolis tourism development. After that, these factors were convoluted into the substantial 15 factors, and the weights of their interconnections were evaluated, and as a result, the cognitive model was made. Then the Big Data analysis helped to verify this model. Then the direct and inverse problems were solved on the cognitive model and an optimal scenario found.

Such a practice can be used in emergencies. In this case, the process of networked decision-making will cover casualties, groups of people, different rescue teams, administration's representatives, headquarters, experts, and groups' leaders. The problem is to find an agreement among all participants very quickly about aims and ways of action.

The convergent approach helps to synchronize the participants' behaviors. But the above-mentioned 4 hours that were spent during the strategic conversation for getting a consensus is an excessively long period of time, which is unacceptable in an emergency, and the decision-making process has to be substantially accelerated.

To further develop a convergent approach for speeding up decision-making in an emergency, appropriate research must be carried out in a classic and non-classical way [28]. Some of these studies are still debatable.

10. Discussion

The most complex and debatable issue that is the subject for further research is the construction of AI logical models' cognitive semantics. This construction is non-formalizable and cannot be described with the programmer's logic. Currently, studies are being conducted in the field of building quantum semantics and optical computing.

The decision-making process with cognitive modeling can be accelerated by using a quantum semantic approach. Quantum semantics make the models' interpretations more holistic. They help consider the phenomenon that every word of the texts of documents interferes with an infinite number of the hidden (shadow) words. It makes the possibility to create as much as possible complex virtual objects. On the other side, attempts to represent the cognitive semantics of words with only a discrete

(digital) computer and logic give zero intersection ones with shadow words [38].

Creating cognitive semantics cannot be exactly reproduced by mathematical logic and discrete computers. They reduce the possibility of the cognitive semantic interpretation of the cognitive model and cannot consider the non-formalized processes of human thinking, emotions, thoughts, etc. For these aims, it will be useful to consider the opportunity of quantum entanglement (non-locality) effects [30]. The Maxwell and Schrödinger equations were theoretically researched for compensating the cognitive semantics gaps. In this context, the potential of optical calculations using lasers, deflectors, and multidimensional rewritable holographic storage devices was theoretically investigated. It could be useful for creating cognitive semantics. [38].

Perhaps the unified mathematical approach may be the group theory, which can cover the convergence of collective decision-making processes, the formation of cognitive semantics using the quantum field theory, and relativistic theory. Apparently, this may require exploiting the transformations' operators in the Lorentz group, where fixed points represent the events, and the intervals between events (points) are preserved, or it may be useful to apply the Poincare group, where the fixation of points is not required.

Such an approach will place the symbolic AI model with its denotative (formalizable) semantics and, most importantly, cognitive (unformalizable) semantics in some single space, and then use unified operators to implement transformations of events of various nature. In this case, events can be attributed to different phenomena, described as by logic, as by quantum electrodynamics' laws and the special theory of relativity. If cognitive semantics allows a quantum interpretation with its representation in the infinite Hilbert space and a relativistic interpretation, with its representation in terms of the transformation of smooth manifolds (for example, to study fields' anomalies), then the application of group theory can be fruitful.

At the same time, the adequacy of the approach to the interpretation of cognitive semantics using group theory is a subject of lengthy verification, especially regarding its application to the possibility of joint research, primarily, of events that interpret these semantics at the subatomic level and the theory of relativity level.

Thus, the important discussion issue to substantially speed up decision-making during the rescue operation in transport emergency based on AI is constructing cognitive semantics of AI models, involving the mechanisms of subatomic and relativistic levels. Some ideas in this way were suggested in this paper.

11. Conclusions

To accelerate the processes of decision-making in transport emergency (air, rail, car, etc.), it is advisable to structure all generated information in a special way: making 3-level arranging the goals, dividing rescue means on finite numbers of components, automatic creating cognitive models with neural network deep learning, and applying an inverse problem-solving method by genetic algorithm on the cognitive model for finding the best decision. For this, the author's convergent approach was proposed that considers the non-formalized cognitive semantics of AI logical models and ensures the sustainability and purposefulness of decision-making processes.

During cognitive modeling and decision-making, different impacts on the same factors can give successful or destructive results. E.g., the order and strength of influence on the factors are critical for getting the desired results of modeling. The successful impact on the situation can be determined comparatively accurately with a genetic algorithm applied to the cognitive model.

The traditional logical approaches and deep learning methods in AI cannot cover cognitive semantics holistically. This semantics requires to be enriched by new approaches, such as quantum and relativistic theories, networked experts procedures, which have to be convergently organized, and different means like situational awareness tools and situational centers systems.

Thirty-five years of practical experience in applying the convergent approach has shown the possibility of drastically reducing collective situational analysis duration and building optimistic strategies. However, the application of a convergent approach for transport emergency requires its further development.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgment

Supported by the Russian Foundation for Basic Research, grant No 18-29-03086 "Methods for identifying the needs of economic sectors in digital platforms and end-to-end technologies based on the analysis of large amounts of texts."

References

[1] M.R. Endsley, "Situation awareness in aviation systems," In: Garland, D.J., Wise, J.A., Hopkin, V.D., eds., Handbook of aviation human factors. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 257–276, 1999.

[2] P.M. Salmon, G.H. Walker, N.A. Stanton, "Pilot error versus sociotechnical systems failure: a distributed situation awareness analysis of Air France 447," Theor Issues Ergon Sci, **17**(1):64–79, 2016, doi: 10.1080/1463922X.2015.1106618.

[3] P. Illankoon, P. Tretten, U. Kumar, "Modelling human cognition of abnormal machine behavior," Hum-Intell Syst Integr, 1:3–26, 2019, https://doi.org/10.1007/s42454-019-00002-x.

[4] A.N. Raikov, "Convergent cognitype for speeding-up the strategic conversation," IFAC Proc, **41**(2):8103–8108, 2008, doi: 10.3182/20080706-5-KR-1001.01368.

[5] D. Gubanov, N. Korgin, D. Novikov, A. Raikov, E-expertise: Modern collective intelligence, Springer, 2014, https://doi.org/10.1007/978-3-319-06770-4.

[6] A. Raikov, "Health care trust space based on collective artificial intelligence and blockchain technologies," J eHealth Technol Appl, **16**(1):37–42, 2018.

[7] A. Raikov, "Accelerating technology for self-organising networked democracy," Futures, **103**:17–26, 2018, https://doi.org/10.1016/j.futures.2018.03.015.

[8] A. Raikov, "Convergent networked decision-making using group insights," Complex & Intell Syst, **1**(1): 57–68, December 2015, https://doi.org/10.1007/s40747-016-0005-9.

[9] A.N. Raikov, "Holistic discourse in the network cognitive modeling," J. Math Syst Sci, **3**(10):519–530, 2013.

[10] B.B. Frey, J.H. Lohmeier, S.W. Lee, N. Tollefson, "Measuring collaboration among grant partners," Amer J Eval, **27**(3):383–392, 2006, https://doi.org/10.1177/1098214006290356.

[11] G. Griffin, A. Rajabifard, and D. Williams, "The Victorian emergency management continuum and the benefits of spatial enablement," In: Spatially enabling government, industry and citizens (Global Geospatial Conference 2012), GSDI, 2012.

[12] M.R. Endsley, "Direct measurement of situation awareness: Validity and use of SAGAT," In: Endsley, M.R., Garland, D.J., eds., Situation awareness

analysis and measurement, Lawrence Elbaum Associates, Mahwah, NJ, 131–157, 2000. https://doi.org/10.4324/9781315087924-9.
[13] J.A. Crowder, J.N. Carbone, "Collaborative shared awareness: Human-AI collaboration," In: International conference on information and knowledge engineering, WorldComp, Athens, 1–6, 2014.
[14] D. Kahneman, "A psychological point of view: Violations of rational rules as diagnostic of mental processes," Behav Brain Sci **23**:681–683, 2000, https://doi.org/10.1017/S0140525X00403432.
[15] A.A. Hadi, A.W. Mohamed, and K.M. Jambi, "LSHADE-SPA memetic framework for solving large-scale optimization problems," Complex Intell Syst. **5**, 25–40, 2019. https://doi.org/10.1007/s40747-018-0086-8
[16] D. Perkins, The Eureka effect. The art and logic of breakthrough thinking, W.W. Norton & Company, New York, London, 2000.
[17] K.R. Hammond, Principles of organization in intuitive and analytical cognition (report 231). Center for Research on Judgement and Policy. University of Colorado, Boulder, CO, 1981.
[18] M.R. Endsley, "Situation awareness: Progress and directions," In: Banbury, S., Tremblay, S., eds., A cognitive approach to situation awareness: Theory, measurement and application, Ashgate, Aldershot, 317–341, 2004.
[19] D.D. Heikoop, J.C.F. de Winter, B. van Arem, N.A. Stanton, "Psychological constructs in driving automation: A consensus model and critical comment on construct proliferation," Theor Issues Ergon Sci **17**(3):284–303, 2004. https://doi.org/10.1080/1463922X.2015.1101507.
[20] H. Ibbara, R. Barbulescu, "Identity as narrative: Prevalence, effectiveness, and consequences of narrative identity work in macro work role transitions," Acad Manag Rev, **35**(1):135–154, 2010. https://doi.org/10.5465/amr.35.1.zok135.
[21] C. Bolstad, M.R. Endsley, A.M. Costello, C.D. Howell, "Evaluation of computer-based situation awareness training for general aviation pilots," Int J Aviat Psychol, **20**:269–294, 2010.
[22] N.A. Stanton, P. Salmon, G. Walker, D. Jenkins, "Is situation awareness all in the mind?," Theor Issues Ergon Sci **11**(1–2):29–40, 2010. https://doi.org/10.1080/14639220903009938.
[23] I. Kotseruba, J.K. Tsotsos, "40years of Cognitive Architectures: Core Cognitive Abilities and Practical Applications," Artif Intell Rev, **53**(1):7–94, 2020. https://doi.org/10.1007/s10462-018-9646-y.
[24] R.M. Pirsig, Zen and the art of motorcycle maintenance. William Morrow and Company, New York, 1974.
[25] R. Fjelland, "Why general artificial intelligence will not be realized," Humanit Soc Sci Commun **7**(10), 2020. https://doi.org/10.1057/s41599-020-0494-4.
[26] N. Ilyin, G. Malinetsky, K. Kolin, A. Zatsarinny, et al., "Distributed situational centres system of cutting edge development," In: 10th international conference management of large-scale system development (MLSD), 2017. https://doi.org/10.1109/MLSD.2017.8109638.
[27] T.-J. Huang, "Imitating the brain with neurocomputer a 'new' way towards artificial general intelligence," Int J Autom Comput, **14**(5):520-531, 2017. https://doi.org/10.1007/s11633-017-1082-y.
[28] A. Raikov, "Post-Non-Classical Artificial Intelligence and its Pioneer Practical Application." IFAC-PapersOnLine. **52**(25):343-348, 2019. https://doi.org/10.1016/j.ifacol.2019.12.547.
[29] B.M. Lake, T.D. Ullman, J.B. Tenenbaum, and S.J. Gershman, "Building machines that learn and think like people," Behavioral and Brain Sciences, Published online by Cambridge University Press, 40:1-70 e253, 2017. https://doi.org/10.1017/S0140525X16001837.
[30] H. Atmanspacher, "Quantum approaches to brain and mind: An overview with representative examples," in: Schneider, S., Velmans, M., eds., The Blackwell companion to consciousness, John Wiley, Chichester, 298–313, 2017. https://doi.org/10.1002/9781119132363.ch21.
[31] A. Dalela, Quantum meaning: A semantic interpretation of quantum theory. Shabda Press, 2012.
[32] C.K. Wong, "Covering properties of fuzzy topological spaces," J. Math Anal Appl, **43**:697–704, 1973. https://doi.org/10.1016/0022-247X(73)90285-0.
[33] A.N. Raikov, A.N. Ermakov, and A.A. Merkulov, "Assessments of the economic sectors needs in digital technologies," Lobachevskii J Math, **40**(11):1837–1847, 2019. https://doi.org/10.1134/S1995080219110246
[34] V.K. Ivanov, "Incorrect problems in topological spaces," Siberian Math J, **10**(5):785–791, 1969. https://doi.org/10.1007/BF00971654.
[35] A.N. Raikov, S.A. Panfilov, "Convergent decision support system with genetic algorithms and cognitive simulation," IFAC Proc, **46**(9):1108–1113, 2013. https://doi.org/10.3182/20130619-3-RU-3018.00404.
[36] T. L. Saaty, The analytic hierarchy process, University of Pittsburg, 1988.
[37] A.N. Raikov, Z. Avdeeva, and A. Ermakov, "Big data refining on the base of cognitive modeling," IFAC PapersOnLine, **49**(32):147–152, 2016. https://doi.org/10.1016/j.ifacol.2016.12.205.
[38] A. Raikov, "Cognitive modelling quality rising by applying quantum and optical semantic approaches," IFAC PapersOnLine, **51**(30):492–497, 2018. https://doi.org/10.1016/j.ifacol.2018.11.309.
[39] J. Goodfellow, Y. Bengio, A. Courville, Deep learning, The MIT Press, Cambrige, Massachusets, London, England, 2017.
[40] A. Raikov, "Megapolis Tourism Development Strategic Planning with Cognitive Modelling Support," In: Yang XS., Sherratt S., Dey N., Joshi A. (eds), Fourth International Congress on Information and Communication Technology (London). Advances in Intelligent Systems and Computing, vol. 1041. Springer, Singapore, 2020. https://doi.org/10.1007/978-981-15-0637-6_12.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 531-537 (2020)

www.astesj.com

Second International Virtual Conference on Multidisciplinary Research 2020

ASTESJ

ISSN: 2415-6698

Determinism of Replicated Distributed Systems–A Timing Analysis of the Data Passing Process

Adriano A. Santos[*,1], António Ferreira da Silva[1], António P. Magalhães[2], Mário de Sousa[3]

[1]*Centre for Research & Development in Mechanical Engineering (CIDEM), Mechanical Department, School of Engineering (ISEP), Polytechnic of Porto, 4249-015 Porto, Portugal*

[2]*Mechanical Engineering Department, Faculty of Engineering, University of Porto (FEUP), 4200-465 Porto, Portugal*

[3]*Electrical and Computer Engineering Department, Faculty of Engineering, University of Porto (FEUP), 4200-465 Porto, Portugal*

ARTICLE INFO

Article history:
Received: 10 August, 2020
Accepted: 07 October, 2020
Online: 20 November, 2020

Keywords:
Distributed Systems
Event-Base Control
Fault-Tolerance (FT)
IEC 61499
Industrial Control
Real-Time (RT)
Replication
Commercial Off-the-Shelf (COTS)

ABSTRACT

Fault-tolerant applications are created by replicating the software or hardware component in a distributed system. Communications are normally carried out over an Ethernet network to interact with the distributed/replicated system, ensuring atomic multicast properties. However, there are situations in which it is not possible to guarantee that the replicas process the same data set in the same order. This occurrence will lead to inconsistency in the data set produced by the replicas, that is, the determinism of the applications is not guaranteed.

To avoid these inconsistencies, a set of Function Blocks has been proposed which, taking advantage of the inherent properties of Ethernet, can guarantee the synchronism and determinism of the real-time application. This paper presents this set of Function Blocks, focusing our action on the development of reliable distributed systems in real-time. This demonstrates that the developed Function Blocks can guarantee the determinism of the replicas and, as such, that the messages sent are processed, in the same order and according to the time in which they were made available.

1. Introduction

Fault-tolerance or replication is implemented, in most cases, by replication of one or more critical components, by replication of the hardware, the software or both. Therefore, regardless of the approach used, the main objective is to ensure that if one of the replicas fails, the remaining replicas will continue to function and therefore mask the existence of the failed replica before the remaining application, making it as transparent as possible. So, we must consider a fault-tolerant distributed system as an interconnection of several unitary components that, in each call of an event, process data, generating new events and/or data. On the other hand, event and data, generated on a different replica, located to different nodes, must be synchronized to assure that the replicas receive the same set of data in the same order. All replicas must have the same perception of the data and this perception will be obtained through a multicast protocol.

To support the Distributed Computer Controller System (DCCS) in real-time and reliably replicated over Commercial Off-The-Shelf components (COTS), it is essential to provide a simple and transparent programming model. So, programmers should be unaware of implementation problems and the details of distribution and replication. However, it is important that the replication mechanism allows us to develop a generic and transparent approach without concerns the inherent requirements of the distributed system and the replication issues. Therefore, to overcome the problems inherent to the development of distributed/replicated systems, we opted for the use of a framework that guarantees not only the required abstraction capacity but also the ability to carry out the distribution and replication of real-time systems as well as, ensure determinism. Thus, taking as a starting point the new standard IEC 61499 [1], we opted for the use of an application that allows perform the development of distributed systems and, consequently, their replication, which, in parallel, supports the requirements of this same standard, i.e., based on the Open Source PLC Framework for Industrial Automation &

[*]Corresponding Author: Adriano A. Santos, Centre for Research & Development in Mechanical Engineering (CIDEM), Department of Mechanical, School of Engineering (ISEP), Polytechnic of Porto, ads@isep.ipt

Control, *Eclipse 4diac™* [2] (infrastructure for distributed Industrial Process Measurement and Control Systems – IPMCS).

This paper is organized as follows: Section 2 presents a literature review, where ideas from different authors for the distributed and replicated systems problems will be exposed. Section 3 presents an overview of IEC 61499 standard. Section 4 presents the proposed implementation for reliable real-time SIFB communication and their decision timing analysis. A numerical example is presented in Section 5 and the conclusions are outlined in Section 6.

2. Related work

Replicated systems is based in the replication of the critical components of software or hardware, connected by a network. So, to use a networks communication to support DCCS applications requires not only a bounded times transmission services, but also ensuring the dependability for the applications with real-time needs.

Fault-tolerance architectures based on software have been proposed by many authors, all of them exploring the diversity of implementation, diversity of data and temporal diversity [3]. Consequently, two approaches can be taken to tolerated or recovery faults: the forward recovery (N-Version Programming [4] and its variations, N-Self Checking Programming [5], N-Copy Programming such as Distributed Recovery Block (DRB) [6] or Extend DRB [7]), and backward recovery (use of checkpoints from which recovery is attempted). So, when using in a replicated system forward recovery it is necessary that all replicas remain synchronized in order to produce the same set of data and events outputs in the same order (replicas will have to be deterministic) [8] and, somewhere, the replicas outputs need to be consolidated. On the other hand, replica determinism it will be possible achieved through the use of clocks synchronization, atomic multicast protocol and consensus agreement protocols [9] but also by timed messages [10]. Base in these concepts, a similar technic for a DEAR-COTS framework based on Ada 95 has already been proposed by Pinho *et al*. [11]. These authors using a Network Time Protocol (NTP) to synchronize replicas and messages transmission time are set offline.

Likewise, considering replication systems, [12] presents an approach based on a Fast and scalable Byzantine Fault-Tolerance protocol (FBFT) using message aggregation technique combined with reliable hardware-based execution environments. Based on a multicasting replication the aggregation reduces the complexity of messages and computation overhead. In this turn, Pinho *et al.* [13] using a Stat Machine approach to develop fault-tolerant distributed system. Replication is based on a priority algorithm (total order and consensus) like Raft (PRaft). Incoming messages are executed according to the priority level, so the processes do not have to wait to the confirmation of the request. Messages are executed at the moment they are received. Hu *et al.* [14] propose a standard for fault-tolerance modulation based on the programming of N-Versions that can be integrated, transparently, into existing applications, improving its operation, maintaining the characteristics of time. The model, developed in C language, consists of a set of components (initiator, member versions and the voter) where they encapsulated several alternative algorithms to obtain the same outputs.

Some works based on the IEC 61499 fault tolerance system have already been tested and presented in [15], where a distributed replication structure, similar to the work presented in this document, is presented. In this case, a timed message protocol is used to ensure synchronization of the internal states of the replicas and was also implemented in *Eclipse 4diac™* and validated in a FORTE runtime multicast environment. In the same line of reliability [16], it presented a formal modeling methodology to validate and evaluate the reliability of IEC 61499 applications applied to critical safety situations. On the other hand, the works developed by Batchkova *et al*. [17] and Dai *et al*. [18], which are related in some way, focus on the development of reconfigurable IEC 61499 control systems. The methodology proposed by these authors allows IEC 61499 applications to be reconfigured during execution, replacing one Function Block (FB) with another. The substitution, done in real time, does not create a significant impact on the execution of the system and as such can be used as an approach to fault-tolerance. However, they focused on reconfiguration of the system and not on a fault-tolerance scenario, where time to bring up the system is not considered.

In the same area of the IEC 61499 Lednicki *et al.* [19] presents a model to calculate the Wast-Case Execution Time (WCET) for the software FB execution. This model works with a set of events, considering the information associated with the inputs, in which the execution is initiated by the arrival of the input event to the function blocks. The WCET value represents the maximum time that a FB needs to execute its functionality, from the entry of an event until its internal activity is completed (exit of the event). The WCET is independent of the internal path or the execution paths. This model gives us the time needed to activate the next FB.

3. Overview of the IEC 61499

IEC 61499 applications are made up of interconnections of FBs that exchange information, data and events, based on a graphical representation. Each of these graphic representations (FB) consists of a rectangular structure that incorporates an upper part, called the head, and a lower part, called the body. The head is the interface for receiving events, initializing the FB and activating the internal algorithms. It is also used for sending events, confirming the initialization of the FB, as well as the execution of the internal algorithm or algorithms. The body is the interface for received and sending data and it is also de base for the internal algorithms. Data and events inputs are on the left and outputs on the right.

An FB network is assumed to be an event exchange process in which each outgoing event is linked to an incoming event, and each outgoing data is linked to an incoming data. FBs are executed when they receive events so, from that moment on data can be processed, reading inputs data. However, it is only after the execution of the internal algorithms that the data is updated, placed on the output link, and one or more output events can be generated. Input and output data are defined according to the type (Int, Real, etc.) so, in IEC 61499 applications, it will only be possible to establish connections between data of the same type, while the events can be considered as the base event type used in the FB activation.

As stated earlier, the central structure of the IEC 61499 is the FB, which interconnected in a network may represent a device, like a Personal Computer (PC) or a Programmable Logical Controller

(PLC), for example, connect in a control node. In this sense, we will be able to fit the FBs in the object-oriented paradigm where each of the FBs can be considered as a single object (Basic Function Block – BFB, with an Execution Control Chart – ECC, consisting of one or more algorithms), as a Composit Function Block (CFB), constituted by an interconnected BFBs or CFBs and Special Interface Function Block (SIFB) used in communication. The program algorithm for the BFB, CFB and SIFB may be developed in any languages defined by the IEC 31131-3 [20] and also in all additional languages supported by the IEC 61499 (e.g., Java, C, C++, C#, etc.). Developers are free to choose language since the standard does not specify a recommended language.

A distributed IEC 61499 or a replicated fault-tolerance system, split among several computer devices (e.g., PC, microcomputer or PLC), needs to send, over the network, events and data to each one of the FB distributed or replicated into the devices. To do this, the developer, insert a SIFB communication which will allow communication between FBs allocated to the same device or distributed by remote devices. The purpose of this article is to define a methodology to guarantee replicas determinism, using standard communication interfaces (SIFB), which will communicate between replicas, on remote devices, ensuring that the replicas process the same data set in the same order. On the other hand, industrial redundancy is typically archived at the hardware level, where the access to physical I/Os is done by communication SIFB Publish/Subscribe pair over UDP/IP – User Datagram Protocol/Internet Protocol (unidirectional communication), or Client/Server pair over TCP/IP (Transmission Control Protocol/Internet Protocol), bidirectional data/event communication [21].

4. Proposed IEC 61499 Implementation

An IEC 61499 application is seen as a combination by several devices, sub-applications or interconnected unitary processing elements (FBs), which at each invocation of events, process the data, generating new events and/or data. So, to tolerate individual faults, ensure the reliability of the application, only the critical components of the application must be replicated. Components is defined as an atomic and indivisible component (FB) which can include tasks and resources replicated on multiple nodes or allocated in just on node. As an example, Figure 1 shows a real-time sub-application "C" with 2 FBs (FB1 e FB2) distributed over nodes 1 and 2 and replicated over nodes 2 and 3 or, alternatively, replicated entirely in a single node (node 4).

So, the communication infrastructure of the proposed replication framework (base on active replication, i.e., all replicas are active and running at all times), based on the same communication structure used by IEC 61499, must guarantee that all messages sent by computer devices, delivery to all receiver, is correctly received. However, it will also be necessary that replicas agree with the order of the data set sent and consolidate data from replicated inputs into a single value that will be propagated to the subsequent FB.

Component replication can be performed using multiple replicas, however, the most common is the use of two ($f + 1$) or three replicas ($2 * f + 1$), to tolerate f failures. Therefore, replication based on these assumptions places us in the presence of several communication/interaction scenarios in which the exchange of messages can be defined according to the following four approaches [22]: *1-to-1* (communication from a nonreplicated FB to another nonreplicated or communication inside of the replica, base of the IEC 61499 communication); *1-to-many* (communication from a nonreplicated FB to a group of replicated FB. An atomic multicast protocol [9] must be used to ensure that all the replicas received the same set of information (data/events) in the same order. Replicas need to maintain internal state synchronized); *many-to-1* (a replicated FB sends data/events to a nonreplicated FB. Nonreplicated FB receives a set of inputs from all replicas and vote, consolidate mechanism [3], on the output value to process continue) and *many-to-many* (a mix of the last two cases where each received replica need to agree on the value to forward process. An atomic multicast protocol is used to disseminate values and the agree decision can be performed by one of the received values or on some value calculate based on majority, average, median, etc.). Voting mechanism can itself be replicated or only a single copy can be executed.

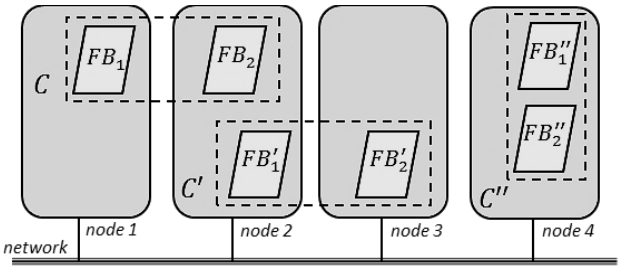


Figure 1: Replicated real-time sub-application

Replication model was implemented using *Eclipse 4diac™*. It is developed using the 4DIAC-IDE (Integrated Development Environment), graphical platform, and the FORTE runtime execution environment [2]. The graphical platform is used to develop the application, perform the interconnections between the instances, create, compile (in C ++) and integrate (in FORTE) the new types of FBs, once FORTE is compiled and executed in each computer device. Communication between replicas was carried out using standard communication FBs, made available by the 4diac repository and in accordance with IEC 61499. Data and events connection, between computer device and all instantiated FBs is supported by the FORTE runtime environment.

4.1. Communication architecture

The communication architecture is based on standards SIFB (Publish/Subscribe or Client/Server) interactions using Internet Protocol (IP) and FORTE runtime [23]. So, each of communication layers needs to be configured by the addressing schema accomplished by the identifier parameter (ID). Communication protocol is implemented in FBDK, inside a multicast group, using an Internet address and a unique port number [IP:port, e.g., 239.0.0.100:61023]. In a multicast communication scenarios (SIFB Publish/Subscribe) any of the subscribe blocks, located in the same network segment as the publish, can receive all published data/events [24]. On the other hand, a real-time industrial control application requires clocks synchronization, use of the timed messages to ensure determinism and the analyses of the worst-case execution times of the replicated FBs (including clock lag and communication delays). In fact, the developer needs to create a structure that supports IEC 61499

replication based on the existent communication SIFB considering scenarios presented above, in other words, it will only be necessary to use only two of the presented layers: *1-to-many* and *many-to-1* communication scenarios. Figure 2 shows the interface of the pair Publish/Subscribe used in multicast protocol communications.

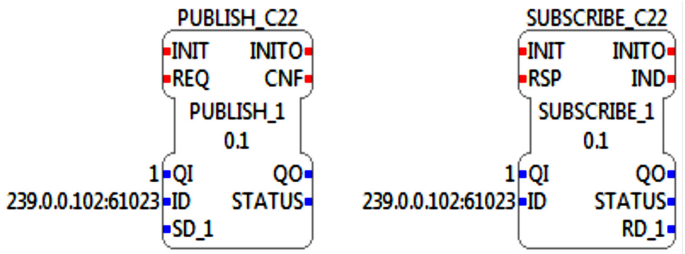


Figure 2: Interface of the Publish/Subscribe communication pair

4.2. Consolidation and voting replicate inputs

Voters can be developed according to the most varied techniques of fault-tolerance. However, these have the ultimate purpose of comparing the results of two or more variables and deciding which is the correct result, if any. There are in fact many types of voters [3] and the decision on which voting algorithm to use will depend on the semantics required by the application. For the voting to be viable, all replicas must send the data in the expected time. So, voting mechanism or consolidate module is built in top of the atomic multicast protocol, to ensure that all replicated FB receives the same set of data in the same order. The subscribe will wait until the set of events and data from the publishes are received. It is only at this point that the data is consolidated (the data is chose) or when it know that you will no longer receive messages at the δ_{voting} specific delay, Figure 3.

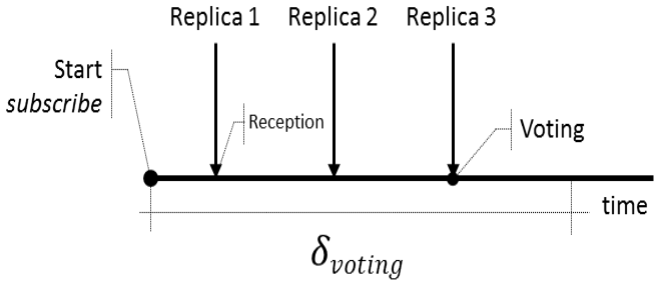


Figure 3: Data consolidate without faults

Note that this procedure must be implemented in a *many-to-1* context and that the decision time will depend on the worst-case response time (WCRT) of the last data received. On the other hand, we should consider that it will not be necessary to use underlying protocols that solve the problems of inconsistency by omitting messages, as it will be enough that only one node delivers a message.

In this approach, as we use a Triple Modular Redundancy (TMR) to mask faults [25] a majority voter was used. In the case of the three replicates, is done by simply comparison of the values received in A and C. The voting mechanism determines which value should be chosen according to the pseudocode shown in Algorithm 1.

Algorithm 1 VOTER algorithm pseudo code

```
1:  /* Initialization */
2:  void FORTE_VOTER
3:  switch(pa_nEIID){
4:  case input event:
5:          if (A() == C())
6:              VOTED() = A();
7:          else
8:              VOTED() = B();
9:          output event;
10:     break;
```

On the other hand, it is also necessary to consider that the determinism of the replicas must be guaranteed by the active replication. Therefore, the concept of timed messages [10] must be implemented to define the correct order of the execution of the received data. Thus, according to the mapping of the replication scenarios presented in [22], the application clocks must be synchronized. Data to be disseminated are associated with the availability times, defined by the execution times of the FBs to which the events/data are linked, that is, immediately after their execution. This validation time will be, in reality, the worst execution time of the FB that makes the data available (since in the IEC 61499 framework, the output data are only available when the algorithm finishes its execution) plus the sending times, which can be determined *offline* [26]. In this sense, when the data manager, a software element that guarantees the achievement of determinism, reads the received values (sent only once), works with the most recent values that have the oldest validation moments associated with the task validation. Figure 4 shows the scheme of how determinism can be obtained in the replicated components depending on the treatment of the timed messages received and treated according to the concept developed in the ordering FB.

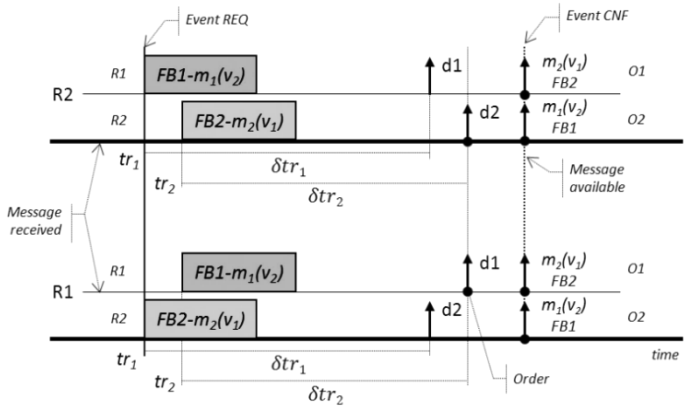


Figure 4: Replicas consistence, adapted to IEC 61499

Each FB is associated with a task, so they will send a message of order m_k where, in this implementation, the k index is associated with the number of the respective FB or task. δt_r is the limit of the predefined time for its execution (time that is activated after receiving the first message). $m_k(v_i)$ is a message of order k made available at the time of validity v_i. The instant tr_i represents the times of recession in processes *P1* and *P2* so, δtr_i is the waiting time, associated with event i received. When a new event is received the δtr_i waiting time is restarted, associated with the new event i, at the end of which the events will be ordered according to

the instant of validation v_i. Message that has the most recent value and that has the oldest validation time associated with its validation will be allocated to O1 (OUT1) while the second message will be allocated to output O2 (OUT2). CNF event confirms the execution of the FB and the availability, at the same time, of data *d1* and *d2*.

4.3. Consolidate time analysis

The analysis of the consolidate protocol execution time, at the receiving replicas, aims to define the delay time in the decision phase ($\delta_{decision}$), necessary for the FB to consolidate the received data, i.e., guarantee that FB will not receive any more messages. This time is dependent on the worst-case response time of the replicated messages as well as their best-case response time (BCRT), message processing, having as reference the initial time common to all sending and receivers nodes. However, we must consider that the time to send messages is common to all nodes (synchronization of local clocks) there were small variations or errors in the clock readings (*jitter*) that, like the *offset*, made the clocks only approximately synchronized.

Knowing the worst-case response time, we can determine the $\delta_{decision}$ assuming that the first message received has the best-case transmission time and the last has the worst-case response time. Therefore:

$$\delta_{decision} = \max_{\forall m \in res(m)} \{W_m\} - \min_{\forall m \in res(m)} \{W_m\} + \varepsilon \qquad (1)$$

where $max\{W_m\}$ is the worst-case response time of message and $min\{W_m\}$ is the best-case response time of message, considering a common time reference. $res(m)$ is the set of replicated messages received and ε is the maximum deviation between nodes synchronized local clocks. Figure 5 shows the time relationships referred.

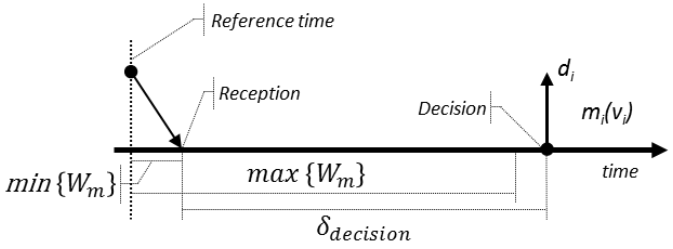


Figure 5: Relationship of times consolidate protocol

5. Numerical example

In order to explain the use of the presented model, a simple example of application is used. In Figure 6 is presented a real-time distributed system replication scheme using in the considered example. System application is constituted by five nodes, connected by TPC/IP network based on multicast protocol.

The application used in this example is constituted by five components (C_1 to C_4), each one with a task (τ_1 to τ_4) which are distributed over the nodes. A simple replication of critical components was performed by interconnecting the distributed and replicated components, with a switch, over an Ethernet TCP/IP network at a rate of 10/100 Mbps. Component C_1 (FB1) encapsulate tasks τ_1 at node 1, component C_2 (FB2) encapsulate task τ_2 at node 2, components C_{3a} and C_{3b} (replicated components, FB3 e FB3', at nodes 3 and 4) encapsulate τ_3 and τ_3', and finally component C_4 (FB4) encapsulated τ_4 (used to synchronize the start of components C_1 and C_2) at node 5.

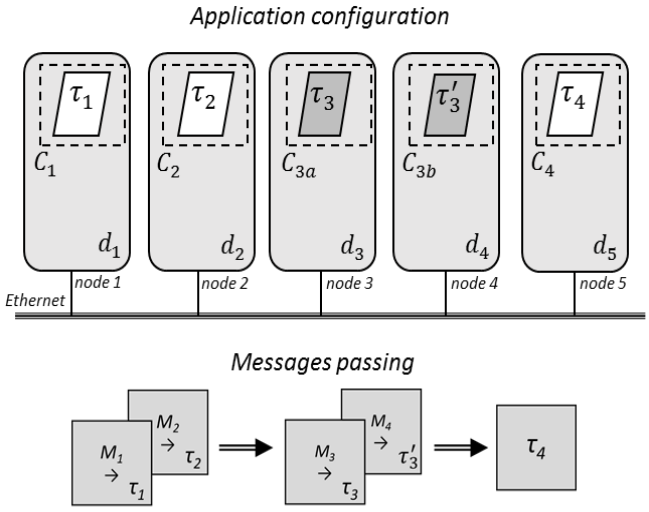


Figure 6: Application structure (replication and messages passing)

Table 1: Tasks characteristics

Tash	Type	Component	Node
τ_1	*Periodic*	*C1*	*1*
τ_2	*Periodic*	*C2*	*2*
τ_3	*Periodic*	*C3a*	*3*
τ_3'	*Periodic*	*C3b*	*4*
τ_4	*Spor/Per*	*C4*	*5*
τ_5	*Periodic*	*C5*	*6*

Table 1 presents each of the task's characteristics of the distributed application, while Table 2 presents the messages exchanged between FB (all values are in milliseconds).

Table 2: Messages passing characteristics

Msg	Bytes	Period (*ms*)	From	To	Protocol
M_5	---	---	τ_4	τ_1, τ_2	*Multicast*
	56(84)	*ping -f*	τ_4	τ_5	*Unicast*
M_1	*24*	*1000/500*	τ_1	τ_3, τ_3'	*Multicast*
M_2	*24*	*1000/500*	τ_2	τ_3, τ_3'	*Multicast*
M_6	*32*	*ping -t*	τ_5	τ_4	*Unicast*

Note that message from component C_1 and C_2 (M_1, M_2) is a *1-to-many* communication (multicast protocol) and also the message of system initialization (event synchronization starts of the C_1 and C_2) M_5. Messages M_5 is also used to increase the network traffic, in the switch, to component C_5 (Windows PC). Message M6 from C_5 to C_4 is the task response and also a contribute to the network traffic, it is a *1–to–1* communication. Messages M_1 and M_2 are messages from nonreplicated components to a replicate's components (C_{3a} and C_{3b}), therefore, they will have to be consolidate in each of the receiving replicas. This consolidation mask node failures of the sender's components.

5.1. Determinism testing

In order to clarify the possible determinism of the application, using only standard elements of the *4diac framework*, Publish/Subscribe communication pairs, 10000 events and data were launched in the network with frequency ranges of 1Hz and 0.5Hz, that is, with 1000 ms and 500 ms intervals. Each of the components (C_1 and C_2) sends messages to the replicas, in the same frequency ranges, associated with the availability time (*tsa - timestamp to availability* in *sec* and *nsec*) in the format [*data, sec, nsec*]. Replicas receives data associating the reception time (*tsr – timestamp to receiving*) in the same format.

The purpose of this example is to test the possibility of guaranteeing the determinism of the replicas according to the proposed and applied communication protocols (Publish/Subscribe pairs, inherent to *4diac*), for both the delivery and response time of messages. Thus, we consider the response time as the time interval between the instant when the message is sent and the instant when it is received by replies. A multicast protocol is used, to propagate messages, assuming that all the messages are delivered and received correctly. Table 3 shows the worst-case response time (WCRT) for a set of messages sent to the network using a 1000 ms trigger frequency. These experiments are carried out considering three conditions of operation: no additional traffic, with daily traffic directed simply to the switch (*tp-link, TL-SF1008D*) and additional traffic directed from a nonreplicated component to a replicated. Table 4 shows the results obtained for the WCRP considering a 500 ms trigger frequency.

Table 3: Messages response time, 1000 ms trigger frequency

Msg	Period	WCRT (*ms*)		Reception order
		C3a	C3b	
M_5	-	-	-	
M_1	1000	6,720	3,979	93,35%
M_2	1000	6,765	3,806	
M_1	1000	6,527	2,602	94,33%
M_2	1000	6,464	2,942	
M_1	1000	6,208	1,556	97,09%
M_2	1000	6,448	1,941	

As can be seen, the WCRP is obtained for M_2 messages. This carryout the messages sending from component C_2 (nonreplicated) to the replicated component C_{3a}, response time of 6.765 ms. The average order value of the received data (*Reception order*) is 94.92%, which demonstrates the absence of the determinism of the replicas (C_{3a} and C_{3b} components) due to the frequency of the experience triggering.

Table 4: Messages response time, 500 ms trigger frequency

Msg	Period	WCRT (*ms*)		Reception order
		C3a	C3b	
M_5	-	-	-	
M_1	500	6,760	6,236	97,94%
M_2	500	7,207	8,045	
M_1	500	5,843	2,457	93,43%
M_2	500	6,248	2,569	
M_1	500	7,196	3,178	94,34%
M_2	500	6,575	3,235	

As can be seen, the WCRP is obtained for M_2 messages. This carryout the messages sending from component C_2 (nonreplicated) to the replicated component C_{3b}, response time of 8.045 ms. The average order value of the received data (*Reception order*) is 95.24%, which demonstrates the absence of the determinism of the replicas (C_{3a} and C_{3b} components) due to the frequency of the experience triggering.

Table 5 shows the WCRT (maxims time) and BCRT (minimums times), *calculated offline*, for a set of messages exchanged between components C_1, C_2 and C_{3a}, C_{3b} characterizing the messages of stream M_1 and M_2. These values are the result of the data and events received in the replicas, considering the experiences defined above. The incidence based is 10 k events processed in each of the experiments. These values are the result of the *offline* analysis of the 60 k records processed.

Table 5: Messages times characteristics, maxims and minimums value

Msg	WCRT (*ms*)	BCRT (*ms*)	from→to	Average order
M_1	7,169	0,170	$\tau_1 \rightarrow \tau_3$	95,08%
M_2	8,045	0,131	$\tau_2 \rightarrow \tau_3'$	

As can be seen, the average order for data received in the replicas has a value less than 100%, which induces the existence of failures in the ordering of the data. On the other hand, since the messages must be consolidated and ordered according to the time validity requirement, to guarantee determinism, it is necessary to determine the $\delta_{decision}$ parameter of the consolidation protocol as defined in (1). The maximums and minimums times for transmission data/events are defined in Table 5 and the maximum deviation between synchronized clocks (ε), also calculated *offline*, is 164 μs. Therefore, using (1):

$$\delta_{decision} = 8.045 - 0.131 + 0.164 = 8.078\ ms \quad (2)$$

The worst-case decision time for consolidation, considering $\delta_{decision}$, can be defined considering all messages received in the set *res(m)* so, that the worst-case decision time will be given by:

$$W^{decision} = \max_{\forall m \in res(m)} \{W_m\} + \delta_{decision} = 16.123\ ms \quad (3)$$

where $W^{decision}$ is the worst time of the consolidation, i.e., time decision when a new event arrives at the end of the $\delta_{decision}$ time.

6. Conclusions

The *Eclipse 4diac*™ tool was developed in accordance with IEC 61499 and it is specially directed to the development of distributed industrial applications portables and modular.

The Publish/Subscribe pair, provides by the 4diac object repository, are a communication FB, fundamental for the interconnection of the distributed components, guaranteeing not only the synchronization of the system but also the implementation

of the atomic multicast protocol. However, the introduction of software or hardware component replication introduces new problems that were not anticipated in IEC 61499 as well as in 4diac. These pairs, using different communication interfaces, guarantee replication synchronization, but cannot guarantee their determinism and, as such, promote the occurrence of failures that must be masked.

Based on the experiences carried out we can conclude that the Publish/Subscribe pair, most used in the interconnection of distribute/replicated computer devices, has a failure rate, independent and identically distributed, of 4.92%. Therefore, it is not possible, by itself, to ensure determinism, so the programmer will need to develop a FB, also replicable, which ensures that all replicas process the same data set in the same order.

Acknowledgment

We acknowledge the financial support of CIDEM, R&D unit funded by the FCT – Portuguese Foundation for the Development of Science and Technology, Ministry of Science, Technology and Higher Education, under the Project UID/EMS/0615/2019 and ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme, and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project SAICTPAC/0034/2015- POCI-01-0145-FEDER-016418.

References

[1] International Standard IEC 61499-1, Function Block-Part 1: Architecture, 2nd Edition. International Electrotechnical Commission, 2012.

[2] Eclipse 4diac, Open Source PLC Framework for Industrial Automation & Control [Online], https://www.eclipse.org/4diac/.

[3] L. Laura, Software Fault Tolerance – Techniques and Implementation, Artech House Publishers, ISBN: 9781580531375, 2001.

[4] L. Chen, A. Avizienis, "N-Version Programming: A Fault-Tolerance Approach to Reliability of Software Operation" in Proceedings of FTCS-8, Toulouse, France, 1978. DOI**:** 10.1109/FTCSH.1995.532621.

[5] J. C. Laprie, J. Arlat, C. Beounes and K. Kanoun, "Definition and analysis of hardware- and software-fault-tolerant architectures," in Computer, **23**(7), 39-51, July 1990. DOI: 10.1109/2.56851.

[6] K. H. Kim. The Distributed Recovery Block Scheme, in Software Fault Tolerance, M. R. Lyu (Ed.), New York: John Wiley & Sons, 1995.

[7] D. Nguyen and Dar-Biau Liu, "Recovery blocks in real-time distributed systems," in Annual Reliability and Maintainability Symposium. 1998 Proceedings. International Symposium on Product Quality and Integrity, Anaheim, CA, USA, 1998, 149-154. DOI: 10.1109/RAMS.1998.653703.

[8] R. Guerraoui and A. Schiper, "Software-based replication for fault tolerance" in Computer, **30**(4), 68-74, April 1997. DOI: 10.1109/2.585156.

[9] V. Hadzilacos, S. Toueg, "Fault-Tolerant Broadcasts and Related Problems" in Distributed Systems (2nd Ed.), ACM Press/Addison-Wesley Publishing Co, USA, 97-145, 1993.

[10] S. Poledna, A. Burns, A. Wellings and P. Barrett, "Replica determinism and flexible scheduling in hard real-time dependable systems" in IEEE Transactions on Computers, **49**(2), 100-111, Feb. 2000. DOI: 10.1109/12.833107.

[11] L.M. Pinho, "Replication Management in Reliable Real-Time Systems", in Real-Time Systems, **26**, 261–296 (2004). https://doi.org/10.1023/B:TIME.0000018248.18519.46.

[12] J. Liu, W. Li, G. O. Karame and N. Asokan, "Scalable Byzantine Consensus via Hardware-Assisted Secret Sharing" in IEEE Trans. on Computers, **68**(1), 139-151, 1 Jan. 2019. DOI: 10.1109/TC.2018.2860009.

[13] R. Paulo, et al., "Replicação de Máquina de Estado Baseada em Prioridade com PRaft" in XXXIV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, Salvador, Brazil, 2016.

[14] H. Tingting, Bertolotti, Ivan C. and Navet, Nicolas, "Towards seamless integration of N-Version Programming in model-based design", in 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA2017), Cyprus, 1-8, 2017. DOI: 10.1109/ETFA.2017.8247678.

[15] M. de Sousa, C. Chrysoulas and A. E. Homay, "Multiply and conquer: A replication framework for building fault tolerant industrial applications" in 2015 IEEE 13th International Conference on Industrial Informatics (INDIN), Cambridge, 1342-1347, 2015. DOI: 10.1109/INDIN.2015.7281930.

[16] L. Yoong, "Modelling and Synthesis of Safety-Critical Software with IEC 61499", PhD Thesis submitted for Electrical and Electronic Engineering, University of Auckland, 2010. http://hdl.handle.net/2292/6691.

[17] I. Batchkova, G. Popov, H. Karamishev and Grigor Stambolov, "Dynamic reconfigurability of control systems using IEC 61499 standard", in 15th Workshop on International Stability, Technology, and Culture the International Federation of Automatic Control, **46**(8), 256-261, 2013. https://doi.org/10.3182/20130606-3-XK-4037.00050.

[18] W. Dai, V. Vyatkin, J. H. Christensen and V. N. Dubinin, "Bridging Service-Oriented Architecture and IEC 61499 for Flexibility and Interoperability," in IEEE Transactions on Industrial Informatics, **11**(3), 771-781, June 2015. Doi: 10.1109/TII.2015.2423495.

[19] Lednicki, Luka, Carlson, Jan e Sandstrom, Kristian, "Model level worst-case execution time analysis for IEC 61499", in Proc. of the 16th International ACM Sigsoft symposium on Component-based software engineering, June, 169–178, 2013. https://doi.org/10.1145/2465449.2465455.

[20] International Electrotechnical Commission, "International Standard IEC 61131-3, Programmable Controllers - Part 3: Programming Languages", 3rd Edition, 2013.

[21] H.-M Hanisch and V. Vyatkin, "Achieving Reconfigurability of Automation Systems by Using the New International Standard IEC 61499: A Developer's View", in The Industrial Information Technology Handbook, CRC PRESS, Section 4, 66, 2004. https://doi.org/10.1201/9781315220758.

[22] A. A. Santos and M. de Sousa, "Replication Strategies for Distributed IEC 61499 Applications", in *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, Washington, DC, 2018, 2225-2230. Doi: 10.1109/IECON.2018.8592737.

[23] M. Hofmann, M. Rooker and A. Zoitl, "Improved Communication Model for an IEC 61499 Runtime Environment", Proc. of IEEE 16th Conference on Emerging Technologies & Factory Automation (ETFA2011), Toulouse, France, 2011, 1-7. Doi: 10.1109/ETFA.2011.6059121.

[24] A. A. Santos, A. F. da Silva, M. de Sousa and P. Magalhães, "An IEC 61499 Replication for Distributed Control Applications", in IEEE 16th International Conference on Industrial Informatics (INDIN), Porto, 2018, 362-367. Doi: 10.1109/INDIN.2018.8471958.

[25] M. Yang, G. Hua, Y. Feng, and J. Gong, "Fault-Tolerance Architectures and Key Techniques", in Fault-Tolerance Techniques for Spacecraft Control Computers (1st. ed.). Wiley Publishing, pp 29-76, 2017. DOI: 10.1002/9781119107392.

[26] K. Mohamed, R. Xavier, S. Françoise, "A Schedulability Analysis of an IEC-61499 Control Application", in 6th IFAC International conference on Fieldbus Systems and their Applications (FeT'2005), **38**(2), 71-78, 2005. https://doi.org/10.3182/20051114-2-MX-3901.00011.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 538-543 (2020)
www.astesj.com
Second International Virtual Conference on Multidisciplinary Research 2020

ASTES Journal
ISSN: 2415-6698

Variation Between DDC and SCAMSMA for Clustering of Wireless Multipath Waves in Indoor and Semi-Urban Channel Scenarios

Jojo Blanza[*,1,2], Lawrence Materum[1,3]

[1]*Department of Electronics and Communications Engineering, De La Salle University, 2401 Taft Ave., Malate, Manila, 0922, Philippines*

[2]*Electronics Engineering Department, University of Santo Tomas, Espana Blvd., Sampaloc, Manila, 1008, Philippines*

[3]*Tokyo City University, 1-28-1 Tamazutsumi, Setagaya, Tokyo, 158-8557, Japan*

ARTICLE INFO

Article history:
Received: 08 September, 2020
Accepted: 07 October, 2020
Online: 20 November, 2020

Keywords:
Channel Models
Data Handling
Radiowave Propagation

ABSTRACT

The performance of Simultaneous Clustering and Model Selection Matrix Affinity (SCAMSMA) and Deep Divergence-Based Clustering (DDC) in clustering wireless multipaths generated by COST 2100 channel model (C2CM) is compared. Enhancing the accuracy of clustering multipaths is an open area of research which the clustering approaches try to improve. Jaccard index is used as the clustering validity metric of the clustering approaches. The results of the clustering approaches are compared using the analysis of variance (ANOVA) toolbox of MATLAB and displayed using the box plots. Results show that the cluster-wise Jaccard index is different between SCAMSMA and DDC for indoor scenarios, while the membership-wise Jaccard index is not. On the other hand, the cluster-wise Jaccard index is not different between the clustering approaches for semi-urban scenarios, while the membership-wise Jaccard index is a little different. The clustering approaches can be used in indoor scenarios based on accuracy.

1 Introduction

The European Cooperation in Science and Technology (COST) 2100 Channel Model (C2CM) [1]–[4] can reproduce the properties of multiple-input multiple-output (MIMO) wireless propagation channels. A multipath component (MPC) is classified based on the delay (τ), angle of departure (Azimuth of Departure (AoD), Elevation of Departure (EoD)), and angle of arrival (Azimuth of Arrival (AoA), Elevation of Arrival (EoA)). Groups of multipath components with similar delays and angles comprise a multipath cluster, which characterized C2CM.

Analyzing wireless multipaths is an important problem where clustering is crucial. The attribute, performance, and efficiency of the communications system can be studied, understood, and improved by the generated channel model. The accuracy and correctness of the channel models significantly affect the precision and exactness of clustering wireless propagation multipaths. Several channel models and measurements reveal the clustering of multipaths. Inaccurate clustering of the wireless propagation multipaths leads to incorrect channel models and thereby degradation in performance.

Clustering finds the underlying structure of the data. It also group similar data together [5]–[16]. In our previous works [17]–[20], Simultaneous Clustering and Model Selection Matrix Affinity (SCAMSMA) [21] and Divergence-Based Clustering (DDC) [22] were used to cluster the dataset [23, 24] generated by C2CM. In this work, the comparison of the clustering accuracy of SCAMSMA and DDC are presented. The main contributions of this study are (1) the paper shows the variation in the performance of SCAMSMA and DDC in clustering the COST 2100 dataset; and (2) SCAMSMA and DDC have a significant difference in their accuracy in clustering the wireless multipaths.

The paper is organized in the following way. Section 2 presents the dataset generated by C2CM. Section 3 describes the clustering approaches. Section 4 explains the ANOVA used. Section 5 defines the Jaccard index. Section 6 discusses the ANOVA results. Section 7 concludes the work.

2 COST 2100 Dataset

COST 2100 channel model (C2CM) [1]–[4] can replicate the stochastic properties of multiple-input multiple-output (MIMO) wireless propagation channels. Multipath clusters characterize

[*]Corresponding Author: Jojo Blanza, Department of Electronics and Communications Engineering, De La Salle University, 2401 Taft Ave., Malate, Manila, 0922, Philippines, +639327882997 & jojo_f_blanza@dlsu.edu.ph

C2CM. A multipath component (MPC) is defined by its delay (τ), angle of departure (AOD) (Azimuth of Departure (AoD), Elevation of Departure (EoD)), and angle of arrival (AOA) (Azimuth of Arrival (AoA), Elevation of Arrival (EoA)). MPCs with similar delays and angles are grouped into multipath clusters, as shown in Figure 1.

A channel impulse response (CIR) from the base station (BS) to mobile station (MS) antennas is characterized by the combination of MPCs from all the active multipath clusters and is given as

$$h(t,\tau,\boldsymbol{\Omega}^{\mathrm{BS}},\boldsymbol{\Omega}^{\mathrm{MS}}) = \sum_{k\in K}\sum_{p}\alpha_{k,p}\delta(\tau-\tau_{k,p})\delta(\boldsymbol{\Omega}^{\mathrm{BS}}-\boldsymbol{\Omega}^{\mathrm{BS}}_{k,p})\delta(\boldsymbol{\Omega}^{\mathrm{MS}}-\boldsymbol{\Omega}^{\mathrm{MS}}_{k,p}) \quad (1)$$

where K is the set of visible cluster indexes, $\alpha_{k,p}$ is the complex amplitude of the pth MPC in the kth cluster, $\boldsymbol{\Omega}^{\mathrm{BS}}_{k,p}$ is the direction of departure (AoD, EoD), and $\boldsymbol{\Omega}^{\mathrm{MS}}_{k,p}$ is the direction of arrival (AoA, EoA) of the MPC.

The dataset [23, 24] is generated by the C2CM, which consist of two indoor channel scenarios at 5.3 GHz and six semi-urban channel scenarios at 285 MHz as follows:

1. Indoor, Band 1 (B1), Line-of-Sight (LOS), Single Link (SL)
2. Indoor, Band 2 (B2), Line-of-Sight, Single Link
3. Semi-Urban, Band 1, Line-of-Sight, Single Link
4. Semi-Urban, Band 2, Line-of-Sight, Single Link
5. Semi-Urban, Band 1, Non-Line-of-Sight (NLOS), Single Link
6. Semi-Urban, Band 2, Non-Line-of-Sight, Single Link
7. Semi-Urban, Band 1, Line-of-Sight, Multiple Links (ML)
8. Semi-Urban, Band 2, Line-of-Sight, Multiple Links

Each channel scenario has thirty sets of data consisting of a different number of multipaths and multipath clusters. The seven features of the datasets are the following: whitened x-component of AOD (X_AOD_W), the whitened y-component of AOD (Y_AOD_W), the whitened z-component of AOD (Z_AOD_W), the whitened x-component of AOA (X_AOA_W), the whitened y-component of AOA (Y_AOA_W), the whitened z-component of AOA (Z_AOA_W), and the whitened delay (delay_W). The reference cluster identification of the data is given by refclusID. It serves as the ground truth in evaluating the performance of the clustering approach. The power component (rel_pow) is not included since it is not needed in clustering the data.

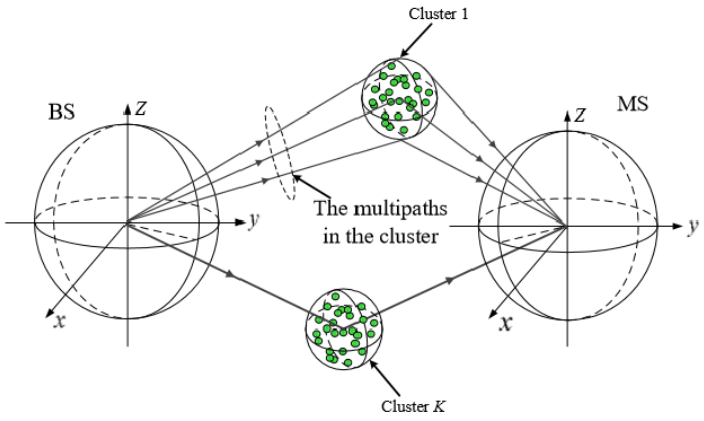


Figure 1: Generated wireless multipath components grouped as multipath clusters in C2CM [11]

3 Clustering Approaches

The dataset generated by C2CM is clustered using SCAMSMA and DDC. SCAMSMA can simultaneously determine the number of clusters and the membership of the clusters. DDC can solve the membership of the clusters and the cluster count can be calculated according to the membership of the multipaths to their clusters. The clustering approaches are used to cluster images and it is the first time that they are applied to cluster multipaths.

SCAMSMA [21] begins by formulating an affinity matrix $\boldsymbol{C}_{\mathrm{calc}}$ using the self-expression method where a given dataset $\boldsymbol{X}$ can be represented as $\boldsymbol{X}\boldsymbol{C}_{\mathrm{calc}}$ as follows

$$\min\|\boldsymbol{C}_{\mathrm{calc}}\|_1 \quad \text{s.t.} \quad \boldsymbol{X} = \boldsymbol{X}\boldsymbol{C}_{\mathrm{calc}}, \quad \mathrm{diag}\,(\boldsymbol{C}_{\mathrm{calc}}) = \boldsymbol{0} \quad (2)$$

where $\|\cdot\|_1$ is the ℓ_1 norm, which returns the sum of the absolute values of all elements and diag$(\cdot)$ are the diagonal entries of the matrix. The solution of (2) corresponds to $\boldsymbol{C}_{\mathrm{calc}}$.

Introducing an ideal affinity matrix such that $\boldsymbol{C}_{\mathrm{ideal}} = \sum_{k=1}^{K}\mathbf{z}_k \circ \mathbf{z}_k$ and $\{\mathbf{z}_k \in \{0,1\}^M\}_{k=1}^K$ where $\mathbf{z}_k = 1$ if the point belongs to the cluster otherwise $\mathbf{z}_k = 0$ and $\circ$ represents the vector outer product. By denoting $\boldsymbol{W} = -\boldsymbol{C}_{\mathrm{calc}}$, the clustering problem can be expressed as

$$\begin{aligned} &\min\langle \boldsymbol{W}, \boldsymbol{C}_{\mathrm{ideal}}\rangle, \quad \text{s.t.} \quad \mathbf{z}_k\{0,1\}^M, \\ &\sum_{k=1}^{K}\mathbf{z}_k = \mathbf{e}_M, \quad \boldsymbol{C}_{\mathrm{ideal}} = \sum_{k=1}^{K}\mathbf{z}_k \circ \mathbf{z}_k, \quad \mathrm{rank}(\boldsymbol{C}_{\mathrm{ideal}}) = K \end{aligned} \quad (3)$$

where $\langle\cdot,\cdot\rangle$ is the Frobenius inner product, $\mathbf{e}_M$ is an all one vector of size R while K is the number of clusters. SCAMSMA simultaneously solves the number of clusters and the membership of clusters.

DDC [22] optimizes a loss function based on information-theoretic measures. The loss function is defined by

$$L = \frac{1}{k}\sum_{i=1}^{k-1}\sum_{j>i}\frac{\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{K}_{\mathrm{hid}}\boldsymbol{\alpha}_j}{\sqrt{\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{K}_{\mathrm{hid}}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j^{\mathrm{T}}\boldsymbol{K}_{\mathrm{hid}}\boldsymbol{\alpha}_j}} + \mathrm{triu}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) + \frac{1}{k}\sum_{i=1}^{k-1}\sum_{j>i}\frac{\boldsymbol{m}_i^{\mathrm{T}}\boldsymbol{K}_{\mathrm{hid}}\boldsymbol{m}_j}{\sqrt{\boldsymbol{m}_i^{\mathrm{T}}\boldsymbol{K}_{\mathrm{hid}}\boldsymbol{m}_i\boldsymbol{m}_j^{\mathrm{T}}\boldsymbol{K}_{\mathrm{hid}}\boldsymbol{m}_j}} \tag{4}$$

where k is the number of clusters, $\boldsymbol{K}_{\mathrm{hid}}$ is the kernel similarity matrix, $\boldsymbol{\alpha}$ is the soft cluster assignment, $\boldsymbol{A}$ is the cluster assignment matrix, $\mathrm{triu}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})$ is the upper triangular of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ and $\boldsymbol{m}$ is the simplex corner assignment.

The clusters are represented by their probability density functions. Divergence measures the dissimilarity between clusters. The divergence builds on two fundamental objectives: the separation between clusters and the compactness within clusters, as shown in Figure 2. DDC explicitly exploits knowledge about the geometry of the output space during the optimization. DDC supports end-to-end learning, does not require hand-crafted feature design, and does not need a pre-training phase.

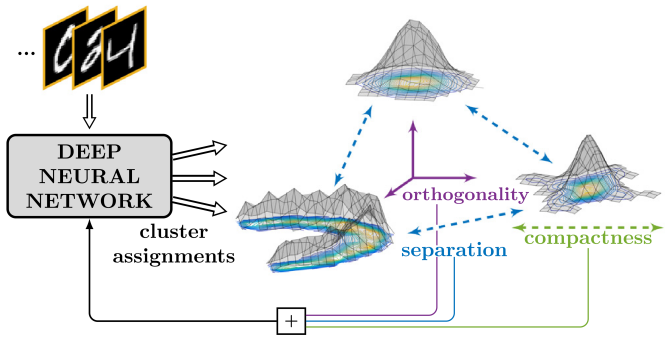


Figure 2: Fundamental objectives of divergence: separation between clusters and compactness within clusters [22]

4 Analysis of Variance (ANOVA)

One-way ANOVA [25, 26] is used to determine if there is a common mean of the data from several groups of a factor. The statistical tool can find out if there are different effects on the response variable of the different groups of an independent variable. The anova1 function of Statistics and Machine Learning Toolbox of MATLAB returns the p-value for a balanced one-way ANOVA. The MATLAB function also displays the box plots of the independent variable. Lastly, the anova1 function tests the hypothesis that the samples in the independent variable are drawn from populations with the same mean against the alternative hypothesis that the population means are not all the same.

One-way ANOVA is a simple, special case of the linear model which can be expressed as

$$y_{ij} = \alpha_j + \epsilon_{ij} \tag{5}$$

where

y_{ij} is an observation, i represents the observation number, and j represents a different group of the predictor variable y. All y_{ij} are independent.

α_j represents the population mean for the jth group.

ϵ_{ij} is the random error, independent and normally distributed, with zero mean and constant variance.

The equality of column means for the data in matrix y is tested using the MATLAB function anova1(y), where each column is a different group and has the same number of observations.

ANOVA tests the hypothesis that all group means are equal versus the alternative hypothesis that at least one group is different from the others as follows:

$$\begin{aligned} H_0 &: \alpha_1 = \alpha_2 = \ldots = \alpha_k \\ H_1 &: \text{not all group means are equal} \end{aligned} \tag{6}$$

ANOVA tests for the difference in the group means by partitioning the total variation in the data into two components: variation of group means from the overall mean and variation of observations in each group from their group mean estimates. It means that ANOVA partitions the total sum of squares (SST) into sum of squares due to between-groups effect (SSR) and sum of squared errors (SSE) given by

$$\underbrace{\sum_i\sum_j(y_{ij}-\bar{y}_{..})^2}_{SST} = \underbrace{\sum_j n_j(\bar{y}_{.j}-\bar{y}_{..})^2}_{SSR} + \underbrace{\sum_i\sum_j(y_{ij}-\bar{y}_{.j})^2}_{SSE} \tag{7}$$

where n_j is the sample size for the jth group, $j = 1, 2, ..., k$.

ANOVA compares the variation between groups to the variation within groups. If the ratio of within-group variation to between-group variation is significantly high, then the group means are significantly different from each other. This ratio can be measured using a test statistic that has an F-distribution with $(k-1, N-k)$ degrees of freedom where

$$F = \frac{\frac{SSR}{k-1}}{\frac{SSE}{N-k}} = \frac{MSR}{MSE} \sim F_{k-1,N-k} \tag{8}$$

where MSR is the mean squared treatment, MSE is the mean squared error, k is the number of groups, and N is the total number of observations. If the p-value for the F-statistic is smaller than the significance level (0.05), then the test rejects the null hypothesis that all group means are equal and concludes that at least one of the group means is different from the others. The p-value is derived by anova1 from the cumulative distribution function (CDF) of the F-distribution. The p-value is correct, if ϵ_{ij} are independent, normally distributed, and have constant variance.

5 Clustering Validity Index

The performance of a clustering approach is measured using a clustering validity index. The study uses the Jaccard index, which compares the similarity between the reference data and the calculated data. For the number of clusters, the Jaccard index is calculated as

$$\eta = \frac{|C_{11}|}{|C_{11} + C_{10} + C_{01}|} \quad \in \quad [0, 1] \tag{9}$$

where $|\cdot|$ refers to cardinality, $C_k \in C$, $K = |C|$ is the number of multipath clusters, C_{11} is the number of clusters that are present in the calculated clusters that are also present in the reference clusters, C_{10} is the number of clusters that are present in the calculated clusters but not present in the reference clusters, and C_{01} is the number of clusters that are present in the reference clusters but not present in the calculated clusters. For the membership of the clusters, the Jaccard index is calculated as

$$\eta = \frac{M_{11}}{M_{11} + M_{01} + M_{10}} \quad \in \quad [0, 1] \tag{10}$$

where M_{11} is the number of members that are present in the calculated clusters that are also present in the reference clusters, M_{10} is the number of members that are present in the calculated clusters but not present in the reference clusters, and M_{01} is the number of members that are present in the reference clusters but not present in the calculated clusters. A Jaccard index of one means that the calculated multipath clusters are the same as the reference multipath clusters, or the membership of the calculated multipath clusters is the same as the membership of the reference multipath clusters. A zero Jaccard index, on the other hand, means that there are no calculated multipath clusters that are equal to the reference multipath clusters, or there is no membership of the calculated multipath clusters that are equal to the membership of the reference multipath clusters.

6 Result and Discussion

The clustering accuracy of SCAMSMA and DDC are shown in Table 1 for the indoor scenarios and Table 2 for the semi-urban scenarios. The performance of the clustering approaches in both indoor and semi-urban scenarios are based on the number of clusters and the membership of clusters. The means of the Jaccard indices are compared using ANOVA. The ANOVA is illustrated using box plots. The box plots and p-values are generated using the anova1 one-way approach function of MATLAB. The box plots display the range of Jaccard indices and can be used to visualize the means. Values of $p < 0.05$ indicate that the means of SCAMSMA and DDC are significantly different.

Table 1: Jaccard index means of SCAMSMA and DDC in indoor scenarios

Clustering Approach	Indoor Number of Clusters	Indoor Membership of Clusters
SCAMSMA	0.6261	0.7444
DDC	0.7499	0.8054

Table 2: Jaccard index means of SCAMSMA and DDC in semi-urban scenarios

Clustering Approach	Semi-Urban Number of Clusters	Semi-Urban Membership of Clusters
SCAMSMA	0.0112	0.1615
DDC	0.0102	0.2172

The box plots of the Jaccard indices of the number of clusters of SCAMSMA and DDC for the indoor scenarios are shown in Figure 3. The p-value is 0.0388, which indicates that there is a significant difference in the means of the clustering accuracies. This difference can be seen in the figure where the box plot of DDC is higher than that of SCAMSMA. DDC is more accurate than SCAMSMA in clustering multipaths by 19.77%.

Figure 4 presents the box plots of the Jaccard indices of the membership of clusters of SCAMSMA and DDC for the indoor scenarios. The p-value is 0.0996, which attests that there is no significant difference in the means of the Jaccard indices. This difference can be visualized in the figure where the box plots are almost on the same level ($\sim$ 0.8). DDC has higher accuracy in clustering multipaths than SCAMSMA by only 8.19%.

The box plots of the Jaccard indices of the number of clusters of SCAMSMA and DDC for the semi-urban scenarios are illustrated

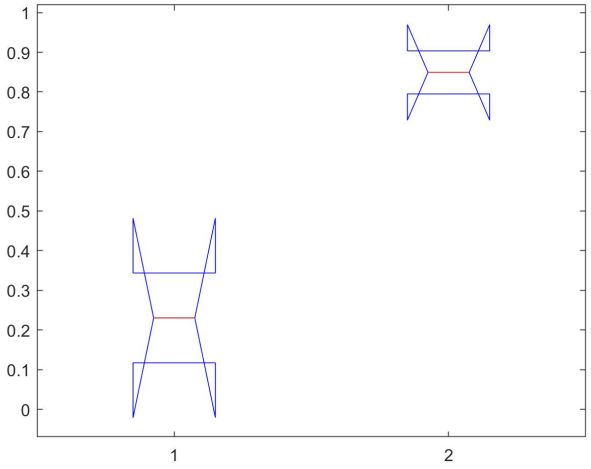


Figure 3: Box plots of the Jaccard indices for the number of clusters in indoor scenarios of SCAMSMA in box plot 1 and DDC in box plot 2

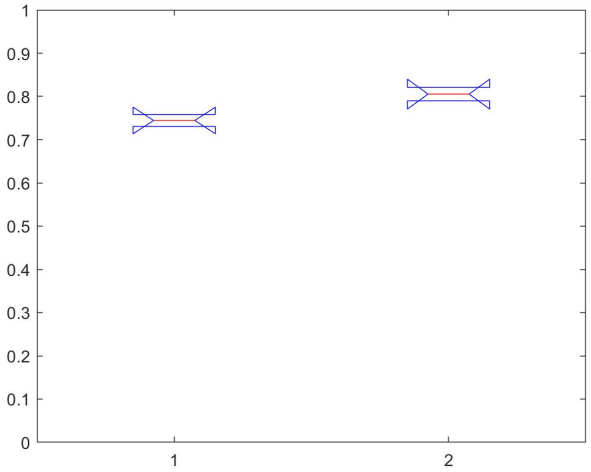


Figure 4: Box plots of the Jaccard indices for the membership of clusters in indoor scenarios of SCAMSMA in box plot 1 and DDC in box plot 2

in Figure 5. The p-value is 0.7419, which suggests that the means of SCAMSMA and DDC are similar (0.0112 vs. 0.0102). The figure indicates that the box plots are on the same level near the horizontal axis (~ 0). SCAMSMA is more accurate this time by 9.80%.

Figure 6 displays the box plots of the Jaccard indices of the membership of clusters of SCAMSMA and DDC for the semi-urban scenarios. The p-value is 0.0139, which proves that the means of SCAMSMA and 3CAM-SCAMSMA are significantly different. The figure shows that the box plot of DDC is higher than that of SCAMSMA. DDC has a higher clustering accuracy of 34.49%.

DDC shows consistency in its clustering performance due to higher accuracy of clustering multipaths in all channel scenarios except for the membership of clusters in semi-urban scenarios where SCAMSMA has a slight clustering advantage of 0.0012. Also, the means of SCAMSMA and DDC are significantly different since the p-values are less than 0.05 except in the membership of clusters in semi-urban scenarios. Lastly, the clustering approaches can be used in indoor scenarios but not in semi-urban scenarios based on accuracy which is validated by the measurements done in indoor environment [27].

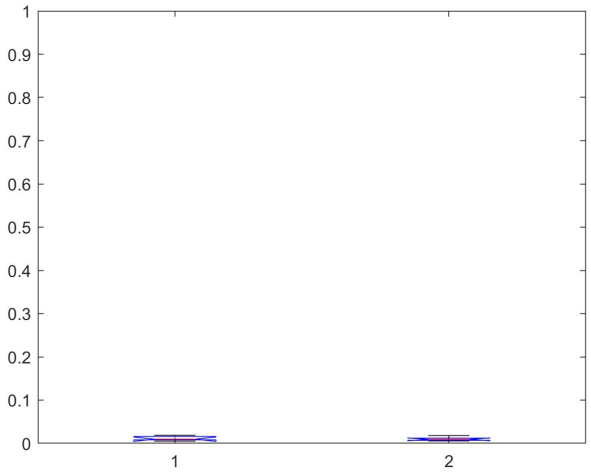


Figure 5: Box plots of the Jaccard indices for the number of clusters in semi-urban scenarios of SCAMSMA in box plot 1 and DDC in box plot 2

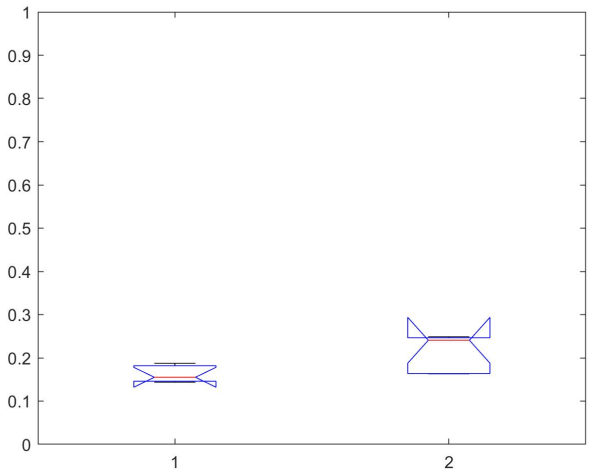


Figure 6: Box plots of the Jaccard indices for the membership of clusters in semi-urban scenarios of SCAMSMA in box plot 1 and DDC in box plot 2

7 Conclusion

This work presents the comparison of the clustering accuracy of SCAMSMA and DDC in clustering wireless propagation multipaths generated by C2CM. Jaccard index is used as the performance metric of the clustering approaches. Results show that there is a significant difference in the cluster-wise Jaccard index between SCAMSMA and DDC for indoor scenarios while the membership-wise Jaccard index is not different. On the other hand, the cluster-wise Jaccard index is not different between the clustering approaches for semiurban scenarios while the membership-wise Jaccard index is a little different. The clustering approaches can be used in indoor scenarios based on accuracy. However, a better multipath clustering method should be used for semi-urban scenarios. For future work, the results will be compared with other clustering approaches to determine the best performance in terms of clustering wireless multipaths in indoor and semi-urban environments.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors would like to thank the Department of Science and Technology-Engineering Research and Development for Technology (DOST-ERDT), Office of Vice-Chancellor for Research and Innovation (OVCRI) of De La Salle University (DLSU), and Tokyo City University (TCU) for the realization of this paper.

References

[1] J. Poutanen, K. Haneda, L. Liu, C. Oestges, F. Tufvesson, P. Vainikainen, "Parameterization of the COST 2100 MIMO Channel Model in Indoor Scenarios," in Proceedings 5th European Conference on Antennas and Propagation, 3606–3610, 2011.

[2] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, P. De Doncker, "The COST 2100 MIMO channel model," IEEE Wireless Commun., **19**(6), 92–99, 2012. https://doi.org/10.1109/MWC.2012.6393523

[3] R. Verdone, A. Zanella, "Pervasive Mobile and Ambient Wireless Communications: COST Action 2100 (Signals and Communication Technology)," London: Springer, 2012.

[4] M. Zhu, G. Eriksson, F. Tufvesson, "The COST 2100 channel model: Parameterization and validation based on outdoor MIMO measurements at 300 MHz," IEEE Trans. Wireless Commun., **12**(2), 888–897, 2013. https://doi.org/10.1109/TWC.2013.010413.120620

[5] N. Czink, P. Cera, J. Salo, E. Bonek, J.-P. Nuutinen, J. Ylitalo, "A Framework for Automatic Clustering of Parametric MIMO Channel Data Including Path Powers," in IEEE 64th Vehicular Technology Conference, 1–5, 2006. https://doi.org/10.1109/VTCF.2006.35

[6] C. Gentile, "Using the kurtosis measure to identify clusters in wireless channel impulse responses," IEEE Trans. Antennas Propag., **61**(6), 3392–3395, 2013. https://doi.org/10.1109/TAP.2013.2253299

[7] B. Li, C. Zhao, H. Zhang, Z. Zhou, A. Nallanathan, "Efficient and robust cluster identification for ultra-wideband propagations inspired by biological ant colony clustering," IEEE Trans. Commun., **63**(1), 286–300, 2015. https://doi.org/10.1109/TCOMM.2014.2377120

[8] S. Cheng, M.-T. Martinez-Ingles, D.P. Gaillot, J.-M. Molina-Garcia-Pardo, M. Liénard, P. Degauque, "Performance of a novel automatic identification algorithm for the clustering of radio channel parameters," IEEE Access, **3**, 2252–2259, 2015. https://doi.org/10.1109/ACCESS.2015.2497970

[9] R. He, W. Chen, B. Ai, A.F. Molisch, W. Wang, Z. Zhong, J. Yu, S. Sangodoyin, "On the clustering of radio channel impulse responses using sparsity-based methods," IEEE Trans. Antennas Propag., **64**(6), 2465–2474, 2016. https://doi.org/10.1109/TAP.2016.2546953

[10] R. He, Q. Li, B. Ai, Y. Geng, A. Molisch, V. Kristem, Z. Zhong, J. Yu, "A kernel-power-density-based algorithm for channel multipath components clustering," IEEE Trans. Wireless Commun., **16**(11), 7138–7151, 2017. https://doi.org/10.1109/TWC.2017.2740206

[11] Y. Li, J. Zhang, Z. Ma, Y. Zhang, "Clustering analysis in the wireless propagation channel with a variational Gaussian mixture model," IEEE Trans. Big Data, **6**(2), 223–232, 2020. https://doi.org/10.1109/TBDATA.2018.2840696

[12] H. Xue, L. Tian, Y. Zhang, C. Wang, J. Zhang, W. Li, H. Wei, "A Novel Clustering Method Based on Density Peaks and Its Validation by Channel Measurement," in Proceedings 12th European Conference on Antennas and Propagation, 1–5, 2018. https://doi.org/10.1049/cp.2018.0700

[13] D. Abinoja, L. Materum, "Evaluation of techniques for identifying multipath propagation clusters in wireless systems by Bayesian information criterion model-based clustering," Lecture Notes Adv. Res. Eng. Inf. Technol. 1, 97–102, 2019.

[14] A. Teologo, L. Materum, "Cluster-wise Jaccard accuracy of Kpower means on multipath datasets," Int. J. Emerg. Trends Eng. Res., **7**(8), 203–208, 2019. https://doi.org/10.30534/ijeter/2019/16782019

[15] O. Bialer, N. Garnett, D. Levi, "A Deep Neural Network Approach for Time-Of-Arrival Estimation in Multipath Channels," in IEEE International Conference on Acoustics, Speech, and Signal Processing, 2936–2940, 2018. https://doi.org/10.1109/ICASSP.2018.8461301

[16] A. Abdallah, Z. Kassas, "Deep Learning-Aided Spatial Discrimination for Multipath Mitigation," in IEEE/ION Position, Location and Navigation Symposium, 1324–1335, 2020. https://doi.org/10.30534/ijeter/2019/16782019

[17] J. Blanza, L. Materum, "Grouping of COST 2100 indoor multipaths using simultaneous clustering and model selection," Int. J. Emerg. Trends Eng. Res., **7**(8), 197–202, 2019. https://doi.org/10.30534/ijeter/2019/15782019

[18] J. Blanza, L. Materum, "Wireless propagation multipath clustering: On simultaneously solving the membership and the number of clusters," Int. J. Adv. Trends Comput. Sci. Eng., **8**(5), 1914–1919, 2019. https://doi.org/10.30534/ijatcse/2019/15852019

[19] J. Blanza, L. Materum, "Joint identification of the clustering and cardinality of wireless propagation multipaths," Int. J. Emerg. Trends Eng. Res., **7**(12), 762–767, 2019. https://doi.org/10.30534/ijeter/2019/057122019

[20] J. Blanza, L. Materum, T. Hirano, "Deep divergence-based clustering of wireless multipaths for simultaneously addressing the grouping and the cardinality," Int. J. Emerg. Trends Eng. Res., **8**(7), 3104–3110, 2020. https://doi.org/10.30534/ijeter/2020/37872020

[21] Z. Li, L. Cheong, S. Yang, K. Toh, "Simultaneous clustering and model selection: Algorithm, theory and applications," IEEE Trans. Pattern Anal. Mach. Intell., **40**(8), 1964—1978, 2018. https://doi.org/10.1109/TPAMI.2017.2739147

[22] M. Kampffmeyer, S. Løkse, F. M. Bianchi, L. Livi, A.-B. Salberg, R. Jenssen, "Deep divergence-based approach to clustering," Neural Netw., **113**, 91–101, 2019. https://doi.org/10.1016/j.neunet.2019.01.015

[23] J. Blanza, A. Teologo, L. Materum, "Datasets for Multipath Clustering at 285 MHz and 5.3 GHz Bands Based on COST 2100 MIMO Channel Model," in 9th International Symposium on Multimedia and Communication Technology, 1–5, 2019. https://doi.org/10.1109/ISMAC.2019.8836143

[24] J. Blanza, A. Teologo, L. Materum, "Datasets for Multipath Clustering at 285 MHz and 5.3 GHz Bands Based on COST 2100 MIMO Channel Model," 2019. [Online]. Available: http://dx.doi.org/10.21227/4cb9-hf81

[25] C. Wu, M. Hamada, Experiments: Planning, Analysis, and Parameter Design Optimization, USA, Willey, John & Sons Incorporated, 2000.

[26] J. Neter, M. Kutner, C. Nachtsheim, W. Wasserman, Applied Linear Statistical Models 4th ed., Irwin Press, 1996.

[27] J. Poutanen, "Geometry-based Radio Channel Modeling: Propagation Analysis and Concept Development," Ph.D Thesis, Aalto University School of Electrical Engineering, 2011.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 544-549 (2020)
www.astesj.com
Second International Virtual Conference on Multidisciplinary Research 2020

ASTES Journal
ISSN: 2415-6698

Interface for Visualization of Wireless Propagation Multipath Clustering Outcomes

Jojo Blanza[*,1,2], Lawrence Materum[1,3]

[1]*Department of Electronics and Communications Engineering, De La Salle University, 2401 Taft Ave., Malate, Manila, 0922, Philippines*

[2]*Electronics Engineering Department, University of Santo Tomas, Espana Blvd., Sampaloc, Manila, 1008, Philippines*

[3]*Tokyo City University, 1-28-1 Tamazutsumi, Setagaya, Tokyo, 158-8557, Japan*

ARTICLE INFO

Article history:
Received: 08 September, 2020
Accepted: 07 October, 2020
Online: 20 November, 2020

Keywords:
Channel models
Data preprocessing
Graphical user interfaces

ABSTRACT

A graphical user interface (GUI) is presented to visualize the multipaths generated by COST 2100 channel model (C2CM) and the results of clustering the wireless propagation multipaths using Modified Simultaneous Clustering and Model Selection (MSCAMSMA). The usual practice of authors is to show their data and results using figures, tables, and graphs which are already sufficient to present their studies. However, the manner of displaying the data and results is static; that is, the user cannot examine them to analyze further the relationship of the different variables involved in the study. The paper presents a step further by showing dynamically the data and results which the user can manipulate according to its needs.

1 Introduction

Data and results are integral parts of a scientific study. They are important in the research process as the data is the basis of the study, and the results show the outcomes of the work done. That is why it is important to present the data and results in a manner that is easily read, seen, and understood by the reader.

Most papers use figures, tables, algorithms, equations, diagrams, illustrations, programs, and graphs to visualize the data and results of the study. The technique of presenting the data and results is sufficient for what the authors want to show. However, this method is static and dependent on what the authors want to present. Researchers cannot manipulate the data and analyze the effects of changing the input parameters. This paper addresses this concern by presenting a graphical user interface (GUI) that is dynamic, modifiable, and suitable to the need of the user. This way, researchers can have a deeper and better experience and understanding of what they are studying and examining.

GUI is one of the key technological advancements due to improvements in computer hardware and software. GUI greatly improved user experience by displaying the information. A GUI is an interactive system with visual components for manipulation. An interface is a useful tool that makes the most of the vision capacity of the user. A GUI allows a person to visually control the flow of information and decide on what is relevant and discard what is not [1]. Some lack analysis platform [2]–[4] unlike those in [5]–[13]. In this paper, a GUI is developed to manage the data and results of clustering European Cooperation in Science and Technology (COST) 2100 dataset [14, 15] using Modified Simultaneous Clustering and Model Selection (MSCAMSMA) [16]. The interface is an effective method in visualizing the COST 2100 dataset and the clustering results. A previous work [17] on GUI can visually analyze the reference data and the calculated clusters but considered only the cluster count. In this work, the GUI can be used by the user to visualize, manipulate, and analyze the multipaths generated by the COST 2100 channel model (C2CM) and the clustering results of MSCAMSMA not only giving the accuracy on the number of clusters but also the membership of clusters. The paper is organized in the following way. Section 2 presents the dataset generated by C2CM. Section 3 describes the clustering approach. Section 4 discusses the GUI developed. Section 5 concludes the work.

2 COST 2100 Channel Model (C2CM) Dataset

C2CM [18]–[21] is characterized by multipath clusters, as shown in Figure 1. C2CM can reproduce the stochastic properties of multiple-

[*]Corresponding Author: Jojo Blanza, Department of Electronics and Communications Engineering, De La Salle University, 2401 Taft Ave., Malate, Manila, 0922, Philippines, +639327882997 & jojo_f_blanza@dlsu.edu.ph

input multiple-output (MIMO) wireless communication channels. A multipath component (MPC) is defined by its delay (τ), angle of departure (Azimuth of Departure (AoD), Elevation of Departure (EoD)), and angle of arrival (Azimuth of Arrival (AoA), Elevation of Arrival (EoA)). Multipath clusters are groups of MPCs with similar delays and angles.

A channel impulse response (CIR) from the base station (BS) to mobile station (MS) antennas is characterized by the combination of MPCs from all the active multipath clusters and is given as

$$h(t,\tau,\boldsymbol{\Omega}^{\mathrm{BS}},\boldsymbol{\Omega}^{\mathrm{MS}}) = \sum_{k\in K}\sum_{p}\alpha_{k,p}\delta(\tau-\tau_{k,p})\delta(\boldsymbol{\Omega}^{\mathrm{BS}}-\boldsymbol{\Omega}^{\mathrm{BS}}_{k,p})\delta(\boldsymbol{\Omega}^{\mathrm{MS}}-\boldsymbol{\Omega}^{\mathrm{MS}}_{k,p}) \quad (1)$$

where K is the set of visible cluster indexes, $\alpha_{k,p}$ is the complex amplitude of the pth MPC in the kth cluster, $\boldsymbol{\Omega}^{\mathrm{BS}}_{k,p}$ is the direction of departure (AoD, EoD), and $\boldsymbol{\Omega}^{\mathrm{MS}}_{k,p}$ is the direction of arrival (AoA, EoA) of the MPC.

For a MIMO system using V and U multiple antennas arrays at BS and MS, respectively, the $U \times V$ MIMO channel matrix $\mathbf{H}(t,\tau)$, under the plane wave and balanced narrowband array assumptions, can be expressed as

$$\mathbf{H}(t,\tau) = \sum_{k\in K}\sum_{p}\alpha_{k,p}\boldsymbol{s}_{\mathrm{MS}}(\boldsymbol{\Omega}^{\mathrm{MS}})\boldsymbol{s}^{\mathrm{T}}_{\mathrm{BS}}(\boldsymbol{\Omega}^{\mathrm{BS}}) \quad (2)$$

where $\boldsymbol{s}_{\mathrm{MS}}$ is the antenna steering vector at MS and $\boldsymbol{s}_{\mathrm{BS}}$ is the antenna steering vector at BS.

The dataset [14, 15] is generated by C2CM, which consist of two indoor and six semi-urban channel scenarios as follows:

1. Indoor, Band 1 (B1), Line-of-Sight (LOS), Single Link (SL)
2. Indoor, Band 2 (B2), Line-of-Sight, Single Link
3. Semi-Urban, Band 1, Line-of-Sight, Single Link
4. Semi-Urban, Band 2, Line-of-Sight, Single Link
5. Semi-Urban, Band 1, Non-Line-of-Sight (NLOS), Single Link
6. Semi-Urban, Band 2, Non-Line-of-Sight, Single Link
7. Semi-Urban, Band 1, Line-of-Sight, Multiple Links (ML)
8. Semi-Urban, Band 2, Line-of-Sight, Multiple Links

Each channel scenario has thirty sets of data. There are a different number of multipaths and multipath clusters for each set of data. The seven features of the dataset are the following: whitened x-component of AOD (X_AOD_W), the whitened y-component of AOD (Y_AOD_W), the whitened z-component of AOD (Z_AOD_W), the whitened x-component of AOA (X_AOA_W), the whitened y-component of AOA (Y_AOA_W), the whitened z-component of AOA (Z_AOA_W), and the whitened delay (delay_W). The reference cluster identifications of the data are given by refclusID, which serves as the ground truth in evaluating the performance of the clustering approach. The power component (rel_pow) is not included since it is not needed by the clustering approach.

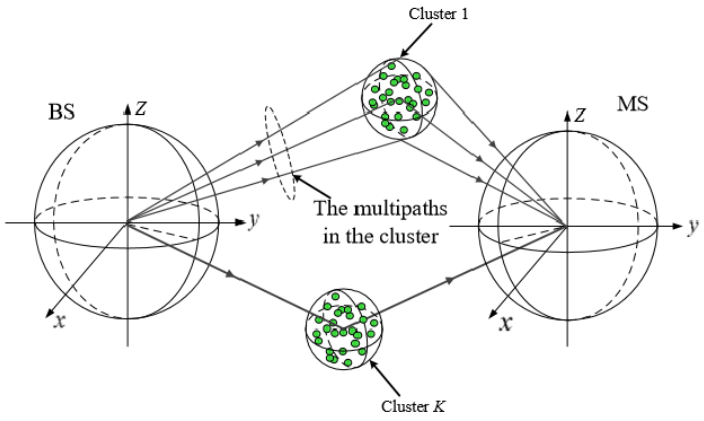


Figure 1: Multipath components and multipath clusters in COST 2100 channel model [22]

3 Modified Simultaneous Clustering and Model Selection (MSCAMSMA)

MSCAMSMA [16] is used to cluster the COST 2100 dataset. The clustering approach is introduced to improve the clustering accuracy of SCAMSMA [23]. The accuracy of SCAMSMA depends on the correct formulation of the data affinity matrix, $\boldsymbol{C}_{\mathrm{calc}}$, so that a 0-1 block diagonal is formed. MSCAMSMA can be described as follows: Let the dataset $\boldsymbol{X} \in \mathbb{R}^{L\times D}$ where L was the number of multipath components or the number of rows and D the number of dimensions, which is seven, or the number of columns. A multipath, $\boldsymbol{x}$, consists of a single row of seven columns. A dataset can be expressed as

$$\boldsymbol{X} = [\boldsymbol{x}_1,\ldots,\boldsymbol{x}_L]^{\mathrm{T}}. \quad (3)$$

MSCAMSMA depends on three constraints: pairwise, binary, and proximity. Pairwise Constraint, $\boldsymbol{P}_{\ell,d}$, depends on the absolute distance, $\boldsymbol{d}_{\mathrm{abs}}$, of the corresponding pair of multipaths for the seven dimensions which are given by

$$\left|\boldsymbol{x}_{\ell,d} - \boldsymbol{x}_{d,d}\right| \quad (4)$$

where $\boldsymbol{x}_{\ell,d}$ and $\boldsymbol{x}_{d,d}$ are multipaths, ℓ is the row number, and d is the column number. The distance parameter, $\boldsymbol{d}_{\mathrm{par}}$, is determined by

$$\frac{600}{L}\left[\max(\boldsymbol{X}(:,d)) - \min(\boldsymbol{X}(:,d))\right] \quad (5)$$

where max( ) is the function that gives the maximum value of a column while min( ) gives the minimum value of the said column, and ":" denotes all rows if it is in the first argument. Factor 600 is used so that MSCAMSMA applies to all of the datasets. A value less than 600 gives at least one all-zero row leading to an incorrect affinity matrix. A value greater than 600 leads to an all-one matrix,

excluding the main diagonal. The resulting affinity matrix is also not correct. The Pairwise Constraint is given by

$$\boldsymbol{P}_{\ell,d} = \begin{cases} 1 & \text{if} \quad \boldsymbol{d}_{\text{abs}} < \boldsymbol{d}_{\text{par}} \\ 0 & \text{if} \quad \boldsymbol{d}_{\text{abs}} \geq \boldsymbol{d}_{\text{par}} \end{cases} \tag{6}$$

Binary Constraint, $\boldsymbol{B}_{\ell,d}$, takes on the sum of the values of the $\boldsymbol{P}_{\ell,d}$ of the seven dimensions and returns a binary value of either zero or one. If the sum is less than four, then the returned value is zero, which means that the multipaths are not of the same cluster. On the other hand, when the sum is greater than or equal to four, then a value of one is returned, which means that the multipaths are within the same cluster. Binary Constraint is given by

$$\boldsymbol{B}_{\ell,d} = \begin{cases} 1 & \text{if} \quad \sum \boldsymbol{P}_{\ell,d} \geq 4 \\ 0 & \text{if} \quad \sum \boldsymbol{P}_{\ell,d} < 4 \end{cases} \tag{7}$$

Proximity Constraint is followed by combining multipaths to form clusters around the main diagonal, thus, creating the 0-1 block diagonal of $\boldsymbol{C}_{\text{calc}}$. $\boldsymbol{C}_{\text{calc}}$ serves as input to Alternating Direction of Method of Multipliers (ADMM) which is used to calculate the ideal affinity matrix, $\boldsymbol{C}_{\text{ideal}}$. Asso Constrained Boolean Matrix Factorization (AssoCBMF) takes $\boldsymbol{C}_{\text{ideal}}$ as input to solve for the number of clusters and the membership of the clusters.

4 Graphical User Interface (GUI)

A GUI is an interactive system with visual components that can be used to manipulate a computer program. The study uses a GUI to manage the data and results of clustering COST 2100 dataset using MSCAMSMA. The introductory page showing the different tabs and the explanations is illustrated in Figure 2. Aside from the Introduction Tab, there are five other tabs, namely: COST 2100 Dataset, Directional Cosine Transform and Whitening Transform, Modified SCAMSMA, Relative Frequencies, and Summary.

The COST 2100 Dataset tab displaying the various channel scenarios is shown in Figure 3. Each chosen channel scenario gives the corresponding data generated by C2CM and the clustering results.

Figure 4 shows the window arising from the chosen channel scenario in the COST 2100 Dataset tab. When the generate button is pressed, the window shows the characteristics of the channel scenario like the channel frequency, mobile station velocity, the distance between the mobile station and the base station, the number of clusters, and the number of multipaths as shown in Figure 5. The parameters of the multipaths are tabulated and shown in a graph. The multipaths are color-coded with lighter colors representing higher relative power, while darker colors represent lower relative power. The default is Trial 1. Choosing another trial, from Trial 2 to Trial 30, by changing the position of the slider and pressing the Generate button, shows the corresponding characteristics, parameters, and graph of the multipaths.

The multipaths are transformed using Directional Cosine Transform (DCT) and Whitening Transform (WT). DCT is performed to the datasets to overcome the problem of the circular nature of the angular domain. At the same time, WT is applied to the dataset to eliminate the issues on units and standardized the dataset [14]. The corresponding graphs of the transformed multipaths and the

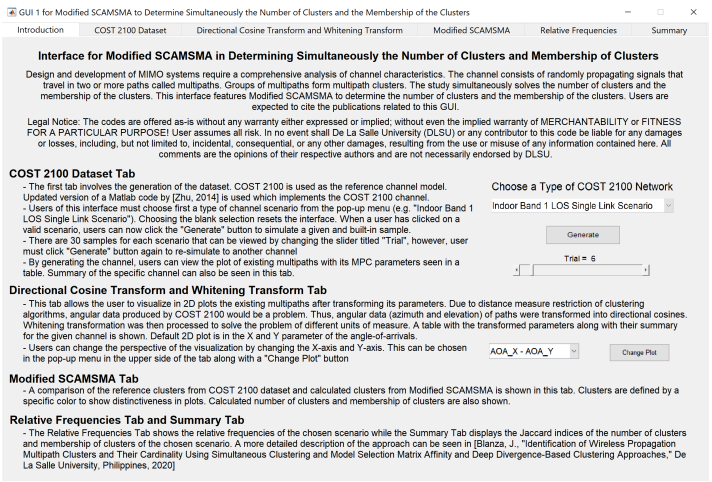


Figure 2: Graphical user interface introductory page containg the different tabs and their descriptions

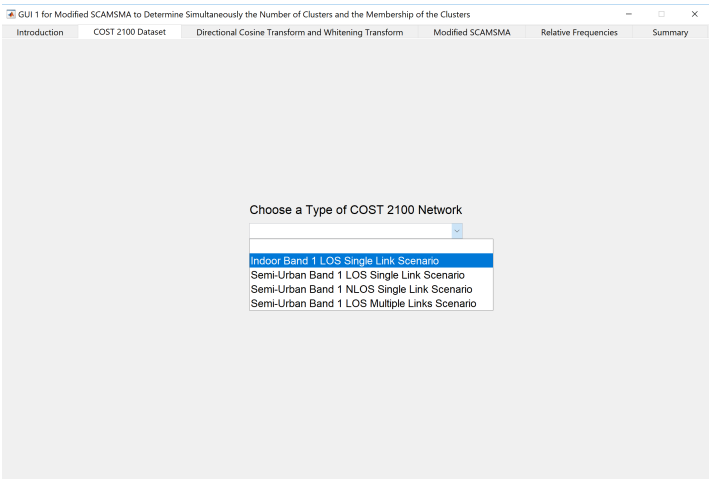


Figure 3: Ensuing window for the COST 2100 dataset tab showing the different channel scenarios

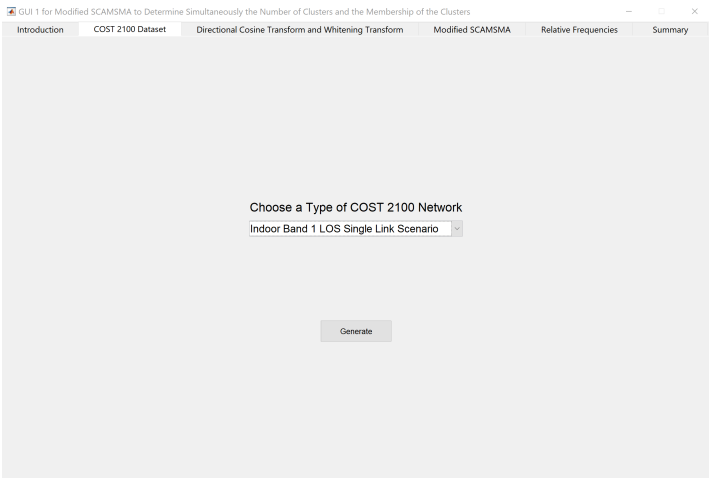


Figure 4: Subsequent window for the chosen channel scenario of the COST 2100 dataset

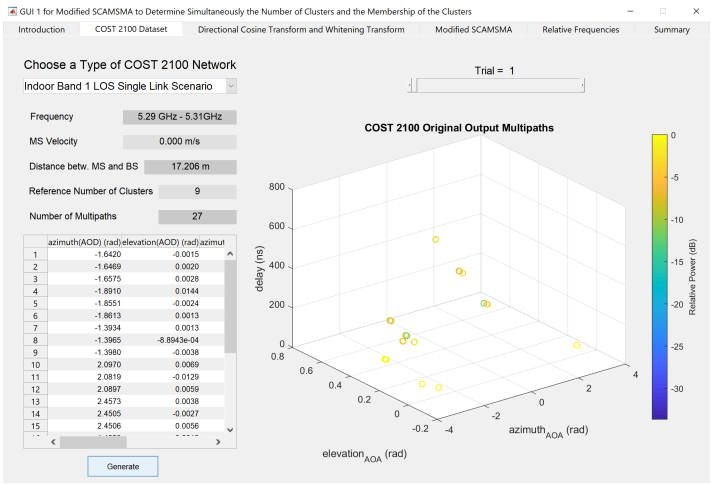


Figure 5: COST 2100 Dataset tab displaying the characteristics, parameters, and graph of the multipaths of the chosen channel scenario

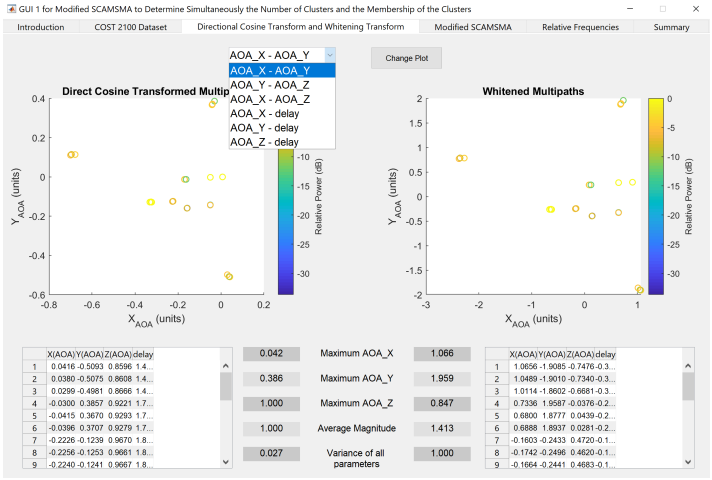


Figure 6: Graphs and tabulated parameters of the transformed multipath components of the chosen combination in the drop down menu

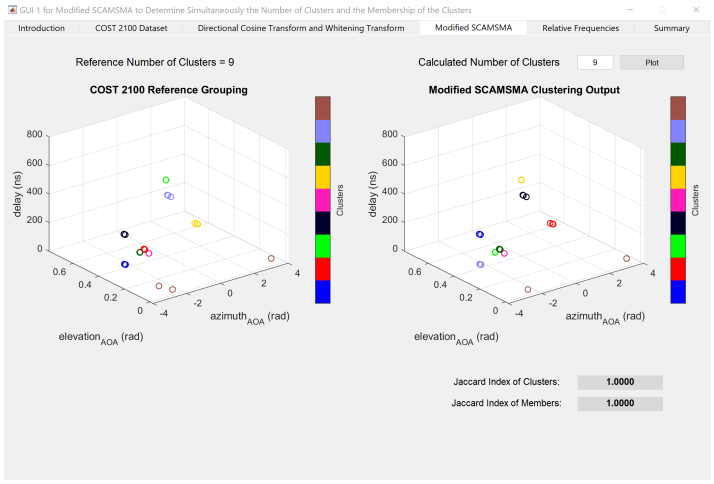


Figure 7: Comparison of the reference multipath clusters and the calculated multipath clusters and the resulting Jaccard indices

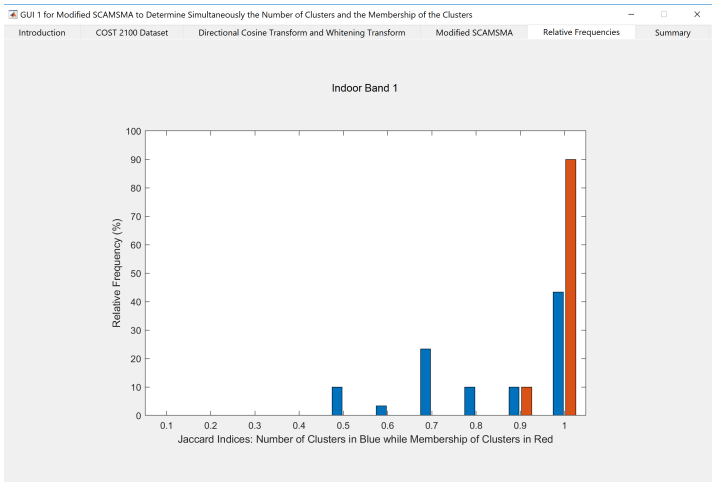


Figure 8: Window for tab four showing the relative frequencies of the Jaccard indices of the number of clusters in blue and the membership of clusters in red

tabulated parameters are shown in the Directional Cosine Transform and Whitening Transform tab in Figure 6. The graphs and tabulated parameters can be changed by choosing a different combination in the drop-down menu and pressing the Change Plot button.

Figure 7 shows the Modified SCAMSMA tab illustrating the grouping of the COST 2100 dataset at the left and the clustering output of MSCAMSMA at the right. The window also shows the reference number of clusters, the calculated number of clusters, the Jaccard index accuracy of the number of clusters, and the Jaccard index accuracy of the membership of clusters. The graphs, number of clusters, and Jaccard indices appear only when the calculated number of clusters is equal to the reference number of clusters. Otherwise, the results will not be shown.

The Relative Frequencies tab in Figure 8 shows the relative frequencies of the number of clusters in blue and the membership of clusters in red. The higher the relative frequency, when the Jaccard index is closer to one means that the clustering accuracy is higher. On the other hand, relative frequencies near zero Jaccard index implies that there is lower clustering accuracy.

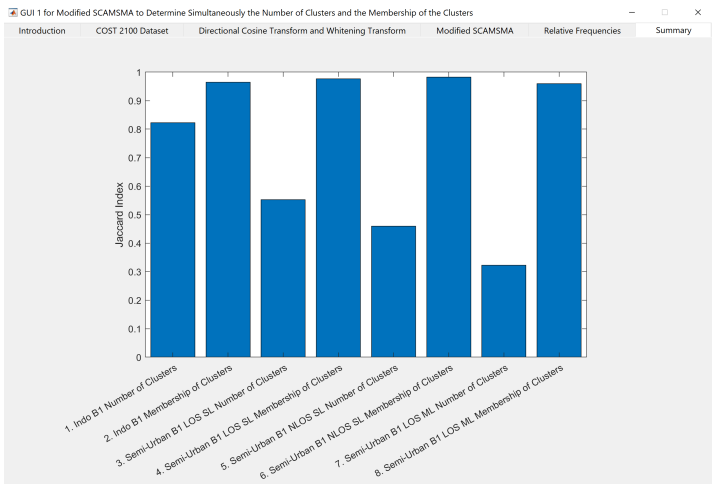


Figure 9: Window for tab five showing the mean Jaccard indices of the number of clusters and the membership of clusters of the channel scenarios

Figure 9 shows the Summary tab displaying the mean Jaccard indices of the number of clusters and the membership of clusters of all the channel scenarios. Higher mean Jaccard index means that the multipaths are clustered more accurately.

The MATLAB codes for the GUI, the reference dataset, and the calculated dataset can be found at [24]. The codes together with the reference dataset and the clustered multipaths can be compiled using the MATLAB version 2019a software to manipulate the dataset and the clustering results.

5 Conclusion

This paper shows a GUI that can be used to visualize the multipaths generated by C2CM and the results of clustering the wireless propagation multipaths using MSCAMSMA. The usual practice of authors is to illustrate their data and results using figures, tables, and graphs, which are already sufficient to present their studies. However, the presentation is static and cannot be manipulated by the readers. This study extended the common practice of showing the data and results by presenting an interface to visualize and manage what is not displayed in the publication. This way, researchers can have a better understanding of the study being presented and explore the portions of the study not shown.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors would like to thank the Department of Science and Technology-Engineering Research and Development for Technology (DOST-ERDT), Office of Vice-Chancellor for Research and Innovation (OVCRI) of De La Salle University (DLSU), and Tokyo City University (TCU) for the realization of this paper.

References

[1] A. Edwards, "The rise of the graphical user interface," Library Hi Tech, **14**(1), 46–50, 1996. https://doi.org/10.1108/eb047980

[2] L. Villanueva, R. Gustilo, "Artificial neural network based antenna sensitivity assignments for chaotic internet service provider network architecture," Int. J. Eng. Technol.(UAE), **7**(2), 14–17, 2018. https://doi.org/10.14419/ijet.v7i2.3.9958

[3] S. Chatvichienchai, Y. Kawasaki, "Spreaddb: Spreadsheet-based user interface for querying and updating data of external databases," Int. J. Adv. Trends Comput. Sci. Eng., **7**(2), 6–10, 2018. https://doi.org/10.30534/ijatcse/2018/01722018

[4] L. Vladareanu, V. Vladareanu, H. Yu, H. Wang, F. Smarandache, "Robot advanced intellectual control developed through flexible intelligent portable platform," Int. J. Adv. Trends Comput. Sci. Eng., **8**(1.1), 202–208, 2019. https://doi.org/10.30534/ijatcse/2019/3881.12019

[5] M. Cabatuan, E. Dadios, R. Naguib, "Computer Vision-Based Breast Self-Examination Palpation Pressure Level Classification Using Artificial Neural Networks and Wavelet Transforms," in Proceedings IEEE Region 10 Annual International Conference, 2012. https://doi.org/10.1109/TENCON.2012.6412282

[6] M. Eman, M. Cabatuan, E. Dadios, L. Gan Lim, "Detecting and Tracking Female Breasts Using Neural Network in Real-Time," in Proceedings IEEE Region 10 Annual International Conference, 2013. https://doi.org/10.1109/TENCON.2013.6718899

[7] J. Jose, M. Cabatuan, E. Dadios, L. Gan Lim, "Depth Estimation in Monocular Breast Self-Examination Image Sequence Using Optical Flow," in 7th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management joint with 6th International Symposium on Advanced Computational Intelligence and Intelligent Informatics co-located with 10th Engineering Research and Development for Technology Conferrence, 2014. https://doi.org/10.1109/HNICEM.2014.7016220

[8] J. Jose, M. Cabatuan, E. Dadios, L. Gan Lim, "Stroke Position Classification in Breast Self-Examination Using Parallel Neural Network and Wavelet Transform," in Proceedings IEEE Region 10 Annual International Conference, 2015. https://doi.org/10.1109/TENCON.2014.7022288

[9] J. Jose, M. Cabatuan, R. Billones, E. Dadios, L. Gan Lim, "Monocular Depth Level Estimation for Breast Self-Examination (BSE) Using RGBD BSE Dataset," in Proceedings IEEE Region 10 Annual International Conference, 2016. https://doi.org/10.1109/TENCON.2015.7372948

[10] R. Masilang, M. Cabatuan, E. Dadios, "Hand Initialization and Tracking Using a Modified KLT Tracker for a Computer Vision-Based Breast Self-Examination System," in 7th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management joint with 6th International Symposium on Advanced Computational Intelligence and Intelligent Informatics co-located with 10th Engineering Research and Development for Technology Conference, 2014. https://doi.org/10.1109/HNICEM.2014.7016244

[11] R. Masilang, M. Cabatuan, E. Dadios, L. Gan Lim, "Computer-Aided BSE Torso Tracking Algorithm Using Neural Networks, Contours, and Edge Features," in Proceedings IEEE Region 10 Annual International Conference, 2015. https://doi.org/10.1109/TENCON.2014.7022300

[12] E. Mohammadi, E. Dadios, L. Gan Lim, M. Cabatuan, R. Naguib, J. Avila, A. Oikonomou, "Real-time evaluation of breast self-examination using computer vision," Int. J. Biomed. Imag., **2014**, 2014. https://doi.org/10.1155/2014/924759

[13] S. Rabano, M. Cabatuan, E. Sybingco, E. Dadios, E. Calilung, "Common Garbage Classification Using MobileNet," in Proceedings 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, 1–4, Nov. 2018. https://doi.org/10.1109/HNICEM.2018.8666300

[14] J. Blanza, A. Teologo, L. Materum, "Datasets for Multipath Clustering at 285 MHz and 5.3 GHz Bands Based on COST 2100 MIMO Channel Model," in 9th International Symposium on Multimedia and Communication Technology, 1–5, 2019. https://doi.org/10.1109/ISMAC.2019.8836143

[15] J. Blanza, A. Teologo, L. Materum, "Datasets for Multipath Clustering at 285 MHz and 5.3 GHz Bands Based on COST 2100 MIMO Channel Model," 2019. [Online]. Available: http://dx.doi.org/10.21227/4cb9-hf81

[16] J. Blanza, "Identification of Wireless Propagation Multipath Clusters and Their Cardinality Using Simultaneous Clustering and Model Selection Matrix Affinity and Deep Divergence-Based Clustering Approaches," Ph.D Thesis, De La Salle University, 2020.

[17] M. Roque, L. Materum, "Interface for the Factor-inclusion Weighting Approach in Determining the Number of Multipath Propagation Clusters," Int. J. Adv. Trends Comput. Sci. Eng., **8**(4), 1768–1776, 2019. https://doi.org/10.30534/ijatcse/2019/107842019

[18] J. Poutanen, K. Haneda, L. Liu, C. Oestges, F. Tufvesson, P. Vainikainen, "Parameterization of the COST 2100 MIMO Channel Model in Indoor Scenarios," in Proceedings 5th European Conference on Antennas and Propagation, 3606–3610, 2011.

[19] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, P. De Doncker, "The COST 2100 MIMO channel model," IEEE Wireless Commun., **19**(6), 92–99, 2012. https://doi.org/10.1109/MWC.2012.6393523

[20] R. Verdone, A. Zanella, "Pervasive Mobile and Ambient Wireless Communications: COST Action 2100 (Signals and Communication Technology)," London: Springer, 2012.

[21] M. Zhu, G. Eriksson, F. Tufvesson, "The COST 2100 channel model: Parameterization and validation based on outdoor MIMO measurements at 300 MHz," IEEE Trans. Wireless Commun., **12**(2), 888–897, 2013. https://doi.org/10.1109/TWC.2013.010413.120620

[22] Y. Li, J. Zhang, Z. Ma, Y. Zhang, "Clustering analysis in the wireless propagation channel with a variational Gaussian mixture model," IEEE Trans. Big Data, **6**(2), 223–232, 2020. https://doi.org/10.1109/TBDATA.2018.2840696

[23] Z. Li, L. Cheong, S. Yang, K. Toh, "Simultaneous clustering and model selection: Algorithm, theory and applications," IEEE Trans. Pattern Anal. Mach. Intell., **40**(8), 1964—1978, 2018. https://doi.org/10.1109/TPAMI.2017.2739147

[24] https://sites.google.com/a/dlsu.edu.ph/jojoblanza/

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 550-554 (2020)

www.astesj.com

Second International Virtual Conference on Multidisciplinary Research 2020

ASTESJ

ISSN: 2415-6698

The Effect of Different Starches in the Environmental and Mechanical Properties of Starch Blended Bioplastics

Adriana C. Neves[*,1], Tew Ming[1], Marta Mroczkowska[1], David Culliton[2]

[1]*Department of Science and Health, Institute of Technology Carlow, R93-A003, Ireland*

[2]*Department of Aerospace, Mechanical and Electronic Engineering, Institute of Technology Carlow, R93-A003, Ireland*

ARTICLE INFO

Article history:
Received: 27 August, 2020
Accepted: 07 October, 2020
Online: 20 November, 2020

Keywords:
Bioplastic
Starch
Biomaterial

ABSTRACT

The problematic of plastic pollution in the world has led to an interest in the research and development of thermoset starch-protein blend bioplastics as a possible alternative for single use non-recyclable plastics. However, these bioplastics lack the physico-chemical characteristics that make them useful as a replacement for the currently used plastics. This work aims to assess what differences in the mechanical and environmental impact can be seen in the thermoset starch-protein blend bioplastics when different starches are used in their formulation. Rice, kuzu, corn, wheat and potato starches were used to generate bioplastics, and these were tested in terms of colour, lightness, roughness, chemical composition, moister content, water solubility and soil toxicity when degraded. Characteristics such as chemical composition, colour and moister content do not change significantly when the different starches are used, however it was possible to identify change in characteristics such as lightness, roughness, and water solubility of the bioplastics. Moreover, not only none of the bioplastics showed to be toxic to the soil when degraded, but also promoted the growth of the plant species tested. It was possible to conclude that the use of different starches in the formulation of thermoset starch-protein blend bioplastics allows the generation of bioplastics with different characteristics. This leads to an increase in applicability of theses bioplastics and consequently a higher positive impact in the environment.

1. Introduction

Plastics usage is increasing exponentially, especially in the packaging industry, which is having devastating environmental impacts, globally. Increasing levels of consumer awareness, in terms of environmental sustainability and plastics-pollution, had created a demand for environmentally-friendly bioplastics, which decompose rapidly into water, carbon dioxide and biomass [1]. However, during events such as the current global pandemic, some members of the plastics industry have used the climate of fear and uncertainty to actively suspend or roll back hard-won environmental measures to reduce plastic pollution. As a result, there is an immediate and urgent need to increase the development and usage of alternative plastics, which have higher biodegradability and are made with fully organic components [2], typically referred to as bioplastics.

The term bioplastic was originally used to describe polymers which naturally biodegrade and compost. However, the term is now erroneously used by affected industry to additionally refer to biobased but either non-compostable plastics, or biodegradable plastics which require industrial processing to fully degrade [1]. In truth, only about 1 % of the current yearly production of plastic belongs to the bioplastic category and of those, only 43.2% are fully biodegradable. Moreover, many of the fully organic bioplastics generated do not have the mechanical properties equivalent to the currently used non-compostable plastics decreasing their applicability and industrial appeal [2]. For that reason, there is a compelling need to further develop bioplastics which are compostable, and which have no negative environmental impact.

Starch blended thermoset bioplastics are promising alternatives to petrol based plastics, due to lower production costs and better

[*]Corresponding Author: Adriana C. Neves, nevesa@itcarlow.ie

degradability [3]. In addition, blending starches with other biodegradable molecules and plasticisers can expand the applicability and improve the characteristics of thermoset starch-based bioplastics to different industries [4, 5]. The increasingly diverse blend-formulations have generated bioplastics with varied, tailorable, physico-chemical and mechanical properties, combined with the desired environmental impacts [3, 6].

These bioplastics will require more sustainable starch sources, as the most widely available are typically food-negative (ie. they derive from, or replace, in terms of agricultural production, product that is a current, critical human consumable). For example, potato, wheat and maize are the most widely produced starches in Europe, with 23.6 million tonnes of these starches produced in the EU in 2017 [7]. While in the United States of America more than 50% of starch production derives from corn [8]. However, alternative, gluten free starches, such as rice and kuzu starches, are increasing in popularity in the western world. The aim of this study is to compare the physico-chemical properties of five starch blended bioplastics as well as their post-degradation environmental impact. For this study, the starches used in the production of bioplastic were rice, kuzu, corn, wheat, and potato starch.

2. Materials and Methods

2.1. Materials

Sodium alginate LB (Duchefa Biochemie B.V., Netherlands), potato starch (MP BIomedicals LLC, Netherlands), food-grade wheat starch (Foo Lung Ching Kee, Hong Kong), food-grade rice flour (Foo Lung Ching Kee, Hong Kong), food-grade kuzu starch (Clearspring Ltd., London) produced in Japan and food-grade corn starch obtained from local grocery store (Aldi, Carlow) were used to generate bioplastics. Glycerol (EMPROVE®bio, Merck KGaA, Germany) was used as plasticiser. The Phytotoxkit was purchased from MicroBio Tests (Mariakerke, Gent, Belgium).

2.2. Preparation of bioplastics

The bioplastic-forming solution was prepared by dissolving 0.5g of sodium alginate LB into 20mL deionised water at room temperature with constant stirring speed of 595 rpm for at least 30 minutes or until a homogeneous aqueous solution was formed. In a beaker, 1g of starch and 1.5mL glycerol were mixed in 10mL of deionised water and added to the alginate solution. The solution was heated at 90°C with constant stirring at low agitation to avoid air bubbles formation for 20 minutes or until complete homogenisation was observed. The prepared solution was then poured into petri dish and kept at room temperature for 72 hours to dry.

2.3. Colour and brightness

The colour and transparency were read with a colorimeter (PCE-CSMI, PCE instruments, Meschede, Germany). The L*, a* and b* values were standardized for absolute white and black as per instrument instructions and then measured from four samples of each formulation and plotted accordingly.

2.4. Roughness test

The roughness of bioplastic samples was determined using SURFCOM130A roughness tester with a TS100 roughness tester sensor. The tester was calibrated with a standard of Ra=1,54µm before preforming the test. To determine the roughness sample was placed under the sensor and the placement of the sensor needle was adjusted with a knob on the tester. The test was performed in the center of five samples of the same material.

2.5. Fourier-transform infrared (FTIR) spectroscopy

The FTIR spectra were recorded as % transmission using a Spectrum 65 FT-IR Spectrometer (Perkin Elmer, Massachusetts, United States)

2.6. Soil toxicity

The soil toxicity was assessed using the Phytotoxkit instructions. In short, the water holding capacity was determined as per the instructions on the kit and the soil was pre-treated by degrading 10g of each bioplastic in soil samples weighing 100g over a 4 weeks period with daily hydration of the soil. The test plates were filled with 100g of test soil in the lower compartment, hydrated with 10ml of deionised water, a black filter paper was placed on top of the soil and 10 seeds of the same test plant were placed on top of the filter paper. The plants species used were garden cress (*Lepidium sativum*), and mustard (*Sinapis alba*). The control soil corresponds to soil where no bioplastics were degraded. The test containers were incubated at 25ºC for 72h in darkness. The root length was measured using ImageJ software (U. S. National Institutes of Health, Bethesda, Maryland, USA).

2.7. Moister content and water solubility

According to Costa, et al. (2018), bioplastics with 2 cm of diameter were dried at 105°C for 24 hours or until ~±0.01g differences in weight were obtained. Four replicates of each formulation were operated in this study. The percentage of water lost in the system (moisture content) was calculated [9]. The samples were further used to determine water solubility based on method proposed by Costa, et al. (2012) and Fakhoury, et al. (2012). Method was slightly modified as samples are fully soluble in water after 1 hour. Instead of 24 hours, samples were immersed in 50mL deionised water for 30 minutes at low agitation (100rpm) then dried until ~±0.01g differences in weight were observed.

2.8. Statistical analysis of data

The results were analysed by one-way or two-way analysis of variance (ANOVA) with a significance level of $P = 0.05$ when appropriate. Independent sample *t*-test and Tukey's multiple comparison tests were used, where applicable. The software employed for statistical analysis was Excel 2016 (Microsoft Excel 2016 MSO, California, USA) and SPSS Statistics (IBM SPSS Statistics 2015, New York, USA).

3. Results and discussion

3.1. Preparation of bioplastics

It was possible to generate five types of thermoset starch-protein blend bioplastic samples using the five different starches investigated (Figure 1). Bioplastics generated with different starches have been widely reported in literature, however, no published studies could be found which compared bioplastic generated from kuzu starch with bioplastics generated from rice,

wheat, corn, and potato starch. The overall appearance of the kuzu-bioplastics was similar to other thermoset bioplastics reported in literature, with no detected starch-related plasticity differences. This was attributed to the same amount and type of plasticiser being used. This plasticiser component, used in the investigated formulations, has been reported to be the critical component in the plasticity of the final product, with no starch - bioplastic plasticity nexus observed [6, 10].

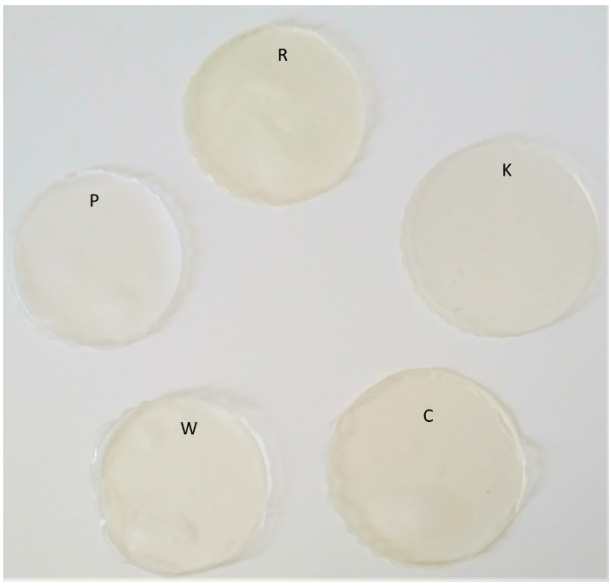


Figure 1: Thermoset starch blend bioplastics generated using rice (R), kuzu (K), corn (C), wheat (W) and potato (P) starches.

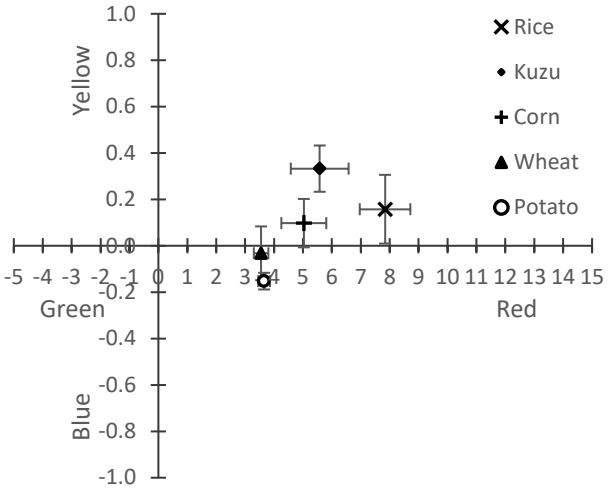


Figure 2:Colour a*; b* values of the bioplastics generated with rice, kuzu, corn, wheat, and potato starches. N=5 ± sd.

3.2. Colour and brightness

Natural based bioplastics often display a yellow tone, though the addition of chemicals, or exposure to oxidation or bleaching processes, have been shown to produce a brighter and whiter tone. Unfortunately the effect of these processes on the environment has yet to be assessed [11, 12]. Alternatively, this intonation can be also affected by starch type, which is a more environmentally friendly way of generating brighter bioplastics with less of a yellow tone. For example, the colour and lightness analysis in this research demonstrated that bioplastics generated with rice, kuzu, and corn, have greater red/yellow components when compared to bioplastics generated from wheat and potato starches. Additionally, visual examination of the wheat- and potato- starch bioplastics indicated a less yellow tone (Figure 1). Moreover, the rice starch derived bioplastic had a significant lower brightness then all the other bioplastic tested (Figure 2). That might be linked with the higher red component in its colour. However, all of the bioplastics generated were considered to be white, according to the instrument guidelines, with the higher values corresponding to the rice derived bioplastic (L*90.73± 1.12; a*0.16± 0.15; b*7.84± 0.88).

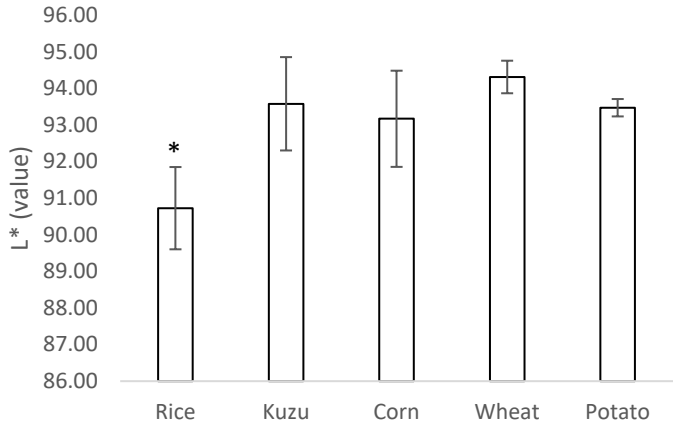


Figure 3: Brightness L* values of the bioplastics generated with rice, kuzu, corn, wheat, and potato starches. N=5 ± sd with * representing significantly different value.

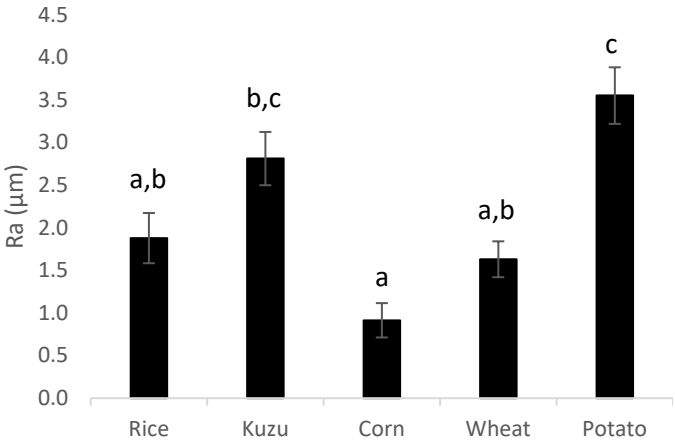


Figure 4: Roughness values of the surface in contact with air of thermoset starch-blend bioplastics generated with rice, kuzu, corn, wheat and potato starches. N=5 ± sd with different lower-case letters representing significantly different values.

3.3. Roughness Test

Polymer samples generated in the reported research were formed using flat moulds. As a result, one surface of the thermoset bioplastic films replicated the roughness of the mould. However, the surface in contact with the air acquired a natural texture. The texture of that surface was tested, and it was observed that the

different starches used in the bioplastic formulation produced different surface roughness (Figure 4). The starches that lead to higher roughness were Potato and Kuzu with 3.55 ± 0.33 and 2.81 ± 0.31 Ra (μm) respectively and the starches that lead to smoother surface roughness are Wheat and Corn with 1.63 ± 0.21 and 0.92 ± 0.20 Ra (μm) respectively.

3.4. Fourier-transform infrared (FTIR) spectroscopy

The purpose of FTIR analysis was to identify the presence of chemical bonds that were formed in bioplastics using the different starches. Figure 5 shows that the spectrums of all the bioplastics tested have similar profiles displaying the presence of O-H, C-H, C=O and C-O absorption peaks. This might be because although structurally the starches might have slightly different structures, their chemical composition is, overall, similar. Moreover, the profile of the bioplastics show similar chemical bonds to others starch-blend bioplastics reported in literature [13, 10, 11].

3.5. Soil toxicity

The effect of the degradation of the bioplastics with each starch in the growth of two plant species was tested with a soil toxicity test (Figure 6). The results show that there was no decrease in the growth of either *Lepidium sativum* or *Sinapis alba* after the degradation of any of the bioplastics tested in their growth soil. Contrarily, the bioplastics generated with kuzu, corn, wheat and potato starch showed a significant increase of growth in both plant species. The increase in plant growth in the soil after bioplastic degradation compared with the control soil might be related with the increase of Carbon (C) added to the soil after the bioplastics degradation. The addition of C to poor soils has been used for soil remediation showing to improve plant growth and crop productivity [14, 15]. The degradation of bioplastics with kuzu, corn, wheat or potato starches in carbon poor soils has the potential to improve their productivity.

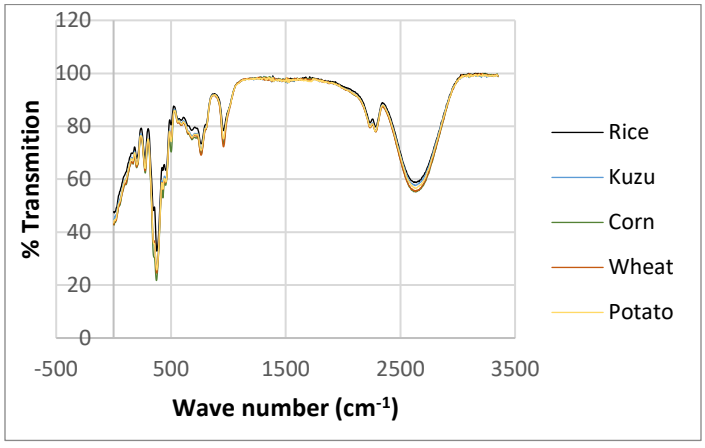


Figure 5: Fourier transform infrared (FTIR) spectrums of the bioplastics generated with rice, kuzu, corn, wheat, and potato starch.

3.6. Moister content and water solubility

The moisture content of bioplastics can influence properties such as their compostability, elasticity, hydrophobicity among others and consequently influence their applicability [16]. For that reason, small changes in composition of the starch-based bioplastics might influence the time in which it fully composts. Composting time is of high importance for the applicability of bioplastics [2].

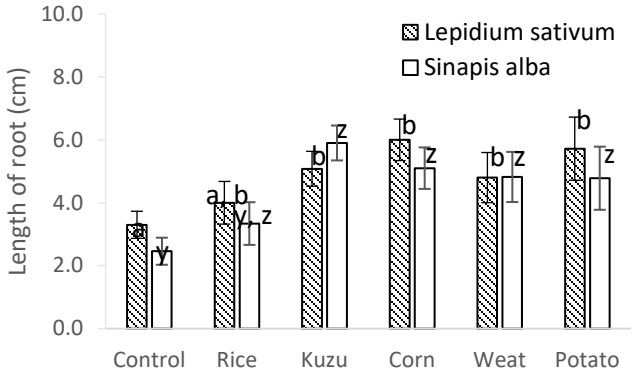


Figure 6: Growth of Lepidium sativum and Sinapis alba in soil where bioplastics with rice, kuzu, corn, wheat and potato starch were degraded. No bioplastic was degraded in the control soil prior to seeding. N=10 ± sd with different lower-case letters representing significantly different values.

Using the different starches tested in the bioplastic formulation does not seem to lead to significantly different moisture contents of the bioplastics, with all the formulations ranging from 41 ± 7 to 49 ± 4 % moisture content in the rice and corn based bioplastic respectively (Figure 7).

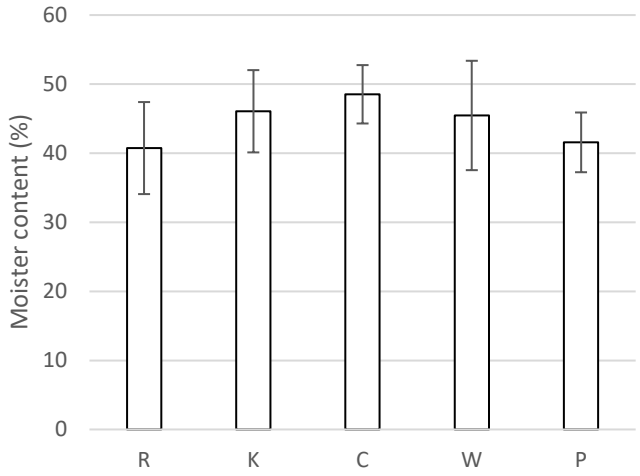


Figure 7: Moister content of thermoset starch-blend bioplastics generated with rice (R), kuzu (K), corn (C), wheat (W) and potato (P) starches. N=5 ± sd.

Similarly, understanding the effect of different starches in the solubility of the bioplastic is not only relevant industrially to assess potential applications of bioplastics but also needed to understand the impact of the bioplastics in the environment [2]. The solubility of the bioplastics tested ranged from 35 ± 6 to 81 ± 1% with the bioplastic sample generated with Kuzu being the least soluble and the one generated with wheat the most soluble. Moreover, there was not a significant difference found in the samples generated with potato, corn and rice starch or between the samples generated with Rice and Kuzu. Previous studies showed the link between starch concentration and solubility of the bioplastics [6], however, no previous studies have looked into a direct comparison of solubility of bioplastics when different starches are used in their formulation.

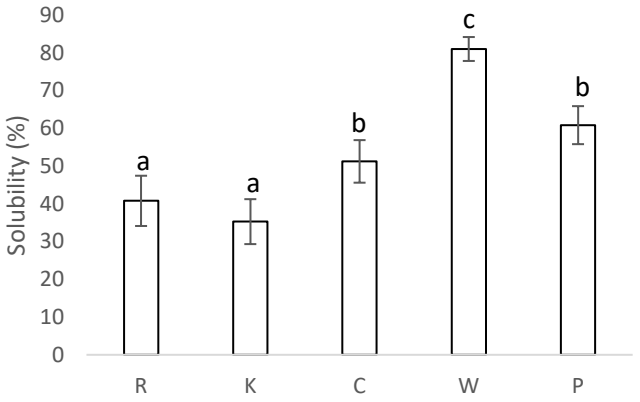


Figure 8: Solubility of thermoset starch-blend bioplastics in deionised water generated with rice (R), kuzu (K), corn (C), wheat (W) and potato (P) starches. N=5 ± sd with different lower-case letters representing significantly different values.

4. Conclusion

The generation of thermoset starch blend bioplastics using the five different starches tested lead to bioplastics that did not cause any toxicity to the growth of the plant species *Lepidium sativum* and *Sinapis alba* when degraded in the soil where they grew. Moreover, in all cases, except the rice-based bioplastic, their degradation lead to a significant increase of growth of these plant species.

The use of different starches led to the generation of bioplastics with similar FTIR profiles, as well as colour and water content. However, other characteristics differed depending on which starch was used for their formulation. Higher solubility was obtained when wheat starch was used. In opposition to rice and kuzu starches that can be used to obtain bioplastics with lower solubility rates.

Although in terms of colour all the bioplastics generated had a near-white colour, the brightness of these were dependent on the type of starch used. The rice, kuzu and corn starches lead to bioplastics with more red/yellow tones with rice derived bioplastic being the only one significantly different in terms of brightness.

In terms of roughness it was also possible to observe a difference between potato and kuzu based bioplastics with rougher surfaces in comparison with wheat and corn with smoother surfaces.

The results of this study show that there is a link between starch in the bioplastic composition and some of the properties tested. This is of high importance to industry when choosing formulations for the generation of organic thermoset bioplastics for different purposes.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. Misra, V. Nagarajan, J. Reddy and A. Mohanty, "Bioplastics and green composites from renewable resources: where we are and future directions.," 2011. In18th International Conference on Composite Materials 2011 August.

[2] A. Neves, M. Moyne, C. Eyre and B. Casey, "Acceptability and Societal Impact of the Introduction of Bioplastics as Novel Environmentally Friendly Packaging Materials in Ireland.," Clean Technologies, **1**, 127-143., 2, 2020. doi: https://doi.org/10.3390/cleantechnol2010009

[3] M. Rahmatiah Al Faruqy and K. Liew, Properties of Bioplastic Sheets Made from Different Types of Starch Incorporated With Recycled Newspaper Pulp, Journal of Transaction of Science and Technology. **3**:257-264, 2016..

[4] N. Azahari, N. Othman and H. Ismail, "Biodegradation studies of polyvinyl alcohol/corn starch blend films in solid and solution media.," Journal of Physical Science, **2**(22), 15-31, 2011.

[5] K. Satyanarayana, G. Arizaga and F. Wypych, "Biodegradable composites based on lignocellulosic fibers—An overview," Progress in polymer science, **9**(34), 982-1021, 2009.

[6] R. Santana, R. Bonomo, O. Gandolfi, L. Rodrigues, L. Santos, A. dos Santos Pires, C. de Oliveira, R. Fontan and C. Veloso, "Characterization of starch-based bioplastics from jackfruit seed plasticized with glycerol.," Journal of food science and technology, **1**(55), 278-286., 2018. https://link.springer.com/article/10.1007/s13197-017-2936-6

[7] Starch Europe, "Starch.EU," 2017. https://starch.eu/the-european-starch-industry/#data.

[8] Grand view research , "North America Organic Starch Market Analysis, By End-Use (Meat, Bakery, Confectionery and Dry Blends), By Ingredient (Corn, Potato, wheat), By Country, And Segment Forecasts, 2018 - 2025," 2017. https://www.grandviewresearch.com/industry-

[9] M. J. Costa, A. M. Marques, L. M. Pastrana, J. A. Teixeira, S. M. Sillankorva and M. A. Cerqueira, "Physicochemical properties of alginate-based films: Effect of ionic," Food Hydrocolloids, **81**, 442-448, March 2018. doi: https://doi.org/10.1016/j.tplants.2018.11.010

[10] N. Ramakrishnan, S. Sharma, A. Gupta and B. Alashwal, "Keratin based bioplastic film from chicken feathers and its characterization.," International journal of biological macromolecules, **111**, 352-358., 2018.

[11] A. Cifriadi, T. Panji, N. Wibowo and K. Syamsu, "Bioplastic production from cellulose of oil palm empty fruit bunch.," In IOP Conference Series: Earth and Environmental Science, **65**(1), 012011, 2017.

[12] A. Low, "Decoloured bloodmeal based bioplastic," (Doctoral dissertation, University of Waikato) Waikato, 2012.

[13] M. Ginting, M. Kristiani and Y. Siagian, "The effect of chitosan, sorbitol, and heating temperature bioplastic solution on mechanical properties of bioplastic from durian seed starch (Durio zibehinus)," Princ. author al. International Journal of Engineering Research and Application, **1**(6), 33-38, 2016.

[14] X. Zhu, B. Chen, L. Zhu and B. Xing, "Effects and mechanisms of biochar-microbe interactions in soil improvement and pollution remediation: a review.," Environmental Pollution, **227**, 98-115, 2017. doi: https://doi.org/10.1016/j.envpol.2017.04.032

[15] R. Lalande, V. Furlan, D. Angers and G. Lemieux, "Soil improvement following addition of chipped wood from twigs.," American journal of alternative agriculture, **13**(3), 132-137., 1998. https://www.jstor.org/stable/44503064

[16] H. Karan, C. Funk, M. Grabert, M. Oey and B. Hankamer, "Green bioplastics as part of a circular bioeconomy.," Trends in plant science,, **24**(3), 237-249, 2019. doi: https://doi.org/10.1016/j.tplants.2018.11.010

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 555-560 (2020)

www.astesj.com

Second International Virtual Conference on Multidisciplinary Research 2020

ASTESJ

ISSN: 2415-6698

Proposal of a New Descriptive-Correlational Model of Population Lifestyle Analysis and Disease Diagnosis

Selene Tamayo Castro[1,*], Kristian Aldapa Salcido[1], Linda García Rodríguez[2]

[1]*Business Planning and Regional Development, Tecnológico Nacional de México, Instituto/I.T. de Los Mochis, 81200, México*

[2]*Strategic Planning and Technology Management, Tecnológico Nacional de México, Instituto/ I.T. de Los Mochis, 81200, México*

ARTICLE INFO

Article history:
Received: 10 September, 2020
Accepted: 07 October, 2020
Online: 20 November, 2020

Keywords:
Strategic planning
Disruptive innovation
Lifestyle
Chronic diseases

ABSTRACT

This document proposes a new methodology for lifestyle analysis and disease diagnosis for young academics using a strategic planning and disruptive innovation approach. Its objective is to consider and study a new form of treatment to improve the quality of life and health of students in parallel with their development, this through a quantitative methodology with a descriptive-correlational scope, being possible to develop a diagnosis and analysis, resulting A high relationship between variables is evident, thus developing a preliminary model based on engineering in the process of improving lifestyles through the Deming cycle and based on the S curve of innovation, hand in hand with technology and motivation.

1. Introduction

Changes in lifestyle in recent years have negatively affected health, mainly in low and middle developed countries [1, 2, 3], poor quality of diet, sedentary lifestyle, and emotional stress can cause inflammation, oxidative stress, sympathetic activation, and contribute to obesity, together these factors can trigger or exacerbate hypertension, diabetes, dyslipidemia, and sleep-disordered breathing, which can lead to atherosclerosis and ultimately heart failure [4]. That is why investigations such as that of Gómez, Roca, Gili, García, and García [5], analyze that in a depressive scenario, an incorrect lifestyle is present; In addition, depression causes sleep disorders, loss of energy, low level of physical activity and with this increases the desire to eat junk food.

Thus, unhealthy lifestyles lead to the development of chronic diseases (CD) and premature death. 70% of deaths worldwide are caused by CD [2]; North America and Latin America are among the first places worldwide in factors that promote CD [1, 2, 3]. Although CDs impact people, it has consequences on the economic development and human capital of countries, mainly low-income countries, ranging from productivity to the costs generated by their treatment [6]

Mexico in particular is in the first places in obesity, overweight and diabetes worldwide and among the OECD countries [7], Sinaloa, Mexico is aligned to national statistics with 70% obesity and overweight in its population with age equal to or greater than 20 years, and Los Mochis, Sinaloa with a 72.5% prevalence of obesity and overweight, this factor being the cornerstone of the development of chronic diseases of greater affectation; The need and lack of effective guidance are evident, one that not only tries to raise awareness but is also effective enough to achieve positive change on a large scale.

There is a wide variety of studies that support the positive impact of exercise, good nutrition, adequate sleep schedules and quality of sleep, as well as the importance of the environment on both physical and mental health [1-5, 7], however, there is a gap in information regarding the application of these factors as a whole, even more so the implementation of strategic planning and disruptive innovation in the improvement of lifestyles.

For all the above, the objective of this article is the diagnosis and analysis of the lifestyle, chronic diseases, strategic planning, and disruptive innovation of young university students from the private sector, for the development of a disruptive strategic model.

The essence of the proposed solution of the problem in question is the diagnosis of the needs of the population and therefore the development of a model with planning and disruptive innovation applied to the prevention and control of chronic diseases, such as overweight, obesity, diabetes, and cardiovascular diseases, which allows young students to have an adequate balance between physical and mental health, helping them to have a better

*Corresponding Author: Selene Tamayo Castro, selene.tamayo@hotmail.com

academic and work performance, but above all, a better quality of life.

2. Theoretical foundations

2.1. Strategic model

According to the Royal Spanish Academy, the word model, coming from Italian, is an archetype or point of reference to imitate or reproduce it, differently, the strategy, the "Latin strategĭa" is the art, traces to direct a matter [8].

On the other hand, strategy, coming from the war trade, is defined as the art of drawing to direct an issue, in an adjustable process, a set of rules that ensure an optimal decision at all times [8].

Deming Cycle, the cycle of continuous improvement for development, through the factors: plan, do, verify, and act that incorporated feedback and observation of uninterrupted practice [9].

2.2. Disruptive innovation

Disruptive innovation was defined for the first time by Christensen in 1997 as a process by which a product or service is brought to the market, through simple applications, to later win that market by displacing other competitors. The contribution of Master Christensen leads to disruptive innovation to break the status quo of companies or any industry. Likewise, Sartor in 2017 commits that disruptive technology is any that can deeply affect a sector of activity and its value chain; Disruptive innovation can be implemented in three areas, such as: disruptors in the improvement of a process, a change of product or business model [10].

2.2.1. S-Curve pattern of innovation

The model S-Curve of innovation Figure 1, helps to evaluate in which stage of the life cycle the technology is and provides an indicator of the maturity of the industry. Method used to locate what has been the behavior of a certain trend in the past and extrapolate it to the possible future as a guide.

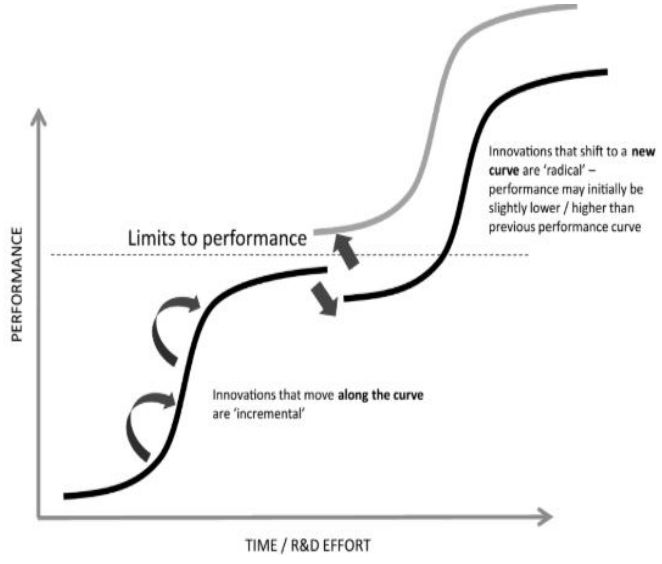


Figure 1: S-Curve of innovation

2.3. Lifestyle

Lifestyle is the set of behaviors that a specific individual puts into practice in a consistent and sustained way in their daily life and that can influence the maintenance of their health or places them at risk for disease: inadequate diet, activity poor physics, disturbed sleep-wake cycles, excessive consumption of substances and toxic habits, poor stress management, etc., are expressed in specific individual behaviors that are interpreted as risk factors, or behaviors that are shared by a group of people in a specific context [11].

Describing the sociodemographic, psychosocial, quality of life and educational variables in young university students and their relationship between them, as well as the ideal diet and habits based on the variables mentioned, are essential in a healthy lifestyle [12].

2.4. Chronic diseases

Chronic diseases (CD) account for new epidemics in developed societies, being link through conditions and lifestyles [3]

In Latin America, the main chronic diseases are obesity, cardiovascular diseases, cancer, chronic respiratory diseases, diabetes and arthritis are the main ones; Along the same lines, in Mexico the most prevalent chronic diseases are: obesity, arterial hypertension, diabetes, metabolic syndrome, fatty liver, chronic kidney disease [1].

3. Methodology

The present study was approached with a methodology focused on Hernández and Mendoza [13], being a quantitative investigation, with a scope according to the level of knowledge obtained with said investigation: descriptive - correlational.in the paper.

3.1. Sample selection

The first step focused on determining the population, defining as population young people from private universities located in Los Mochis, Sinaloa, 3,898 young people according to figures from the interactive educational statistics consultation program by the Ministry of Public Education.

To determine the size of the sample, the STATS 2.0 program was used, obtaining the amount of 350 as the sample size, results shown in Table 1. The sample is considered simple random probabilistic, since it can be answered by any student who is in the unit of analysis. The unit of analysis was represented by university students from the private sector, located in Los Mochis, Sinaloa.

Table 1: Characteristics of the sample

Characteristics	Number value/ percentage
Universe size	3,898
Maximum acceptable error	5%
Estimated percentage of the sample	50%
Desired level of confidence	95%
Sample size	350

3.2. Instrument design

A quantitative measurement instrument was applied through the use of a structured questionnaire. The information collected was explicitly and standardized. The questionnaire used was of our elaboration. On the other hand, the degree of validity was determined through content and construct, the first based on impact literature, and the second through Pearson's correlation.

The scale used in the questionnaire was of the Likert type [14], made up of 5 points, which range from totally agree to totally disagree, determining its degree of reliability using Cronbach's Alpha.

The instrument consists of a total of 65 items divided by variables and dimensions. It should be noted that the degree of reliability of the instrument was measured through the application of the questionnaire, whose results were tabulated, and the calculation of the degree of reliability was generated using Cronbach's Alpha, using SPSS in Version 27, resulting 0.944, consequently demonstrating its excellent reliability to be applied.

Structured questionnaire in 5 sections, as follows:

- Section 1, diagnosis to identify the respondent's framework, 5 items.
- Section 2, lifestyle (EV), 13 items.
- Section 3, chronic diseases (CD), 16 items.
- Section 4, disruptive innovation (ID), 9 items.
- Section 5, strategic planning (SP), 22 items.

The scale used in the questionnaire consists of 5 points that go from totally disagree (1) to totally agree (5), as shown in Table 2.

Table 2: Instrument coding

Value	Meaning
1	Strongly disagree
2	Disagree
3	Neither agree nor disagree
4	Agree
5	Strongly agree

It should be noted that two of the items are open questions.

4. Hypothesis

Applying strategic planning and innovation to the model potentiates the improvement of lifestyles from the health perspective of young university students.

5. Results

5.1. Descriptive results

Below are the results obtained from the structured survey, 350 responses received per application to the unit of analysis; 75% of the respondents were women (264), the remaining 25% were men (86).

Other research results are introduced below; The information is of a descriptive statistical nature of the set of variables extracted from the questionnaire, which allows us to know some characteristics of the sample.

Table 3: Responses with the most impact in affirmation regarding the strategic planning block

Strategic planning	Percentage
	Strongly agree/ agree
I am interested in an initiative that motivates young people in the vocational and professional education stage to lead a healthier life.	86.86%
An initiative that aims to be a leader in the country in terms of reducing chronic diseases and improving lifestyles interests me.	87.43%
I need to be advised by experts in nutrition, physical and psychological health, who offer me plans that adapt to my objectives and goals.	76.57%
Including innovation and creativity in my lifestyle enhances my development	84.57%
A clean and organized infrastructure format makes it easier for me to choose whether or not to join the initiative either online or in person.	76.57%
Having employees trained and frequently evaluated in quality aspects, leadership fosters excellent services.	89.14%
Constant updating of specialists to know and implement the most innovative and innovative work methods of the organization.	86.86%
It is desirable to find clean and tidy areas when attending training.	91.43%
Having certifications in health and organizational hygiene generates a substantial attraction to the initiative.	87.43%
Knowing that employees are properly compensated let me know that I am part of an organization with values.	84.57%
Having a collaborative environment within the organization generates security for staff.	90.29%
A safe work environment encourages a diversity of thoughts, thus generating creative ideas.	91.43%
Having teams for collaboration motivates a mindset of social contribution.	85.71%
Knowing the strengths and weaknesses of the organization allows us to foresee errors or failures of the plans, under or to be developed.	88.57%
In the application of high-level programs, the value of the organization is perceived.	86.86%
Establishing controls and administration generates flexibility to adapt to the organization.	86.29%
Offering a periodic evaluation to collaborators allows maintaining constant development and verification of alignment to the objectives.	85.14%
Having relationship programs with different communities generates cultural enrichment, therefore a range of ideas for the development of new training and motivation methods.	90.86%
Establishing responsibilities and objectives for each position correctly empowers each employee in their area of expertise.	89.71%

The table above represents the frequencies and percentages of the responses with an opinion of "Totally agree and agree" it could be observed in the first instance, referring to the strategic planning block, that there is a statement by 90% of the participants regarding interest for belonging to an organization or initiative with a safe work environment, which is aligned with the generation of ideas and offering excellent customer service, in turn, they claim to want a clean physical and online format, constant evaluation, and training of collaborators with the most novel and innovative in order to apply it to work methods and improvement of results, this hand in hand with relationship programs with various communities

generating a range of ideas for the development of new methods of training and motivation, within a collaborative work environment, generating in turn security to staff.

Regarding the block of disruptive innovation Table-4, 90.86% of the participants in the survey affirm that a change in lifestyle is possible through smart spaces and health monitoring, hand in hand with the accessibility of support from experts and price range for a greater reach and impact, in turn, more than 80% affirm their interest in prevention and control computer sessions in terms of the health approach and the extension of these extensive advantages both to face-to-face format and online (in- place / at-home).

Table 4: Responses with the most impact weight in affirmation regarding the disruptive innovation block

Disruptive innovation	Percentage
	Strongly agree/ agree
I would like to be part of an organization with new ideas to achieve my physical and psychological goals.	72.00%
Smart spaces and health monitoring allow generating a lifestyle change.	90.86%
In-place/at-home training products or features create an extension of possibilities for club/gym members	74.86%
Having the possibility of improving my lifestyle in an accessible way allows me to prevent chronic diseases.	90.86%
Having options for different ranges opens the possibilities for a substantial number of people to join the initiative to improve their lifestyle.	82.29%
I feel identified with an organization that sees the human being as a whole and enhances their development from the psychological, physical, and spiritual approach.	72.57%
It is of interest to me to have at my disposal computer sessions on prevention and health control.	80.57%

From the lifestyle block, it is resolved that the participants recognize there is a relationship between mind and body to achieve a state of well-being, 90% identify the consequences of an unhealthy lifestyle, however, only 56% affirm taking care of what eats and schedules, in addition to this 40% do not achieve their training objectives and consistency in them, on the other hand, there is a high rate of interest in doing physical activity (76%), their interest is liable to the training of various or innovative disciplines methods that motivate them to achieve goals and improve their lifestyle.

Table 5: Responses with the most impact weight in affirmation regarding the lifestyle block

Lifestyle	Percentage
	Strongly agree/ agree
I consider that a healthy mind inhabits a healthy body.	89.14%
I identify the consequences of not maintaining a healthy lifestyle.	90.86%
I am interested in doing physical activity that improves my quality of life.	76.57%
I identify the relationship between my mood and my diet.	73.71%
Recognition of my efforts and achievements through rewards generates motivation to leave my comfort zone.	72.57%

In Figure 2, the percentages of those participants who attend the gym, club, or classes for physical activity, and those who do not attend are presented, which is the main reason.

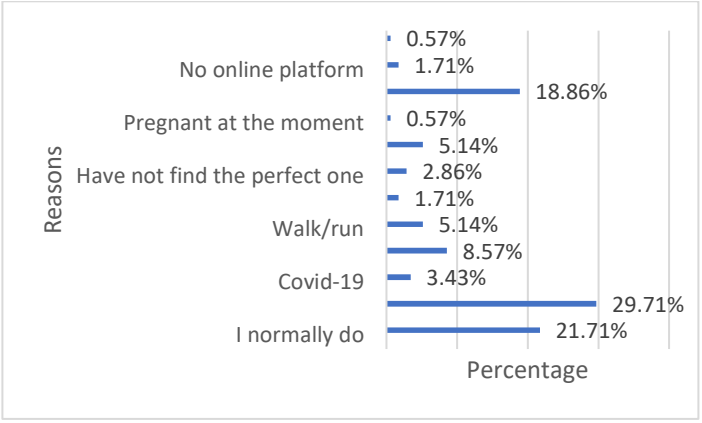


Figure 2: Percentages question EV-13

Regarding the chronic diseases block, the participants claim to be informed about the impact of obesity and overweight on their wellbeing, and later its reflection in a variety of chronic diseases, 92% consider it important to be informed of news and innovations against chronic diseases and maintain a health monitoring at least once a year.

Table 6: Responses with the most impact in affirmation regarding the block, chronic diseases

Chronic dieases	Percentage
	Strongly agree/ agree
It is my knowledge that chronic diseases are generally slow to develop, therefore my daily habits make a big difference.	82.86%
I consider it important to be informed of news and innovations concerning chronic diseases.	92.00%
Those who are overweight or obese have a higher risk of developing chronic diseases.	93.14%
I am informed of the causes and consequences of being overweight or obese.	93.71%
I consider it important to carry out at least one annual health check-up.	86.29%

It is important to highlight that the results of the open question CD-16 are presented in Figure 3; Of the 350 participants 94.4% affirm that one or more of their relatives suffer from one or more chronic diseases; it is important to highlight that this question had the possibility of selecting multiple answers; 590 answers were received, around 40% of the participants selected more than one chronic disease, placing diabetes in the first place, this information is of great value since those relatives of patients with type 2 diabetes mellitus tend to suffer from hypertension and overweight, the last one being the most recurrent in 63% of cases, providing strong signals of major diabetes awareness gaps among family members of patients living with type 2 diabetes mellitus [13].

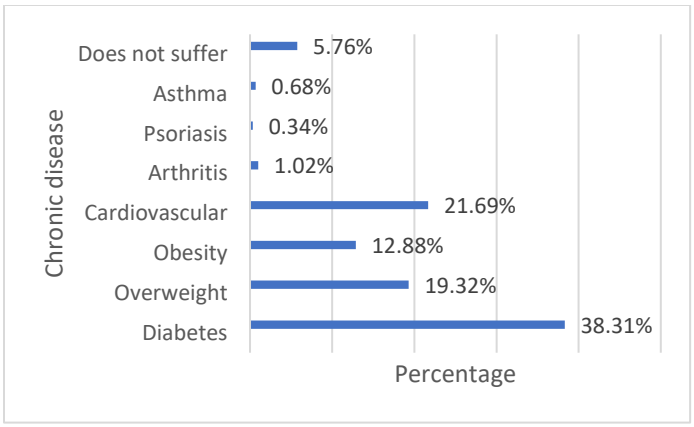


Figure 3: Percentages question CD-16

5.2. Correlational results

The reason for considering the correlation coefficient was to identify the strength of the relationship between variables strategic planning, disruptive innovation, lifestyle, and chronic diseases, in the positive sense; The results of Pearson's correlation are presented below in Table 7.

Table 7: Correlation of variables

		VI	VD1	VD2	VD3
VI	Correlación de Pearson	1	,502**	,592**	,550**
	Sig. (bilateral)		,000	,000	,000
	N	350	350	350	350
VD 1	Correlación de Pearson	,502**	1	,557**	,559**
	Sig. (bilateral)	,000		,000	,000
	N	350	350	350	350
VD 2	Correlación de Pearson	,592**	,557**	1	,783**
	Sig. (bilateral)	,000	,000		,000
	N	350	350	350	350
VD 3	Correlación de Pearson	,550**	,559**	,783**	1
	Sig. (bilateral)	,000	,000	,000	
	N	350	350	350	350

Once the correlation between variables is carried out, the result is a moderate to high correlation, according to Tables 7 and 8.

Table 8: Null-Correlation of variables

Value	Meaning
0	Null correlation
0.01 a 0.19	Very low positive correlation
0.2 a 0.39	Low positive correlation
0.4 a 0.69	Moderate positive correlation
0.7 a 0.89	High positive correlation
0.9 a 0.99	Very high positive correlation
1	Large and perfect positive correlation

Although the literature affirms the relationship between good habits and their impact on lifestyle improvement, the correlation found between innovation and strategic planning is key to promoting said improvement. It should be noted that university students do give importance to lifestyle care and chronic diseases, besides, they affirm feeling related to a well-established company and an excellent work environment, promoting development in the personal and social development of well-being, focused on their lifestyle improvement.

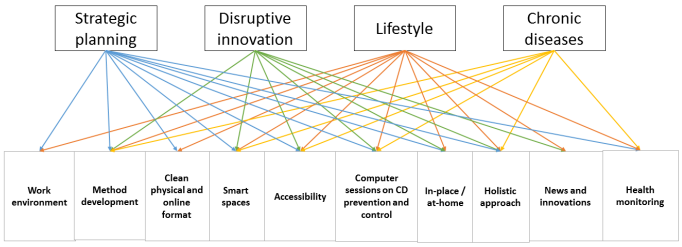


Figure 4: Correlation

Having said the above, what San Román indicates is confirmed, by saying that describing the sociodemographic, psychosocial, quality of life, and educational variables in young university students and their relationship between them, as well as diet and suitable habits, are essential in a healthy lifestyle [12].

6. Disruptive strategic model

After applying the instrument and collecting the data, a model was structured, which offers a holistic service for the detection of defects in the way in which the client interacts with the environment, causing health problems. After establishing a diagnosis, the strategic plan confirms that it generates the discipline to produce individual patterns that can be monitored by specialists, either in person or online in their respective areas, such as psychology, nutrition, and physical preparation, providing constant updating, training, and study of the innovations in their main areas of expertise.

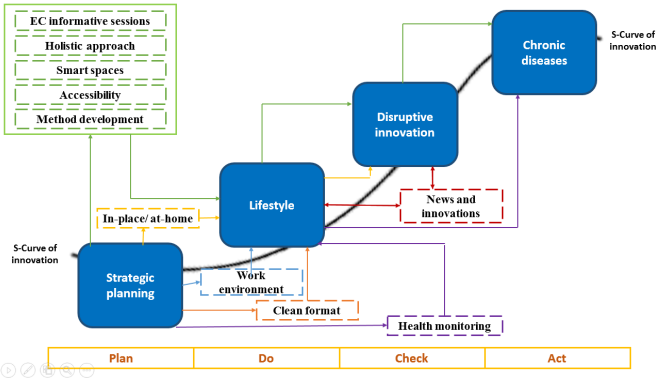


Figure 5: Correlation of variables model

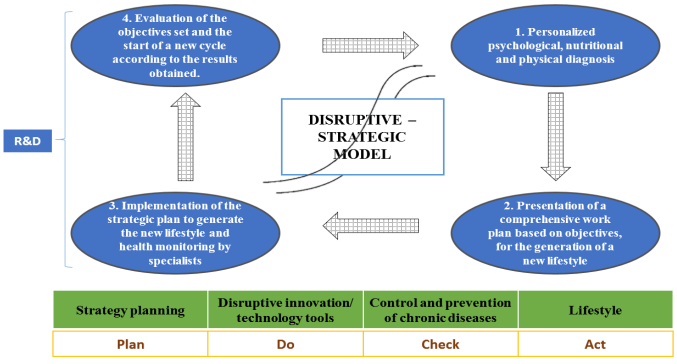


Figure 6: Model

Figures 5 and 6, presents the proposed model based on the diagnosis and correlation previously carried out, implementing both the Deming cycle and the "S" of innovation throughout the "Strategic-disruptive model", in which those factors are present of greater correlation and need on the part of university students in terms of organizational structure, lifestyle, innovation and chronic diseases, generating a sequence of uninterrupted continuous improvement.

7. Discussion

An inclination for a clean format has been found, a collaborative environment which encourages security when generating ideas and proposing models or work methods based on research and innovation, this from the planning point of view and referring to the collaborators; on the other hand, inclusive programs, both speaking of cultures, as well as financial accessibility and outreach, highlighting the interest in staying informed of news and innovations in terms of lifestyle and chronic diseases, added to this a high rate of interest in stay motivated and generate consistency in lifestyle improvement; the above treating the human being from a holistic approach.

Aligned with Anderson and Fowers, it is confirmed that lifestyle behaviors, related to physical and mental health, generate the well-known hedonic well-being, which is based on satisfaction with life, positive emotions, and low negative emotions; that is, well-being and social interaction are intimately related, affecting the lifestyle of individuals, those who are more sociable tend to perform more physical activity and mental stability, they point out [15].

Likewise, what Leyton, Batista, and Castuera [6] analyze, the effects that motivational processes have on the adoption of lifestyle habits that enhance health in adolescents are confirmed; Describing the sociodemographic, psychosocial, quality of life and educational variables in young university students and their relationship between them, as well as the ideal diet and habits based on the aforementioned variables, are essential in a healthy lifestyle [7].

8. Conclusions

The objective of the study in this first phase of the research was to diagnose and generate a preliminary model with strategic planning and disruptive innovation applied to the prevention and control of chronic diseases for university students from the private sector.

The diagnostic and correlation study was carried out in the Los Mochis area, Sinaloa; The limitations of the study rely on reduced access to students due to the ongoing pandemic Covid 19, it is suggested for future studies to evaluate a greater number of areas, to measure the possibility of replication of the model in various geographical areas and/or its adaptability, either within or outside the country, extending the structured questionnaire referring to the current routine of the participants, current lifestyle, objectives, and the reason for them.

The strengths of the study lie in the generation of information regarding the needs of university students in the face of lifestyle improvement and the recognition of the correlation of the variables, this among all the four variables under study.

The hypothesis, applying strategic planning and innovation to the model, potentiates the improvement of lifestyles from the focus of youth health, it is demonstrated in terms of literature and diagnosis.

It is relevant to note that the result of this research is a preliminary model, which gives the guideline to develop the following phases of the present study, to implement and evaluate the final model, in consequence allowing a high rate of improvement in lifestyle at an early age in a broad social spectrum, positively impacting personal and social development at the moment and in the long term, shaping a more productive society.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] B. Legetic, A. Medici, M. Hernández-Ávila, G. Alleyne, A. Hennis, Economic Dimensions of Noncommunicable Diseases in Latin America and the Caribbean, Pan American Health Organization, 2016.

[2] M. H Forouzanfar, A. Afshin, L. T Alexander, H R. Anderson, Z. A Bhutta, S. Biryukov, M. Brauer, R. Burnett, K. Cercy, F. J Charlson, et al. ., “Global, regional, and national comparative risk assessment of 79 behavioral, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015” The Lancet, **388**(10053), 1659-1724, 2016. doi.org/10.1016/S0140-6736(16)31679-8

[3] J. M. Mayoral, N. A. Sanz, P. Godoyc, M. J. Sierra, R. Cano, F. González, Á. Pousa, "Chronic diseases as a priority for public health surveillance in Spain" Gac Sanit, **30**(2), 2016. doi.org/10.1016/j.gaceta.2015.12.008

[4] M. Aggarwal, B. Bozkurt, G. Panjrath, B. Aggarwal, RJ Ostfeld, ND Barnard, H. Gaggin, AM Freeman, K. Allen, S. Madan, et al., "Lifestyle Modifications for Preventing and Treating Heart Failure ”Journal of the American College of Cardiology, **72**(19), 2018. DOI: 10.1016 / j.jacc.2018.08.2160

[5] R. Gómez, M. Roca, M. Gili, J. García, M. García, "Healthy lifestyle: an undervalued protection factor against depression" Biological Psychiatry, vol. 24, Issue 3, 2017, https://doi.org/10.1016/j.psiq.2017.10.004

[6] M. Leyton, M. Batista, R. Jiménez, "Prediction model of healthy lifestyles through the Theory of Self-determination of Physical Education students" Revista de Psicodidáctica, **25**(1), 2020. doi: 10.1016 / j.psicod.2019.05.002

[7] F. Torres, A. Rojas, "Obesity and public health in Mexico: the transformation of the hegemonic pattern of food supply-demand", Problems of development, **49**(193), 2018. doi.org/10.22201/ iiec.20078951e.2018.193.63185

[8] Royal Spanish Academy, "Dictionary of the Spanish language", edition 23, 2019.

[9] H. Schmidt, "Explosive precursor safety: An application of the Deming Cycle for continuous improvement," American Chemical Society, **26**, 31-36, 2019. doi.org/10.1016/j.jchas.2018.09.005

[10] M. Borghino, Disruption: Beyond innovation, Grijalbo, 2018.

[11] J. Grau, “Non-communicable chronic diseases: an approach from psychosocial factors”, Salud & Sociedad, **7**(2), 2016. https://www.redalyc.org/pdf/4397/439747576002.pdf

[12] S. Roman, "Analysis of the profile of the university student in relation to healthy habits and lifestyles, motivations and socio-educational expectations", Doctoral thesis in education, University of Granada, 2018.

[13] R. Hernández-Sampieri, C. Mendoza, Research Methodology. The quantitative, qualitative, and mixed routes, Mc Graw Hill Education, 714, 2018.

[14] N. Urrunaga, JE Montoya, J. Jaime, M. Moscoso, MK Cárdenas, F. Diez, RH Gilman, A. Bernabe, “Attitudes, healthy lifestyle behaviors and cardiometabolic risk factors” Primary Care Diabetes, 2020. doi.org/10.1016/j.pcd.2020.07.003

[15] A. R. Anderson, B. J. Fowers, "Lifestyle behaviors, psychological distress, and well-being: A daily diary study", Social Science & Medicine, **263**, 2020. doi.org/10.1016/j.socscimed.2020.113263

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 561-570 (2020)

www.astesj.com

Second International Virtual Conference on Multidisciplinary Research 2020

ASTESJ

ISSN: 2415-6698

sharpniZer: A C# Static Code Analysis Tool for Mission Critical Systems

Arooba Shahoor[1,*], Rida Shaukat[1], Sumaira Sultan Minhas[1], Hina Awan[1], Kashif Saghar[2]

[1]*Software Engineering Department, Fatima Jinnah Women's University, Rawalpindi, 46000, Pakistan*

[2]*NESCOM, Centre for Excellence in Science and Technology, Rawalpindi, 46000, Pakistan*

ARTICLE INFO

Article history:
Received: 10 September, 2020
Accepted: 07 October, 2020
Online: 20 November, 2020

Keywords:
Software Testing
Software Quality
Verification and Validation,
Safety-Critical Programming

ABSTRACT

Until recent years, code quality was not given due significance, as long as the system produced accurate results. Taking into account the implications and recent losses in critical systems, developers have started making use of static code analysis tools, to assess the eminence of source code in terms of quality. Static code analysis is conducted before the system is sent into production. The analysis aims to identify the veiled defects and complex code structures that result in the decrement of code quality or are likely to become a cause of malfunction during execution. To address this line of work, this research paper presents a static code analyzer for C#, named as sharpniZer.

The key purpose of this tool is to verify the compliance of the source code written in C#, in congruence with the target set of rules defined for analysis as per the accepted industry standards set particularly for the development of mission-critical systems. sharpniZer efficiently figures out the lines of source code that hold probable concern appertain to the category of design rules, usage rules, maintainability rules, inefficient code, complexity, object model and API rules, logical rules, exception, incomplete code, and naming conventions. Each violation encountered in source code is ranked by the severity level as: critical, major, and minor. The tool shall prove to be worthwhile, especially if utilized in critical systems.

1. Introduction

The prevalence of software has raised serious concerns in the software industry to think of ways in which software quality can be ensured. How the quality of the software system or the underlying source code can be quantified may differ from system to system. However, the system must be optimized and proficient in terms of parameters including maintainability, testability, reusability, resource consumption, etc [1].

Most of the aspects of software quality largely depend upon the skills and expertise of the development team. As the requirements from enterprises regarding system functionality are becoming critical, the quality of source code is likely to diminish. A bad-written code leads to increased resource consumption and decreased efficiency and productivity [2, 9].

The hidden defects in source code, if not identified and addressed appropriately, can ultimately lead to unreliable system behavior during execution. The systems need to be thoroughly tested for the detection of loopholes left during development. The extensively used techniques include static code analysis and dynamic code analysis. However, the practice under consideration for this work is that of static code analysis.

1.1. Current Scenario

As previously stated, static code analysis has become indispensable particularly for safety (or mission) critical systems where there is no room for even trivial bugs at the static or runtime stage. History testifies several incidents that stress the need of using code analysis tools. Below we will briefly discuss 3 such incidents where ignoring trivial coding flaws cost the developers heavy losses.

1.1.1. The Ariane 5 Explosion

Ariane 5, a heavy-lift launcher, made to launch a 3-ton payload into orbit, cost a total of 10 years and $8 billion to be manufactured. On its first launch in July 1996, the rocket soon diverted from its path of flight due to a diagnostic produced as a result of a software exception. The exception was caused during

*Corresponding Author: Arooba Shahoor, arooba.shahoor@gmail.com

a data conversion [11]. Specifically, a 16-bit value of type float was being converted to a 16-bit signed number, i.e. the value being converted was much greater and could not be possibly represented by a 16-bit number, consequently leading to an operand error [12]. Since it had reached an angle of attack of more than 20 degrees, the launcher started to disintegrate, and finally obliterated itself through a self-demolishing method, along with 4 satellites that it carried within itself [13].

1.2. Patriot Missile Error

At the time of gulf war (2 August 1990 – 28 February 1991), 28 American soldiers lost their lives, and yet 100 others severely wounded when the US missile defense system failed to detect the Scud missile that the Iraqi forces launched [5]. The root cause of the failure lies in the conversion of time from integer to floating value, which consequently resulted in a major shift in the range gate [6, 28].

1.2.1. AT & T (American Telephone & Telegraph) Network Goes Down

About 75 million call requests were not answered and approximately 50 % of the network of AT&T went down on January 15, 1990, when a bug in a single line of code incurred the network failure for several hours [7]. Specifically, the problem occurred where a break statement within a switch case, was nested within an if clause. Companies linked to its services faced fiscal setbacks, among which was "American Airlines", whose incoming calls were reduced by two-third owing to the network crash [8, 29].

The incidents mentioned above, stress upon the need and significance of thorough analysis of the code before the runtime state. This paper presents a static code analysis tool for C#, named as sharpniZer. The rest of the paper is organized as follows: part B of this section will delineate the basic aim and purpose of the utilization of static code analyzers in the software development process. Section 2 reveals the solution proposed in this paper i.e. a static code analyzer for C# that is designed to detect such seemingly non-destructive but critical bugs usually ignored in the development of mission-critical systems. This section will also discuss the categories in which the rules implemented by the proposed analyzer can be generally segregated. Part 2 of section 2 will present the methodology adopted for the development of the tool. It also indicates the flow of the application. Section 3 outlines the system features and characteristics of sharpniZer. Each module is briefly presented; the figures attached to each module shall assist the reader in gaining better insight into the tool. Section 4 gives a comparison of the proposed tool with other existing tools in a tabular format. The textual summary of the comparative analysis is given in section 5. Section 6 presents the direction of future work i.e. discussion of the ways the tool can be extended or modified, to cater to the rapidly changing market needs. Section 7 concludes the research presented in the paper.

1.3. The Goal of Static Code Analyzers

Static code analysis chiefly focuses on examining the source code, without actually executing it, to help ensure that the code is abiding by the established coding standards. This helps recognize such flaws and code constructs that may decrease code quality, or in the worst-case scenario, incur a costly disaster month or years later. The article an overview on the Static Code Analysis approach in Software Development [30] talks about and objectives for developers to introduce static code analyzers in the testing phase of the development life cycle and here we summarize six of those factors

- Saves time and money

Reviewing code manually can take up much time, in comparison to which automated static code analyzers are much faster. These analyzers can take large sets of known bugs (common as well as uncommon) and combine them with special algorithms to track them anywhere in the code in hand, which enables the bugs to be detected in a matter of seconds which would otherwise take hours or even days.

- In-Depth Analysis

With manual testing, it is highly likely to overlook some code executions paths, which is not the case in automated testing. Automated testing is performed as we build our code, so we can get an in-depth analysis of where potential issues might lie, under the rules applied by the user.

- Accuracy

It is always possible for manual reviews to be inaccurate or error-prone. This is another aspect where code analyzers come in handy. They can specify exactly where and when the error occurs without any ambiguity. Issues such as stack overflows or race conditions, that only show up when code goes into production, are hard to figure out if not resolved beforehand (through static code analysis) since the scenarios are less likely to be able to be recreated.

- Ease of use

A clear benefit the analyzers provide is their ease of access and usability. These analyzers come in stand-alone as well as integrated form, you can easily integrate analyzers with your code editing tool. Moreover, everything is performed automatically by the analyzer, it does not require any profound or in-depth knowledge or expertise on the part of the user.

- Comprehensive Feedback/Overview

At the end of the analysis, the user is provided with a detailed overview of the code and the violations (often with visualizations), which helps the user to get to the action and work on fixing the violations there and then.

- Uniform Code

Code analyzers analyze the code against set coding standards, which means by introducing a particular code analyzer for code testing, the corporate is enforcing the use of best coding practices which will consequently set a uniform coding practice among all coders and programmers. This will facilitate developers in an easier and quicker understanding of each other's code.

2. Proposed solution

Taking into account the concerns about the quality of code and recognizing that the cause of recent losses associated with the critical systems is the lack of appropriate code analysis, a static

code analysis tool, named as sharpniZer, has been developed. sharpniZer efficiently analyzes the source code written in C# against the target set of (150) rules defined for analysis as per accepted industry standards. As soon as the analysis completes, the user is presented with the results of the analysis in multiple forms so that the user can easily gain a thorough insight into the quality of the code at hand.

Before diving into the implementation of the tools, let us first present the 10 categories in which the selected 150 rules are divided. The rules have been accumulated from multiple industry-wide accepted standards such as MISRA, JPL, CERT, CWE, etc. The rationale behind the selection of these 150 rules is the frequency of their violations along with their pertinence in the development of mission-critical systems.

2.1. Description of categories of rules

1. Usage Rules: This is the category of rules that ensures the proper utilization of the .Net Framework and correct usage of the common types provided by it. Failure in complying with these rules may have diverse repercussions; they can be complexity, maintainability, or performance issues.

This category contains rules such as "Dispose() & Close(): Always invoke them if offered, declare where needed." and "Do not omit access modifiers".

The former rule recognizes the problem that Class instances often have possession of unmanaged resources, such as database connections. Not disposing of these resources implicitly or explicitly after they are no longer needed, may lead to memory leakage. Similarly, not closing SQL connections after utilizing them, may result in a lack of connections available in case of connecting to the database again.

2. Design Rules: Every programmer or developer is unique and has his style when it comes to coding or implementing logic. Therefore, it is imperative that some standard of coding style is maintained that will be followed by all the developers. Design rules will thus outline certain guidelines for the way logic is to be implemented.

As an example, consider the following rule: "'out' and 'ref' parameters should not be used". 'Out' and 'ref' are required when variables have to be passed by reference. This requires above-average experience and skills in working with pointers and more than one return value as well as an understanding of the difference in the concepts of 'out' and 'ref'. Not all developers are expected to be of the same competency, and therefore to bring all developers to a level playing field, it is recommended to rather shun the use of out and ref.

3. Object Model and API Design Rules: These are the rules that are geared towards object-oriented features such as inheritance, encapsulation, etc. An example rule for this category is: "Always prefer interfaces over abstract classes." Since interfaces are not coupled with other modules, it can be independently tested, unlike abstract classes, that can be only tested during integration testing. Moreover, since in C# a class can inherit from one only class but multiple interfaces, it is highly recommended to prefer interfaces when the developer has a choice between them and abstract classes.
4. Exceptions: Exceptions are yet another part of the code that needs to be catered for with much caution. Consider the following rule: "If re-throwing an exception, preserve the original call stack by omitting the exception argument from the throw statement".

This rule guides and warns us about the case when we want to re-throw an exception. It tells us that if we re-throw the exception with the exception (ex) argument, the compiler will not consider it a "re-throw" of the original exception rather it will consider it as a new exception that occurred where the throw (ex) statement was written. Likewise, many such guidelines need to be considered when using exceptions in our code.

5. Complexity Rules: These rules deal with common coding practices that may seem harmless at the time of implementation but in the long run result in what has become known as spaghetti code, where the flow of control is hard to keep track of. Particularly for critical systems, spaghetti code proves to be extremely risky, since in such systems one can't afford to overlook the validation and verification of every possible path of the control flow [4].

An example rule for this category is "Continue statement should not be used" or "go to statement should not be used". Since goto and continue statements provide an alternative way to exit from control structures, for large code files, it becomes inevitably difficult to keep track of and verify all the paths of the control flow.

6. Inefficient code: Rules of this category cater to the coding practices that negatively affect the performance of the system that may either be the result of greater time or memory consumption.

For example, a rule for this category of issues is "Avoid string concatenation in loop". Since string is an immutable type, every time a string is concatenated using '+' or using 'Concat()', a new memory location is reserved for the newly formed string. If the number of iterations is unknown or is very large, such practice would highly deteriorate the performance of the system due to unnecessary consumptions of memory.

7. Incomplete code: As is suggestive of the title, rules that fall in this category cater to the sections of code that give the impression that some part of code is left out, be it intentionally or mistakenly. Such coding faults, above all things, raise the issues about the understandability of code [10].

As an example, to this category of rules, consider this rule: "Ensure that a 'switch' includes cases for all 'enum' constants", this rule suggests that if a switch statement is based on a variable of type enum, then leaving any of the constants in its cases would be considered as a coding error. To avoid this blunder as well as to enhance the understanding of the switch statement, it is advised to handle all constants of the enum.

8. Logical errors: Rules in this category cater to situations when the coder might feel that the logic or the approach used for implementing a code section is correct, but actually, that code section ends up serving some other purpose, as opposed to the one the coder had for it in mind. Such blunders result in

unexpected results, confusion in the understanding of the code, or redundancy and decreased performance [3].

Example rules for such scenarios are as follows: "Do not perform self-assignment" or "Do not compare identical expressions". Looking at these rules, it is obvious that they prevent redundancy overhead by warning the user not to waste time and memory on something that will have no effect or use whatsoever. Therefore, such an implementation should be regarded as a coding error.

9. Maintainability Rules: This category of rules chosen for our system is the one that deals with the issue of maintainability. These rules ensure that the size of the code or modules and values used within, do not exceed a particular limit, for better performance and efficiency of the system. One example of this category is "Avoid creating files that contain many lines of code". Very large files will not only create problems of maintainability but will lead to many other issues such as:

- Merge conflicts; if more than one developer needs to work on the code at the same time.
- Network traffic; if the system uses a version control system to which the entire file needs to be transmitted for every small change.
- Poor organization of code; since large file size might be the result of appending everything to the same file, rather than building related things separately in respective files.

10. Naming Conventions: This category outlines the conventions set for naming code files and identifiers. The identifiers include properties, variables, methods, fields, parameters, namespaces, interfaces, and classes. Certain guidelines for naming have a much more critical purpose than mere consistency, take for example this rule: "Avoid using the same name for a field and a variable". Since the field is just a local variable of the class, declaring any other local variable with the same name as the field will cause confusion between the two, therefore it is strongly recommended to use different names for fields and variables in general [16].

2.2. Implementation Details

The tool, sharpniZer, is developed in WinForm on .NET Framework of Visual Studio. For development, Waterfall model is chosen, whereas the implementation language is C#. The implementation process and logic is discussed in detail below:

In creating the analyzer, we made use of SDK called Roslyn, which provides the user with useful API's used for C# code parsing and analysis of language constructs. An analyzer created with the Roslyn SDK, inherits Microsoft's CodeAnalysis base class. Given a file as a string, Roslyn can parse it and create a syntax tree from it, using the CSharpSyntaxTree Class of the Microsoft.CodeAnalysis package. It can then access the root of the syntax tree using the following statement.

```
var root = (CompilationUnitSyntax)tree.GetRoot();
```

Once we get the root of the syntax tree, it is fairly easy to access the descendant elements (declarations, expression, etc.) in the code. Microsoft.CodeAnalysis library enables us to access various elements of a node. For e.g Microsoft.CodeAnalysis.CSharp. Syntax enables us to access any type of declaration in C# code.

To show how our analyzer detects the different categories of problems and how checks are implemented, we will present the snippets of code demonstrating the implementation of 2 of the 150 rules implemented by sharpniZer.

Rule: Do not assign to local variable in return statement

When the assignment is made to a variable in a return statement, the variable goes out of scope and the value assigned is never read. Hence resulting in code redundancy. Redundant source code is one that is bloated, less reliable, and difficult to maintain. Any expert programmer would agree that the harder it is to maintain the code, the more likely it is to contain bugs.

Moreover, in a highly knitted team setting where multiple developers are involved, any such code written by a programmer will create much confusion for another to read or understand, resulting in a great deal of wasted time and mental energy. In the following section, we present and explain the snippets from our program which detects and warns about such redundant assignment statements in C# code.

First, all the method declarations in the code are stored in a list.

```
IEnumerable<MethodDeclarationSyntax>
methodsdecroot.DescendantNodes().OfType<MethodDeclarati
onSyntax>();
```

Next, all the variable declarations within each method are stored.

```
foreach (var method in methodsdec) {
   foreach (var VarDec in
method.DescendantNodes().OfType<VariableDeclaratorSyntax>
()) {
        VarDecsList.Add(VarDec.Identifier.ToString());
   }
```

Then, all the return statements within each method are looped through.

```
foreach (var returnStatement in
method.DescendantNodes().OfType<ReturnStatementSyntax>()
) {
```

After which we loop through all the assignment expressions within each method are looped through.

```
foreach (var assignment in
method.DescendantNodes().OfType<AssignmentExpressionSyn
tax>()) {
```

The following code then checks if any of the assignments are made to the variables declared within the method (local variables) and is done in a return statement.

```
if (returnStatement.Contains(assignment) &&
VarDecsList.Contains(assignment.Left.ToString())) {
```

If so, then the line number of that particular assignment is added to the warnings list

```
warnings = "Do not assign to local variable in return statement";
lineno = assignment.GetLine()+1;
WarningsList.Add(lineno + "@" + warnings + "@" + level);
```

In the above example, the rule was implemented by working on merely syntactic level. However, if the rule to be implemented requires some more information about the nodes/syntaxes (such as types, members, namespaces and variables which names and other expressions refer to), we have to be able to retrieve the symbols of the expressions or declarations.

We can retrieve symbols once we have the semantic model for the syntax tree. The method SemanticModel.GetDeclaredSymbol() is used to get the symbol of a given declaration syntax, whereas SemanticModel.GetSymbolInfo() returns the symbol of an expression syntax. The implementation of the second rule presented here makes use of this.

Rule: "Use throw instead of throw e (e for exception) whenever rethrowing the exception"

Throw statements are used so that if, during the execution of a software program, an unexpected condition occurs, the system is unable to process the next statement and instead throws an 'exception' error that specifies the line where the problem occurred, along with the line the throw statement was specified at. Now, when that exception is caught, we can choose to re-throw it. However, when re-throwing the exception, most programmers make the mistake of using the throw statement with the exception object (e); doing so will make the stack trace information within the exception restart at the current location, such that it will then point to the line where exception was thrown, rather than where the problem, causing the exception, occurred.

Though it might not seem like a critical issue per se, any programmer would acknowledge that incorrect stack trace information can lead to much confusion and thereby potential disruption of any logic that would be based on it. Below we present our code implementation that would detect such statements (if any) within C# source code.

First, all the catch clauses in the code are stored through the descendants of the root node (as in the previous example).

```
IEnumerable<CatchClauseSyntax> catchcaluses =
root.DescendantNodes().OfType<CatchClauseSyntax>().ToList(
);
```

Next, the throw statements within a catch clause are stored.

```
foreach (var catchclause in catchcaluses) {
  var throws = catchclause.DescendantNodes(n => n ==
catchclause ||
!n.IsKind(SyntaxKind.CatchClause)).OfType<ThrowStatementS
yntax>() .Where(t => t.Expression != null);
```

Then, all the throw statements, that were saved in the list, are looped through, to get the expression symbol of each throw.

```
foreach (var @throw in throws) {
  var thrown =
model.GetSymbolInfo(@throw.Expression).Symbol as
ILocalSymbol;
```

If the expression symbol matches the exception identifier, it means the throw statement specifies the exception, which violates the rule. Hence the line number of that throw statement is added to the warnings list.

```
if (Equals(thrown, exceptionIdentifier)) {
   warnings = MessageFormat;
   lineno = ( @throw.GetLine()+1).ToString();
   mylist.Add(lineno + "@" + warnings + "@" + level);
}
```

3. System features and usage

Upon launching the tool, the user is asked to browse a .CS file or folder containing.CS files. Once the files are loaded, users can select the categories of rules upon which the analysis is to be conducted. The tool allows users to selectively enable the categories of rules (by default, all categories of rules are enabled). The analysis result presents the defiance and violations in source code in congruence with the underlying set of rules. An analysis summary is presented in the form of a dashboard, presenting the results in tabular form. The graphical representation adds to the visualization of analysis and assists users in gaining deeper insight into the results. An overview of the flow of the analysis process of sharpniZer can be seen in Figure 1.

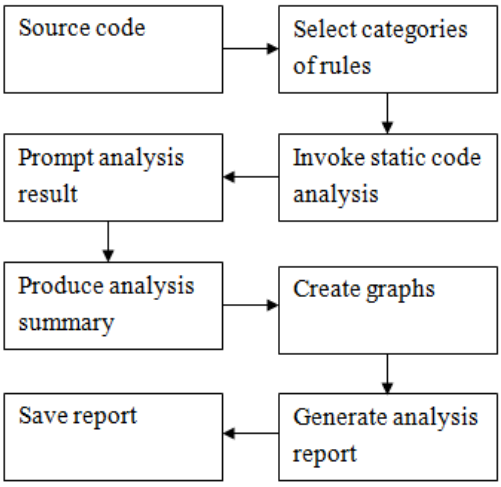


Figure 1: Methodology for conducting analysis

The above mentioned features of the system will now be elaborated per module below

3.1. Browsing C# file or project folder

User may CHOOSE C# file(s) to be analyzed in two ways:

Using the "Choose C# File" file option: allowing the user to choose any code file with the extension of .CS. After the file is selected, the content of the file will be displayed to the user in the Analyze tab (Figure 4). Clicking the Start Analysis button in the Analyze tab will start the analysis of the selected file.

Figure 2: Home page of sharpniZer

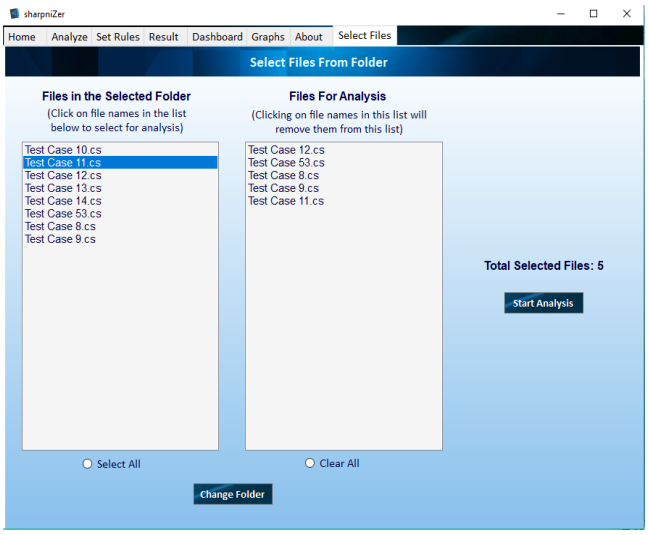


Figure 3: Selection of files for analysis

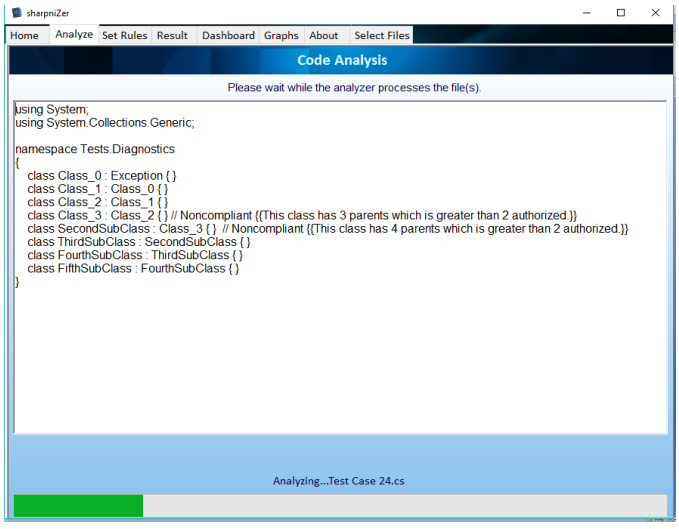


Figure 4. Analysis progress

Using "Choose C# Project" option: The user may choose any folder containing .CS files in the folder itself or its subfolders. If the user wants to deselect a file within that folder, they can click on that file's name in the "Files for Analysis" list (Figure 3), that file will be removed from the list of the files to be analyzed. The selected files will be analyzed sequentially and the progress of the analysis for each file will be displayed (Figure 4).

3.2. Select categories of rules for conducting analysis

Users can select or deselect any of the 10 categories of coding standard rules provided. The available categories can be seen in Figure 5.

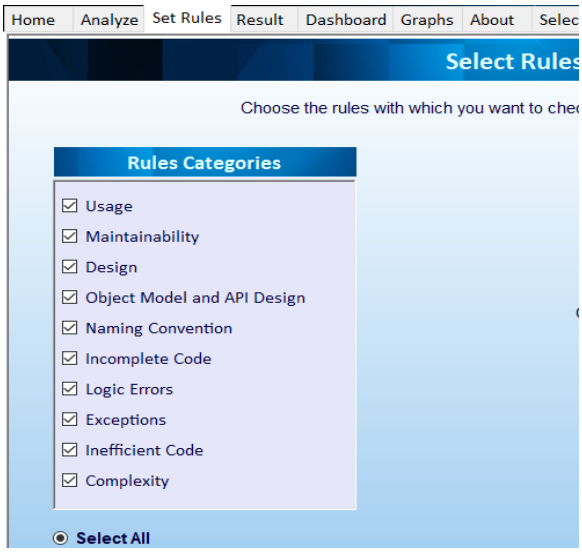
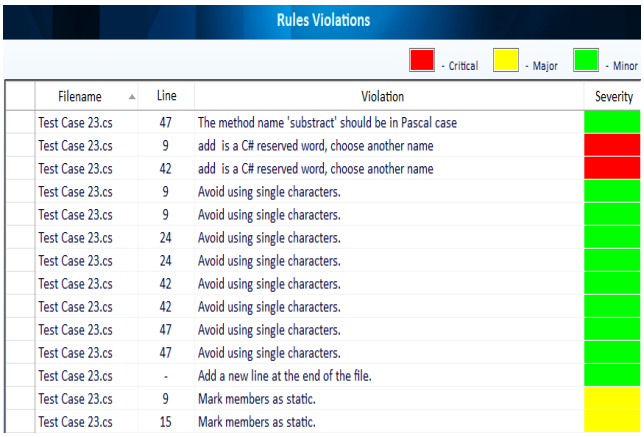


Figure 5. Selection of categories

3.3. Viewing analysis results

As soon as the analysis of the selected files(s) finishes, the tool switches to Results tab (Figure 6), which displays the result of the analysis. The analysis result specifies the filename(s), line no. at which the violation has been encountered, the violated rule, and the severity of the violation.



Filename	Line	Violation	Severity
Test Case 23.cs	47	The method name 'substract' should be in Pascal case	
Test Case 23.cs	9	add is a C# reserved word, choose another name	
Test Case 23.cs	42	add is a C# reserved word, choose another name	
Test Case 23.cs	9	Avoid using single characters.	
Test Case 23.cs	9	Avoid using single characters.	
Test Case 23.cs	24	Avoid using single characters.	
Test Case 23.cs	24	Avoid using single characters.	
Test Case 23.cs	42	Avoid using single characters.	
Test Case 23.cs	42	Avoid using single characters.	
Test Case 23.cs	47	Avoid using single characters.	
Test Case 23.cs	47	Avoid using single characters.	
Test Case 23.cs	-	Add a new line at the end of the file.	
Test Case 23.cs	9	Mark members as static.	
Test Case 23.cs	15	Mark members as static.	

Figure 6: Analysis result

3.4. Viewing Dashboard

The Dashboard presents the analysis summary. The numerical/tabular form of the results allows the user to gain deeper insight into the outcome (Figure 7). It also allows the user to generate an analysis report in a .pdf file that can be saved/downloaded and shared with team members (Figure 8).

Dashboard presents the following content:

- Total File(s) Selected
- Total Lines of Code
- Total number of violation(s)
- Number of violation(s) per file
- Category Violation(s) for all file(s)
- Category of violation(s) per file
- Severity of violation(s) for all file(s)
- Severity of Violation(s) per file

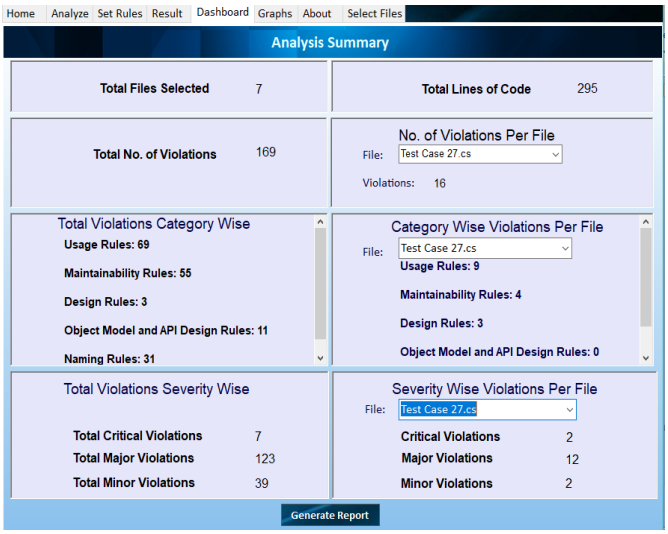


Figure 7: Dashboard

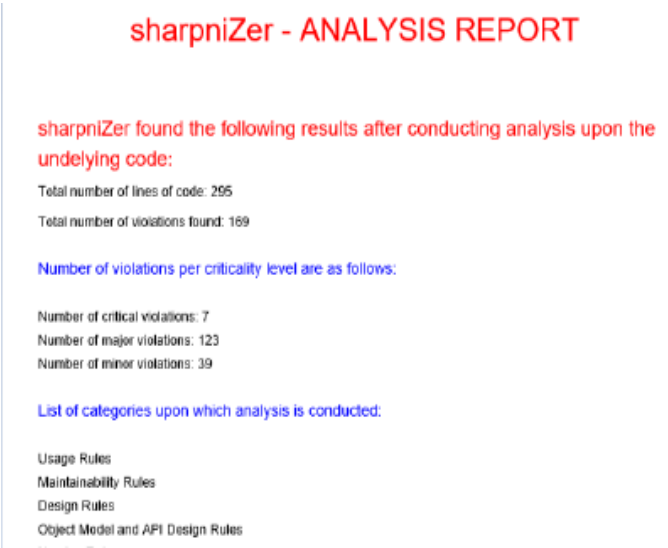


Figure 8: Analysis Report

3.5. Visual representation of Analysis Result

The Visual representation of results is often considered to be more comprehensive and elaborate, therefore, sharpniZer generates graphs by utilizing the numerical values of analysis results (Figure 9). The graphs can be generated for total violations, as well as for each file (if the analysis is conducted upon multiple files).

- The bar chart represents the total violations per category of rules chosen.
- The pie chart represents the total violations of rules per severity level.

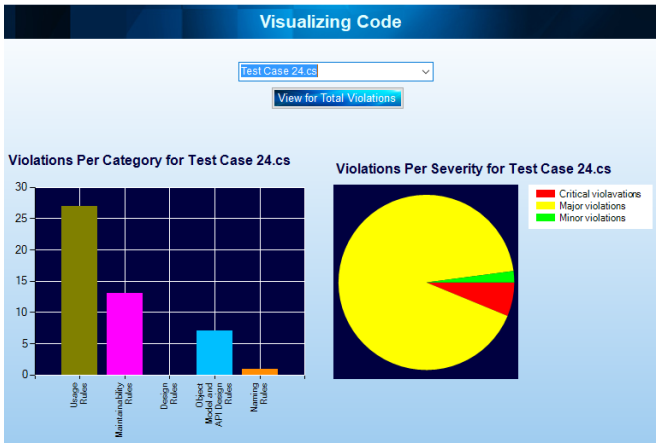


Figure 9: Graphical representation of analysis results

4. Review and Comparison of the Existing Tools with sharpnizer

Below we provide a review of other existing 9 tools that analyze code written in C# (a detailed comparison of C# code analyzers could be found in our previously published review article Probing into code analysis tools: A comparison of C# supporting static code analyzers [27]).

The 9 tools chosen for comparison in this research were selected based on their overall ranking resulting from numerous surveys combined with the fact that prior literature has proved these tools to be the most popular among researchers. The majority of the values for our evaluation are attained by gathering information from surveys and public reviews of users. Rest are procured by self-executing the tools and observing the performance in light of the mentioned 10 parameters.

Table I defines the parameters taken into consideration for comparison. These parameters were selected based on the survey of the code analyzer features most commonly considered by companies when selecting one to assess their mission-critical systems. Table II AND III present the value for each tool against these parameters. Whilst keeping the comparison authentic and unbiased, Tables II and III manifests the benefits sharpniZer has over some other quite decently known tools in use today.

5. Results

From the results of the comparative analysis presented in Tables 1 and Table 2, we can see that sharpniZer's installation process and usability is easier than some of the tools available. Though the number of rules implemented is yet lower than most in comparison, it supersedes many in providing customization of rules' selection, categorizing the nature of violations and covering a wide spectrum of the coding standards. In addition, unlike some of the contemporary tools, sharpniZer does not require code to be compiled first to perform analysis.

Table 1: Comparison Parameter

S. No.	Parameter	Definition
1	Usability	How user-friendly is the tool for a layman (scalable from easy to hard) (The classification was done based on the pubic reviews of the users of the tools listed and those of our client) [13]-[26]
2	Installation	How convenient is the process of installation/integration of the application (scalable from easy to hard) (Classification was done based on published surveys and our client's reviews) [13]-[26]
3	No of Defined Rules	The number of rules/metrics defined within the tool, so that the tool might assess compliance of these rules, during the analysis of codes.
4	No. Of Categories of Defined Rules	Number of types of coding standard violations defined by the tools
5	Requires compiled code	Does the tool require the code to be compiled before it can perform analysis on it?
6	Graphical representation	Does the tool display the graphical representation of the analysis result?
7	Selection of Desired Categories of Rules to be Applied	Does the tool allow the user to select specific categories of rules to check the code compliance with?
8	Severity categorization based on the nature of the violation (Blocker, critical, major, minor)	Does the tool specify the severity of violations in results (minor, major, or critical)?
9	No. of Standards for Code Compliance	Number of coding standards with which the tool checks the compliance of the code at hand
10	Is the Tool Optimized for Mission Critical Systems?	Does the tool predominantly check the compliance with rules defined for the development of mission-critical systems?

Table 2: Comparative Analysis

S. No.	Tool	Usability	Installation	No of Defined Rules	No. of Categories of Defined Rules	Requires Compiled Code?
1.	Ndepend	Normal [13, 14]	Easy [13, 26]	More Than 150 Default Code Rules	14	No
2.	PVS-Studio	Hard [23, 25]	Easy [23, 25]	450 Diagnosis Rules	29	No [23]
3.	Resharper	Easy [21, 26]	Easy [21, 26]	1700+ Code Inspections	10	No [21]
4.	Fxcop	Easy For GUI, Hard for Command Line [19, 25]	Easy [19, 25]	More Than 200	9	Yes

5.	Visual Code Grepper	Easy [18, 26]	Easy [17, 26]	Unknown	3	No
6.	Nitriq	Easy [20, 25]	Difficult [20, 25]	Over 40 Pre-Written Rules Defined	Unspecified	Yes
7.	Parasoft Dot Test	Easy [23, 26]	Untested/ Not Reviewed	Above 400	20	No
8.	Coverity Scan	Hard [16, 25]	Easy [16, 25]	Unknown	30	No
9.	sharpniZer	Easy	Easy	150	10	No

Table 3: Comparative Analysis

S.No.	Tool	Graphical Representation	Selection of Desired Categories of Rules to be Applied	Severity Categorization Based on Nature of Violation (Blocker, Critical, Major, Minor)	No. of Standards for Code Compliance	Is the Tool Optimized for Mission Critical Systems?
1.	Ndepend	Yes [14]	Yes	Yes [13]	Unspecified	No
2.	PVS-Studio	No [24]	No	No	3	Yes[24]
3.	Resharper	No [21]	No	No	Unspecified	No
4.	Fxcop	No	Yes	No	1	No
5.	Visual Code Grepper	Yes [26]	No	Yes	2	Yes [18]
6.	Nitriq	Yes [20]	Yes	No	Unspecified	No
7.	Parasoft DotTEST	Yes [24]	Yes	Yes	3	Yes
8.	Coverity Scan	Untested/ Not Reviewed	No	No	2[15]	Yes [15]
9.	sharpniZer	Yes	Yes	Yes	5	Yes

Due to successful realization of chief requirements for testing static code of mission-critical systems, sharpniZer is currently employed by NESCOM (a military research organization of Pakistan) to test the software (written in C#), embedded in some of the mission-critical and ammunition systems developed by the organization.

6. Future Work

The proposed tool currently encompasses 150 coding rules, obtained from standards (established for the development of mission-critical systems) such as MISRA, CERT, CWE, and JPL and also from Microsoft. The rules list can be extended to cover more rules and standards. Moreover, custom rules can be made

part of the tool that will allow users to define their own rules through a query language.

Refactoring is yet another feature that can be incorporated into the system. Through Refactoring, the violation will automatically be corrected in the code as the user clicks on the violated rule in the Results tab. Also, the tool can be expanded to cater to source codes of other languages along with C#

7. Conclusion

This paper discusses the burgeoning issue of overlooked bugs in static code of critical systems, and proposes a static code analysis tool, sharpniZer, that analyzes the code specifically written in the C#, and precisely identifies all the discrepancies and deficiencies in the source code as per the coding standards set for the development of mission-critical systems, enhancing the overall efficiency and reliability of the code.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research paper is a product of the effort of 5 authors, the references to whom are specified at the beginning of the article. All authors have read and agreed to the published version of the manuscript. The detailed roles of each were as follows:

Conceptualization: Rida Shaukat and Arooba Shahoor Data curation: Arooba Shahoor Writing—original draft preparation: Rida Shaukat Writing—review and editing: Dr. Sumaira Sultan Minhas Supervision, Ms. Hina Awan Funding acquisition: Dr. Kashif Saghar.

References

[1] A. Costin, "Lua Code: Security Overview and Practical Approaches to Static Analysis," 2017 IEEE Security and Privacy Workshops (SPW), San Jose, CA, 132-142, 2017.

[2] S. A. Fatima, S. Bibi and R. Hanif, "Comparative study on static code analysis tools for C/C++," 2018 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, 465-469, 2018.

[3] T. Delev and D. Gjorgjevikj, "Static analysis of source code written by novice programmers," 2017 IEEE Global Engineering Education Conference (EDUCON), Athens, 825-830, 2017. DOI: 10.1109/TII.2016.2604760

[4] A. S. Novikov, A. N. Ivutin, A. G. Troshina and S. N. Vasiliev, "The approach to finding errors in program code based on static analysis methodology," 2017 6th Mediterranean Conference on Embedded Computing (MECO), Bar, 1-4, 2017.

[5] "The Patriot Missile Failure," Soundwaves, 2016. http://www-users.math.umn.edu/~arnold/disasters/patriot.html.

[6] "Lethal Software Defects: Patriot Missile Failure « Barr Code," Barr Code, 2014. https://embeddedgurus.com/barr-code/2014/03/lethal-software-defects-patriot-missile-failure/.

[7] "AT&T Corp. (American Telephone & Telegraph)," Ad Age, 15-Sep-2003. http://adage.com/article/adage-encyclopedia/t-corp-american-telephone-telegraph/98327/.

[8] "AT&T's History of Invention and Breakups," The New York Times, 13-Feb-2016. https://www.nytimes.com/interactive/2016/02/12/technology/att-history.html.

[9] R. Kirkov and G. Agre, "Source Code Analysis – An Overview", Cybernetics and Information Technologies, **10**(2), 2014.

[10] R. Plosch, H. Gruber, C. Korner and M. Saft, "A Method for Continuous Code Quality Management Using Static Analysis," 2010 Seventh International Conference on the Quality of Information and Communications Technology, Porto, 370-375, 2010.

[11] "The Explosion of the Ariane 5," Soundwaves. "The Patriot Missile Failure," Soundwaves, 2019. http://www-users.math.umn.edu/~arnold/disasters/patriot.html

[12] "ARIANE 5 Failure - Full Report," Safeware: System Safety and Computers, 2018. http://sunnyday.mit.edu/nasa-class/Ariane5-report.html

[13] "NDepend In Review," NDepend In Review - CraigTP's Blog, 2018. https://blog.craigtp.co.uk/Post/2017/12/17/NDepend_In_Review.

[14] C. N. D. W. www.ndepend.com, "NDepend Reviews 2018 | G2," G2 Crowd, 01-Sep-2018. https://www.g2.com/products/ndepend/reviews

[15] "27. Coverity Scan," 27. Coverity Scan - Python Developer's Guide, 2019. https://devguide.python.org/coverity/.

[16] Synopsys, "Coverity Reviews 2018 | G2," G2 Crowd, 22-Jul-2018. https://www.g2.com/products/coverity/reviews.

[17] Algaith, P. Nunes, F. Jose, I. Gashi, and M. Vieira, "Finding SQL Injection and Cross Site Scripting Vulnerabilities with Diverse Static Analysis Tools," 2018 14th European Dependable Computing Conference (EDCC), 2018.

[18] Nccgroup, "nccgroup/VCG," GitHub, 03-May-2016. https://github.com/nccgroup/VCG.

[19] FxCop, 2017. https://msdn.microsoft.com/en-us/library/bb429476(v=vs.80).aspx.

[20] M. Valdez, "Home," Home, 2017. http://marcel.bowlitz.com/.

[21] "Nitriq Code Analysis for .Net," Home : Nitriq Code Analysis for .Net, 2017. http://www.nitriq.com/.

[22] "ReSharper: Visual Studio Extension for .NET Developers by JetBrains," JetBrains, 2017. https://www.jetbrains.com/resharper/

[23] "Service Virtualization, API Testing, Development Testing," Parasoft, 2017. https://www.parasoft.com/.

[24] "PVS-Studio: Static Code Analyzer for C, C and C#," PVS-Studio: Static Code Analyzer for C, C and C#. https://www.viva64.com/.

[25] J. Novak, A. Krajnc and R. Žontar, "Taxonomy of static code analysis tools," In The 33rd International Convention MIPRO, 418-422. IEEE, 2010.

[26] H. Prähofer, F. Angerer, R. Ramler and F. Grillenberger, "Static Code Analysis of IEC 61131-3 Programs: Comprehensive Tool Support and Experiences from Large-Scale Industrial Application," in IEEE Transactions on Industrial Informatics, **13**(1), 37-47, Feb. 2017. DOI: 10.1109/TII.2016.2604760

[27] R. Shaukat, A. Shahoor and A. Urooj, "Probing into code analysis tools: A comparison of C# supporting static code analyzers," 2018 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, 2018, 455-464, doi: 10.1109/IBCAST.2018.8312264.

[28] E. Ogheneovo, "Software Dysfunction: Why Do Software Fail?. Journal of Computer and Communications, **02**(06), 25-35, 2014. DOI:10.4236/jcc.2014.26004

[29] M. Faizan & P. Dhirendra. "software testing, fault, loss and remedies. **6**. 553-568, 2019. DOI: 10.1109/52.382180

[30] I. Gomes, et al. "An overview on the Static Code Analysis approach in Software Development." 2009.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 571-576 (2020)

www.astesj.com

Second International Virtual Conference on Multidisciplinary Research 2020

ASTESJ

ISSN: 2415-6698

Experimental Study on Mechanical Behavior of Polypropylene-based Blends with Talc Fillers

Pham Thi Hong Nga*, Van-Thuc Nguyen

HCMC University of Technology and Education, Ho Chi Minh City, 71307, Vietnam

ARTICLE INFO

Article history:
Received: 08 July, 2020
Accepted: 07 October, 2020
Online: 20 November, 2020

Keywords:
Polypropylene
Talc powder
Blend
Tensile strength
Flexural Strength

ABSTRACT

In this report, polymeric composites made from Polypropylene (PP) and talc powder were studied. The talc powder with different portions of 10%, 20%, 30% by weight was used to create samples. The samples were examined by the tensile test and the flexural test according to ASTM D638 and ASTM D790. The surface morphology of the samples were investigated by scanning electron microscope (SEM). Increased the talc powder portion led to a decrease in the tensile strength and flexural strength. Among the added talc powder samples, the 10% talc powder sample presented the highest mechanical tensile strength of 25.91 MPa and flexural strength of 47.99 MPa, while the 30% talc powder sample shown the lowest mechanical tensile strength and flexural strength. The results of SEM analysis indicated the existence of talc plates and porosity between talc plates and PP substrate compared to the neat PP sample. The surface morphology of the samples was observed by SEM, showing higher porosity in samples with increasing filler content. Moreover, increasing the talc powder portion resulted in a higher chance of brittle fracture. The research indicated and explained the effect of talc power on the characteristic of PP.

1. Introduction

Neat Polypropylene (PP) plastic resins are preferred for plastic products requiring a high cost of the material bill. Therefore, with conventional products, manufacturers can use recycled PP to reduce costs and reduce environmental pollution [1-3]. Normally, the products made from recycled PP consist of neat PP, recycled PP, and filler. Fillers help reduce costs as well as improve the functionality of the product [4-6]. They are usually insoluble minerals, adding them will increase the volume of polymers. They can be broadly classified as reinforcing or non-reinforcing fillers. Among them, $CaCO_3$ is a popular choice for filler roles [7-9]. $CaCO_3$ fillers have been used in the form of fine particles with a size of less than 10 µm. $CaCO_3$ content can be added from a few percent to up to 30 percent by weight ratio. Adding $CaCO_3$ could greatly improve the thermal and mechanical properties of polymer materials [10-12].

For example, adding 20% $CaCO_3$ will improve thermal stability by about 27.5°C and in case of 30%, the tensile modules will increase by 97% when compared to pure PP [13]. Similarly, with 5, 10 and 20% $CaCO_3$ by weight, the results of the study have shown an increase in the elastic modulus, yield to break, and elastic stress. It has also been confirmed that viscosity during flow increases with fortifier addition and reaches the highest value with 15% talc [14]. The effect of kaolin on the mechanical properties of PP/Polyethylene through mixing kaolin fillers with the content of 1, 3, 5, 10, 14% by weight ratio is also considered [15]. The results showed that the tensile strength increased simultaneously with the increase of the filler concentration from 2 MPa of non-padded PP/Polyethylene and up to 9 MPa with the appearance of 14% kaolin. Increase the concentration of kaolin from 0 to 14% helps increase from 40 HD (Shore D Hardness) to 70HD.

Talc is also commonly added to increase hardness and improve the properties of polymers [16-18]. Therefore, more than 200 thousand tons of talc powder is mixed with PP annually. Talc has many effects of increasing hardness, durability, thermal resistance, and anti-damage of ultraviolet rays, good scratch resistance [19-21]. Talc is widely used in plastic products for machine parts such as cars and motorcycles, as well as household plastic products such as nylon bags and plastic pots. The current trend in the automotive industry is to produce thin, light, and precisely sized parts, which requires higher rheological resins that talc fillers can meet. China is the key world talc and steatite producing country with an output of about 2.2 million tonnes (2016), which accounts for 30% of total global output.

Several authors showed that adding talc could rise the mechanical properties of pure PP [22-25]. For instance, Urmila el al. pointed out that adding 10, 20, 30, 40, and 50% weight ratio

Corresponding Author: Pham Thi Hong Nga: hongnga@hcmute.edu.vn

results in the improvement of mechanical properties such as tensile strength from 23,49 MPa with primary PP up to 26,65 MPa [22]. Silva reported that adding talc with the ratio of 5, 10, 15, and 20% of the weight ratio leads to an increase in the tensile strength, from 28 ± 5MPa of pure PP up to 130 ± 4 MPa [14]. Mechanical properties of materials such as Charpy impact resistance and tensile strength can also be improved. Pan et. al indicated that tensile strength reaches its maximum value at 33.9 MPa in the presence of a 10% talc compared to 30.8 MPa of pure PP [23]. Moreover, the flexural strength also increased from 22,86 MPa to pure PP up to 32,98 MPa with 50% talc [22]. Too much talc powder is not good for the properties of the polymer. The tensile strength of the composite has 25% Talc is 24 MPa and 30% Talc is 22.5 MPa [24]. Composite PP with Talc shows an increase in elastic modulus when the amount of talc powder is up to 20% compared to pure PP from 85,72 MPa up to 150,0 MPa and decreases at higher concentrations [25].

The hardness of the mixture depends on the particle size and has been observed through microscopic images and the results of the Izod impact test give the result. With the larger particle size, the hardness decreases, and vice versa. For the impact toughness, with the particle size of 30 µm, the impact toughness is 6.5 (ft-lb/in) and with the particle size of 10 µm, the impact toughness is 12 (ft-lb/in) [26]. The hardness of the mixture increases with increasing Talc concentration [24]. Moreover, the impact toughness can also be improved by adding a suitable amount of talc. The impact toughness increased from 2 (KJ/m^2) of primary PP increased to 4 (KJ/m^2) with the appearance of 10% Talc, and then decreased to 3.5 (KJ/m^2) [23].

When comparing directly the two most common fillers in PP, $CaCO_3$, and Talc, there are still some mechanical weaknesses when adding $CaCO_3$ (tensile strength, flexural strength, elastic modulus, stress to break, and elastic stress), while talc is somewhat more dominant. The aims of this study is investigating the effect of talc composition on the tensile strength, flexural strength and microstructure of PP polymer.

2. Materials and methods

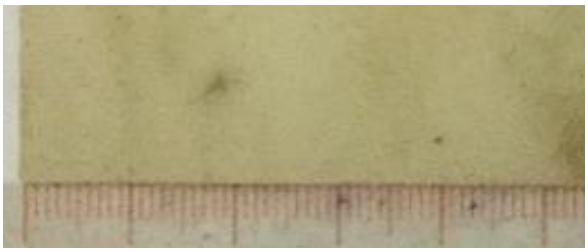
2.1. Materials and samples preparation

PP (Moplen HP500N type) was supplied by Lyondell Basell Industries (Saudi Arabia). Talc (TMD type) particle size was from 28 to 44 µm, whiteness ranged from 88-95%, chemical composition comprises Fe_2O_3 2-3%, MgO 25-27%, CaO < 0.6%, SiO_2 52-54%.

Table 1: Compositions of the samples (%wt.)

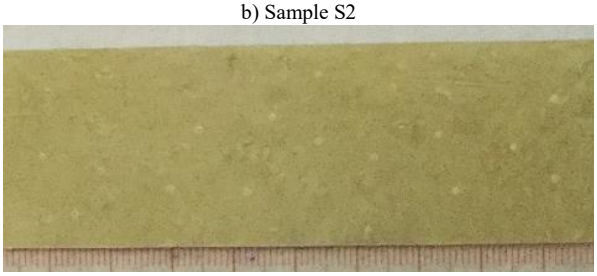
Materials	Compositions of the samples (%wt.)			
	S1	S2	S3	S4
PP	100	90	80	70
Talc	0	10	20	30

Table 1 presents the compositions of the samples S1-S4. The samples making process uses computer systems mixing and extrusion Polylab research OS - Haake (Germany), which has the basic parameters of the system such as mixing chamber contained 120 cm^3, and extrusion equipment L/D = 25, D = 16 mm. Single screw extrusion equipment connected to blown film system (blown film) and sheet extrusion system (0.2 to 1.2 mm x 100mm). Mixed samples were melt-blended during the period from 6-7 minutes with the temperature of the extruder zone temperature was 180°C. After extruding, the resulting pellets were dried at 80°C for 4h and then molded using injection molding machine into plates. The mold-temperature profiles were set at 180ºC. The molding process was carried out in 5 minutes and then cooled for 20 minutes. Figure 1 presents the size of plate samples after molding, it is 145 x 145 mm in length and 2 mm in thickness.

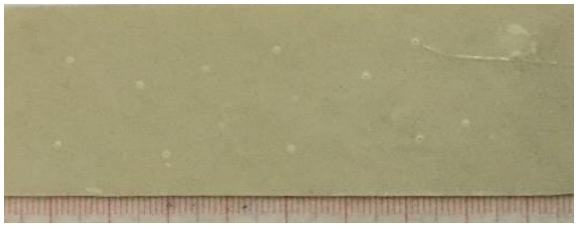


a) Sample S1

b) Sample S2



c) Sample S3



d) Sample S4

Figure 1: Material sheet for making samples: (a) Sample S1, (b) Sample S2, (c) Sample S3, and (d) Sample S4

2.2. Methods for mechanical properties and surface morphology analysis of samples

The tensile property was defined by method ASTM D638-02 experiments with the conditions required, as shown in Figure 2. Testing samples must be cleaned of oil, grease, and other impurities, then put into the environmental conditions of 23 ± 2°C direct heat, humidity 50 ± 5%, for 40 hours prior to the test. The experiments were conducted by Universal testing machine Shimadzu Autograph AG-X Plus 20kN (Japan). Some basic specifications are described as following: load capacity (maximum) 100 kN, traction speed 0.0005 to 1000 mm/min ±

0.1%, tensile capacity (maximum) 600 mm. This paper using speed was 55mm/min

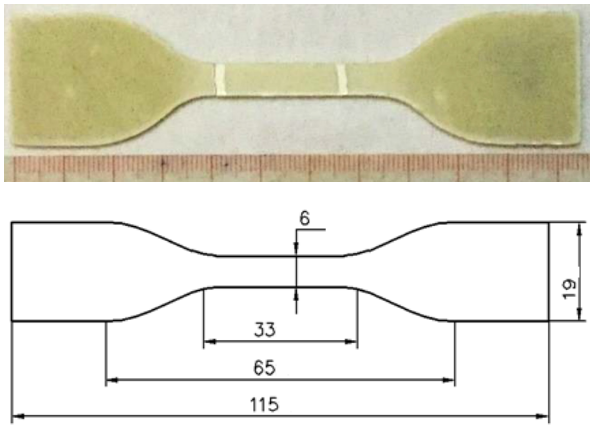


Figure 2: Specimen for tensile test

Flexural properties were defined by method ASTM D790 experiments. Samples are the composite material beams of rectangular cross-section placed 3-point bending under load. The sample size was shown in Figure 3. The experiments were conducted by Universal testing machine Shimadzu Autograph AG-X Plus 20kN.

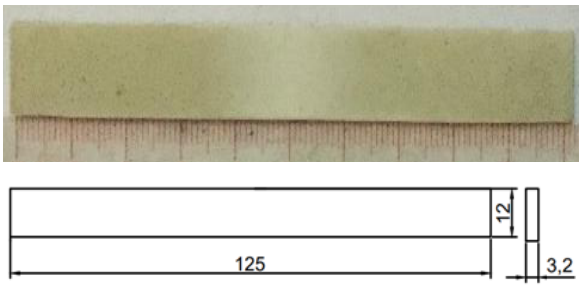


Figure 3: Sample for flexural test

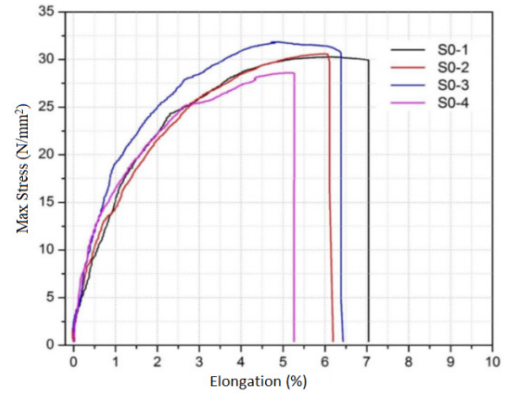
The fracture surface of each specimen in the bending strength test was observed by scanning electron microscope HITACHI S - high resolution - 4800 (SEM), with acceleration 5.0 kV. The surface of the samples used for SEM all was platinum-sputtered with a conductive layer before observation.

3. Results and discussions

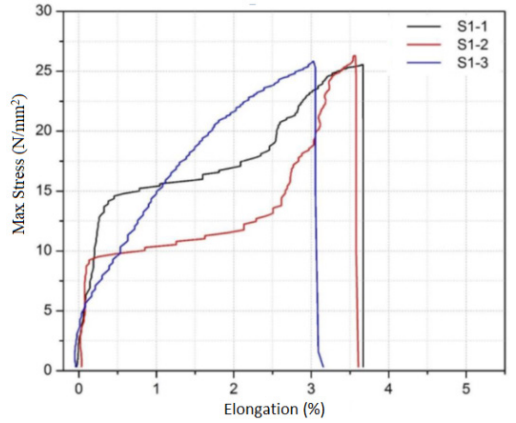
3.1. Tensile strength

The number of replicates is mostly four. Despite there is some differences between replicates, they varied only a few percent. Figure 4 shows the stress-strain curves of S1-S4 samples. Figure 5 shows the average values of the tensile strength of these samples. The tensile strength values are 30.73 MPa, 25.91 MPa, 23.68 MPa, and 23.29 MPa corresponding to S1, S2, S3, and S4 samples, respectively. In general, increase the talc content leads to decrease in tensile strength, a similar result of Zihlif report [24].

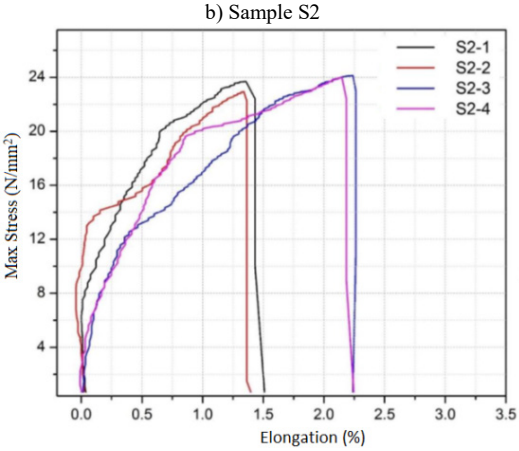
Table 2 presents the elongations of S1, S2, S3 and S4 samples. The average elongation values are 6.09%, 3.45%, 1.89%, and 2.99% corresponding to S1, S2, S3, and S4 samples, respectively. Increase the talc content mostly decrease the elongation value of the samples.



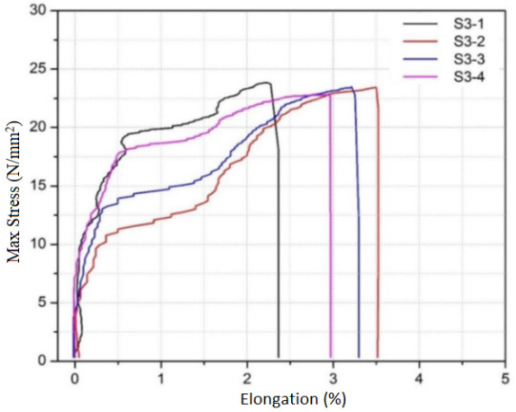
a) Sample S1



b) Sample S2



c) Sample S3



d) Sample S4

Figure 4: Stress-strain behavior of tensile strength: (a) Sample S1, (b) Sample S1, (c) Sample S3, and (d) Sample S4

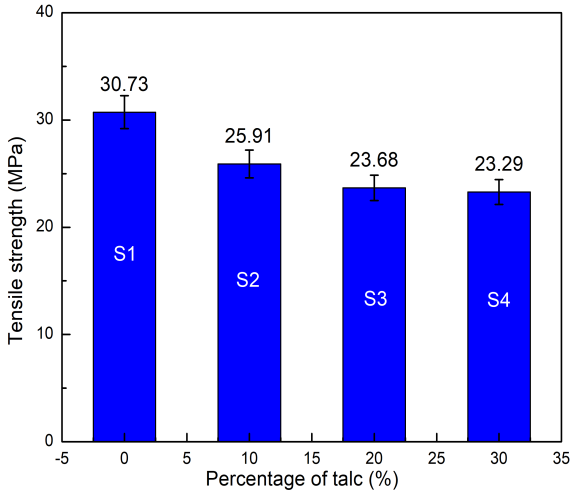


Figure 5: Tensile strength comparison of samples

Table 2: Elongations of the samples

Samples	S1	S2	S3	S4
Elongation (%)	6.09	3.45	1.83	2.99

This result may be due to the function of talc. When mixing talc into the PP, the aggregation of talc affects the crystallization process whereby talc acting as a nucleating agent uniformly distributed in the matrix resin PP and filling the matrix. However, an uneven talc particle size and the excessive incorporation of this filler may lead to filler agglomeration in the polymer matrix leading to the formation of micro-filter due to the difficulties in achieving a homogeneous dispersion of fillers. This phenomenon leading to weakening the interfacial bonding with the PP matrix, and creating some voids and defected structure of the composites, therefore, causing a reduction in tensile strength.

3.2. Flexural strength

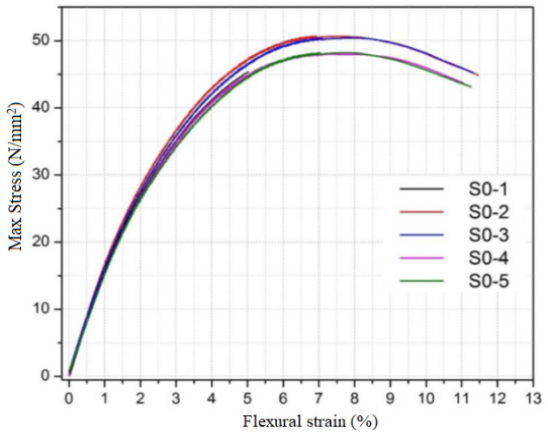
Figure 6 illustrates the stress-strain curves of the flexural test of S1-S4 samples. Figure 7 shows the average values of the flexural strength of these samples. The flexural strength values are 49.31 MPa, 47.99 MPa, 46.06 MPa, and 47.62 MPa corresponding to S1, S2, S3, and S4 samples, respectively. In general, increase the talc content leads to decrease the flexural strength, a similar result of Zihlif report [24].

Table 3 presents the elongations of S1-S4 samples. The average elongation values are 6.09%, 3.45%, 1.89%, and 2.99% corresponding to S1, S2, S3, and S4 samples, respectively. Increase the talc content mostly decrease the elongation value of the samples.

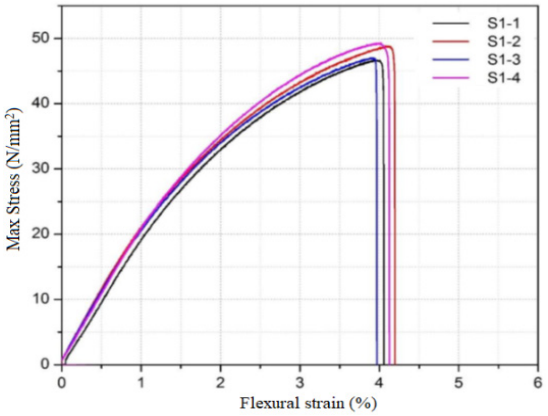
Table 3: Flexural strain of the samples

Sample	S1	S1	S3	S4
Flexural strain (%)	7.24	3.99	2.56	2.73

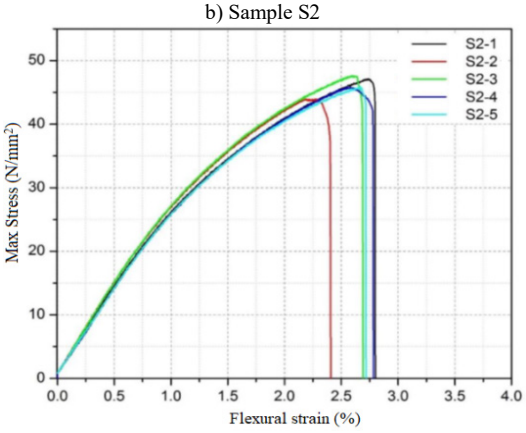
The reductions of the flexural strength and flexural strain are the result of the toughness and hardness increase. The introduction of rigid particles in a ductile matrix leads to lower elasticity/ deformability of the resulting compound.



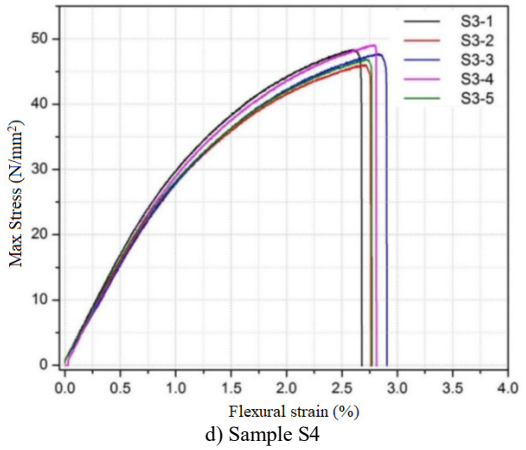
a) Sample S1 (100% PP)



b) Sample S2



c) Sample S3



d) Sample S4

Figure 6: Stress-strain behavior of flexural strength: (a) S1, (b) S2, (c) S3, and (d) S4

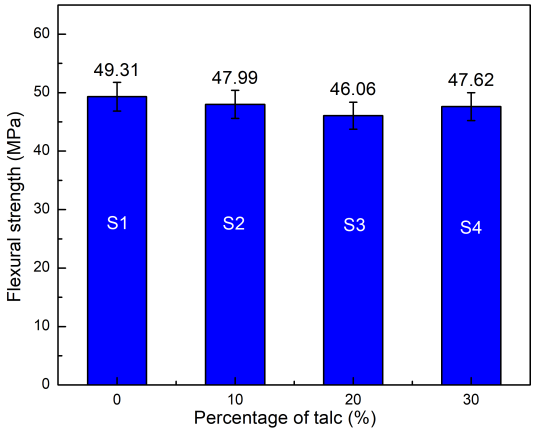
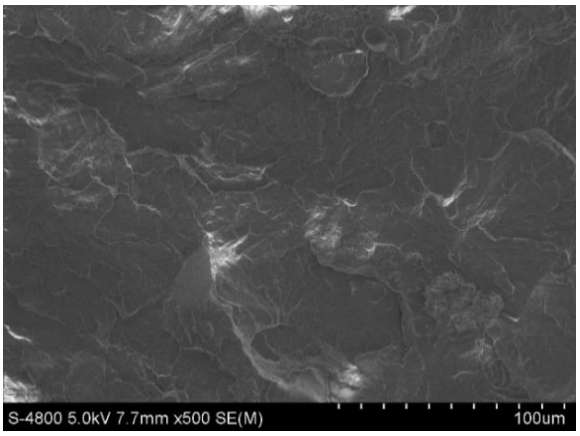
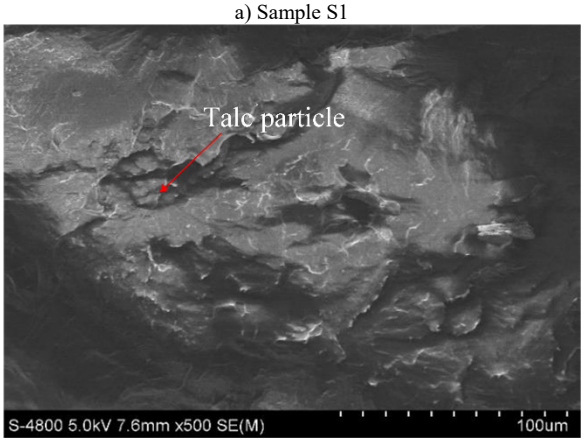


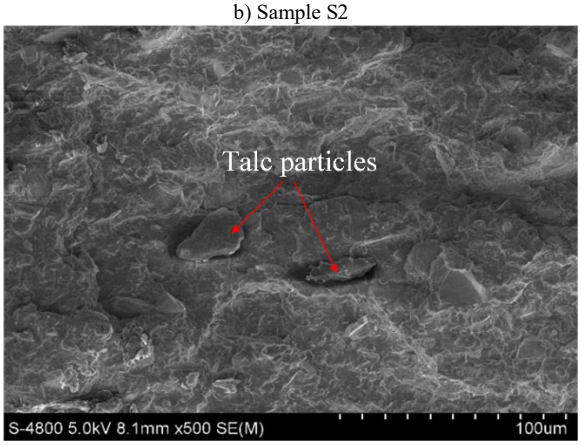
Figure 7: Flexural strength comparison of samples



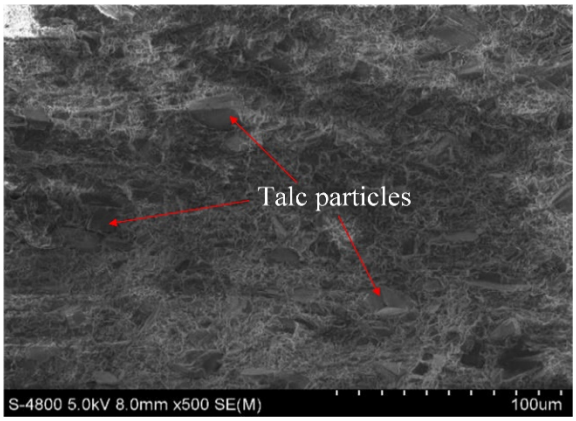
a) Sample S1



b) Sample S2



c) Sample S3



d) Sample S4

Figure 8: Fracture surface microstructure of samples: (a) Sample S1, (b) Sample S2, (c) Sample S3, and (d) Sample S4

3.3. Microstructure

The fracture surface microstructure of the samples is analyzed by SEM to have a more insightful view, as shown in Figure 8. Figure 8(b), (c) and (d) indicates the presence of talc powder in the polymer matrix. These figures also show the porosity between the talc powder and the PP matrix. Increase the talc will lead to an increase in the existence of this powder and porosity in the matrix.

All these factors contributed to the transfer of stresses from the matrix to the filler leads to reduce the bending properties. Furthermore, it was cleared that the fracture surface of PP matrix was clear and the number of the craze was limited, so the fracture was a brittle fracture. But the fracture surface of PP/Talc blends was rough, and the number of crazes increased with the increasing dosage of talc. Therefore, in this case, the fracture of PP matrix was a brittle fracture as the crystallization of the higher talc contain sample was higher, a similar result to Pan et al. [23].

4. Conclusions

In this study, the effect of 10%, 20%, and 30% talc to the tensile strength and flexural strength was studied. Some important points can be withdrawn as follow:

- The 10% talc powder sample posses the highest mechanical tensile strength of 25.91 MPa and flexural strength of 47.99 MPa among the talc powder samples. The 30% talc powder sample presents the lowest mechanical tensile strength, while the 20% talc powder sample represents the lowest flexural strength.
- The SEM results indicates the existence of talc powder and porosity between talc plates and PP substrate compared to the pure PP sample. Increasing the talc powder leads to a higher rate of porosity and a higher chance of brittle fracture.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We are acknowledged HCMC University of Technology and Education, Dong Nhan Phat Co., Ltd and Material Testing

Laboratory (HCMUTE). They gave me an opportunity to join in their team, accessed to the laboratory and research machines. Without their appreciated support it would not be possible to conduct this research.

References

[1] S. Yin et al. "A life cycle assessment of recycled polypropylene fibre in concrete footpaths". Journal of Cleaner Production, **112**, 2231-2242, 2016. https://doi.org/10.1016/j.jclepro.2015.09.073

[2] J.H. Lin et al. "Study on the processing technology and mechanical properties of nonwoven fabric composited by recycled PP selvedges". Advanced Materials Research, **287**, 2748-2751, 2011. https://doi.org/10.4028/www.scientific.net/AMR.287-290.2748

[3] A.L. Inacio et al. "Recycled PP/EPDM/talc reinforced with bamboo fiber: Assessment of fiber and compatibilizer content on properties using factorial design". Polymer Testing, **61**, 214-222, 2017. https://doi.org/10.1016/j.polymertesting.2017.05.022

[4] P.D. Bloom et al. "Development of novel polymer/quasicrystal composite materials". Materials Science and Engineering: A, **294**, 156-159, 2000. https://doi.org/10.1016/s0921-5093(00)01230-2

[5] W. Gacitua, "Polymer nanocomposites: synthetic and natural fillers a review". Maderas. Ciencia y tecnología, **7**(3), 159-178, 2005. http://dx.doi.org/10.4067/S0718-221X2005000300002

[6] R.N. Rothon, "Mineral fillers in thermoplastics: filler manufacture and characterisation". In Mineral Fillers in Thermoplastics I, 67-107, 1999. https://doi.org/10.1007/3-540-69220-7_2

[7] N. Shimpi, "Dispersion of nano CaCO3 on PVC and its influence on mechanical and thermal properties". Journal of composite materials, **44**(2), 211-219, 2010. https://doi.org/10.1177/0021998309344637

[8] T. Lam et al. "Effect of nanosized and surface-modified precipitated calcium carbonate on properties of $CaCO_3$/polypropylene nanocomposites". Materials Science and Engineering: A, **501**(1-2), 87-93, 2009. https://doi.org/10.1016/j.msea.2008.09.060

[9] S. Ozay "Preparation of organically modified $CaCO_3$ and its use in the manufacture of polypropylene composites". In ECCM15: 15th European Conference on Composite Materials, 24-28, 2012.

[10] L. Xie et al.. "Rheological and mechanical properties of PVC/CaCO3 nanocomposites prepared by in situ polymerization". Polymer, **45**(19), 6665-6673, 2004. https://doi.org/10.1016/j.polymer.2004.07.045

[11] B. Zhou et al.. "Mechanical and thermal properties of poly-ether ether ketone reinforced with $CaCO_3$". European polymer journal, **40**(10), 2357-2363, 2004. https://doi.org/10.1016/j.eurpolymj.2004.05.019

[12] P. Yang et al. "Design of self-healing microcapsules containing bituminous rejuvenator with nano-$CaCO_3$/organic composite shell: Mechanical properties, thermal stability, and compactability". Polymer Composites, **39**(S3), E1441-E1451, 2018. https://doi.org/10.1002/pc.24343

[13] H. Essabir, M. Ouadi, D. Rodrigue, "A comparison between bio - and mineral calcium carbonate on the properties of polypropylene composites". Construction and Building Materials, **134**, 549-555, 2017. https://doi.org/10.1016/j.conbuildmat.2016.12.199

[14] A.L.N. da Silva. "Mechanical and rheological properties of composites based on polyolefin and mineral additives". Polymer Testing, **21**, 57-60, 2002. https://doi.org/10.1016/s0142-9418(01)00047-2

[15] C. Onuoha, O.O. Onyemaobi, C.N. Anyakwo. "Effect of filler loading and particle size on the mechanical properties of periwinkle shell filled recycled polypropylene composites" Indian Journal of Materials Science, 2017.

[16] Katiyar J. K., Sinha S. K., Kumar, A.. "Friction and wear durability study of epoxy-based polymer (SU-8) composite coatings with talc and graphite as fillers". Wear, **362**, 199-208, 2016. https://doi.org/10.1016/j.wear.2016.06.011

[17] M. Ü. N. İ. R. Taşdemir, Ersoy S.. "Mechanical, morphological and thermal properties of HDPE polymer composites filled with talc, calcium carbonate and glass spheres", 2015.

[18] H.G. Premalal "Comparison of the mechanical properties of rice husk powder filled polypropylene composites with talc filled polypropylene composites". Polymer Testing, **21**(7), 833-839, 2002. https://doi.org/10.1016/s0142-9418(02)00018-1

[19] M. Abdouss et al. "Effects of compatibilization of oxidized polypropylene on PP blends of PP/PA6 and PP/talc". Journal of Applied Polymer Science, **92**(5), 2871-2883, 2004. https://doi.org/10.1002/app.20207

[20] R. Browning et al. "Effects of slip agent and talc surface-treatment on the scratch behavior of thermoplastic olefins". Polymer Engineering & Science, **46**(5), 601-608, 2006. https://doi.org/10.1002/pen.20507

[21] J.K. Katiyar "Friction and wear durability study of epoxy-based polymer (SU-8) composite coatings with talc and graphite as fillers". Wear, **362**, 199-208, 2016. https://doi.org/10.1016/j.wear.2016.06.011

[22] S.K. Urmila, "Study of talc filled polypropylene a concept for improving mechanical properties of polypropylene". International Journal of Research in Engineering and Technology, **2**(4), 441-445, 2013.

[23] W. Pan, X. Ding, H. Gu. "Effect of Talc on Crystallization and Properties of Polypropylene". American Journal of Materials Research, **2**(04), 35-43, 2015.

[24] A.M. Zihlif, G. Ragosta. "Mechanical properties of talc polypropylene composites". Materials Letters, **ll**, 10-12, 1991. https://doi.org/10.1016/0167-577X(91)90136-T

[25] M.A. Khan, M. Arif Siddiqui. "Effect Of Talc On Bending Strength Of Polypropylene Composites". International Journal of Mechanical and Production Engineering Research and Development, **3**(3), 41-46, 2013.

[26] H.G. Karian (eds). Handbook of polypropylene composites. Marcel Dekker INC, Second Edition, 295-323, 2003.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 577-583 (2020)
www.astesj.com
Second International Virtual Conference on Multidisciplinary Research 2020

ASTES Journal
ISSN: 2415-6698

Ontology-based Data Management Tool for Studying Radon Concentration

Félix Fernández-Peña[1,2], Alex Maigua-Quinteros[3], Pilar Urrutia-Urrutia[1], Diana Coello-Fiallos[*,1]

[1]*GIIA - SciTeM, Universidad Técnica de Ambato, 180103, Ecuador*

[2]*ESAI, Universidad Espíritu Santo, 092301, Ecuador*

[3]*MIVILSOFT s.a., 180103, Ecuador*

ARTICLE INFO

Article history:
Received: 09 August, 2020
Accepted: 10 October, 2020
Online: 20 November, 2020

Keywords:
Radon level measurement
Data Semantics
Web ontology
data management

ABSTRACT

Interpreting data for solving information needs demands an understanding of the semantics of row data. In this paper, an ontology-driven management of data of the presence of radon gas in soil is presented. The main contribution of this research relies on the formal definition of the semantic of the presence of radon in the dimensions of space and time. As a result, a dynamic data view generator for the observation of radon measurements was deployed without actual software programming but with the proper adjustment and instantiation of the ontology ViewOnto, a web ontology which was created for the formal description of data views. An experimental assessment of the proposal was carried out with data obtained from stone aggregate mines and civil constructions in Tungurahua. The usability of the resulting data management system was successfully validated with the criteria of experts on usability and researchers with experience in radon measurement.

1 Introduction

According to United States Environment Protection Agency estimates, radon is the number one cause of lung cancer among non-smokers and the second one overall [1]. Recent studies in the detection of radon have been focused on the definition of alternative methods for detecting radon presence in soil and air [2, 3], making the radon detection cheaper and accessible.

The study of collected radon data is important not only for warning about the humans' health risks because of its inhalation [4, 5] but for the prediction of earthquakes [6] and spontaneous combustion scenarios in coal mines [7]. In these quite different scenarios, time and space are recognised as fundamental dimensions of collected data in order to depict an actual throbbing picture of the presence and potential impact of radon [4]–[8].

Meanwhile, other variables like meteorological parameters and soil characteristics are intrinsically related because of their influence in radon concentration [3, 5, 7, 9]. This kind of studies is characterised by a challenging task: to derive meaningful insights from collected data [6, 7].

In the literature review that was carried out, it was not possible to find a common data model for characterising the presence of radon. On the contrary, depending on the research, different variables and even different measure units characterise each research. This lack of data normalization makes harder to work with retrieved data from the field. Requiring such a flexibility level on data management tools, it is important to note that developing adaptive collective structures of knowledge by ontology-based techniques has been successfully applied to different research areas [10]–[12].

This paper presents an ontology-based dynamic tool for the observation of radon measurements. The decision for an ontology-based solution is twofold. Firstly, a flexible solution is designed, adaptable to different objectives of research and changing mechanisms of measurement as the research goes on. Secondly, the tool does not need changes on code in order to adapt to changes on data to collect and data views required by researchers for analysing data in the space and time dimensions. Other important issues as user authentication and data access protection are out of the scope.

The proposed solution is based on *ViewOnto*, a web ontology which was defined for the formal description of the semantics of data views in different scenarios [10, 11]. Many domain ontologies have been created until now [11, 14, 17] but *ViewOnto* is the only one with this purpose of those available in the consulted literature. In this work, this ontology is taken into the dimensions of space and time. Its usability was validated by a usability test of the data views generated on the use case of the study of the presence of radon in

[*]Corresponding Author: Diana Coello-Fiallos, Carrera de Ingeniería Civil - UTA. Av. Los Chasquis y Río Payamino, Ambato, dc.coello@uta.edu.ec

mines and civil constructions in Ecuadorian highlands.

This paper is structured as follows. Section 2 depicts related works. *ViewOnto* is described in section 3 and the general architecture of the ontology-based data management tool for studying radon concentration is presented. Section 4 presents the results of the experimental evaluation of the tool in the scenario of the measurement of radon diffusion rate of building materials in stone aggregate mines and civil constructions in Tungurahua, in the Ecuadorian highlands. Finally, in section 5 conclusions and future work are discussed.

2 Related works

Tareen et al. proposed a method for predicting earthquakes based on the automatic detection of anomalous behaviour in radon time series data [6]. As result of their research, anomalous behaviour seen in radon time series data has been determined as an indicator of seismic activity [6].

Meanwhile, surface-based radon detection is a technique used for spontaneous combustion prediction in coal mines by studying the radon migration mechanism of overlying strata [7]. These studies show the importance of tools for supporting the understanding of data series of radon measures.

Other studies determine the acceptance of the influence of meteorological variables in radon concentrations [3, 7]. This actually leads to the estimation of soil radon gas with meteorological parameters, determining the necessity of taking into account meteorological parameters when making an analysis of radon influence on human life.

The results of the research of Verdonck et al. show that ontology-driven conceptual modeling has advantages for achieving higher quality models when compared to traditional conceptual modeling. Likewise, they concluded that it is advantageous to apply an ontology-driven concept modeling technique when having to model the more challenging and advanced facets of a certain domain or scenario [14].

Formalising the semantic of data is fundamental for its automatic processing [15]. Semantic aspects of data determines logic rules of knowledge base systems [12, 13, 15] whilst web ontologies are accepted as formal mechanisms for structuring data conceptualizations [13].

In another study, Verdonck et al. have studied the pragmatic quality of ontology-driven models and they have concluded that the notion of time is easier to comprehend in ontology-driven models where an individual endures through time and is regarded totally present at any moment in its lifetime. The same was confirmed when analysing how easy to understand was the formation of semantic relationships [13]. Meanwhile, Baumgartner et al. highlighted that making timely identification, resolution, and prevention of critical situations is fundamental for having a semantically richer knowledge model [15].

Wagih et al. focused on defining the semantic aspects of user's mobility for location recommendation systems by defining Ridology, an ontology model for exploring human behaviour trajectories in ridesharing applications [12]. Meanwhile, Sorokine et al. proposed the use of ontologies in the generation of scenarios for the purpose of testing the algorithms used to detect nuclear materials [16].

Another study proposing an ontology-driven representation of knowledge is the research carried out by Mantovani et al. for managing geological maps [17]. In their approach, a domain ontology in the context of geological knowledge named *OntoGeonous* was used in order to constrain the behaviour of a dynamic data entry tool to refer to the terminology conveyed by the taxonomic-axiomatic nature of the ontology [17]. None of related work analyzed during this research take into account the semantics of data in the space and time dimensions. Nevertheless, all these research results support the idea of choosing an ontology-driven concept model technique for carrying out the analysis of the impact of radon in human life both, on the space and time dimensions.

3 Description of the ontology-based data management tool

The software architecture of this solution is structured in three layers. Fig. 1 illustrates a general view of the interaction among layers of the proposed solution and with the user. The data entry and data views generators automatically generate a user interface by using the semantics of the instance of the *ViewOnto* ontology for radon observation.

Changes in the instance of the ontology automatically generate changes in the user interface. The user interacts with the data entry interface for inserting, updating and deleting data stored in the the structured database whilst he interacts with the data views generator for interpreting and learning from annotated data.

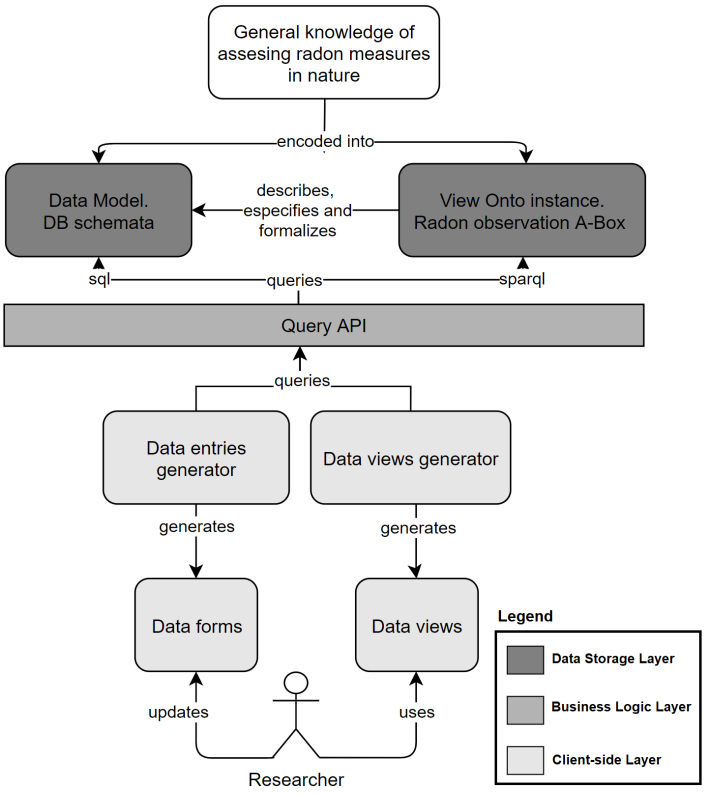


Figure 1: General view of the proposed solution.

The most important outcome with this implementation is that there is no need of updating the encoding of the application for changing the behaviour of the application. Those changes are automatically generated in user interfaces by changing the instance of *ViewOnto* that describes the semantics of data stored in the relational database. Below, the three layers of the architecture are described.

3.1 Semantics of the measurement of the presence of radon

The semantics of the database design in relational database models define the ultimate restrictions of datatype for each data item, unique index restrictions and the kind of relationships existing among data entities. Fig. 2 shows the relationships among the main data entities defined in the use case of this research for measuring radon's presence in the highlands of Ecuador. Other entities, like the User entity for specifying which users that have access to these data, were ignored in the diagram for improved clarity of the concept.

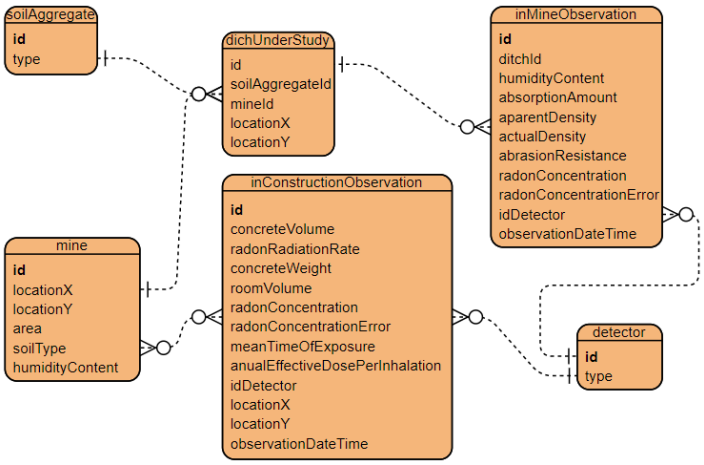


Figure 2: Main entities in use case's database design.

For this use case, in-mine and in-constructions observations were recorded. Different detector mechanisms were used and different soil aggregates were analysed. Both, time and space dimensions of data are fundamental for the study since it is required to be able to analyze the evolution of radon presence on observations. In this scenario, the semantic restrictions of the designed database were not enough but considered as a baseline guarantee for the integrity of managed data. Defining further semantic restrictions was necessary for building a data entry and report generator tool which adapts to changes in the semantics of data. Otherwise, the automatic generation of proper interfaces for end-users would be intractable.

For this purpose, modeling the semantics of data within a custom instantiation of the ontology *ViewOnto* was considered. Figure 3 shows the conceptual model of *ViewOnto* for formalising a database schemata.

Next, the main semantic definitions reused and extended from [10] in the instantiation are formally encoded using the description logic of the OWL language. For deeper insights about *ViewOnto*, the extended definition of the T-Box of *ViewOnto* and the specific A-Box of *ViewOnto* for radon observation are available int the online repository that accompanies this paper at Open Science Framework[1].

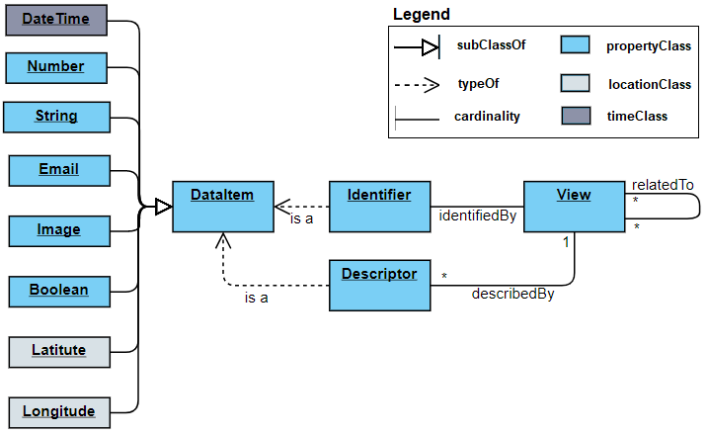


Figure 3: Conceptual model of *ViewOnto*.

$$V, D, I, S, R, F \sqsubseteq \top; V \sqcap S \sqcap R \sqcap D \sqcap I \sqcap F \equiv \bot \quad (1)$$

View V, *Descriptor* D, *Identifier* I, *DataItem* S, *Relationship* R and *Filter* F are the basic classes in the terminology of ViewOnto. Those classes are disjoint sets and this terminology corresponds to the original definition of *ViewOnto* by Fernández et al. [10].

$$\begin{aligned} S_{dt}, S_n, S_s, S_e, S_i, S_b, S_{la}, S_{lo} \sqsubseteq S; \\ S_{dt} \sqcap S_n \sqcap S_s \sqcap S_e \sqcap S_i \sqcap S_b \sqcap S_{la} \sqcap S_{lo} \equiv \bot \end{aligned} \quad (2)$$

Semantic datatypes are represented by the classes ***DateTime*** S_{dt}, *Number* S_n, *String* S_s, *Email* S_e, *Image* S_i, *Boolean* S_b, ***Latitude*** S_{la} and ***Longitude*** S_{lo}. All these data items are subclasses of *DataItem* S. Those data items in bold were added to the original definition of *ViewOnto* in order to take into account the space and time dimensions of data. The appropriate *DataItem* is decidable by the corresponding datatype for each data item in the database.

$$\begin{aligned} V \equiv & = 1 title \sqcap \\ & = 1 mappedTo \sqcap \forall isIdentifiedBy.I \sqcap \\ & = 1 isIdentifiedBy \sqcap \forall isDescribedBy.D \sqcap \\ & \forall isRelatedBy.R_v \end{aligned} \quad (3)$$

A *View* V is defined as the class with one, and only one, *title*, *mappedTo* and *isIdentifiedBy* properties. The property *title* is used for generating labels and menu options in the automatically generated data entry forms and report views. Meanwhile, the property *mappedTo* is used to link the *View* with the corresponding dataset in the database. V is identified by an *Identifier* I. Further, V may be related (or not) with one or more *Descriptors* D and *Relationships* between *Views* R_v by instantiating the *isDescribedBy* and *isRelatedBy* properties, respectively.

[1] https://osf.io/wj46b/

$$
\begin{aligned}
D \equiv &= 1title \sqcap \forall describedBy.S \sqcap \\
&= 1describedBy.S \sqcap \\
&= 1isAlternative \sqcap \forall isRelatedBy.R_D
\end{aligned} \tag{4}
$$

A *Descriptor* is defined as the concept D, which instantiates once, and only once, the roles *title*, *describedBy* and *isAlternative*. The properties *describedBy* and *isRelatedBy* only take value in the range of S and R_D instances, respectively. These restrictions are mandatory for a *Descriptor*. Figure 4 shows the connections among *DataItems* and *Descriptors* in the spatial dimension of the use case. This diagram was generated with the OntoGraf plugin for Protégé using the instance of *ViewOnto* that was built for studying radon concentration.

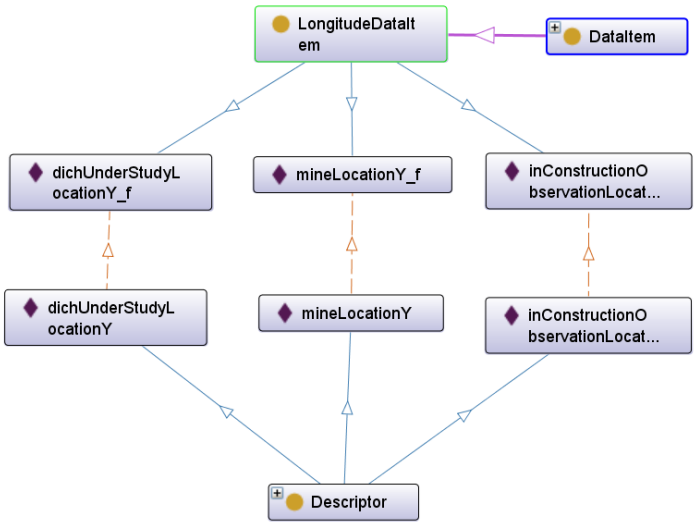


Figure 4: Graphical representation of *Descriptors* and *Dataitems* in the space dimension of the use case.

3.2 Implementation of the business logic layer

The business logic layer is exposed as a REST-API. A json encoder/decoder was integrated for making possible the communication with the client-side layer of the data management tool. Inside, the design of the business logic layer was threefold. One component was built for querying and updating the database whilst another component allowed to query the *ViewOnto* A-Box of the use case. Finally, an inference component that uses the Jena inference engine was implemented. This component inferences semantic rules based on the OWL encoding of *ViewOnto*. As Jena is an inference engine implemented in Java, all features of the business logic were exposed as web services in a TomCat Server. Figure 5 shows the structure of the proposed business logic layer.

Four main API-REST methods were proposed that conform to the management of the data of a data view. With different parameters, these methods allow to retrieve and to trigger CRUD actions over the data of a data view. Other eight API-REST methods allow to retrieve the list of available data views and data of inter-related data views, their descriptors and identifiers. Table 1 shows the structure of possible requests to the API-REST.

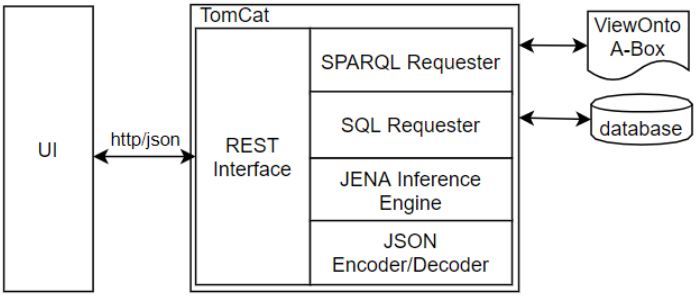


Figure 5: Graphical representation of the business logic layer.

Table 1: Examples of requests to the API-REST

Method	Path	Action
POST	/api/view /inMineObservation	records an observation.
POST	/api/view /inMineObservation/1	updates an observation.
GET	/api/view /inMineObservation	retrieves data of observations.
DELETE	/api/view /inMineObservation/1	deletes record *1* from observations.

With this implementation of the business logic layer, it is possible to have more than one client-side layer implementation. Without changes in the code, just by changing the specification of the *ViewOnto* A-Box, the behaviour of the business logic may adapt to changing scenarios. Data retrieved by the business logic flow to the client-side layer in an efficient way.

3.3 Implementation of the client-side layer

The implementation of the client-side layer is twofold. A generator of data entry interfaces and a report generator were implemented. An automatically generated interface for data entry allows the user to register annotations of radon measurements. Meanwhile, the report generator allows the user to navigate through data views for analysing collected data. All these interfaces are generated on request, based on the information stored in the instance of *ViewOnto*.

Figure 6 shows a data entry interface generated from the semantics stored in the *ViewOnto* A-Box that was defined for supporting the study of radon concentration. The definition of descriptors of longitude and latitude (*mineLocationX* and *mineLocationY*, as appear in figure 4), determines the rendering of a map component for choosing a location. The implementation of this map component allows to mark a point, a set of points, and a path. These functionalities are activated or not, taking into account the information that the user will enter in the data entry form. How to store data in the database is specify in the *mappedTo* data property of *ViewOnto*.

Meanwhile, other data entry controls (text fields and a combobox) are rendered for entering corresponding values. The rules for validating the values entered in each of these controls come from the instance of ViewOnto. The kind of control to render is chosen taking into account the *DataItem* that matches to each *Identifier* and *Descriptor* in the instance of *ViewOnto*. Other restrictions are

defined in the instance of *ViewOnto* using specific data properties. This is the case of mandatory fields in the data form, minimum and maximum values of each numeric field and the set of valid values (if this is the case).

The instance of *ViewOnto* includes rules for validation and error messages, just in case the rules are violated. The contribution with this work is that the semantic definition of form fields, and how to deal with them, is not mixed up with HTML and JavaScript. Instead, they are defined using the declarative syntax of OWL as they are stored in a web ontology.

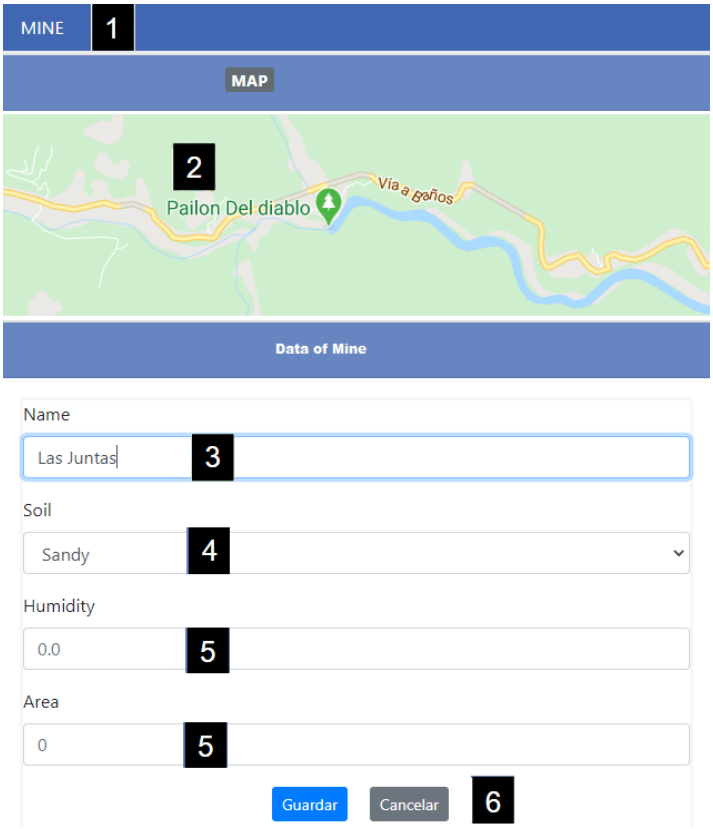


Figure 6: Sample of the prototype of data entry interface. (1) Title of the data view as defined in the instance of *ViewOnto*. (2) Interactive map component for the input of the location of new mine record. (3) Text fields for the input of the name of the mine. (4) Combobox field for the input of the type of soil of the mine. Valid values are specified in the instance of *ViewOnto*. (5) Text field for the input of humidity and area of the mine. Automatic validation of these fields restrict user input to numbers. (6) Buttons for saving and cancelling user data entry.

On the other hand, figure 7 shows the interface of the prototype when rendering information about observed mines. Location of the mines are rendered over a map canvas and the descriptors of the data view are shown in a tabular view. Each data view does not match necessarily with a table in the database. The amount of data views depends on the interests of the user. A list of related views is shown in the upper right corner of the interface when the options button is tap. The user may filter data per descriptor and he can show/hide descriptors, as well. Navigating to most related data views is possible by tapping buttons on the left side of each data row. In this case, it would be navigating to the ditches under study on a specific mine.

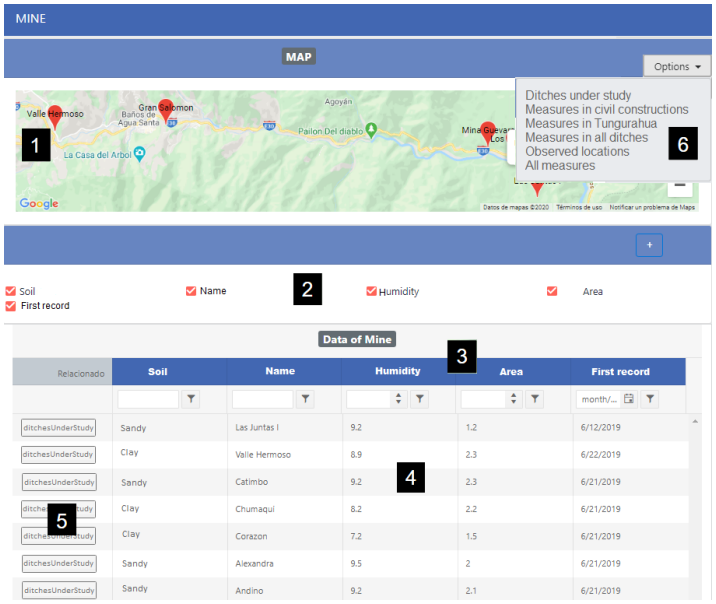


Figure 7: Sample of the prototype of report generator interface. (1) Interactive map component. (2) List of visible/hidden properties of the active data view. (3) Ordering criteria and filters applicable to each property. (4) Tabular data component. (5) Navigation access into related data views. (6) Links to other data views described in the instance of *ViewOnto*.

4 Analysis of Results

The experimental evaluation of the results show that it was possible to build a generator of data management tools. Whenever changes in the user interfaces would be needed, the code of the generator would not change. Only the instance of *ViewOnto* would be rewritten and, if necessary, the structure of the database would be modified. This way, the role of the programmer was changed. Instead of prescriptive programming, descriptive programming for coding the instance of the ontology and abilities on SPARQL and SQL are now on required for the maintenance of a generated data management tool.

This software was used for the deployment of a system for studying radon concentration. The user interfaces are automatically generated each time that the user requires to interact with the system. From the data model of figure 2, three data entities were managed in the space dimension, and two other data entities were managed in the time dimension. The navigation among *data views* was automatically generated taking into account the semantics of the *relationships* among data *views*.

The results were also assessed on an usability evaluation of the data management tool. Nielsen and Budiu suggest to have from 5 to 11 participants in a usability study [18]. Two experts in usability and three experts in the study of radon concentration participated in the assessment. All of the participants in this study have more that 15 years in the industry and 8 years of academic experience. In a first phase, the participants in the study were called for a virtual presentation on how to use the software. Then, all of the experts quantified the level of usability of the interfaces that the software would be able to generate. Their evaluation was made using Nielsen principles of usability [18].

Each expert determined if each usability principle was accom-

plished, not accomplished or not relevant. Their answers were anonymously collected. Then, feedback of the answers was given to the experts and a new qualitative analysis was carried out. A consensus on usability principles that were not relevant was found. A total of seven usability principles were considered as relevant for an ontology-based data management tool. Results were debated in a final public session of work and experts were asked to give a weight to each relevant usability principle. A Likert scale of 1-5 was used.

As a result of the study, it was found that experts agree that the interfaces generated with the software comply with high standards of usability. Table 2 shows a resume of relevant usability principles and the mean of the weight that experts gave to each of them.

Table 2: Results of the validating survey

Criteria	Mean value
I would use this tool	5.0
This tool seems easy to use	4.6
Using this tool, the understanding of data in the real world gets easier	4.8
This tool facilitates the control of the user over the data management process	5.0
This tool facilitates the user to freely manage data upon his needs	4.6
The options of this tool are clearly understandable	4.8
This tool is flexible and efficient for data management	4.8

5 Conclusions

This paper presented the implementation of an ontology-based data management tool. The semantic restrictions of the designed database were not enough but considered as a baseline guarantee for the integrity of managed data. Defining further semantic restrictions was necessary for building a data entry and report generator tool which adapts to changes in the semantics of data.

This data management tool 1) adapts to changes in the understanding of data, 2) dynamically integrates data in different measure units and 3) interprets the semantics of data for deciding its appearance in the user interface.

The space dimension of data views is expressed by its location on a map. On the other hand, the time dimension of data views is expressed by describing its nature -beginning, enduring and lasting of events-.

Both querying languages, SQL and SPARQL were necessary for communicating with the database and the ontology, respectively. Mapping the semantics with actual data in the data storage was possible by using the mappedTo property of the *ViewOnto* ontology.

Results were experimentally validated in the management of radon observations in the Ecuadorian highlands and experts expressed their acceptance of the usability of the tool. Using this generator in other scenarios is theoretically possible but, as future work, further research is needed in order to assess the capabilities of *ViewOnto* for expressing the semantics of managed data in other scenarios.

Acknowledgment This research was funded by the Research Project #PFICM019 "Arquitectura de Servicios Web Semánticos para la Integración de Fuentes de Datos Estructuradas de Proyectos Urbanísticos" of Universidad Técnica de Ambato.

References

[1] US Environmental Protection Agency: Health Risk of Radon (2020), https://www.epa.gov/radon/health-risk-radon.

[2] S. Elisio, L. Peralva, "Development of a low-cost monitor for radon detection in air" Nucl. Instrum. Meth. A, **969**, 164033, 2020. https://doi.org/10.1016/j.nima.2020.164033

[3] S. Li, H. Liu, G. Yang, S. Liu, R. Liu, Ch. Lv, "Detection of radon with biosensors based on the lead(II)-induced conformational change of aptamer HTG and malachite green fluorescence probe" J. Environ. Radioactiv., **195**, 60-66, 2018. https://doi.org/10.1016/j.jenvrad.2018.09.021

[4] V. Jobbágy, H. Stroh, G. Marissens, M. Hult, "Comprehensive study on the technical aspects of sampling, transporting and measuring radon-in-water" J. Environ. Radioactiv., **197**, 30-38, 2019. https://doi.org/10.1016/j.jenvrad.2018.11.012

[5] P. S. Miklyaev, T. B. Petrova, A. M. Marennyy, D. V Shchitov, P. A. Sidyakin, M. A. Murzabekov, M. N. Lopatin, "High seasonal variations of the radon exhalation from soil surface in the fault zones (Baikal and North Caucasus regions)" J. Environ. Radioactiv., **219**, 106271, 2020. https://doi.org/10.1016/j.jenvrad.2020.106271

[6] A. Tareen, K. Asim, K. Kearfott, M. Rafique, M. Nadeem, T. Iqbal, S. Rahman, "Automated anomalous behaviour detection in soil radon gas prior to earthquakes using computational intelligence techniques" J. Environ. Radioactiv., **203**, 48-54, 2019. https://doi.org/10.1016/j.jenvrad.2019.03.003

[7] B. Zhou, J. Wu, J. Wang, Y. Wu, "Surface-based radon detection to identify spontaneous combustion areas in small abandoned coal mine gobs: Case study of a small coal mine in China" Process Saf. Environ., **119**, 223-232, 2018. https://doi.org/10.1016/j.psep.2018.08.011

[8] D. Huxtable, D. Read, G. Shaw, "Measuring radon-222 in soil gas with high spatial and temporal resolution" J. Environ. Radioactiv., **167**, 36-42, 2017. https://doi.org/10.1016/j.jenvrad.2016.11.003

[9] H. Phong Thu, N. Van Thang, L. Cong Hao, "The effects of some soil characteristics on radon emanation and diffusion" J. Environ. Radioactiv., **216**, 106189, 2020. https://doi.org/10.1016/j.jenvrad.2020.106189

[10] F. Fernández-Peña, P. Urrutia-Urrutia, R. Cañete, R. Acosta-Sánchez, C. Yañez-Márquez, J. Nummenmaa, "A conceptual data model for the automatic generation of data views" Applied Mathematics and Information Sciences, **10**(4), 1331-1342, 2015. https://doi.org/10.18576/amis/100412

[11] A. Maigua-Quinteros, W. Perez-Nata, F. Fernandez-Pena, P. Urrutia-Urrutia, "Fostering Exploration of Georeferenced Data Semantics for Supporting E-government Initiatives" in 5th International Conference on eDemocracy and eGovernment, Ambato Ecuador, 2018. https://doi.org/10.1109/ICEDEG.2018.8372363

[12] H.M. Wagih, H.M.O. Mokhtar, "Ridology: An Ontology Model for Exploring Human Behavior Trajectories in Ridesharing Applications" Studies in Systems, Decision and Control, **295**, 553-567, 2021. https://doi.org/10.1007/978-3-030-47411-9_30

[13] M. Verdonck, F. Gailly, S. de Cesare, "Comprehending 3D and 4D ontology-driven conceptual models: An empirical study" Studies in Systems, Information Systems, **93**, 101568, 2020. https://doi.org/10.1016/j.is.2020.101568

[14] M. Verdonck, F. Gailly, R. Pergl, G. Guizzardi, B. Martins, O. Pastor, "Comparing traditional conceptual modeling with ontology-driven conceptual modeling: An empirical study" Information Systems, **81**, 92-103, 2019. https://doi.org/10.1016/j.is.2018.11.009

[15] N. Baumgartner, W. Gottesheim, S. Mitsch, W. Retschitzegger, W. Schwinger, "BeAware!—Situation awareness, the ontology-driven way" Data Knowl. Eng., **69**(11), 1181-1193, 2010. https://doi.org/10.1016/j.datak.2010.07.008

[16] A. Sorokine, B. G. Schlicher, R. C. Ward, M. C. Wright, K. L. Kruse, B. Bhaduri, A. Slepoy, "An interactive ontology-driven information system for simulating background radiation and generating scenarios for testing special nuclear materials detection algorithms" Eng. Appl. Artif. Intel., **43**, 157-165, 2015. https://doi.org/10.1016/j.engappai.2015.04.010

[17] A. Mantovani, F. Piana, V. Lombardo, "Ontology-driven representation of knowledge for geological maps" Comput Geosci., **139**, 104446, 2020. https://doi.org/10.1016/j.cageo.2020.104446

[18] J. Nielsen, R. Budiu, Mobile Usability, New Riders Press, 2012.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 584-587 (2020)

www.astesj.com

Second International Virtual Conference on Multidisciplinary Research 2020

ASTESJ

ISSN: 2415-6698

Strategic Plan for the Achievement of the Competitiveness of Small Companies with Respect to Large Ones

Alan Guadalupe Ochoa Navarro*, Juan De Dios Cota Apodaca, Dario Fuentes Guevara

National Technological Institute of Mexico / I.T. Los Mochis, 81259, Mexico

ARTICLE INFO

Article history:
Received: 08 July, 2020
Accepted: 07 October, 2020
Online: 20 November, 2020

Keywords:
Strategic plan
Competitiveness
MSMEs
Large companies

ABSTRACT

The main objective of this research is to analyze the impact that the implementation of a strategic plan would have on achieving the competitiveness of the micro, small and medium enterprises (MSMEs) that are dedicated to bottling, marketing and distributing water in the city of Los Mochis. It is intended to offer a tool to this type of company that allows them to be competitive against large transnational companies, which currently dominate the market. For this, a mixed methodology was used, presenting a descriptive-correlational scope with a non-experimental cross-sectional design; In which a population of 21 MSMEs was analyzed, through the application of a measurement instrument, to quantify the levels of competitiveness that they currently present. The results of the investigation showed a low performance in the competitiveness of these businesses, exposing similarities between their deficiencies, highlighting the mismanagement and ignorance of the economic environment in which they are found.

1. Introduction

Talking about the importance of the micro, small and medium enterprises (MSMEs) is essential to understand the economic growth and development they generate in the world and the positive impact they provide to the global economy, since they generate jobs, pay social security contributions, and also boost local economies; In the opposite case, the large transnational companies that, although they contribute the aforementioned, also return their profits to their country of origin. Likewise, according to [1], more than 99.8% of the companies in Mexico are micro, small or medium-sized, while in [2] it is indicated that MSMEs contribute 60% of the GDP of that country.

However, the low level of competitiveness that these small businesses present, represent into a relatively short life span, (since according to [3], only 25% of these companies survive the first 2 years after opening, while the average life span is around 7 years), in which they fail to compete with the large transnational companies. In other words, the vast majority of these small businesses fail and subsequently disappear from the market. That is why the need arises to offer tools and methodologies that, through interdisciplinary actions, allow MSMEs to achieve competitiveness in the area.

For the above, it is essential to understand the low competitiveness of MSMEs and the context in which they currently compete, as well as the critical factors that directly influence their low performance, and based on this, determine which strategies are the most appropriate in the design of the strategic plan, given the imminent crisis that threatens these small businesses with their disappearance.

2. Method and Materials

2.1. Type of investigation

According to the problem detected, it was determined that the research pursues a mixed route since, through both quantitative and qualitative tools, it aims to interpret the scenarios or contexts in which MSMEs are currently, understanding the importance they represent for the economy from any country.

2.2. Design and scope of the investigation

The situation for many MSMEs is critical, as they do not show any signs of competitiveness in the market in which they participate, compared to their direct competitors, therefore they do not represent any risk for large companies.

Therefore, this research presents a descriptive-correlational scope with a non-experimental cross-sectional design, on the implementation of a strategic plan in the region's MSMEs, since as in [3], it will be carried out in a period of unique time, trying to shorten the gap between the demand presented by the large brands of bottled water that operate in the region, and the micro and small companies that are dedicated to marketing a similar product. Also,

*Corresponding Author: Alan Ochoa, alanochoa1115@outlook.com

only those MSMEs that are dedicated to the bottling and marketing of water in the region will be taken into account.

2.3. Hypothesis

The information to formulate the hypothesis arises from the diagnosis obtained, as well as from the reviewed literature, which helps to suppose scenarios in relation to the results that will be obtained, and the significant contributions that the strategic plan will have to the competitiveness of the companies. MSMEs, as well as the adaptability that these businesses and their possible contexts may present.

That said, the following research hypothesis is established:

- Hi. The design of a strategic plan will allow small companies to become competitive with large companies.

2.4. Design of the instrument

The instrument used to measure competitiveness in MSMEs in the region was a questionnaire with a Likert-type scale, using the dimensions and indicators of the competitiveness map of the Inter-American Development Bank (IDB) for their reliability, in addition to being a tool widely used by other researchers according to the literature reviewed.

This instrument was used during the development of this research and according to [4], this tool was designed to specifically identify the deficient areas that directly influence the low level of competitiveness of MSMEs, and with this, develop corrective strategies that allow to significantly improve the competitiveness of the diagnosed company. The application consisted of a structured questionnaire used by [5] with a Likert scale, with a total of 103 questions focused on each of the dimensions and indicators established by the IDB competitiveness map.

In addition, according to [6], the Inter-American Development Bank has always contributed to the socioeconomic development of Latin America and the Caribbean, promoting research and supporting companies, which is why it is considered to be a fairly reliable tool, based on the experience of such an organization.

It should be mentioned that as in [7], the dimensions and their respective indicators according to the IDB Competitiveness Map are shown in the Table 1.

Table 1: Dimensions and indicators to measure competitiveness according to the IDB Competitiveness Map:

Variables	Indicators
1. Strategic planning	1.1. Strategic planning process 1.2. Strategy implementation
2. Production and purchases	2.1 Planning and production process 2.2. Process capacity 2.3 Maintenance 2.4 Research and development 2.5 Provisioning 2.6 Inventory management 2.7 Location and infrastructure
3. Quality Assurance	3.1 General aspects of quality 3.2 Quality systems
4. Marketing	4.1 National market: marketing and sales 4.2 National market: services 4.3 National market: distribution
5. Accounting and finances	5.1 Cost monitoring and accounting 5.2 Financial administration 5.3 Legal and tax regulations
6. Human resources	6.1 General aspects 6.2 Training and promotion of personnel 6.3 Organizational culture 6.4 Industrial health and safety
7. Environmental management	7.1 Environmental policy of the company 7.2 Strategy to protect the environment 7.3 Awareness and training of staff on environmental issues 7.4 Waste management
8. Information systems	8.1 System planning 8.2 Inputs 8.3 Processes 8.4 Outputs

Source: Own elaboration, based on Saavedra, Milla and Tapia [1].

Subsequently, once the vulnerable areas of these small businesses have been identified, a strategic plan will be made based on the comprehensive strategic management model [8], which allows them to take the necessary actions, so that they achieve competitiveness with respect to their major competitors.

2.5. Population

The population obtained for this research is 21 of MSMEs that are dedicated to bottling and marketing of water in the city of Los Mochis. Likewise, the unit of analysis for this research consists solely of the owners of said MSMEs, due to their knowledge in all areas and their authority to make decisions.

3. Results

The results obtained yielded some of the following data:

- Of the 21 MSMEs diagnosed, only 2 businesses acknowledged having used some type of strategic planning in the last 2 years, which represents 9.52% of the total companies analyzed, while 66.67% said they have never used strategic planning, running your business solely empirically, as shown in Figure 1.

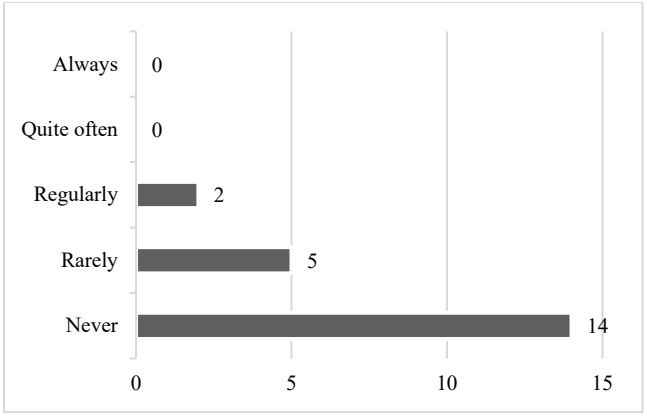


Figure 1: MSMEs that have used strategic planning in the last 2 years.

This indicates that the MSMEs considered for this research are usually empirically guided, so there is no basis in the decisions they make. This indicates that a series of actions can be offered with analytical support, on what actions to take and find the advantages of implementing them.

- From the same diagnosis, it was found that 61.90% never implemented a SWOT analysis to identify their strengths and weaknesses, or to better understand the environment in which they compete and their direct competition. A graph with percentages is shown in Figure 2.

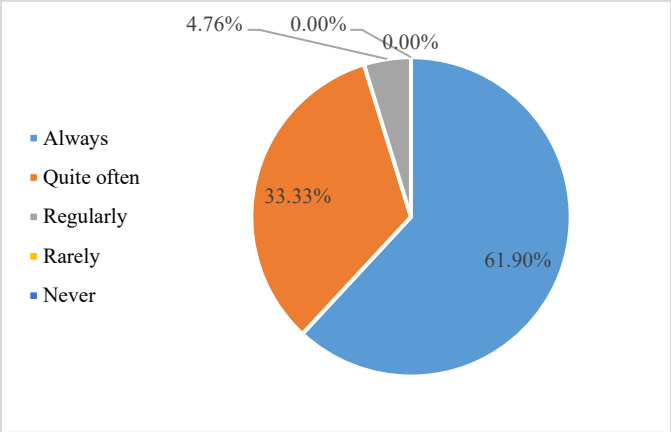


Figure 2: MSMEs that have used SWOT

This result shows that 13 of the 21 MSMEs do not know what their opportunities and threats are in the environment in which they currently compete, and they do not know the strengths they have, or at least not in an analytical and conscious way.

- A very important factor is the care of the work equipment, and in this sense, 17 of the 21 MSMEs admitted not to carry out preventive care to the production teams, which represent into 80.95% of the total of the companies studied. This percentage is expected to decrease to at least 50% since this point is considered to be extremely relevant, since without the production equipment in optimal conditions, the operation of the business may be affected in the short term. It can be seen in Figure 3.

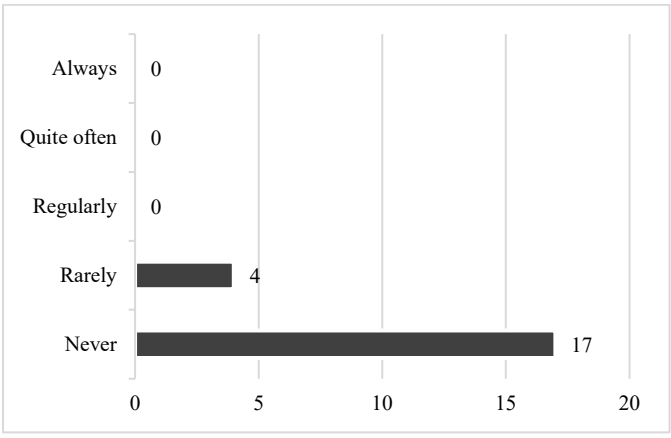


Figure 3: Preventive maintenance program for all equipment and machinery in MSMEs

Based on the previous graph, it can be seen that most MSMEs do not take optimal care of work tools, which will affect these businesses a lot, since the operation and production of the product depends on this.

- There is mismanagement of inventories by diagnosed MSMEs, since 76.19% of these businesses do not have a defined storage system, presenting rotation and shortage problems at very specific times, which significantly affects the image of these businesses. Likewise, the percentages of this item can be seen in Figure 4.

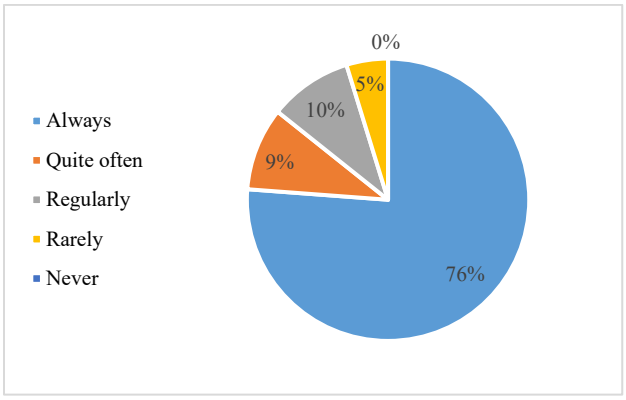


Figure 4: MSMEs that use management and administration systems in their inventories

One of the alarming results is that 16 of the 21 MSMEs diagnosed do not use or consider any inventory system to safeguard their products, which has led to shortages at some point. This can generate a negative image that severely affects MSMEs, so it is evident that it is necessary to generate a methodology for the storage of products, without having excess inventories.

- It was determined that the design of the strategic plan will follow the structure of the comprehensive strategic management model of Fred David and Forest David [2], since it is considered a fairly complete and adaptable tool, which can generate quite significant results for the achievement of competitiveness in MSMEs in the region. In Figure 5, each of the factors that make up the comprehensive strategic management model is shown in detail.

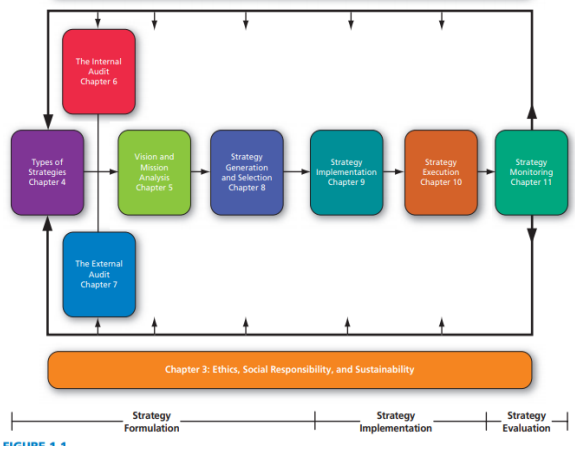


Figure 5: Fred David's Comprehensive Strategic Management Model
Source: Obtained from the book "Strategic Management A Competitive Advantage Approach, Concepts and Cases" [3].

In this model, it is possible to obtain a vision with much more clarity about the stages of the strategies to develop; starting from the type of strategy required according to the needs detected, to then formulate a series of actions and methodologies, on which an analysis of each will be carried out and select the most appropriate according to the context in which each of MSMEs.

Once the strategies have been selected, they are carried out, monitoring the results to obtain an assessment and adjust the strategies implemented according to the evaluation obtained.

4. Discussion

Based on the obtained results, the relationship determined by the hypothesis is accepted, which establishes that the design of a strategic plan will make a significant contribution to achieving the competitiveness of MSMEs with respect to their large competitors. Likewise, the data found so far, show an important relationship with respect to the results shown by [6] who point out that strategic planning has a significant influence on the organizational management of MSMEs and that it can increase their competitiveness substantially.

On the other hand, in contrast to what indicate in [1], who mention that the three strategies most used by family business entrepreneurs are to improve product quality, provide a pleasant service or through the development of a new product or service that is related to the image of the business that promotes a better perception among its consumers; However, the results of this research do not agree with the previously mentioned strategies according to the diagnosed analysis unit, who do indicate that it is imperative to improve the inventory storage and control system, since considerable losses are generated in certain periods of time.

Likewise, given the integrity of the strategic management model of Fred David and Forest David, this research is expected to be too beneficial for all those MSMEs that implement the strategic plan and are in search of achieving competitiveness.

5. Conclusions

The lack of competitiveness is something that has affected MSMEs throughout their existence, however, not having an action plan represents a disadvantage compared to their large competitors, which does not allow them to make better use of their own resources in a more adequate way, since the low performance of competitiveness is generated mainly by the ignorance of the owners themselves, to make smarter decisions or initiatives on strategies that allow them to have greater control of their business, and in turn, help them to extend their life time in the market and allow them to grow achieving greater competitiveness.

Currently, MSMEs in the region show almost zero competitiveness and with many areas of opportunity, since they present quite a few deficiencies, among which stand out the lack of care for their equipment, the mismanagement of inventories, in addition to the worrying ignorance they have of the environment in which they compete; However, if they manage to adapt effectively to the strategies developed, they will make better use of their resources and considerably improve their processes.

For local economies, MSMEs are extremely important since they generate jobs and improve the quality of life of many people, so it is essential to promote a more equitable market, in which MSMEs can compete against large companies, achieving a greater participation in the market and generating a significant positioning among consumers in the region.

Acknowledgments

A mention of gratitude is made to all MSMEs in the region, for their willingness to carry out this research, for their approval to carry out the measurement instrument, as well as the facilities provided to know their processes and their admission to implement the relevant strategies that are developed from the diagnosis obtained; Also, a special thanks is offered to all the people who influenced the development of this research.

References

[1] M. Valenzuela, M. Monserrat., T. López, T. Virginia Guadalupe & Moreno, M. Luis Ramón. La planeación estratégica como factor de competitividad en las empresas familiares del sector comercial del Valle de Mexicali, (article in Spanish with an abstract in English). Universidad Autónoma de Baja California, México. European Scientific Journal, 2015.

[2] G. Saavedra, M. Luisa, Milla, T. Sindy Orieta & Tapia, S., Blanca. Determining the competitiveness of SMEs in the micro level: The case of the Federal District, Mexico, (article in Spanish with an abstract in English). Revista FIR, FAEDPYME International Review, **2**(4), 38–52, 2013.

[3] F. David, & R. David, "Strategic Management A Competitive Advantage Approach, Concepts and Cases. Sixteenth Edition. Editorial Pearson, 2017.

[4] H. Roberto & M. Christian, "Metodología de la investigación, las rutas cuantitativa, cualitativa y mixta. Primera edición. Mc-Graw Hill, 2018.

[5] L. Montoya, S. Eduardo. "Aplicación del Mapa de Competitividad del BID a empresas del fondo emprender, (thesis in Spanish with an abstract in English). Universidad Tecnológica de Pereira, 2008.

[6] E. Mora-Rapieri, M. Vera-Colina, Z. Malgarejo-Molina "Strategic planning and competitiveness levels of SMEs in the business sector in Bogotá, (article in Spanish with an abstract in English). Revista Estudios Gerenciales, pág. 79-87, 2015. Doi: https://doi.org/10.1016/j.estger.2014.08.001.

[7] G. Saavedra, L. María & T. Milla, "Sindy Orieta. The MSME competitiveness at the micro level: The case of Querétaro, Mexico, (article in Spanish with an abstract in English). En-Contexto Revista de Investigación en Administración, Contabilidad, Economía y Sociedad, Institución Universitaria Tecnológico de Antioquia, Colombia, 2017.

[8] S. Javier, P. José & Uzcátegui, Carolina de los Ángeles. Metodologías de Medición de la Competitividad. II Congreso Internacional Virtual sobre desafíos de las empresas del siglo XXI, 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 588-593 (2020)

www.astesj.com

Second International Virtual Conference on Multidisciplinary Research 2020

ASTESJ

ISSN: 2415-6698

"Traffic Congestion Triangle" Based on More than One-Month Real Traffic Big Data Analysis in India

Tsutomu Tsuboi*

Global Business Development Office, Nagoya Electric Works, 490-1294, Japan

ARTICLE INFO

Article history:
Received: 27 April, 2020
Accepted: 07 October, 2020
Online: 20 November, 2020

Keywords:
Traffic Flow Analysis
Traffic Congestion
Traffic Occupancy

ABSTRACT

This research describes new traffic congestion analysis method "Traffic Congestion Triangle" based on more than one month real traffic big data analysis in India. The location of the research is one of typical economical growing cities Ahmedabad in Gujarat state of India. The traffic congestion becomes more serious issues in most developing countries and the congestion causes so called negative impact of environment destruction, un-necessity fuel consumption, health problem, economical loss and fatality by traffic accidents. Therefore there are lots of challenges for transportation under many projects focused on "smart city" development. The traffic analysis and theory have been developed after 1950s and it has helped many countries. However there are still lot of unknown condition in traffic analysis of developing countries.

This research started from 2015 and it enables to get real traffic flow data. This manuscript is described not only the typical traffic flow characteristics such as daily traffic volume but also introduces unique traffic congestion characteristics by using "occupancy" which has triangle shape characteristics as traffic congestion daily trend.

1. Introduction

This study shows new traffic congestion analysis in one of typical mega city in India. In general, it is hard to make traffic analysis especially in development countries where there is economical difficulty for infrastructure improvement under higher demand of transportation by strong economic growth. Author has a chance to proceed traffic management system deploy in Ahmedabad city of Gujarat state of India since 2015 by Japan International Agency support program for the reduction of traffic congestion. The traffic management system consists of 14 traffic monitoring Cameras along their main streets and traffic condition indication by the 4 electric sign boards so called Variable Message Sign or VMS board in the city. In terms of Ahmedabad city, the population of the Ahmedabad is about 6 million [1] and the number of four wheelers is 2.6 million, that of two wheelers is 15.8 million in 2018 [2]. The four wheelers growth rate is 65% compared with in 2010. And Ahmedabad is an economic and industrial hub of India and is the largest city in Gujarat stat. Therefore there are heavy traffic congestion and local government Ahmedabad Municipal Corporation faces sever condition for urban transportation management.

The traffic congestion analysis itself has been established in traffic theory by several report recently based on real traffic flow observation [3, 4]. From those related study, the traffic congestion is annualized by traffic flow parameter such as traffic volume, traffic density, traffic speed, and traffic occupancy. In general, the traffic congestion is detected from the traffic speed because the drivers feel slow traffic by their driving speed itself. But there is some cases traffic slow movement under slow traffic speed but it is not under congestion. Therefore it is difficult to define its traffic congestion between human feeling and physical vehicle occupation on the driving road. From those related studies, it is not enough to use just one traffic parameter like vehicle speed, or traffic density etc. In another word, it is necessary to manage multiple traffic parameter for understanding traffic congestion and it becomes more important especially for the developing countries. In [5], the author has analysis traffic congestion by using traffic density and space headway parameter but the measurement is only four days in Chennai in India. And M.Goutham and B.Chanda shows vehicle probe data in terms of traffic service by traffic volume and speed in Hyderabad of India based on Indian Road Standard IRC-106-1990 [6].

In this study, the fundamental traffic flow characteristics from one month measurement in the city of India is shown compared with the traffic flow theory and how it is difficult to analyze for its traffic congestion condition. Then it focuses on traffic volume and traffic occupancy relationship as the traffic congestion parameter

*Corresponding Author: Tsutomu Tsuboi, 29-1 Mentoku Shinoda Ama Aichi, +81-902738-3496 & t_tsuboi@nagoya-denki.co.jp

and proposes "traffic congestion triangle" from this relationship as the new traffic congestion parameter.

2. Traffic Congestion Theory and Measurement

In the traffic flow theory, there are three major characteristics as the fundamental traffic flow characteristics. One is traffic density (k) to vehicle speed (v)—$k-v$ curve. The $k-v$ curve shows the typical traffic condition instinctively. When the traffic becomes crowded, the vehicle speed is reduced and one of typical observation model is Greenshields equation [7]. The second one is traffic density (k) to traffic volume (q)—$k-q$ curve. The $k-q$ curve explains the traffic flow relationship between physical road length (vehicle number / km) and time scale (vehicle number / hour). The third one is traffic volume (q) to vehicle speed (v)—$q-v$ curve. Those three characteristics are so called the traffic flow fundamental characteristics of the traffic flow. The next section explain the traffic congestion condition by using this theoretical traffic flow characteristics. And then it is compared with the traffic theory and the result of actual measurement in the city of India.

2.1. Traffic Flow Theory

Based on the traffic theory, Greenshields equation is given by Equation (1).

$$v = v_f\left(1 - \frac{k}{k_j}\right) \tag{1}$$

v_f : free speed

k_j : jam traffic density

And the relationship between traffic volume (q) and traffic density (k) is provided by Equation (2) by traffic flow conservation law [8].

$$q = k \times v \tag{2}$$

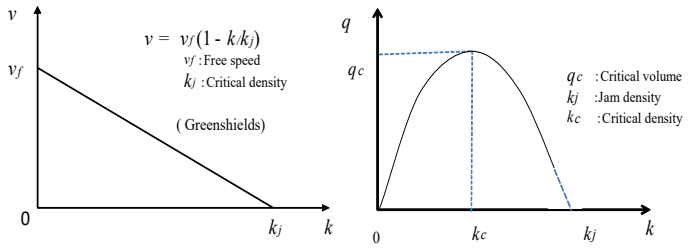
From Equation (1) and (2), Equation 3 is taken,

$$q = v_f\left(1 - \frac{k}{k_j}\right)k \tag{3}$$

Then Equation (4) is taken by transforming Equation (3).

$$q = -\frac{v_f}{k_j}\left(k - \frac{k_j}{2}\right)^2 + \frac{v_f k_j}{4} \tag{4}$$

From Equation (1) and (4), the $k-v$ curve and $k-q$ curve are shown in Figure 1 (a) and (b).



(a) $k-v$ curve (b) k- q curve

Figure 1: Fundamental Traffic Flow Curve

In terms of the traffic congestion from the traffic flow curve such as $k-q$ curve, the traffic congestion area in the graph is shown in Figure 2.

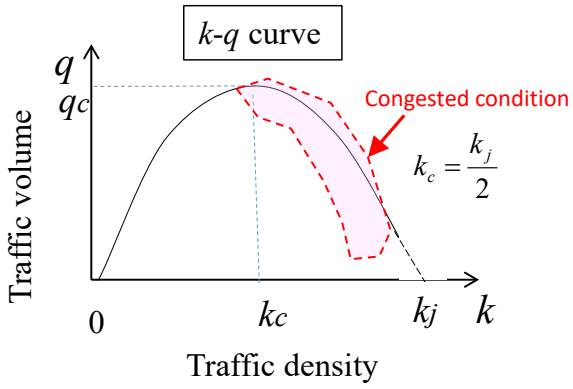


Figure 2: The traffic congestion area in $k-q$ curve

The traffic congestion occurs after the critical traffic density (k_c) and eventually its traffic volume declines from this point. After reaching traffic jam density (k_j), the traffic volume becomes theoretically zero. In case of the $k-v$ curve, the traffic congestion is located between k_c and k_j as shown in Figure 3.

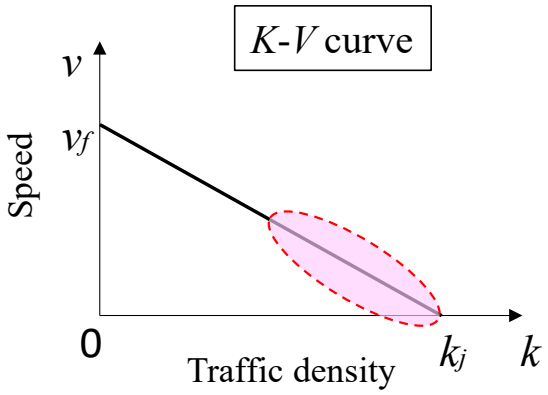


Figure 3. The traffic congestion are in k – v curve

2.2. Traffic Flow Measurement

In this research, the measurement field is used in Ahmedabad City of Gujarat State in India, where is one of big city in the state as mentioned in introduction. The traffic data is collected by 14 CCTV monitoring Cameras which are installed in the west side of the city as shown in Figure 4. In Figure 4, Camera#1 means CCTV number 1 and Camera#1 has one CCTV with pole on the street. The VMS#1 means VMS (Variable Message Sign board) number 1 and it also has CCTV with Traffic Sign board. The traffic data is measured by the CCTV and measures traffic flow data such as number of vehicles, average vehicles speed, traffic density. The traffic flow data was collected in June 2015 one month data by every minutes. The total traffic flow data for each CCTV becomes more than 40,000 points and author took 11 Camera data through Camera number 1 to 10 and VMS number. In this paper, we take eleven CCTV data because VMS#1 and VMS#2 data are missing during measurement by communication network error. In this traffic flow observation by traffic monitoring camera, we use

major road traffic flow monitoring because the condition of intersection is different traffic flow with traffic signal and pedestrians. The intersection analysis is our next study.

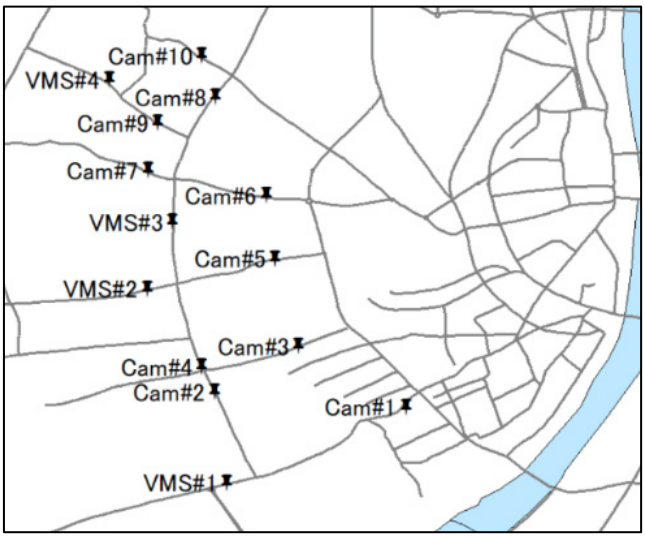
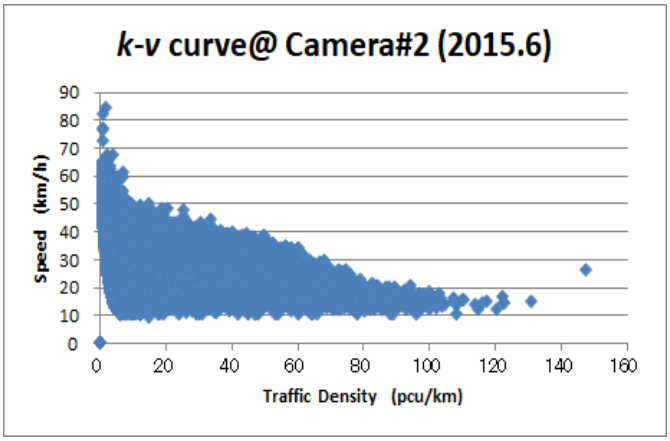
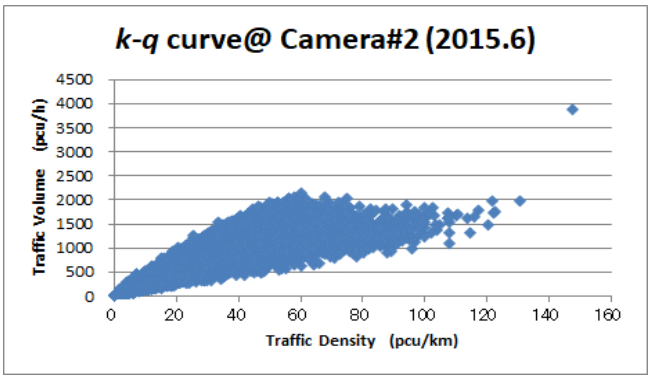


Figure 4: Traffic monitoring Camera location in Ahmedabad city.

Among the result of measurement point, it takes the traffic flow data from Cameraera#2 in June 2015. The $k - v$ curve and $k - q$ curve are shown in Figure 5. Each data in Figure 5 shows generalized data by lane. In case of Cameraera#2 road, there are two lanes.



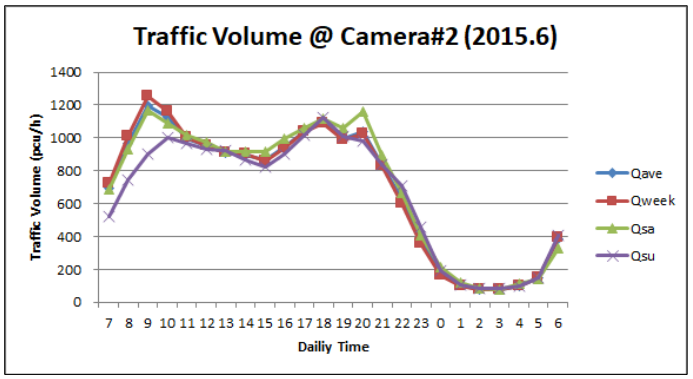
(a) k – v curve



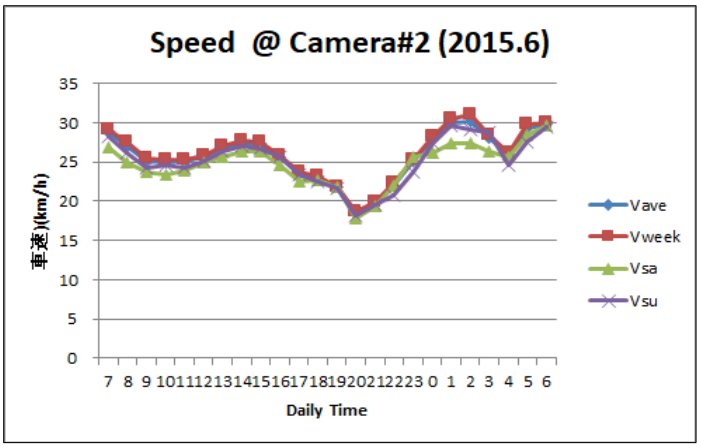
(b) $k - q$ curve

Figure 5: The measurement result of the traffic flow characteristics of Camera#2

From those two graphs, it is hard to define where traffic congestion point is. Therefore, the daily traffic condition observation is shown in Figure 6 as daily basis traffic volume change and vehicle speed change from 7:00 am to 6:00 am next day.



(a) Traffic Volume daily basis change at Camera#2



(b) Vehicle Speed daily basis change at Camera#2

Figure 6: Traffic Flow Daily Basis Change at Camera#2

From Figure 6, it is clear that the evening time from 19:00 to 21:00 is most congested time zone because the vehicle speed is quite low (10km/h) than that of other time zone. But it is hard to know this condition from the traffic flow characteristics. In the next section, one of other traffic flow parameter is introduced to define traffic congestion condition.

3. Traffic Congestion Parameter

There are several traffic flow parameter such as occupancy, headway, and gap in the traffic monitoring data. Among those parameter, the occupancy is one of parameter which shows traffic condition. In this section, the occupancy parameter is described and it shows the relationship between occupancy and traffic volume. And after this, it is introduced new traffic congestion parameter "Traffic Congestion Triangle".

3.1. Occupancy (OC)

The occupancy is one of traffic flow parameter to explain how vehicle is congested on the road how much vehicle occupies their road. From traffic flow theory, (OC) is defied by Equation (5).

$$OC = \frac{1}{T}\sum_i t_i \times 100\ (\%) \tag{5}$$

where T is time of measurement, t_i is detected time of vehicle i.

When number of existing vehicles a certain section is N, average length of vehicle is $\bar{l}$, Equation (6) is given.

$$OC = 100\frac{q}{v}\bar{l} = 100\ k\bar{l} \tag{6}$$

Therefore occupancy (OC) is proportional to traffic density (k) and traffic volume (q).

3.2. Relationship between Occupancy and Traffic Volume

From the one month measurement data of Camera #2 in June 2015, traffic volume (q) to occupancy (OC) relationship are shown in Figure 7. According to Figure 7, the relationship between (q) and (OC) is proportional and the number in the graph shows daily O'clock time such as 22 means 22:00. All the data is used from average from all time zone.

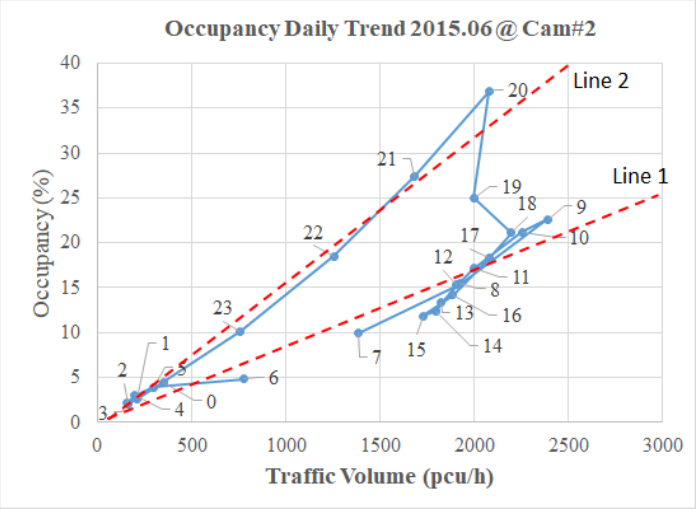


Figure 7: The Relationship between Traffic Volume and Occupancy at Camera#2

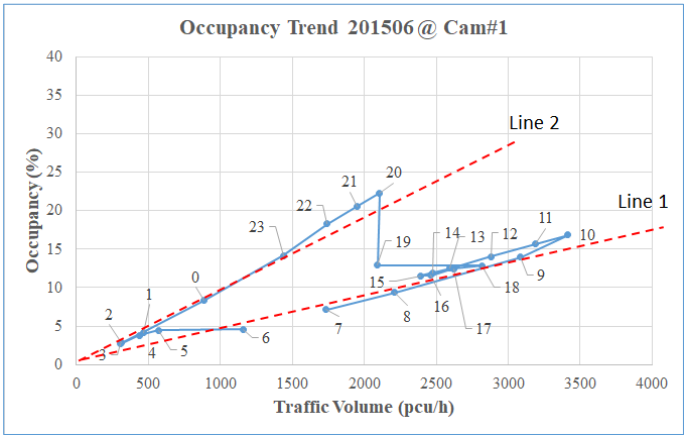
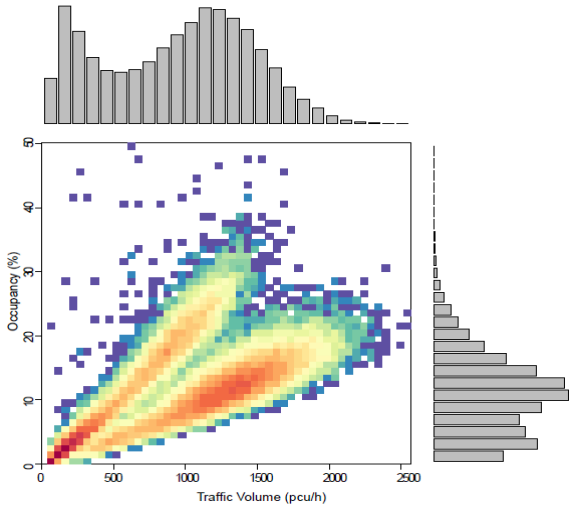
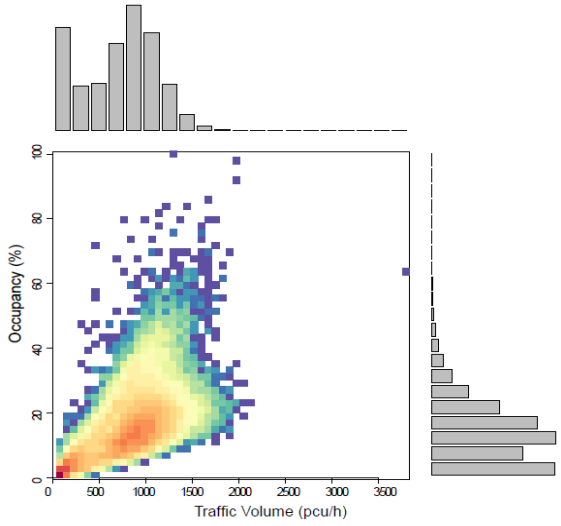


Figure 8: The Relationship between Traffic Volume and Occupancy at Camera#1

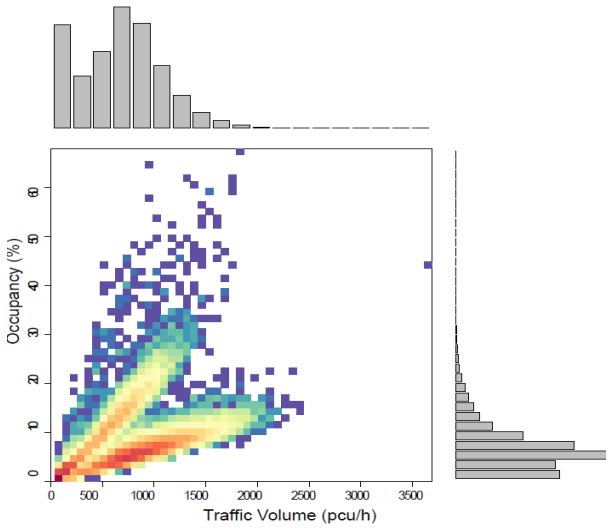
From Figure 7, there are two types' lines which indicate the different growth ratio of the occupancy to the traffic volume. The Line 2 is large growth ration than that of the Line1 and the Line 2 starts from congested time zone from 20:00. The occupancy level



(a) 2D Histogram between traffic volume and occupancy at Camera#1



(b) 2D Histogram between traffic volume and occupancy at Camera#2



(c) 2D Histogram between traffic volume and occupancy at Camera#3
Figure 9: 2D Histogram between traffic volume and occupancy

is more than 30%, which means heavy congested condition as seen in Figure 6 (b). From this observation, there are two types of traffic congestion scenario like light traffic congestion by Line 1, which occurs normally in the morning and heavy traffic congestion by Line 2, which occurs normally in the evening. From our measurement records in June 2015, the traffic condition at Camera#2 is most congested than that of other locations. So, when it takes the case study of Camera#1, Figure 8 shows the relationship between the traffic volume to the occupancy. The occupancy value is lower than that of Camera #2, which means the traffic congestion at Cameraera#1 is smaller than at Cameraera#2.

3.3. "Traffic Congestion Triangle"

In the section 3.2, it takes the average value of each traffic volume and occupancy in June 2015 record. In case of the analysis from the total measurement, 2D histogram graph shows that of Camera#1, Camera#2, and Camera#3 in Figure 9. From Figure 9, each distribution of traffic volume and occupancy of Camera are different. In Camera#1 and Camera#3, it is clear two batches, which means there is clear traffic congestion condition between Line 1 and Line 2 in Figure 8 and Figure 10. On the other hand, there is not clear data batch between traffic volume and occupancy in Camera#2, which means Camera#2 area is relatively congested compared with Camera#1 and Camera#3. This is also able to explain by daily bases traffic volume and vehicle speed characteristics in Appendix Figure A. The vehicle speed of Camera #1 and Camera #3 is not so much drop during most congested time frame between 19:00 to 21:00.

As the observation of the relationship between the traffic volume and the occupancy, it is able to define the new traffic congestion parameter which illustrates as "Traffic Congestion Triangle" in Figure 10. The graph in Figure 10 is based on the measurement analysis at Camera #3. We have similar "Traffic Congestion Triangle" characteristics in other camera locations. Therefore this "Traffic Congestion Triangle" characteristics is common traffic flow characteristics in Ahmedabad city and it shows two types of traffic congestion trends in the morning and in the afternoon.

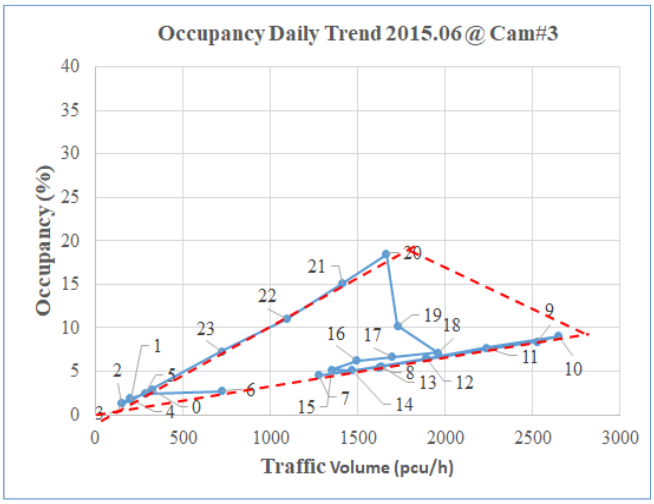


Figure 10: Traffic Congestion Triangle Concept

When we collect the traffic volume and the occupancy at each road in particular month, it is able to create the "Traffic Congestion Triangle" from the relationship between their average traffic volume and the occupancy daily basis. Then the "Traffic Congestion Triangle" line shows the occupancy growth ratio in daily basis traffic flow condition. According to our other observation analysis, we have similar trend of characteristics.

4. Conclusion

This study focuses the traffic congestion analysis in the developing countries, especially in India. And based on one month actual traffic data collection in one of typical major city of India—Ahmedabad—where its population is about 6 million, author creates new traffic congestion analysis parameter "Traffic Congestion Triangle". This "Traffic Congestion Triangle" is able to indicate when and how the traffic congestion occurs in the city. And it shows there are two types of traffic congestion—high traffic destiny congestion(line1 in Figure 7,8) and low traffic density congestion (line2 in Figure 7,8). When we have this traffic congestion triangle at each road, it is easy to understand what kind of congestion occurs. However from this characteristics, it still does not provide solution how to reduce its traffic congestion. We only know its current condition.

We use one month data in June 2015 and the each item data becomes 43,200 (= 60 min use 10 traffic monitoring Cameras at main roads in the city and its traffic data is collected by very minute as traffic density, veh. × 24 hours × 30 days) points. After the detail traffic congestion observation, it is successful to show the unique relationship of traffic parameter between the traffic volume and occupancy. There are two patterns of occupancy growth ratio to the traffic volume in the morning and in the evening. According to all location traffic data analysis in Ahmedabad, the growth ratio in the evening is larger than that in the morning, which means its traffic congestion occurs in the evening time from 19:00 to 21:00. In the traffic volume to occupancy graph, those slope of two lines and the line between these two lines build the triangle shape, which it is defined as

"Traffic Congestion Triangle" parameter. The members of "Traffic Congestion Triangle" indicates each traffic congestion condition. Therefore when the traffic volume and occupancy are detected, it is able to estimate how traffic congestion occurs and when its congestion goes to more crowded or smooth.

This study is based on June 2015 traffic data. So it is necessary to continue collecting and observing at all location such as whole year traffic data measurement. It is also necessary to check other major city case study at least in India and confirm that the "Traffic Congestion Triangle" is valid to other developing countries in future work. It is also important to analyze the intersection point in terms of traffic congestion in future.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

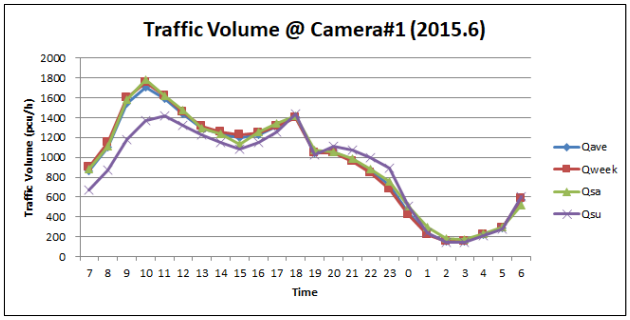
This research is part of SATREPS program 2017 (ID: JPMJSA1606) between India and Japan.

References

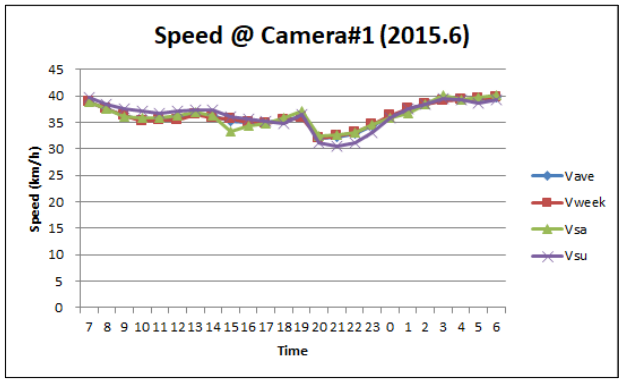
[1] Ahmedabad Municipal Corporation official website information: https://ahmedabadcity.gov.in/portal/jsp/Static_pages/demographics.jsp
[2] Commissionerate of Transport Department of Port and Transport, Government of Gujrat data at http://rtogujarat.gov.in/statistics_vehicle.php
[3] N. Gartner, C.J. Messer,A.K.Rathi, Traffic Flow Theory A State-of-the-Art Report: Committee on Traffic Flow Theory and Characteristics (AHB45), 2001.
[4] C. Millikarjuna, K.R.Rao, Area occupancy characteristics of heterogeneous traffic, Journal Transportmetrica **2**(3), 2006. Doi: https://doi.org/10.1080/18128600608685661
[5] A. Salim, L. Vanajakshi, C. Subramanian, Estimation of Average Space Headway under Heterogeneous Traffic Conditions, International. of Recent Trends in Engineering and Technology, **3**(5), 2010
[6] M. Goutham and B. Chanda, Introduction to the selection of corridor and requirement, implementation of IHVS (Intelligent Vehicle Highway System) In Hyderabad, International Journal of Modern Engineering Research, **4**(7), 49-54, 2014
[7] B.D. Greenshields, A Study of Traffic Capaci, Proc.H.R.B., **14**, 448-477, 1935.
[8] W. Hole, 75 Years of Fundamental Diagram for Traffic Flow Theory Greenshields Symposium, Traffic Flow Theory Characteristics Committee, Transportation Research CIRCULAR, Number E-C149, **6**, 2011.

Appendix

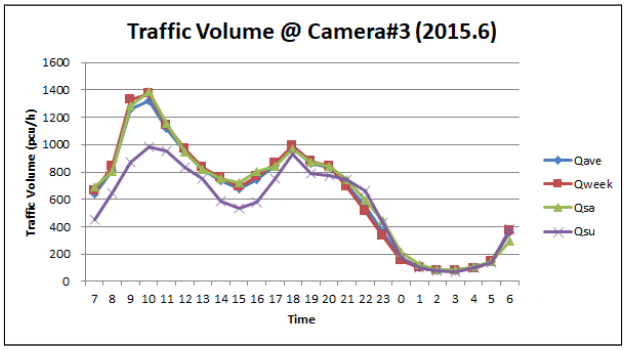
The daily basis traffic flow characteristics at Camera#1 and Camera#3 are shown in Figure A. From Figure A, there is not so much traffic congestion occurred in both case at Cameraera#1 and #3.



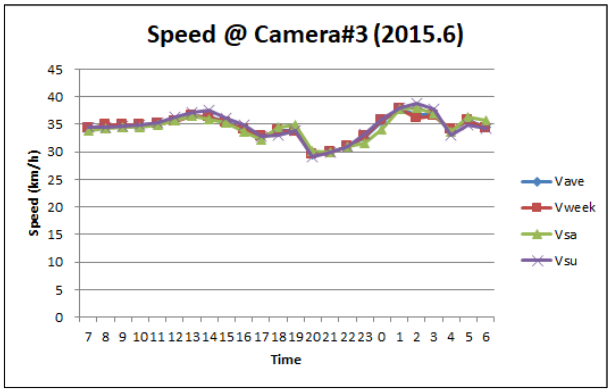
(a) Traffic Volume daily basis change at Camera#1



(b) Vehicle Speed daily basis change at Camera#1



(c) Traffic Volume daily basis change at Camera#3



(d) Vehicle Speed daily basis change at Camera#3

Figure A. Traffic Flow daily basis change

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 594-605 (2020)
www.astesj.com
Second International Virtual Conference on Multidisciplinary Research 2020

ASTES Journal
ISSN: 2415-6698

Effective Application of Information System for Purchase Process Optimization

Pearl Keitemoge*, Daniel Tetteh Narh

School of Engineering Management, Beihang University, Beijing, 100191, China

ARTICLE INFO

Article history:
Received: 09 August, 2020
Accepted: 10 October, 2020
Online: 20 November, 2020

Keywords:
purchasing management
information technology
supply chain ERP System
Cloud-based
Purchase Process Optimization
Internet

ABSTRACT

This paper comprehensively focused on the information systems for purchase process optimization (IS) that are currently being used. It elaborates four crucial and important areas in relation to purchasing and supply management. These includes; Identifying the process involved in purchasing, how information systems are implemented and used in enhancing the supply process, understanding the key features that constitute enterprise resource planning, purchasing database and electronic communication between customer and supplier, how Enterprise Resource Planning (ERP) system facilities and improves effective communication in managing supplier selection process and lastly describing the future requirements of electronic purchasing system that will enable transparency, visibility and accessibility of information across the overall supply process. The parameters effectively enable optimization of purchase process. In today's era where Internet is accessible, the most thriving firms are those that are in-step with the fast pace of technological revolution. Currently, there are more high-end devices at the finger-tip of the customers, which has significantly enabled customers and many business firms to prefer executing transactions on their portable devices instead. The available technologies such as ERP system has opened an effective growth path for firms to stay competitive and relevant in the market.

1 Introduction

Technology is evolving rapidly and purchasing products are ongoing around companies/industries, therefore it is key that supply management is enhanced to develop business around the globe with and through the help of Information System (IS). Information Technology (IT) systems have exponentially increased and enhanced marketing technique and purchasing process for buyers (companies and individuals) and seller of products. Users of knowledge systems have progressively found themselves as customers of multiple technology vendors, this IT procurement method assumes bigger and vital management. In light of the upward trend of technology, IT system has or will become must-have for every company that is planning intending to compete for customers, this because we are in an era in which customers are constantly demanding fast service from service providers. Therefore, for companies to thrive in this ever growing technology age, they must change with rhythm to meet user demand.

In recent years, the world's economy at large has become increasingly competitive. It has practically become an ineffective and unsustainable approach for firms to sustain the upward competitiveness in the market and further returning profits margins for investors through a frequent increase in product cost. Instead, firms must integrate and harness the premise of product innovation to achieve and deliverer higher quality, and quicker response times or performance, while continually maintaining the minimum unit cost [1, 2].

Suppliers are very crucial in making firms competitive, by providing firms with products that meet user demands [1]. The impact of integrating IT in purchasing process or supply management has a great influence on every firm's competitive ability. In [3], a piece empirical information was collected from a total of 108 Taiwanese firms was studied and interviews with four active managers were carried to collect the sensible insights of the theoretical framework of e-procurement and provide chain performance.

According to [3–5], information sharing and supply chain integration allow a strategic viewpoint when implementing e-procurement systems intended to improve supply chain performance. This project will be sharing new ideas on how a new technology channel is making a difference in purchasing and supply management. Therefore, more discussion and discoveries will be made in relation to the information system in place used for purchasing

*Corresponding Author: Pearl Keitemoge, & Email: pkeitemoge@gmail.com

process currently. It is with the new technologies that new strategies are emerged to impact both organization and individuals on their ways of purchasing and supply of goods. The business can simply grow spontaneously from their lowest profits to a higher one with the aid of the information technology systems in place.

1.1 Purchasing Process

Entails the stages that any organization or business that want to obtain services and goods for the productivity of the business defined as the purchasing process. The author in [6] concisely explained procurement practices as principal activities designed and undertaken by an organization to aid promote and facilitate effective management across the organization's supply chain.

The stage differs based on the specific need of every business or organization. The process becomes completed after specifications have been developed, a buying team monitored and cooperating with the supply manager will select suppliers, generate the request for proposals, evaluate the proposals and select a supplier based on established selection criteria [1, 7].

It is therefore extremely important that all necessary supply chain information should be readily available to all companies involved, and the business process should be structured in a way that allows the full access to the available information. This information will be utilized in the purchasing process. There main aim of every firm is to maximize profit and minimize loss, therefore procurement performance very important dependent on many factors as indicated in [8,9] by Figure 1. Assessment of procurement performance helps organizations to reduce cost, enhance profitability, assured supplies, quality improvements and competitive advantage [6]

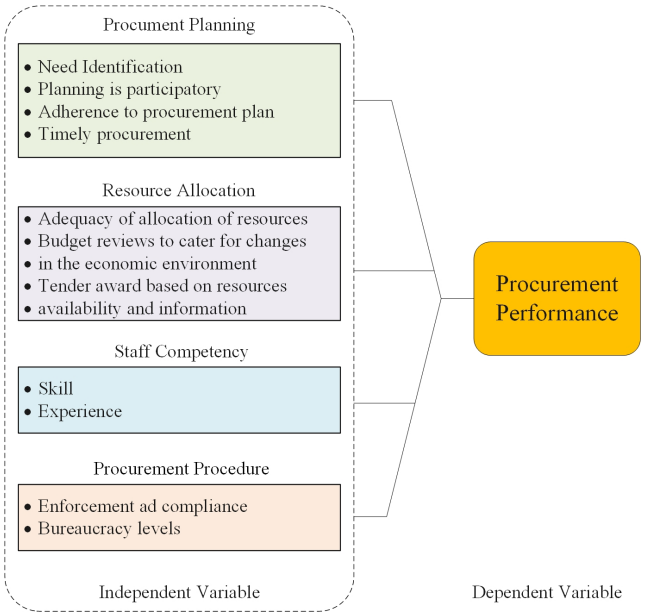


Figure 1: Framework for Procurement Performance [9]

Figure 1 shows and indicates that procurement performance is very sophisticated and it linked to many other parameters such as procurement planning, resource allocation, staff competency, and procurement procedure.

The practice-based commentary outlined both cases of success and failure of firms in relation to the implementation of IS in the purchase process. In light of success, Wal-Mart adopted the implementation of IS in some sections of purchase process, namely, materials management, product ordering, and inventory-tracking through Radio Frequency Identification (RFID) Tag, this adoption enabled the firm to have real-time information in light of demand forecasting as a result of RFID-based inventory management, this eventually lead effective inventory turnover and mitigated inventory costs. The firms' adoption IS applications have thus supported and made it possible for the company's cut cost on the purchase process.

Concerning failure, Nike invested $100 million in the deployment of purchase process IS software but failed handle or prevent significant inventory shortages leading to larger unmet demand, this is due inadequate and inefficient planning for application integration with its existing supply chain objectives [10]. These examples clearly outline the importance of firm's purchase process managers to pay close attention in aligning information systems that are deployed in the purchase process system with the goals and objectives of the purchase process itself [11].

In this section, we examined in detail the Purchasing processes, which includes the product requirement, good delivery, payment, supply management, management of purchasing effectively, suppliers evaluation and selection, review specifications and the awarding of the purchase contracts.

1.2 Purchasing types

Purchasing is defined to be a company/organization or individual paying for goods or services in order to accomplish business or personal defined goals [12].

Personal: This is the system of individual buying products or services for their personal use. They do not engage any third party to negotiate the prices and make selections on the supplier.

Mercantile: The purchasing that is facilitated by the middleman. Retailers and wholesalers are found under this type. They arrange their own channel of distribution of products to their customers.

Industrial: This is the system of buying raw materials to later produce finished products. The materials and necessary tools needed to produce products are also purchased here.

Government or institutional: Buying in bulk for government departments and institutions. Governmental purchases have a high level of bureaucracy and decision making concerning the product.

Table 1: Economic Contribution and Dependency

Purchasing type	Contribution on the economy	Level of dependency on the economy
Personal	Low contribution	Highly dependent
mercantile	Medium contribution	Low dependency
Industrial	High contribution	High Independent
Government	High Contribution	High independent

The Table 1 shows the contributions and how dependent each different purchasing types have on the economy of a country.

The Figure 2 shows the bureaucracy prism for all the abovementioned types of purchasing from bottom to top. The bureaucracy

exists because of the complex leadership structure embedded in each type.

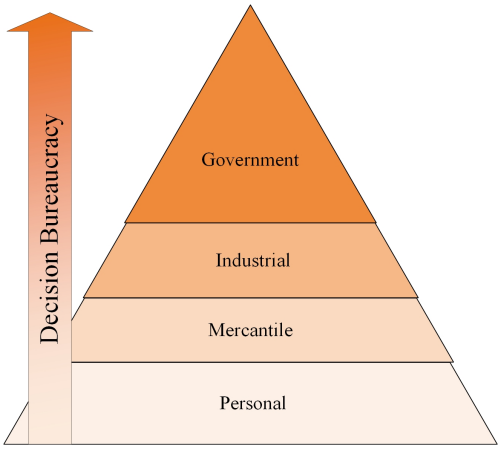


Figure 2: bureaucracy prism

Much of the procurement bureaucracy can actually be eliminated by the implementation of the information system and enhance effective firm data handling and management, especially simple decision which will directly increase productivity.

2 Identifying the Process Involved in Purchasing

Knowing the processes involved in purchase process is crucially important to establishing an effective and productive supply chain management across any firm. These are stages that defines the product requirement, product size, product safety, performance and necessary maintenance requirement, payment and delivery method and supplier selection strategy and requirement.

2.1 Product Requirements (specification)

Before defining or selecting product requirement, the business firm must identify the task product that they will be handling, usage environment, workload and lifespan of the product.Defining some of these parameters can be very daunting and almost impossible, task type can easily be specified but not future environmental conditions. Information technology is helpful in the defining of product specification, the future workload can be approximately specified or predicted using current machine learning algorithms such as neural network, multi-linear regression and other existing predictive analysis algorithms. Product requirement is vast includes functionality, computer application user interfaces, communications, reports, customer services, user help manuals & self-help tutorials, service delivery staffing licensing/certifications Product requirement parameter can also vary according to the kind of product that is required. The basic but most important parameter are discussed as in Figure 3;

A well-defined product requirement is very important. If these parameters are not defined properly it subsequently slows the product design and development phases by having to define and re-clarify objectives and requirements and this also increases production costs due to change requests caused by having to fix execution errors or omissions.

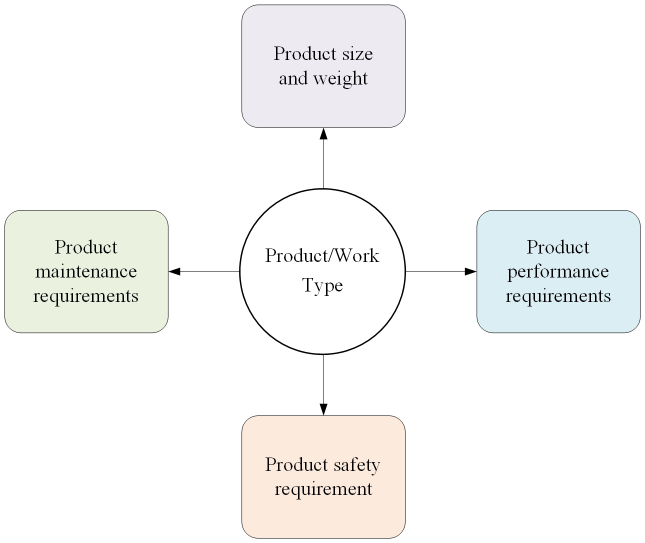


Figure 3: Basic Product Requirements

2.1.1 Product size and weight

The customer will provide the firm with the size of the product they need. In an organization size also is needed to add to the specification of the products.

2.1.2 Product performance requirements

The performance requirement provides the firm with operational details. How fast, often and slow the product will be when used. Other things to consider are the shelf life of the product, the power requirements, and the service life.

2.1.3 Product safety requirement

The safety requirements of a product are the potential sources of liability. Identifying these in the specification process enables suppliers to a provide possible solution to the sources such as potential operational hazards, manufacturing and assembly hazards. These hazards directly have an effect on the product life span.

2.1.4 Product maintenance requirements

This sort to find out if the product can be maintenance-free? If routine maintenance is required, can it be done by the owner/operator? Will professional maintenance be required?

Well defined product requirement is very important. If these parameters are not defined properly it subsequently slows the product design and development phases by having to define and re-clarify objectives and requirements and this also increases production costs due to change requests caused by having to fix execution errors or omissions.

Technology is making the requirements available to the supply via the systems that are availed to the supply to aid the collection of information on the product. If companies employ and integrate information system into these phase of product procurement and facilities quick and fast exchange of this information between the suppliers and customers.

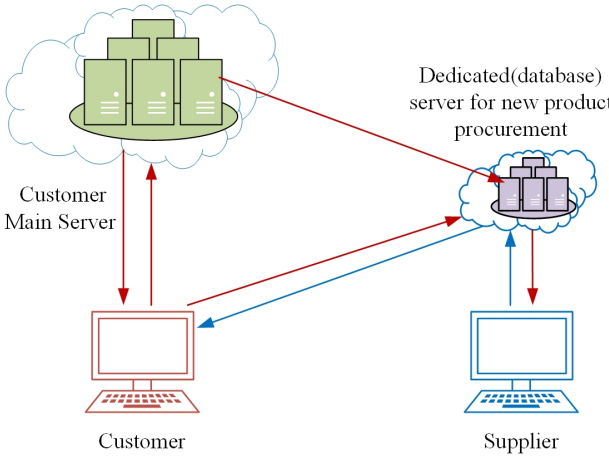


Figure 4: Product Requirement and Information System Workflow

Figure 4 shows a quick and simplified view of information flow when the information system is integrated into the procurement process. The customer can grant the supplier access to a dedicated database server for a specified project or product server. With this access, the supplier can have quick access to the product requirement information, the supplier can also give suggestion and comments (as indicated by the blue arrow) on the product quick before product manufacture commence. Any suggestion indicated by the supplier are reviewed by customer and if suggestions are viable, the changes are then updated to the main server and confirmed to the dedicated product databases server. Using the information system, process speeds up thus avoiding unnecessary delay in the procurement process of the product. The first aspect of purchasing is for the business to identify the product that they will need and provide a detailed specification.

The product or service type needed by any business entity can either originate from the manager's recommendation or collective customers demand. Technology is used in the stage by collecting the specifications of service or goods electronically and the information is later captured into a computer for processing and used in the next stage. Frequent communication between the customers and suppliers is done by the emails. The detailed specifications are compiled and captured into forms in the computer application programs for easier access by other key role players. Hence, improving communication and reduces the design complexities that add more cost to building cost specification [1, 13]. Sharing of information real-time is also crucial because the shrinking product lifestyle and the competitive advantage often reduces time to market [1, 4, 7].

2.2 *Evaluation and Suppliers Selection*

In this purchasing process stage, the team supplier receives the proposals and prices and selection is made based on who is able to guarantee quality assurance and competitive price of the products. The team also review the product requirement specified by the suppliers to cross check with the stipulated need of the company and the individuals. A review of different requisitions may also reveal that different users actually require the same material [14]. By combining the purchase requirements, purchasing can often achieve a lower total cost. The team also evaluate the suppliers based on the capability of the supplier. Comparing the place of the distribution of the supplier, global versus regional and local.

The supplier selection factors considered by firms include but not confined to the supplies team size, the supplier public evaluation or rating, supply risk level, manufacturer versus distributor. Supplier evaluation can be a rigorously daunting task due parameters factors come to play, such as price competitiveness, ability to fulfill expected product delivery time, budget optimization. All procurement expert or manager must know the thorough or effective evaluation will result in the proper selection of suppliers. Making a decision requires a wide consideration of all possible alternatives and then choosing the appropriate and optimal supplier (OpS). According to [15, 16] selection process considerably relies on both the subjective and objective judgment in different areas. The different areas include (1) inter-connected criteria definition: for both qualitative and quantitative (2) criteria conflict: leading to conflicting objectives of the criteria (3) involvement of several alternatives: due high competition (4) internal and external constraints: established and imposed on buying process [16].

In [17], through extensive review and professional interviews, the author established three key factors for supplier selection measure. He further proposed that underlying the documented suppliers criteria is the reason rate supplier effectiveness in terms of quality and repair capabilities and control alignment with the purchasing firm [17]. Another author in [18] also established an integrated model by merging the analytical hierarchy process and grey relation analysis for a single evaluation model. Unavailability of precisely needed or complete information intrinsically increases the difficulty of any decision making. Therefore, most experts in the procurement arena agree that there is no one-for-all confined strategy or method to perform evaluation and selection of suppliers, and organizations use several different approaches in order to obtain optimal selection [19] such as cost, maintenance, product life cycle as illustrated in [20, 21].

Regardless of the selected approach, the overall objective of the evaluation process should aim to reduce know or perceive procurement risk and maximize overall value to the purchasing firm. IT system can be utilized by purchasing firm to ease the supplier selection process and evaluation criteria that cannot be biased. Accessing all needed information from all the potential suppliers using the traditional method is not an easy task and this subsequently delays the supplier selection process. Information system integration aid fasts forward this stage of purchasing process instead of using the

traditional method of scheduling and arranging several long hours in meetings with all potential before deciding whether the supplier meets firm's requirement or not.

2.2.1 *Effective Data Collection*

Firms that considerably invest and implements modern technology (information system) is at an added advantage of accessing all necessary information from all potential suppliers ($s1, s2 \ldots sN$) data center as in Figure 4 .

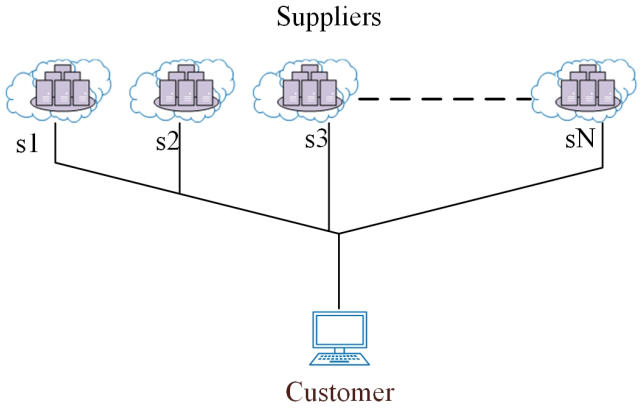


Figure 5: Firms/Customer Accessing Information from *N* Suppliers

An Information System can virtually enable firms to collect and gather important supplier related data such as qualification information through internal and external questionnaires and self-service profiles, risk factor, and policy compliance to aid choose best suppliers. IT also makes firms proactively more effective in achieving business alignment when planning and creating negotiation documents, it also helps create and alerts for monitoring the availability of new information that might affect and influence supplier management, capabilities, or risk factors. Suppliers, as well as sourcing teams, gain adequate and simultaneous real-time visibility of the information on transactions as well as current and upcoming sourcing activities.

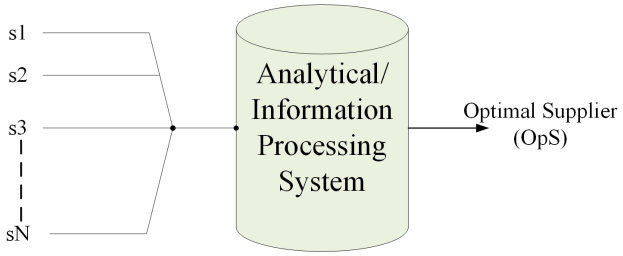


Figure 6: Effective Processing of all *N* Supplier Data

2.2.2 *Effective Data Processing*

With is the acquired data, firms would be able implement machine learning algorithms to process, analyze, narrow down and select optimal supplier that deliver product that meets the defined requirement In simple term, OPTIMAL SUPPLIER (OpS) means suppliers who can supply be able to supply the correct materials with the best price and the right quality that adheres to all defined constraints.Use of technology in the procurement process will ensure quick flow information with the supplier which in turn ensure and results in high productivity since many human factors or errors would be eliminated.

Strategic procurement with information system before initiating negotiations will provide and compare the collection of supplier data monitored to ensure that supplier complies and adhere to all the well-defined corporate policy. These policies include;

Process and design capabilities: Suppliers ought to have up-to-date and capable merchandise, yet as method technologies to provide the fabric required. as a result of completely different producing and repair, processes have varied strengths and weaknesses, the shopping for the organization should remember those characteristics up front. One way to cut back the time needed to develop new merchandise is to use qualified suppliers that square measure able to perform product style activities.

Quality assurance and reliability: Quality levels of the procurement item may be a vital issue and influences provider choice. Product quality ought to be in line with the lay-down product needs since it will directly have an effect on the standard of the finished merchandise.

Cost: The unit cost of a product's material is not deemed the most important aspect of supplier selection, rather the total cost of ownership is considered as the crucial factor. The total cost of product ownership includes the material unit price, the agreed payment terms, the cash discount, the cost of ordering, the cost of transportation or shipment, all logistical costs, projected maintenance costs, and many more qualitative costs that may not be easily assessed.

Service: providers should be in a state to back up their goods by providing the best services to their clients when requested. Buying firms always need to evaluate after purchase service incentives awarded by potential suppliers. This is very important especially when the buying firms have enough expertise concerning the product maintenance.

Capacity and geographical location: The buying organization needs to consider whether a potential supplier has the capacity to fill orders to meet all defined requirements and the ability to fill large orders if needed. Geographical location is another important factor in supplier selection because it inevitably has a huge impacts delivery lead time, transportation, and other logistical costs.

Factors to consider also include; system planning and control, firm's management capability, financial commitment and cost structure, longer-term relationship potential. The implementation of smart IT systems build for effective negotiation and supplier selection process would aid purchasing firms leverages all available negotiation best strategies and practices to achieve bottom-line or optimal results [19]. IT assist in inconsistent monitoring of supplier progress. Using IT gives the firms access to embedded social collaboration makes easy to harness and combine team effort to create

more appropriate and effective negotiation documents. Firms the uses IT in procurement process mostly have access to well-proven analysis tools that allow the?firm to quickly evaluate supplier bids based on predetermined factors—such as price, quality, and value allowing the firm to arrive at the best possible award based on your unique business requirements.

2.3 Approval/ award contract

The supplier that has been selected to provide the product will then be awarded the contract that carefully is having the conditions and terms that will be agreed by two parties. The fixed prices and the cost of the services will reflect on the contract. The organizations are expected to use the product requirements for the selection process. The approved supplier has the mandate to deliver their products as stipulated in the agreement of the contract. Reliability and efficiency are the core elements of the supplier.

2.4 Supply chain management process

This is another part of the purchasing process that links all the necessary information to other systems to enable both buying and capturing of what is needed. Information is always flowing bidirectionally between the supplier and the customer in Figure 7.

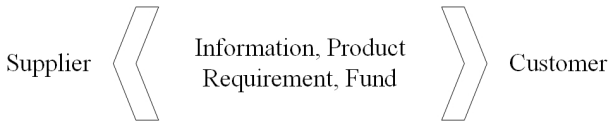


Figure 7: Supply chain fund management flow chart

Supply management has other activities that enhance the purchasing process, that can be breaking down into the supply chain macro process as shown in the Figure 7. The customer's relationship management system is there to capture and manages the details about what is available on the market, the prices, the selling of the products, creating a route for the firm to get updates on what is available. This guides the majority of firms to practice strategic planning and fulfill the demand and supply. The Figure 8 indicates the link in the supply chain macro process.

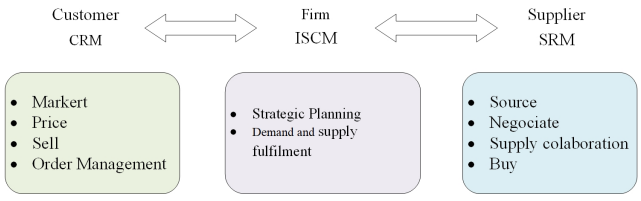


Figure 8: supply chain macro process

2.5 Product delivery, payment review, and process evaluation

2.5.1 Product delivery and payment

The other stage of purchasing is product delivery, this stage collects and compiles the information that is needed for arrangements of how the product will be delivered to the specific firm or an individual. In Order to survive in the industry and gain competitive advantages, improvements such as reducing inventory and dependable delivery must be made, the introduction of data exchange systems that facilitates and communicate demand to suppliers and production progress data to customers within the network [14].

The products are stored to be delivered either in the warehouse or to individuals. Information technology system used in capturing the details in forms, others can be labeled electronically. Payment of the products is processed immediately after selecting the product and placing the order of it.

2.5.2 Process evaluation

After payment and delivery of the product, the systems are reviewed to map a way of eliminating errors and shortfall that occurred during the interaction of the purchasing process. At this stage, crucial decisions about changing the suppliers can be made based on the outcome of the services provided by the previous supplier. Recommendation and amends are drawn for future processing.

3 Key Elements of Enterprise Resource Planning, Purchasing Databases, and Electronic Communication Systems

The business market is constantly revolving and enterprises must revolve with it in order the maintain the necessary market presence. As firms hierarchy decides to invest in ERP system, they must ensure that the targeted system functionalities aligns with the overall firm's goal. This section will sort to elaborate on information system for procurement, crucial factors to consider when an enterprise is planning on purchasing and information system.

3.1 Enterprise Information System Purchasing

Business information systems are generally termed as ERP (Enterprise Resourcing Planning) system. It is a software package largely used by firms to effectively manage and handle day-to-day business activities or transaction, activities such as accounting, procurement, project management, risk management and compliance, and supply chain operations.

A complete ERP suite as an enterprise performance management software package helps firms to arrange, budget, predict, associate degree report on an organization's monetary results. Basically the key elements in the ERP include; One centralised location either onsite or cloud-based, Integrated Reporting Analytics, Efficient Data Import and Data Handling, Customer Relationship Management, Materials Requirement Planning, Materials Management, Engineering Change Order, Recall Management and Tracking, Production Planning & Execution, Document and workflow management, Mobile ERP and Automation.

3.2 ERP System Evolution

The implementation ERP systems by the firm and can be dated back the mid of 20th-century various departments within many organizations begun to use to function independently and these departments are often referred as functional silos [22]. The building blocks of ERP system basically comes from manufacturing resource planning systems (MRP) [23]. Inevitably, those systems were ineffective because they were deprived of many benefits of integrative coordination. It brought a huge financial burden on the firm because most of the systems at the time were not reusable since it needs massive rework and designs needs to be overhauled when the product was completely manufactured. A Framework of MRP is shown in Figure 9

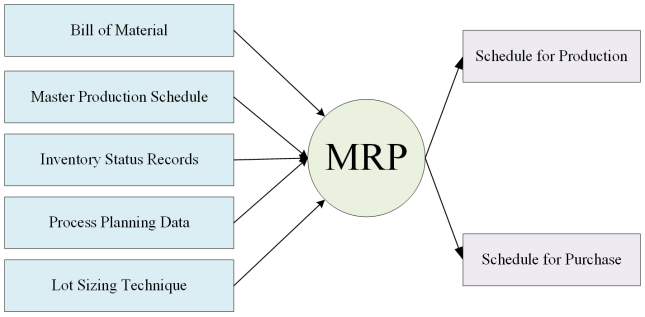


Figure 9: Framework MRP Software [22]

In the subsequent decade, MRP systems were evolved. Material Requirement planning is the software architecture that facilitates material requirement and planning as in Figure 9 [22]. It receives the following user input:

- Bill of material stating details of the materials, components, and sub-assemblies needed to form every product.
- Master Production Schedule. Once the quantities are needed to satisfy demand.
- Shelf life: that is period of keeping materials.
- Inventory status records such as Manufacturing lead time data, Purchase lead time data
- Process Planning Data. This includes a whole range of data required for process planning like routings, process plans, quality testing methods, labour standards, machine standards

This MRP software generates the following output [22]:

- Details report about the product to be produced and production timeline. That involves everything from start to finish each stage is in the process is monitored. It also allows quantities and generates necessary reports.
- Detail report for the schedule of purchase. material to be purchased and date on which purchase must be made both questions are answered here.

Above MRP programs were not flexible enough if something goes wrong, like break down of machines or delay in delivery of purchased items. This eventually created inconveniences for users. In 1990, Gartner created the term ERP to explain the evolution of material requirement planning (MRP) and manufacturing resource planning (MRPII) because of the need expanded beyond manufacturing into other integral parts of the enterprise, typically finance and HR [24–26].

The ERP system implementation began in the 1990s until somewhere the start of the 21st century where the usage system grew exponentially. Later it absolutely was discovered that the prices of implementing ERP system began to rise. The hardware needed to run the software package was usually on-site, with massive machines in a very server area. The capital is required to purchase hardware and software licenses. In addition, nearly all organizations always prefer to customize their ERP systems to fit their specific needs instead of default settings and packages that come with the software and this also generates an additional expense of software consultants and training.

As technology grew up and limits of business crossed regional barriers want for tightly integrated programs which may seamlessly integrate all aspects of business in real time manner was necessary. This was extremely since it does not only to integrate all functionalities in one unit of a firm but also to integrate and unifies different parts of firms having a national or global presence. The modern ERP systems find ways to take handle these issues since they could b joint with the old systems. Though the transition from the legacy system to the modern system was involved with a lot of challenges, firms were able to transit the same [27]. with this, information was available to a specific user at the right time with the right permission to a specific user in a real-time environment. Availability of real-time data across each functional modulation with simulation capabilities were some of the important characteristics of these systems.

The rapid growth of the use internet made it easy for ERP to expand in the business section, such as the customer relationship management CRM, human capital management (HCM), supplier relationship management (SRM), supply chain management (SCM), inventory management, accounting, business intelligence and e-commerce and this exponential enabled firms to increase performance in marketing [28–30].

4 Future requirements of electronic purchasing systems that will enable information visibility

This section address the necessary future requirement or features to sort when purchasing an ERP system. The advantages and disadvantages of ERP system is fully discussed. Recommended solutions are for the possible disadvantages are presented. Further more, a short list ERP software packages are presented according the firm size.

4.1 Cloud Integrated ERP System

The growth of knowledge in ICT has significantly influenced all sides of computing applications across organizations. Additionally, the business setting is changing into more and more complicated with useful units requiring additional and more inter-functional information flow for higher cognitive process, timely and economical acquisition of product components, management of inventory, accounting, human resources and distribution of products and services.

Meanwhile, the continuous evolution of ERP technology has completely embraced the internet, with new options and practicality, like embedded analytics has been added. Many firms are quickly migrating to cloud-based ERP system from on-site ERP because firms have discovered the difficulty in maintaining and keeping up with the trend of security demands or rising technologies, like smartphones whiles using on-site ERP systems. Currently, cloud computing has effectively released firms form overly burdening themselves with ERP security demand, the maintenance cost of the on-site ERP system. Usage of cloud-based ERP system, has for example, improve inventory management capabilities for enclosed technology like barcode scanning and storage or RFID-tags; nowadays, the explosion and rapid development of the interconnected web of devices which is also known as Internet of Things (IoT) [31], has directly made cloud-based ERP even more attractive.

Also, the continued digital revolution and progressively cheap laptop process power has brought changes to nearly every space of business operations [32]. ERP systems have become a cloud-based hosted software technology [33]. In a cloud (server in a remote location), the users are granted access to a system by the respective or selected service provider, this includes full and complete access to hardware, and storage capacity, where system and hardware management are highly transparent to the users [34]. The major reason why firms moved to ERP system is typical to reduce costs, improve overall decision making through better reporting generated by provided capabilities, improve customer-supplier relationships, meet market and legal requirements, and to increase process efficiency [35]. With a cloud-based ERP system, firms have access to the ERP system that is hosted at a chosen service provider's site via the Internet. In addition, the responsibility regarding updates, managing servers, maintenance, and performing backups lies with the service provider instead of the firm which inherently cuts cost [33, 36]. In other words, the cloud ERP service providers ensure that maximum security, constant stability, and effortless operation of the system on behalf of its client users. Currently, cloud-based ERP systems have gained increased attention and investment from all around the globe.

Since the emergence of cloud-based ERP systems, the large ERP system implementations are focused on the cloud [34]. Cloud service model gives the client the opportunity to have quick access to the on-demand network and to share a bundle of resources. These resources will embrace, servers, networks, storage, applications (e.g. ERP), and others [34] Software as a Service (SaaS) technology is implemented as a delivery mode for cloud-based ERP system [37, 38]. SaaS eliminates the requirement to physically install and run the server-side applications on the customer own premises, and eliminates the requirement for back-end hardware and data-centers needed to run the system [27], that successively simplifies the application's maintenance and support operations. This makes it possible for small organizations to take advantage of ERP systems, as they require neither the substantial resources nor the skills needed for successful on-premises implementation [32, 33].

4.2 Advantages and Disadvantages of ERP System

The ERP system comes numerous benefits that cannot be realized in firms that still use traditional purchase process. The positive impact of modern ERP system significantly out weights the negatives, it does comes with its related opportunities and challenges as described in [39]. The advantages and disadvantages are presented in the Table 2 and Table 3 respectively.

Table 2: Advantages of Cloud-Based ERP System

Advantages	**Aspects**
Reliable and easy information availability	Common DBMS permits and allows consistent, accurate data and improved reports.
Redundancy mitigation for data and operations	Modules access the same data through a centralized database, and this helps avoids and remove multiple inputs and update operations.
Less delivery and low cycle time	Minimizes data retrieval and reporting delays.
Cost-effective	Time savings are achieved due to improved control by firm-wide analysis of all purchase decisions making processes.
Easy adaptability	Relay for information about changes in current business processes is easy and swift to adapt and restructure.
Improved scalability	Availability of structured and modular add-ons.
Improved maintenance	Vendor provides long-term supports as part of system procurement.
Global outreach	Availability of module extension like CRM, SCM, etc.
E-Commerce and e-business	Excellent Internet commerce, collaborative culture.

Table 3: Disadvantages of Cloud-Based ERP System

Disadvantages	**Suggested Solution**
Time-consuming	Sensitive issues minimization, reduce internal politics and call for general consensus
Expensive	Cost varies from thousands to millions of funds. Re-engineering existing business process could be extremely high.
Modules conformity	Architecture and related system components should match to current business processes, culture as well as strategic goals of the firm.
Vendor dependence	Multi-vendor vs. single-vendor consideration and long-term support commitment.
Features and complexity	ERP system is packed many modules so firms must do careful consideration and implement only the needful.
Scalability and global presence	Look for vendors with long-term product and services commitment, Internet-enabled systems and global investment.
ERP capability extention	Consider ERP "add-on" modules extension such as Customer Relationship Management (CRM), Supply Chain Management (SCM), etc

4.3 On-Premise ERP versus Cloud-Based ERP

Firms operate on-premise or on-site ERP system is responsible to updating and maintaining their system and keeping it up-to-date, this firms also need to undertake extensive user re-training efforts to correspond with the system upgrade [40]. The cloud-based ERP updates, testing is undertaken by the services provider or vendor. This significantly reduces the burden and unnecessary delay in the overall project actualization. Therefore, if the vendor has thoroughly conducted tests prior to an update, the client organizations would largely be shielded from these any negative effects experienced by the on-site ERP users [41]. Updating a cloud-based ERP system is provided below:Firms operate on-site ERP system is fully responsible in updating, their system maintenance and keeping it up-to-date, this firms also need to engage in extensive user re-training program in an efforts to successfully push new system upgrade [40]. While cloud-based ERP platform or system update and testing are facilitated and undertaken by the services provider or vendor. This significantly reduces the effect of any unnecessary project delay. Typically, vendors conduct enough tests prior to any new update to ensure that new functionalities integrate seamlessly and this largely shields client organizations from any negative effects experienced by the on-site ERP users due to pushed updates] [41]. Cloud-based ERP is equipped with the following benefits:

- It is easier to introduce and use a new functionality introduced by vendors [34];
- Less planning, testing, maintenance, and configurations from the customer's side since everything is fully automated by the vendor [33, 34];
- Lower costs, both with maintenance and IT staff, and updates are free of costs [33, 34, 37];
- The vendor has the possibility to expand its potential customer base, and let the users focus on their core competencies [34, 37];
- It is the vendor who provides a help-desk and maintenance support [42];
- The firms are not the owners of the system, as they do not own the infrastructure or run the applications and thereby not responsible for maintenance [43];
- Cloud-based ERP vendors regularly release new versions [41];
- Cloud-based ERP systems usually get frequent updates and new functionalities as compared on-site ERP systems [34].

As stipulated above, the main advantages of cloud-based ERP are that updates are free. There are enough good reasons why the vendor should invest time and resources in updating the system, even when it is free for users as technology updates are an inescapable process with dealing with any man-made creation utilized for productive gain [40]. A frequent update is done by vendors?ERP system to allow their customer firms to constantly evolve and optimize their business processes, and improve decision-making [40].

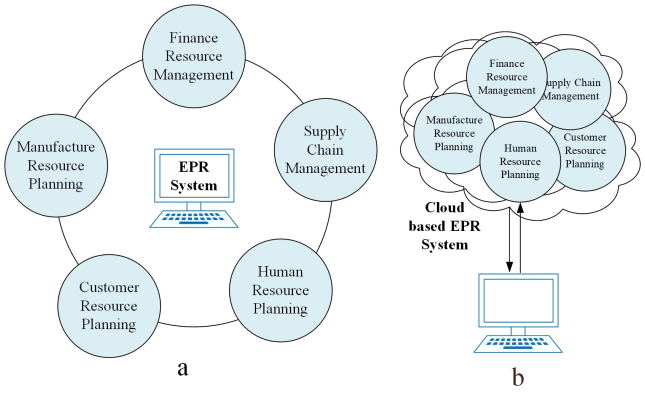


Figure 10: On-Site ERP versus Cloud-Based ERP System

The Figure 10 (a) shows an on-site ERP system which can only be accessed when on-site. Figure 10 (b) shows a cloud-based ERP system that demonstrates the evolving technologies and how efficient it is to managing purchasing and supply chain within organizations.

4.4 Available Cloud-Based ERP System

Purchasing ERP system is greatly dependent on firm size. Therefore, careful and objective decision must be made by firms that are considering purchasing ERP system. Selecting the correct and best ERP system for any firm operation can prove very challenging. this because many parameters are considered such as precision, flexibility, technologically integral. Supply chains are revolving at a very fast pace. With many production facilities and related suppliers gravitating more toward technologically-oriented method of doing business. Embracing cloud-based ERP systems definitely comes with positive impacts

Table 4: Suggested Cloud-Based ERP System to Company Size

Enterprise	Medium-Sized	Small Business
SAP	Netsuite	Deltek
Oracle	Sage	Work(etc)
Microsoft Dynamics	Infor	Syspro
IFS Applications	Macola	Intacct

The Table 4 presents a list few ERP System fits the three categories of business types. ERP software varies widely from one system to the other systems because different vendors have different market focus and thereby offer features that attract targeted markets.

However, most systems offer numerous modules: Human resources, CRM, Finance/Accounting, IT Helpdesk, e-commerce, Supply Chain Management, Order Processing, Inventory, and Procurement. However, most systems can provide many of those modules: Human resources, Finance/Accounting, IT service, e-commerce, provide Chain Management, Order process, and Inventory.

Additionally, department specific modules, several ERP platforms unit are specifically designed for specific industries like producing, field service, or technology. Industry-focused ERP systems package useful options along and should provide a helpful start line

for those new ERP computer code. With ERP system customers will access any data at any time from anyplace victimization devices like a pill, smartphones.

4.5 Factors to Consider When Purchasing (Cloud-Base) ERP System

It is important for firms to migrate to an ERP system in order to remain relevant in the current market trend. Before migration firms must ensure that the preferred ERP system has the following features; It is important for firms to migrate to an ERP system in order to remain relevant in the current market trend. Before system migration firms must ensure that the preferred ERP system is fully equipped with the following features;

4.5.1 Upgrade vs. Replacement

Evaluate first if the firm needs to get a brand new ERP system or just an upgrade. Many current ERP is modularized and allows integration of new modules such as payroll with current applications. With this approached, the firm minimizes disruption and costs. With this also, a firm can leverage the cloud-based ERP for mobility, effective integration, increased scalability, deployment options, machine learning, and predictive analysis.

4.5.2 Training and Setup

On-premise ERP solutions need technical knowhow personnel to install but cloud-based ERP removes that barrier which in turn reduce cost. The installation often has charged independent from the cost of license but using SaaS or cloud-based ERP only requires an account activation to enable access all necessary service from the selected providers' server.

4.5.3 Reporting and Dashboards

Current ERP solutions feature an advanced reporting interface that can generate complete financial statements depending on the configuration. Check available tools such as reporting tools, in-system query, smart filters, and real-time data visualization. Dashboards include functions such as quantitative vs. qualitative data at a user, role, and department levels.

4.5.4 Integration

Desired ERP should seamlessly work with your currently existing business applications; this is essential when firm want to upgrade their existing system. Firms should consider the following integration points; system-system (such as CRM, HRMS), module-system (ERP payroll with existing HRMS) and also have varieties of export/import capabilities (PDF, JPG, CSV extension). The selected ERP system should have flexible integration process in order to ensure safe and smooth records and files migration.

4.5.5 Customization

This depends on every firm's priorities and working culture. Therefore, firms research and for an ERP System with customization tools such as departmental dashboard localization, and configurable workflows, among others. This allows departments such as marketing, accounting, production, sales to effectively execute their define goals and organize the ERP environment based on their parameters.

5 Conclusion

In conclusion, the study defined as purchasing is acquiring goods to satisfy organizational needs, individuals or departments. The key roles of purchasing goods are buying the inputs competitive price and ensuring reliable and consistent flows of goods. The use of an information technology system in purchasing help enhanced the operation and process in an organization by eliminating errors.

The purchasing process has become more efficient because the vast volume of paperwork is now reduced and cross-checking of records from the supplier and the customer's product specifications. This process is underlined with sufficient procurement planning, strategic resource allocations, staff competency, and procurement procedures respectfully. Thus enabling the organization to utilize the information processing features that collect analyze and coordinate in order to establish productive operational, effective and strategic decisions [12]. Recent studies have based on effects associated with the use of IS in supply chains.

There is evidence of an upwards productivity for a well-planned with supply chain coordination and integration that is facilitated by the use of integrated information technologies [44]. The suggested way to promoting effective Purchasing Process for future requirements to use tools that enhances the online requisitions. The use of electronic data interchange and online ordering through electronic catalogs are the new technology routes that individuals and organizations use to purchase their products.

Therefore, the on-site ERP system is being fully replaced by the cloud-based ERP, because it utilizes up-to-date and available cutting-edge technology platforms to reach a vast number of users and business around.

Conflict of Interest The authors; Pearl Keitemoge and Daniel Tetteh Narh certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Acknowledgment We are grateful for the opportunity that was granted to us, and to be able to share more insight on Effective Application of Information System for Purchase Process Optimization. Also grateful for the support that we receive from the family and friends, thank you.

References

[1] W. D. Presutti, "Supply management and e-procurement: creating value added in the supply chain[J]," Industrial Marketing Management, **32**(3), 219–226, 2003, https://doi.org/10.1016/S0019-8501(02)00265-1.

[2] M. Ben-Daya, E. Hassini, Z. Bahroun, "Internet of things and supply chain management: a literature review," International Journal of Production Research, **57**(15-16), 4719–4742, 2019, https://doi.org/10.1080/00207543.2017.1402140.

[3] H. T. Y. C. Chang, H., C. H. Hsu, "E-procurement and supply chain performance[J]," Supply Chain Management An International Journal, **18**(1), 34–51, 2013, https://doi.org/10.1108/13598541311293168.

[4] S. Qrunfleh, M. Tarafdar, "Supply chain information systems strategy: Impacts on supply chain performance and firm performance[J]," International Journal of Production Economics, **147**(2), 340–350, 2014, https://doi.org/10.1016/j.ijpe.2012.09.018.

[5] D. Nelson, P. Moody, J. Stegner, "The purchasing machine," New York:The Free Press;, **184**, 2001.

[6] B. Hamza, A. Gerbi, S. H. Ali, Senait., "Factors Affecting Procurement Performance in the Case of Awassa Textile Share Company," Global Journal of Management And Business Research, **16**(3G), 2017, https://www.journalofbusiness.org/index.php/GJMBR/article/view/2085.

[7] A. Barros, . D. P, C. Ishikiriyama, P. S., C. R., "Processes and Benefits of the Application of Information Technology in Supply Chain Management: An Analysis of the Literature[J]," Procedia Computer Science, **55**, 698–705, 2015, https://doi.org/10.1016/j.ijpe.2012.09.018.

[8] F. Sollish, J. Semanik, The procurement and supply managers' desk reference, Wiley & Sons, USA New Jersey, 2nd edition, 2012.

[9] P. Triantafillou, "Benchmarking in the Public Sector: A Critical Conceptual Framework," Public Administration Journal, **8**(4), 17–32, 2007, https://doi.org/10.1111/j.1467-9299.2007.00669.x.

[10] T. S. Mclaren, M. M. Head, Y. Yuan, "Supply chain management information systems capabilities. An exploratory study of electronics manufacturers," Information Systems and E-Business Management, **2**(2-3), 207–222, 2004, https://doi.org/10.1007/s10257-004-0035-5.

[11] R. Shah, S. M. Goldstein, P. T. Ward, "Aligning supply chain management characteristics & inter-organizational information system types: an exploratory study," IEEE Transactions on Engineering Management, **49**(3), 282–292, 2002, https://doi.org/10.1109/TEM.2002.803382.

[12] S. Qrenfleh, M. Tarafdar, "Supply chain information systems strategy: Impacts on supply Chain performance and firm performance," international Production Economics, **147**(2014), 340–350, 2012, https://doi.org/10.1016/j.ijpe.2012.09.018.

[13] N. E. W. T. Gunasekaran, A., "Information systems in supply chain integration and management," European Journal of Operational Research, **159**(2), 269–295, 2004, https://doi.org/10.1016/j.ejor.2003.08.016.

[14] C. M. Madhusudhana, K. Prahlada, "Delivery Performance Management Supply Chain Management: case study in Batteries manufacturing firm," Serbian Journal of Management, **6**(2), 205–220, 2011.

[15] R. Vera, P. ME., "An analysis of the supplier selection process," Omega International Journal of Management Science, **26**(6), 739–750, 1998, https://doi.org/10.1016/S0305-0483(98)00023-1.

[16] S. K. Ruth M, "Factors Affecting the Selection of Optimal Suppliers in Procurement Management," International Journal of Humanities and Social Science, **2**, 14, 2012.

[17] S. Li, B. Ragu-Nathan, T. S. Ragu-Nathan, S. S. Rao, "The impact of supply chain management practices on competitive advantage and organizational performance," Omega, **34**(2), 107–124, 2006, https://doi.org/10.1016/j.omega.2004.08.002.

[18] K. C. Tan, V. R. Kannan, R. B. Handfield, "Supply chain management: supplier performance and firm performance," International journal of purchasing and material management, **34**(3), 2–9, 1998, https://digitalcommons.usu.edu/manage_facpub/278/.

[19] P. Baily, D. Farmer, B. Crocker, D. Jessop, D. Jones, Procurement Principles and Management, Prentice-Hall, Pearson Education Limited, England, 10th edition, 2008.

[20] L. L. Stanley, J. D. Wisner, "Service quality along the supply chain: implications for purchasing," Journal of operations management, **19**(3), 287–306, 2001, https://doi.org/10.1016/S0272-6963(00)00052-8.

[21] S. E. Fawcett, G. M. Magnan, Achieving world-class supply chain alignment: benefits, barriers, and bridges, Center for Advanced Purchasing Studies Tempe, AZ, 2001.

[22] N. B. Patil, M. Samel, P. Tilak, D. Boban, "Evolution of Modern Enterprise Resource Planning (ERP) Systems on Technological Background," International Journal of Science and Research, **4**(8), 2319–7064, 2015, https://www.ijsr.net/search_index_results_paperid.php?id=SUB157605.

[23] J. C. Wortmann, "Evolution of ERP Systems, Strategic Management of the Manufacturing Value Chain," The International Federation for Information Processing, **2**, 1998, https://doi.org/10.1007/978-0-387-35321-0_2.

[24] Y. B. Moon, "Enterprise Resource Planning (ERP):a review of the literature," International Journal of Management and Enterprise Development, **4**(3), 2007, https://doi.org/10.1504/IJMED.2007.012679.

[25] G. Group, "A Vision of Next Generation MRP II," Scenario S-300-339, **4**, 1990.

[26] T. Beleţ, A. A. Purcărea, "The Evolution of Enterprise Resource Planning Systems Eng," International Journal of Advanced Engineering, Management and Science, **3-Issue-12**, 2017, https://dx.doi.org/10.24001/ijaems.3.12.1.

[27] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, "A view of cloud computing," Commun. ACM, **53**, 50–58, 2010, http://doi.acm.org/10.1145/1721654.1721672.

[28] S. Sadagopan, The World of ERP in ERP: A Manegerial perspective, Tata Publishing Company Limited, McGraw-Hill, New Delhi, 1999.

[29] S. H. Chung, C. A. Snyder, "ERP Adoption:A Technological Evolution Approach," International Journal of Agile Management Systems, **2 (1)**(2), 24–32, 2000, https://doi.org/10.1108/14654650010312570.

[30] C. Yen D. C., "A Synergic Analysis for Web-Based Enterprise Resource Planning Systems," Computer Standards and Interfaces, **24**, 337–346, 2002, https://doi.org/10.1016/S0920-5489(01)00105-2.

[31] M. A. Vouk, "Cloud Computing Issues, Research and Implementations," Journal of Computing and Information Technology, **16**, 235–246, 2008, https://doi.org/10.2498/cit.1001391.

[32] E. Bjelland, M. Haddara, "Evolution of ERP Systems in the Cloud: A Study on System Updates," Systems, **6**, 2, 2018, https://doi.org/10.3390/systems6020022.

[33] M. Haddara, A. Fagerstrøm, B. Mæland, "Cloud Erp Systems Anatomy of Adoption Factors and Attitudes," journal Enterp Resour. Plan Stud, **22**, 2015, https://doi.org/10.5171/2015.521212.

[34] J. Duan, P. Faker, A. Fesak, T. Stuart, "Benefits and drawbacks of cloud-based versus traditional ERP systems," Proceedings of the 2012-13 course on Advanced Resource Planning, 2013, https://www.academia.edu/2777755/Benefits_and_Drawbacks_of_Cloud_Based_versus_Traditional_ERP_Systems.

[35] A. Elragal, M. Haddara, "The Future of ERP Systems: Look backward before moving forward," Procedia Technol, **5**, 21–30, 2012, http://dx.doi.org/10.1016/j.fcij.2017.03.003.

[36] Z. Al-Mashari.M, "Supply chain re-engineering using enterprise resource planning (ERP) systems: an analysis of a SAP-R/3 implementation case," International Journal of Physical Distribution & Logistics Management, **30**(34), 296–313, 2000, https://doi.org/10.1108/09600030010326064.

[37] G. Juell-Skielse, H. Enquist, Implications of ERP as Service in Re-Conceptualizing Enterprise Information Systems, Springer, Berlin, Germany, 2012, https://doi.org/10.1007/978-3-642-28827-2_10.

[38] G. F. H. Raihana, "Cloud ERP-A solution model," International Journal Computer Science Information Technology Security, **2**, 76–79, 2012, iSSN: 2249-9555.

[39] M. A. Rashid, L. Hossain, J. D. Patrick, "The Evolution of ERP Systems: A Historical Perspective," in I. Global, editor, Enterprise Resource Planning: Global Opportunities and Challenges, 1–16, IGI Global, 2002, 10.4018/978-1-931777-06-3.ch001.

[40] C. Calvert, P. B. Seddon, "The importance of ongoing ERP training and support," AISeL, **91**, 2006, https://aisel.aisnet.org/acis2006/91.

[41] C. C. Claybaugh, K. Ramamurthy, W. D. Haseman, "Assimilation of enterprise technology upgrades: A factor-based study," Enterp Inf. System, **11**, 250–283, 2017, http://DOI: 10.1080/17517575.2015.1041060;.

[42] A. Elragal, M. Haddara, "The Future of ERP Systems: Look backward before moving forward," Procedia Technol, **5**, 21–30, 2012, https://doi:10.1016/j.protcy.2012.09.003.

[43] R. Seethamraju, "Adoption of software as a service (SaaS) enterprise resource planning (ERP) systems in small and medium-sized enterprises (SMEs)," Inf Syst. Front, **17**, 475–492, 2015, https://DOI 10.1007/s10796-014-9506-5.

[44] S. K. Vickery, J. Jayaram, C. Droge, R. Calantone, "The effects of an integrative supply chain strategy on customer service and financial performance: and analysis of direct versus indirect relationships," Journal of Operations Management, **21**(5), 523–C539, 2003.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 606-612 (2020)

www.astesj.com

Special Issue on Innovative Research in Applied Science, Engineering and Technology

ASTESJ

ISSN: 2415-6698

Fast and efficient Maximum Power Point Tracking controller for photovoltaic modules

Khalid Chennoufi*, Mohammed Ferfra

Department of Electrical Engineering, Mohammadia School of Engineers, Mohammed V University in Rabat, Rabat, PB765, Morocco

ARTICLE INFO

Article history:
Received: 31 July, 2020
Accepted: 06 November, 2020
Online: 24 November, 2020

Keywords:
Artificial Neural Network
Backstepping
MPPT
Double diode model
SEPIC

ABSTRACT

This paper presents an efficient Maximum Power Point Tracking (MPPT) controller for photovoltaic modules. The MPPT technique consists of a combination between backstepping controller and artificial neural network (ANN).The (ANN) has been employed to generate the optimum voltage, which corresponds to the maximum power voltage delivered by photovoltaic modules, while the backstepping controller is developed to track the generated voltage, by computing the duty cycle of the Single Ended Primary Inductor Converter (SEPIC). The control of the boost converter is based on Lyapunov stability analysis, and an integral action is added to increase system robustness. In order to prove the accuracy of the developed control a comparison between the proposed method and sliding mode was carried out. In addition the stability was evaluated under sudden variation of environmental conditions. The simulation was carried out in MATLAB software, the results shows that the proposed controller tracks the reference voltage within 25 ms, in addition the systems reacts to sudden environments change with no oscillations, which demonstrate the robustness of the proposed method.

1. Introduction

In recent years, the demand of electrical energy increases, which reduces the reserves of the conventional resources and increases the global warming, therefore the use of renewable energies is the only way to reduce the dependence on fossil fuels and ensure the energy security. In fact, photovoltaic energy is the most popular renewable energy source, however, the output power relies on weather conditions and the load. This led the researchers to present various algorithms that make photovoltaic modules operates in the maximum power point. Thus a various Maximum Power Point Tracking algorithms with different degrees of complexity and precision have been proposed in the literature [1-3]. The perturb and the observe, and incremental conductance [4], methods are often used due to their simplicity and ease of implementation, however their main drawback is the fluctuations around the point of maximum power.

Thus to overcome the oscillations issues and increase the performance of the tracking a combination between direct control methods and the nonlinear commands has been employed by the researchers.

In [5], Authors propose a maximum power point tracking controller for hybrid Photovoltaic - Wind energy system. The controller is developed by the combination of perturb & observe and Sliding Mode. P&O loop is used for the computation of the reference voltage, while Sliding Mode is designed to track this voltage by generating the control signal. The proposed method exhibits good convergence speed, but the main disadvantage is the oscillations around maximum power point.

In [6], Authors present a control of a single-phase Photovoltaic inverter using backstepping algorithm which track the desired voltage generated by Perturb and Observe method and adjust the duty cycle in order to operate in maximum power point. The error between the output and reference voltage was eliminated quickly, but some fluctuations have been observed in the desired operating point.

In [7], Authors propose a maximum power point tracking controller for Photovoltaic system. The controller is developed by the combination of Sliding Mode and Artificial Neural Network (ANN). The ANN is employed to generate the reference voltage, while Sliding Mode is designed to track this voltage by generating the control signal. The proposed method presents good

*Corresponding Author: Khalid Chennoufi, Tel: +212 668 01 15 90, khalidchennoufi@research.emi.ac.ma

performance; however, some oscillations is noticed around maximum power point.

In order to overcome fluctuation issues around maximum power point, this paper proposes a hybrid method, which consists of a backstepping controller combined to the Artificial Neural Network (ANN). The ANN is employed to generate the reference voltage, while backstepping controller is designed to adjust the duty cycle of the converter in order to track the desired voltage, which corresponds to the voltage of the Maximum Power Point, (Fig1).

In contrast of previous methods which adopt single diode model, the photovoltaic modelling was carried out using double diode model, as the modelling using single diode degrade at low irradiance. In addition the present work doesn't rely on any simplification which can compromise the accuracy of the controller modelling, furthermore the SEPIC converter was employed since it offer more advantages instead of the traditional converter.

The present paper is organized as follows. In Section 2, the modelling of the PV system is presented. The Section 3, presents the design of the proposed controller. The simulations and discussions are illustrated in Section 4.

2. Modelling of PV system

The Photovoltaic system proposed in this paper consists of a photovoltaic module connected to a resistive load through a Single ended primary converter as shown in Figure 1.

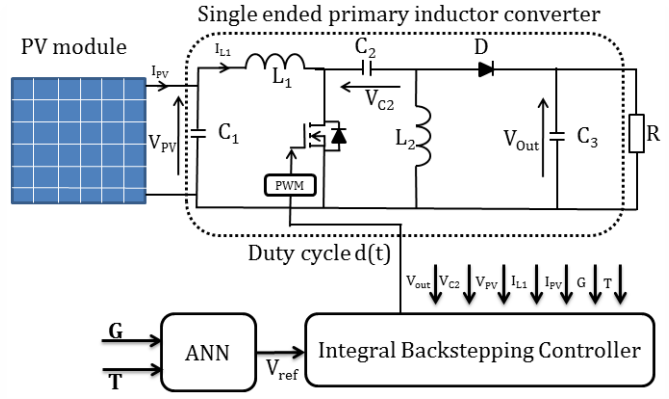


Figure 1: Illustration of the proposed Photovoltaic system.

Where V_{out}, V_{c2}, V_{PV} are the output voltage, the voltage across the capacitor C_2, and the photovoltaic voltage respectively, I_{pv} and I_{L1} are the photovoltaic and inductor 1 currents, While G and T are respectively the irradiation and the Temperature.

The datasheet parameters of the polycrystalline photovoltaic module KD 245 GH are presented in Table 1.

Table 1: PV module datasheet

Parameter	Value
Short-circuit current (I_{sc})	8.91(A)
Open circuit voltage (V_{oc})	36.9 (V)
Maximum current (I_m)	8.23 (A)
Maximum voltage (V_m)	29.8 (V)
Number of cells in series (N_s)	60
Temperature coefficient of current (k_i)	5.35×10^{-3} (A/K)
Temperature coefficient of voltage (K_V)	-0.133(V/K)

2.1. Modelling of the SEPIC converter

The SEPIC is an acronym of (single ended primary inductor converter), which is a DC-DC converter capable to operate as boost or buck mode [8], that offer the possibility to raise or decrease the voltage at the output [9]. In addition the SEPIC converter makes the PV system operate at its maximum power point with 'weak current ripples [10]. Furthermore it can work on any operating point by changing the duty cycle [11]. The converter is modelled mathematically based on these states [12] as shown in figure 2.

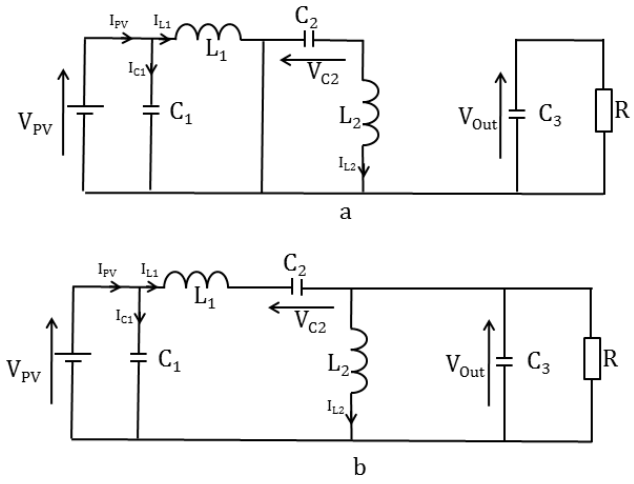


Figure 2: SEPIC converter circuit, (a) Equivalent circuit switch is ON (b) switch is OFF

When the commutator is on (u=1) as shown in figure 2-a. The capacitor C_3 supplies energy to the load and the energy stored in the inductances increases, the inductor L1 is alimented from the input while the inductor L_2 is feeded from the capacitor C_2. The representation of state space can be described as

$$\frac{dI_{L1}}{dt} = \frac{V_{PV}}{L_1} \quad (1)$$

$$\frac{dI_{L2}}{dt} = \frac{-V_{C2}}{L_2} \quad (2)$$

$$\frac{dV_{c2}}{dt} = \frac{I_{L2}}{C_2} \quad (3)$$

When the commutator is off (u=0) as shown in figure 2-b. The two inductors are discharged and the energy stored in the capacitors increases. The representation of state space can be written as:

$$\frac{dI_{L1}}{dt} = \frac{V_{PV} - V_{C2} - V_{out}}{L_1} \quad (4)$$

$$\frac{dI_{L2}}{dt} = \frac{V_{out}}{L_2} \quad (5)$$

$$\frac{dV_{C2}}{dt} = \frac{I_{L1}}{C_2} \quad (6)$$

The combination between the two state spaces can be expressed as:

$$\frac{dI_{L1}}{dt} = \frac{(u-1)V_{C2}}{L_1} + \frac{(u-1)V_{out}}{L_1} + \frac{V_{PV}}{L_1} \quad (7)$$

$$\frac{dV_{C2}}{dt} = \frac{(1-u)I_{L1}}{C_2} + \frac{uI_{L2}}{C_2} \quad (8)$$

$$\frac{dI_{L2}}{dt} = -\frac{dV_{C2}}{L_2} + \frac{(1-u)V_{out}}{L_2} \quad (9)$$

$$\frac{dV_0}{dt} = \frac{(1-u)I_{L1}}{C_3} + \frac{(u-1)I_{L2}}{C_3} - \frac{V_{out}}{RC_3} \quad (10)$$

where u=(0 or 1)

2.2. Modelling of the Photovoltaic module

In this work the modelling of a solar module is developed based on double diode model which is illustrated in figure 3, where the first diode describes the saturation currents due to diffusion mechanism, while the second diode describes the current due to recombination in the depletion zone.

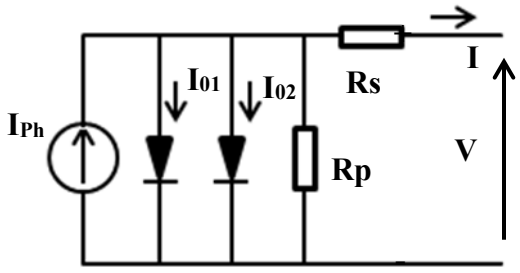


Figure 3: Illustration of the double diode model.

The output current of a solar cell is represented by:

$$I = I_{ph} - I_{01}\left(exp\left(\frac{V+R_sI}{a_1V_T}\right) - 1\right) - I_{02}\left(exp\left(\frac{V+R_sI}{a_2V_T}\right) - 1\right) - \frac{V+R_sI}{R_p} \quad (11)$$

The unknown parameters such as: a_1,a_2,I_{01},I_{02},I_{ph},R_s and R_p. are extracted using a combination between the equation of three main operating points which are, the open circuit voltage, the short circuit current and the maximum power point. The objective is to find the equations of I_{01}, I_{02}, R_p and I_{ph} which depend only of Rs, a1 and a2. The value of series resistance (Rs) and ideality factors (a_1,a_2) are obtained using a fast iteration by adjusting simulated and datasheet powers [13]. The extracted parameters are illustrated in Table 2.

The extracted parameters are integrated in a MATLAB algorithm to plot V-I and V-P curves as shown in Figs. 4, the simulation allow to predict the behaviour of the photovoltaic modules in different meteorological condition, thus when the temperature increases the open circuit voltage decrease while the short circuit current increase slightly which result a loss of generated power. On the other hand, when the solar irradiance increases both short circuit current and the open circuit voltage increase which raise the generated power.

Table 2: Extracted Parameters of KD 245GH module

Parameter	Value
a_1	1.005879
a_2	1.994120
I_{ph} (A)	8.931164
I_{01} (A)	3.906668×10^{-10}
I_{02} (A)	1.094485×10^{-06}
R_s (Ω)	0.286199
R_p (Ω)	120.495147

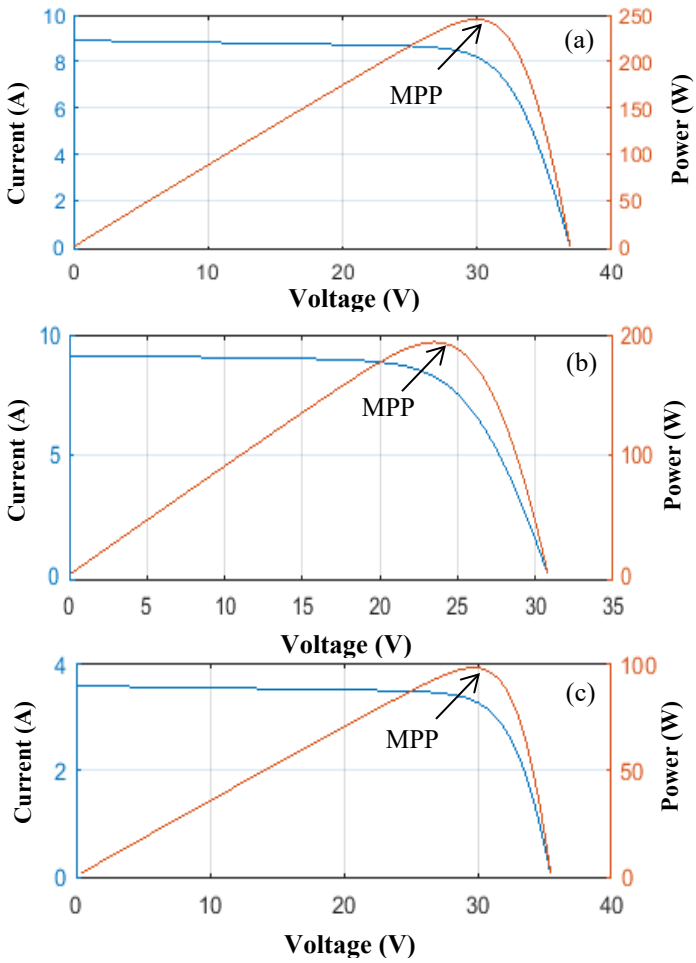


Figure 4: Generated current and power at 25°C - 1000 W/m² (a), 75°C - 1000 W/m² (b) and 25°C - 400 W/m² (c)

3. Proposed controller:

Figure 1 presents the schematic drawing of the proposed controller, it consists of a combination between Artificial Neural Network and backstepping controller. The ANN loop generates a voltage reference for every solar irradiation and temperature, which correspond to the maximum power point voltage, while the backstepping controller tracks the generated reference by adjusting the duty cycle.

3.1. Artificial Neural Network

Artificial neural network is a computational processing model based on artificial intelligence that is meant to simulate the functioning of a human brain , which consists of processing units interconnected by nodes [14]. The ANN architecture consists of three layers as shown in Figure 5, The input layer receives various forms and structures of information based on an internal weighting system, the hidden layer is where artificial neurons get an ensemble of weighted inputs and generates an output via an activation function, while and the output layer receives data from hidden layer and gives the results of the ANN [15, 16].

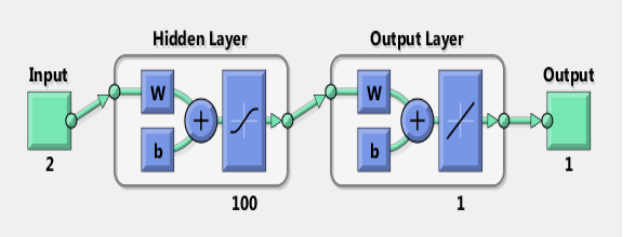


Figure 5: Proposed Neural network hierarchy

In this work the input layer was designed with two neurons, irradiation and the temperature, a database of 158 cases was created by increasing the irradiance by a step of 5 K from 278 K to 333 K and changing the irradiance for each temperature from 200 to 1050 W/m². The hidden layer was created with 100 neurons as shown in Figure 5. The output layer has only one neuron which corresponds to the voltage at maximum power of the 158 pair irradiance temperature.

3.2. Integral backstepping

The backstepping controller is designed to track the reference voltage generated by the artificial neural network by adjusting the duty cycle of the converter. The control law is generated based on Lyapunov stability analysis. The design of the controller is as follows:

The tracking error is defined as the difference between photovoltaic and reference voltages:

$$e_1 = Vpv - Vref \tag{12}$$

An integral action is added in order to improve the tracking precision as follow:

$$e = e_1 + \lambda e_{Int} \tag{13}$$

The tracking error becomes:

$$e = Vpv - Vref \quad + \lambda \int (Vpv - Vref)dt \tag{14}$$

Where λ is a positive constant

The modified Lyapunov function can be written as:

$$V_1 = \frac{1}{2}e_1^2 + \frac{\lambda}{2}e_{Int}^2 \tag{15}$$

The derivative of the (15) with respect to time is:

$$\dot{V}_1 = e_1 \left[(\dot{V}_{pv} - \dot{V}_{ref}) + \lambda \int (Vpv - Vref)dt \right] \tag{16}$$

To guarantee the asymptotic stability, the Lyapunov function has to be negative definite derivative:

$$\dot{V}_1 = -K_1 e_1^2 \tag{17}$$

where K_1 is a positive constant

The following equation is obtained:

$$\left[(\dot{V}_{pv} - \dot{V}_{ref}) + \lambda \int (Vpv - Vref)dt \right] = -K_1 e_1 \tag{18}$$

The photovoltaic current is given by:

$$I_{PV} = I_{C1} + I_{L1} \tag{19}$$

The derivative of photovoltaic tension can be written as:

$$\dot{V}_{pv} = \frac{I_{PV} - I_{L1}}{C_1} \tag{20}$$

The desired value of inductor current would be:

$$I_d = I_{PV} + C_1 \left[(K_1 e_1 - \dot{V}_{ref}) + \lambda \int (Vpv - Vref)dt \right] \tag{21}$$

where (I_d=I_{L1})

The second error represents the difference between the inductor and the desired current:

$$e_2 = I_{L1} - I_d \tag{22}$$

The derivative of Lyapunov function with respect to the time can be expressed as

$$\dot{V}_1 = e_1 \left[\left(\frac{I_{PV} - (e_2 + I_d)}{C_1} - \dot{V}_{ref} \right) + \lambda \int (Vpv - Vref)dt \right] \tag{23}$$

By using equation (18) and (20) the above equation can be written as

$$\dot{V}_1 = -k_1 e_1^2 - \frac{e_1 e_2}{C_1} \tag{24}$$

The time derivative of (22), using equation (9) can be written as:

$$\dot{e}_2 = \frac{1}{L_1} V_{pv} - \frac{1}{L_1}(1 - u)(V_{out} + V_{c2}) - \dot{I}_d \tag{25}$$

Where:

$$\dot{I}_d = \dot{I}_{PV} + C_1 \left[(K_1 e_1 - \ddot{V}_{ref}) + \lambda (Vpv - Vref) \right] \tag{26}$$

The derivative of photovoltaic current with respect to the time can be expressed as:

$$\dot{I}_{PV} = \frac{dI_{PV}}{dV_{PV}} \frac{dV_{PV}}{dt} \quad (27)$$

By differentiating the equation (11) with respect to the voltage the derivative of photovoltaic current is obtained:

$$\dot{I}_{PV} = \left[-\frac{I_{01}\exp\left(\frac{V + R_s I_{PV}}{a_1 V_T}\right)}{a_1 V_T} - \frac{I_{02}\exp\left(\frac{V + R_s I_{PV}}{a_2 V_T}\right)}{a_2 V_T} - \frac{1}{R_P} \right] \dot{V}_{PV} \quad (28)$$

In order to guarantee the asymptotic stability of the system and the convergence of the errors e_1 and e_2 to zero, a composite Lyapunov function V_2 is defined as:

$$V_2 = V_1 + \frac{1}{2} e_2^2 \quad (29)$$

The derivative of V_2 with respect to the time is:

$$\dot{V}_2 = \dot{V}_1 + e_2 \dot{e}_2 \quad (30)$$

By using equation (24) the derivative of V_2 can be expressed as:

$$\dot{V}_2 = -K_1 e_1^2 + e_2 \left[\frac{1}{L_1} V_{pv} - \frac{1}{L_1}(1 - u)(V_{out} + V_{c2}) - \dot{I}_d - \frac{e_1}{C_1} \right] \quad (31)$$

The time derivative of Lyapunov function has to be negative definite:

$$\frac{1}{L_1} V_{pv} - \frac{1}{L_1}(1 - u)(V_{out} + V_{c2}) - \dot{I}_d - \frac{e_1}{C_1} = -K_2 e_2 \quad (32)$$

where K_2 is a positive constant

The control law, which guarantees e_1, e_2 converges asymptotically to 0, is given by:

$$u = 1 - \left[V_{PV} + L_1 \left(K_2 e_2 - \frac{e_1}{C_1} - \dot{I}_d \right) \right] \frac{1}{V_{C2} + V_0} \quad (33)$$

4. Simulation results

In order to study the performance of the proposed method, a simulation was carried out in Matlab software. The parameters of the simulation are illustrated in Table 3.

Table 3: Parameters of the SEPIC

Parameter	Value
Inductor 1 (L_1)	2.5 mH
Inductor 2 (L_2)	2.5 mH
Capacitor 1 (C_1)	4500 uF
Capacitor 2 (C_2)	79 µF
Capacitor 3 (C_3)	2400 µF
Constant (λ)	1
Constant (K_1)	10^4
Constant (K_2)	10^4
Load resistor (R)	50Ω
Switching frequency (f)	7 KHz

The system being simulated into different temperature and irradiation condition as shown in Figure 6, During the first case the [0s, 20s], the PV module is consider operating under the irradiation of the maximum power point while the temperature was raised to 318 K to simulate the temperature effect. In the second case [20s, 40s], the irradiation drop smoothly from 1000 W/m² to 600 W/m² while the temperature decrease between two consecutive levels from 318 k to 307 K. In the last case [40s, 60s], the temperature is considered constant in 298K while the irradiation increase among two successive stage to reach 1000 W/m² thus the PV module is considered operating in the maximum power point.

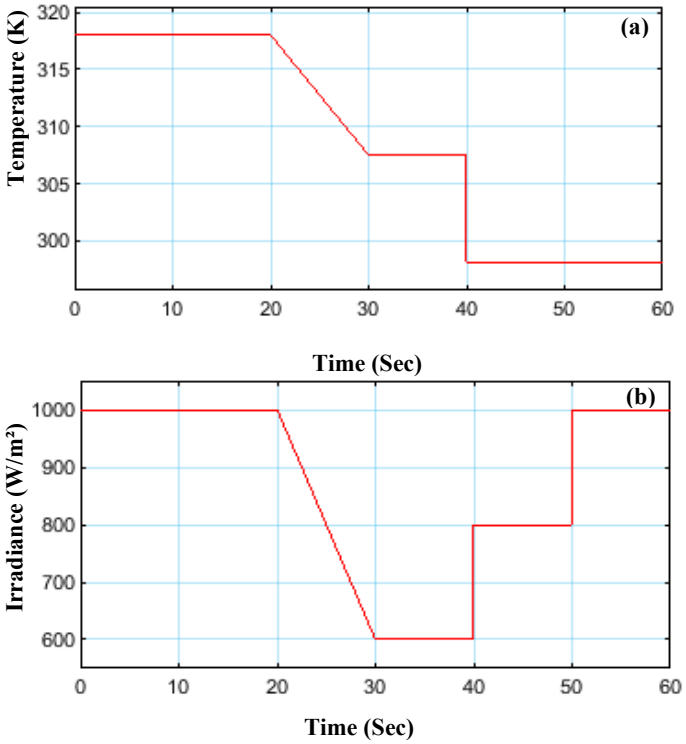


Figure 6: Temperature (a) and Irradiation (b)

Figure 7 shows that the variations in temperature and irradiation cause a slight variation in voltage and a large variation in current and power. By comparing the proposed method with the method proposed by [7] it can be seen that the proposed method bring reach the reference voltage in 25 ms, Figure (b), with high accuracy, in contrast the proposed method in [7] reaches the desired voltage with a delay of 60 ms with large oscillations. In addition the obtained curves exhibit fast and good reactions when sudden variation in temperature and irradiance occurs.

Figure 8 shows the output current, voltage and power of the PV module generated at the output of the SEPIC converter using the proposed method with the result proposed by [7]. As can be seen from the figures (a), (b) and (c), the obtained curves using sliding mode controller present large oscillations and the use of the proposed method allow reducing significantly the magnitude of the fluctuation, which guarantees good performance in the maximum point tracking.

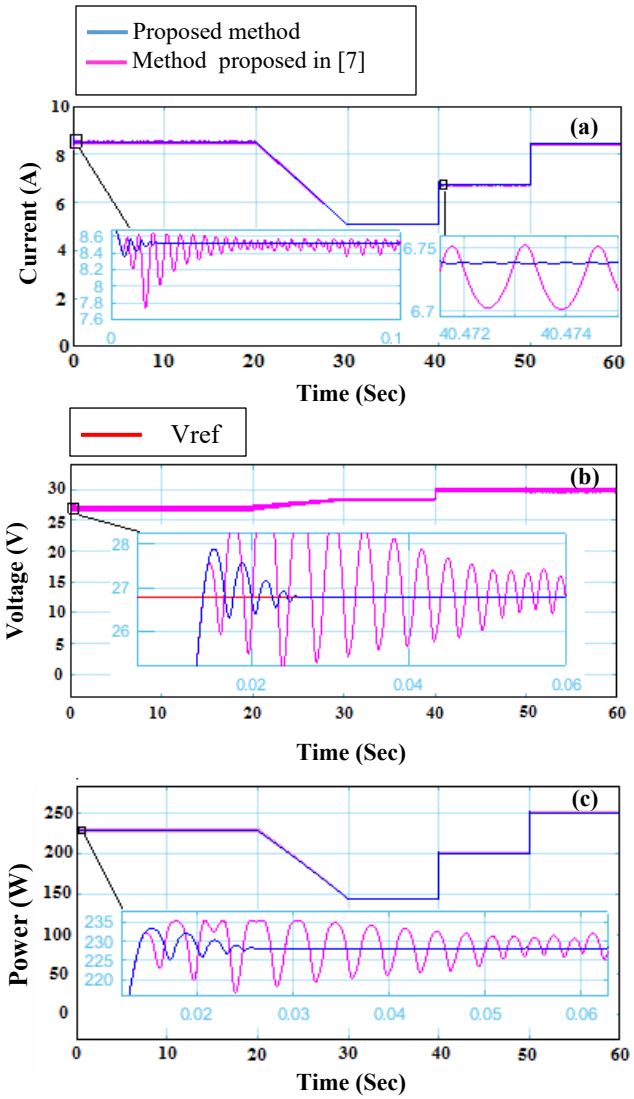
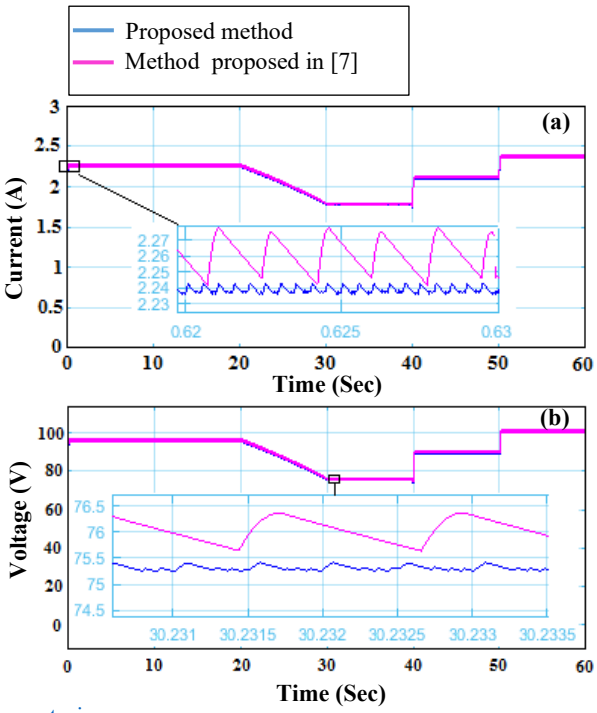


Figure 7: Photovoltaic Current (a), Voltage (b) and Power (c),



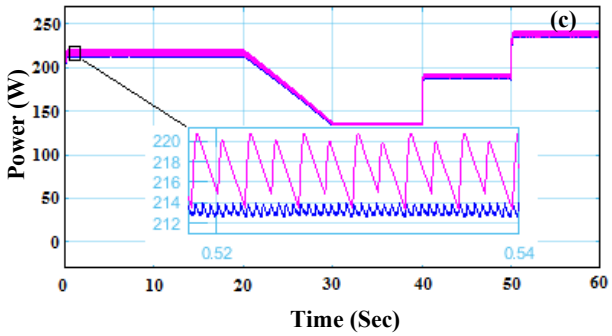


Figure 8: SEPIC output Current (a), Voltage (b) and Power (c).

5. Conclusion

In this paper an integral backstepping controller for maximum power point tracking is proposed. In order to increase the performance of tracking the controller was developed by combining backstepping controller and artificial neural network (ANN). The photovoltaic modelling was carried out using double diode model as the modelling using single diode degrade at low irradiance, and in order to reduce the current ripples the single ended primary inductor converter (SEPIC) has been employed. The simulation has been done in Matlab/Simulink software, the results show that the proposed method tracks the reference voltage with good accuracy and fast convergence speed, in addition the photovoltaic power reaches maximum power in less than 25 ms. Furthermore the obtained curves show that there are no oscillations in the output voltage, current and power. Thus the proposed method presents a fast and stable reaction when the environmental conditions changes, which confirms the effectiveness and the stability of the proposed method.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] T. Esram, P.L. Chapman, "Comparison of photovoltaic array maximum power point tracking techniques," IEEE Transactions on Energy Conversion, **22**(2), 439–449, 2007, doi:10.1109/TEC.2006.874230.

[2] M.K. Dr S.R.Kapoor, "Comparison between IC and Fuzzy Logic MPPT Algorithm Based Solar PV System using Boost Converter," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, **04**(06), 4927–4939, 2015, doi:10.15662/ijareeie.2015.0406007.

[3] J.P. Ram, T.S. Babu, N. Rajasekar, "A comprehensive review on solar PV maximum power point tracking techniques," Renewable and Sustainable Energy Reviews, 67, 826–847, 2017, doi:10.1016/j.rser.2016.09.076.

[4] M. Al-Dhaifallah, A.M. Nassef, H. Rezk, K.S. Nisar, "Optimal parameter design of fractional order control based INC-MPPT for PV system," Solar Energy, 159(June 2017), 650–664, 2018, doi:10.1016/j.solener.2017.11.040.

[5] O. Zebraoui, M. Bouzi, "Robust sliding mode control based MPPT for a PV/Wind hybrid energy system," International Journal of Intelligent Engineering and Systems, **11**(5), 290–300, 2018, doi:10.22266/IJIES2018.1031.27.

[6] O. Diouri, N. Es-Sbai, F. Errahimi, A. Gaga, C. Alaoui, "Modeling and Design of Single-Phase PV Inverter with MPPT Algorithm Applied to the Boost Converter Using Back-Stepping Control in Standalone Mode," International Journal of Photoenergy, **2019**, 2019, doi:10.1155/2019/7021578.

[7] S.A. Touil, N. Boudjerda, A. Boubakir, K. El Khamlichi Drissi, "A sliding mode control and artificial neural network based MPPT for a direct grid-connected photovoltaic source," Asian Journal of Control, **21**(4), 1892–1905, 2019, doi:10.1002/asjc.2007.

[8] W.M. Moussa, "Modeling and Performance Evaluation of a DC/DC SEPIC Converter," Proceedings of 1995 IEEE Applied Power Electronics Conference and Exposition, 1995, 1995, doi: 10.1109/APEC.1995.469096

[9] D. Adar, G. Rahav, S. Ben-Yaakov, "A unified behavioral average model of SEPIC converters with coupled inductors," Power Processing and Electronic Specialists Conference **1997**, 1997,doi: 10.1109/PESC.1997.616761

[10] E. Durán, M. Sidrach-De-Cardona, J. Galán, J.M. Andújar, "Comparative analysis of Buck-Boost converters used to obtain I-V characteristic curves of photovoltaic modules," PESC Record - IEEE Annual Power Electronics Specialists Conference, 2036–2042, 2008, doi:10.1109/PESC.2008.4592243.

[11] A.R. Kashyap, R. Ahmadi, J.W. Kimball, "Input voltage control of SEPIC for maximum power point tracking," 2013 IEEE Power and Energy Conference at Illinois, PECI 2013, 30–35, 2013, doi:10.1109/PECI.2013.6506030..

[12] R.W. Erickson, D. Maksimovic, Fundamentals of Power Electronics, 2nd Edition, 2001.

[13] K. Chennoufi, M. Ferfra, M. Mokhlis, "An accurate modelling of PV Modules based on two-diode model," Renewable Energy, 2020, https://doi.org/10.1016/j.renene.2020.11.085.

[14] N. Nambiar, R.S. Palackal, K. V. Greeshma, A. Chitra, "PV fed MLI with ANN based MPPT," 4th IEEE Sponsored International Conference on Computation of Power, Energy, Information and Communication, ICCPEIC 2015, 293–300, 2015, doi:10.1109/ICCPEIC.2015.7259478.

[15] S.Z. Hassan, H. Li, T. Kamal, J. Ahmad, M.H. Riaz, M. A. Khan, "Performance of Different MPPT Control Techniques for Photovoltaic Systems" in 5th International Conference on Electrical Engineering, UET Lahore, Pakistan from Feb. 15-16th, 2018, doi: 10.1109/ICEE.2018.8566713

[16] S.Z. Hassan, H. Li, T. Kamal, L. Khan, S. Mumtaz, "Neuro-Fuzzy Wavelet Based Adaptive MPPT Algorithm for Photovoltaic Systems" Energies, **10**(3), 2017, doi: 10.3390/en10030394

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 613-626 (2020)

www.astesj.com

Special Issue on the Economic Impact of COVID-19 on Companies and Consumers

ASTESJ

ISSN: 2415-6698

Inventory Management Practices during COVID 19 Pandemic to Maintain Liquidity Increasing Customer Service level in an Industrial Products Company in Mexico

Ignacio Alvarez-Placencia*, Diana Sánchez-Partida, Patricia Cano-Olivos, José-Luis Martínez-Flores

Department of Logistics and Supply Chain Management, UPAEP University, 17 Sur 901, Barrio de Santiago, CP 72410 Puebla, Puebla, México

ARTICLE INFO

Article history:
Received: 01 September, 2020
Accepted: 26 October, 2020
Online: 24 November, 2020

Keywords:
Customer Service
Inventory Control
Delivery Performance
Lean Manufacturing

ABSTRACT

This work presents the development and improvement obtained by implementing inventory managing practices during the COVID 19 pandemic in a Mexican company, a leader in the industrial sector to overcome the effects on the Company's economies due to constrained market to maintain liquidity and preserve employment by increasing customer service level ensuring the availability of the finished product and improving the delivery time in the distribution center, from the receipt of the customer's order, order picking, shipment preparation, invoicing and final delivery to the customer. The areas involved are Commercial, Demand and Materials Planning, Purchasing, Manufacturing, Logistics, and Operations of the Distribution Center. The deployment process considered an ABC study, analysis of the historical demand, product inventory coverage days, analysis of the lead time components of the distribution center, to implement a minimum and maximum inventory control policy for the product portfolio A and manufacture products classified as B and C on a make to order basis, an audit was performed of its distribution center, to identify opportunities for improvement in their processes to reduce the time of delivery to the customer.

The Company was not in compliance with the level of customer service with an 89% performance due to the low availability of the finished product inventory, affecting the customer's delivery time in the distribution center with a time of 5 hours versus the goal of 90 minutes per event.

The implementation of the improvement achieved an increase in the level of service up to 5% starting the fourth quarter of 2019, the savings in inventory optimization at an annual base represented USD 530,785.00 (Five hundred thirty thousand seven hundred eighty-five dollars), and the Company has implemented this good practice nationwide in its other Manufacturing Plants and Distribution Centers as part of the strategic corporate policy deployment.

1. Introduction

The value chain is all actions to make a product to flow from the raw material to the customer's hands. Within the value chain, the primary processes include the transformation of the product, inbound logistics that is the delivery of suppliers of the raw material to a warehouse of the manufacturing Company, the logistics of operations whose purpose is the proper administration of the raw material in the transformation process to ensure the availability of raw material inventory or inventory in process in the work centers and finally the outbound logistics from the storage, distribution and perfect delivery of the finished product to the end customer [1].

The secondary processes in a value chain are the ones that give support to the primary processes for the transformation of the product. The timeline determines the response time of the value chain, and the flow of information in the value chain is essential because it includes the release of the sales estimate, placement of the customer order, release of the purchase orders and manufacturing orders to the delivery of the product to the end customer [1].

The efficiency and optimization of the distribution center's functions is a priority for the achievement of the goal of perfect delivery to the customer and inventory availability assurance to ensure the desired service level [2].

* Corresponding Author: Ignacio Alvarez-Placencia, ignacio.alvarez@upaep.edu.mx

This case study's scope is a company of 100% Mexican capital, market leader, is dedicated to the production and marketing of industrial products located in Mexico with ten manufacturing plants and three Distribution Centers with 800 employees.

This case study helps the Company to face COVID 19 pandemic with a constrained market to contain the economic effects of the pandemic that we are facing helping to preserve employment as much as possible and provide liquidity to create the conditions for the recovery of economic growth by ensuring the availability of the inventory of its products to meet the level of delivery service on time and improve the delivery time to the customer in the distribution center through an audit in the distribution center, establish a baseline of product inventory coverage days, analysis of the lead time components of the distribution center, value chain mapping, ABC classification of finished products, inventory control implementation with the maximum and minimum methodology, production balancing, measurement of customer's delivery time as days of inventory coverage in the distribution center.

Masaaki Imai, in his work, "Kaizen, The Key to Japanese Competitiveness," mentions that Japanese companies follow a defined strategic planning scheme; the administration develops and communicates the rules of the game: transparent processes, that is, with a standard procedure for Operation. It translates into top management shall first establish policies and procedures for all critical operations, then communicate and ensure that they are understood by all levels [3].

The rest of the research work consists of describing the case study, theoretical framework, and research methodology as its conclusions and results obtained.

2. Literature Review

According to the customer's perspective, the value chain's purpose is to determine which activities are value-added and non-value-added [4]. All activity that transforms the product should be understood as an added value; in contrast, all the activities that do not transform the product are no value-added. They are known as waste: Reworks, defects, unnecessary movements, inventories, waiting times, overproduction, unnecessary space, and transportation. This waste shall be eliminated to improve the value chain [5].

A distribution center is a vital component of the value chain for any company to achieve perfect delivery to the product customer as to the quantity, time, price, quality, and correct information.

The distribution center functions include receipt, inspection, storage, picking, preparation of shipment, dispatch, consolidation of orders, and final delivery to the customer. The efficiency of these activities is critical to ensure speed and optimize the flow of the product to reduce the response time and operating costs, from receipt of the customer's order to delivery, to send the products with the correct specification, quality, price with proper documentation to the customer. Always searching for continuous improvement to reduce waste as mentioned earlier and add only value in the supply chain.

The correct management of the distribution center supports the strategic objectives of the Company: Generate more sales, profit margin, cash flow, stability in the jobs satisfying customer demand with the correct level of service.

Therefore, storage and distribution activities for the product should be regarded as a vital component of the Company's strategic planning for it to be coupled with the expectations raised in the business plan [6].

The management of the Company must consider several factors to maintain optimal inventory levels, such as the financial liquidity of the Company, sales behavior such as seasonality, reliability, and availability of suppliers, available financial resources. An adequate level of inventory reduces the potential risks of lost sales due to the lack of inventory availability as well as an increase in the cash flow of the Company, such as the release of warehouse capacity [7].

It is strategic for the companies to achieve a high level of customer service with a minimum level of inventory by implementing the Lean Pull Replenishment methodology by having product's supermarkets for the products that represent 80% of the Company's revenues which inventory is replenished based on the consumer's signal at the time of consumption and made to order for those products that represent the other 20% of the Company's revenue. The Lean Pull Replenishment methodology should be coupled with the Process Capability analysis to reduce variation in processes and determine the processing capabilities of the Company through the commitment of the delivery date of the customer's order [8].

Inventory in the lean philosophy is viewed as a waste given that it impacts the cash flow of the Company when it is not in the correct amount; besides, the Company requires an expense to manage it, at the same time, it hides problems such as downtime of machinery, long periods of changeover of products, quality problems, lack of reliability of suppliers' performance among others. Therefore, it is essential to focus on reducing the longest process lead time and reduce inventory. It is vital to take into consideration synchronization between processes with a correct balanced production plan as well as reduce the change over time between products for each of the processes and always work with small batch sizes and avoid overproduction [9].

In the implementation of a continuous flow, the reduction of equipment change over times is vital, which consequently translates into smaller lot sizes aligned with a *KanBan* signaling system with an adequate plant layout allowing to achieve improvements in the reduction of manufacturing delivery times, thereby reducing work in process and finished goods inventories, eliminating unnecessary movements of materials resulting in having the correct level of inventory to meet customer demand and thus increase the customer service level [10].

It is essential to align the demand planning by reducing the variation of the demand estimate and leveling the demand, establishing EOQ to determine the optimum lot size to produce to minimize the ordering and logistic costs of maintaining the inventory, also optimizing equipment change over times allows reducing the number of Kanban cards which results in a reduction of storage space such as the transport of materials by creating a

continuous flow in the value chain and making timely deliveries to the customer [11].

Within the investigation of inventory management, the optimization strategy stands out; the correct coordination of prices and inventory management technology is vital. Many companies begin to improve logistics by applying inventory control technology to reduce costs. The opportunity lies in ensuring having enough inventory to meet demand by minimizing inventory costs in the right place and time [12].

Any improvement in inventory performance has an impact on the quality of the product as well as the financial performance of the Company, so management must pay attention to proper inventory management. Companies without sufficient financial liquidity to obtain better financial performance can improve product quality by improving inventory management [13].

The correct inventory management practices have a direct impact on working capital, production, and the level of customer service and sales. Goods inventory management practices include having the correct levels of raw material and in-process inventory to ensure the continuous flow of supply to production, to avoid stoppages of the production line, to anticipate the correct prices of materials, ensure the correct levels of inventory of finished product to complete customer orders, follow the customer's demand to increase the level of service provided to the same [14].

The ABC Inventory Classification usually considers the annual volume multiplied times the ordered price in a descending manner classified into three categories A, B, and C. According to the ABC classification, the same service level is assigned for each part number according to the classification. Regularly, the articles generate the most considerable profit and are considered higher levels of service. To classify each of the part numbers, a new approach that takes into account four parameters is proposed: the cost of having a shortage times the rate of demand, divided into the cost of maintaining the inventory (purchase price) times the amount to be ordered for each of the articles; the result is sorted in descending order and is classified on ABC, which considers three classes: 50%, 30% and 20%, where A must have the highest level of service, followed by B and C [15].

Once certain products classified as type A, for the calculation of finished product inventory level, known as supermarket Figure 1, the following formula is considered [16]:

Average daily demand x Replacement delivery period		Stock for the cycle
+	Demand Variation % (stock for the cycle)	Buffer stock
+	Security Factor % (stock for the cycle + buffer stock)	Security stock
=		Supermarket stock

Figure 1: Approach to the calculation of the supermarket inventory level

The supermarket inventory is equal to the stock for the cycle (it is the average demand of each product classified type A, which considers the replenishment delivery time. Replenish time is the period elapsed from the date the order is received, manufactured, and made available at the warehouse) plus buffer stock (which is the amount of each product to be placed to cover any supply variation, such as quality defects, equipment stoppages, personnel missing, raw materials shortage). This supermarket shall be revised daily to ensure finished product availability to comply with service level upon any variation in demand or supply.

Inventory management with a maximum and minimum policy established on this case study for the industrial products industry has also been applied as the policy to health services, which have considered the importance of ensuring the supply of medicine, as well as a level of safety stock inventory to have continuity in the service in the event of supply shortages. By having a correct level of service and reducing delivery times, the decrease in inventory level is favored. The factors contributing to a shortage in medication shall be known, and evaluation is to be made from the demand pattern and all processes of the value chain intervening up to the delivery of the medication [17].

3. Problem Statement

The Company is committed to a quality management system that seeks to continuously improve profitability, customer satisfaction, internal processes, and staff development through the fulfillment of quality objectives and has considered strengthening its continuous improvement process in the distribution center located on the main plant solving the next opportunity for improvement: Position the correct level of finished product inventory under A classification, to increase the level of service to 95% as the preparation of shipment is interrupted due to the shortage of finished products for direct sale to or transferred to other manufacturing plants of the Company, having to reschedule shipments impacting the Company's sales, improve response time in the distribution center from the receipt of the finished product, collection of the product according to the customer's order, preparation of shipment, invoicing and clearance to 90 minutes.

Change management for continuous improvement as outlined in his Theory of Constraints (TOC) concepts and principles for improving companies' performance by focusing on some restrictions in the system. These influence points are system restrictions [18] [19].

4. Methodology

It is necessary to validate if the processes are oriented to serving the customer and achieve a high standard of service level to understand the value chain's baseline of performance. For this purpose, a mapping of the value chain is performed, which is a tool of lean manufacturing that is carried out by product family which outlines the value chain, customers, primary and secondary processes, suppliers, flows of information, and timeline, to display the waste, non-value-added activities and take actions to eliminate them. First, the map of the value chain in current status is carried out to identify the waste, which is transportation, inventory, movements, waiting, over-processing, overproduction, and defects; improvement events known as Kaizen are proposed to eliminate waste and streamline the value chain and therefore map the ideal state [2].

ABC analysis for inventory management is a method of inventory classification to book value (cost or acquisition) of the materials stored that considers the use of the article by its book value, classified from highest to lowest [20].

The principle of ABC analysis in inventory management applies:

1) In the classification of inventory items, based on their relative importance.
2) By setting different administrative controls for different classifications of materials and setting different degrees of control according to each classification's different ranges.

The criterion of classification should reflect the difficulty of controlling an article and the impact of this on the cost and in profit.

The ABC analysis determines the importance of an article in the Company's inventory. To perform this classification, companies could consider some of these no limitation criteria:

1) The annual transaction volume of an article (valued monetarily).
2) Shortages of material used in the production of an article.
3) Delivery time.
4) Storage requirements of an article.
5) Risk of theft.
6) Useful life.
7) Cost of a shortage of inventory, among other criteria.

The criterion used in this case study for the ABC classification is the volume of sales valued monetarily considering the last six months of current sales.

For classified items A, inventory control of maximum and minimum inventory method is implemented, which is a system specially developed to achieve up-to-date warehouses control and achieve optimal inventory; it sets the maximum and minimum levels of stock for each product, as shown below:

Definition of Variables:

$\bar{x}$ = Average sales in three months.

T_{Info} = Time since the manufacturing plant receives the request from the distribution center.

T_{mfg} = time to manufacture considers the days the manufacturing plant requires to produce the finished product.

$T_{transit}$ = time in transit from the plant warehouse to the distribution center.

Var.$T_{transit}$: Time the shipment may be delayed from the main plant warehouse to the distribution center.

Frec.$_{Emb}$: Shipment frequency. Several days between each shipment.

Var.$_{Dem}$. = Demand variation in days needed as additional coverage in the event of any demand fluctuation.

Equations for the calculation of maximum and minimum inventory:

The minimum inventory is calculated based on the average daily demand times the replenish lead time [16].

$$Min.Inventory = \bar{x} * (T_{Info} + T_{mfg} + T_{transit} + Var._{Ttransit} + Frec._{Emb}) \quad (1)$$

The maximum inventory is calculated based on the average daily demand times the replenish lead time in days plus the variation of the demand in days to have additional coverage for any demand fluctuation [16].

$$Max.Inventory = \bar{x} * (T_{Info} + T_{mfg} + T_{transit} + Var._{Ttransit} + Frec._{Emb} + Var._{Dem}) \quad (2)$$

$$Optimal\ inventory = \frac{(Min\ Inventory + Max\ Inventory)}{2} \quad (3)$$

The reorder point is the point when the inventory drops to a certain level, an order for a predetermined amount to replenish the inventory within the supplier's time of delivery is placed. This type of system can be used effectively for articles with independent demand. The reorder point is established to cover the average use within the delivery time plus variation in demand or variation in delivery time. The inventory to absorb this variation is known as safety stock. The amount of variation to cover depends on the level of customer service desired [20].

The safety stock (SS) required is a function of the random variation in demand during the estimated period, the desired level of customer service, and the radius of the delivery time during the estimated period.

$$SS = z\ s\ \sqrt{LT} \quad (4)$$

Where SS is the safety stock in a warehouse, z is the level of service sought for the Company represented by the value of the area under the curve, s is the standard deviation of the product demanded in the warehouse per unit of time.

Once inventory levels are set, next is the leveling of manufacturing within a period. It allows avoiding fluctuations to make the consumption of raw material, labor, and production capacity more predictable, ensuring finished product availability, minimizing inventory, working capital, and lead time throughout the value chain [16].

For this proposal, it is transcendent to determine the control condition of the process before the proposed improvements, as well as the determination of process capability indexes before the improvements and the associated sigma level thereof, to have transcendent measurements

To establish the central distribution center's baseline of performance, the Auditing Warehouse Performance, Ohio: Ackerman Publications checklist was applied [21].

Table 1: Summary of results of the Audit Central Distribution Center

AREA	WEIGHT	RANK	MAX RANK	GAP
External	0.05	0.117		
Internal	0.05	0.126		
Warehouse IT systems	0.1	0.3	0.4	0.1

Mechanical handling equipment	0.1	0.264	0.4	0.136
Contingency planning	0.05	0.067		
Inbound operation	0.1	0.211	0.49	0.278
Put-away and storage	0.05	0.14		
Stock control	0.15	0.417	0.6	0.183
Picking	0.05	0.15		
Dispatch	0.15	0.48	0.72	0.24
Staff and housekeeping	0.05	0.123		
Warehouse - performance management	0.05	0.1		
Legislative & Health and Safety requirements	0.05	0.154		
Overall Warehouse Rating		2.648	3.585	

1	2	3	4	0
No	Poor	Good	Excellent	N/A

Overall, the general grade of DC was 2.648, which indicates performance with an opportunity for improvement. When considering the ascribed areas as the most important in the auditing process, the maximum range was calculated. It denotes three areas with an opportunity for significant improvement:

1) Inbound Operation with a gap of 0.278.
2) Dispatch with a gap of 0.240.
3) Stock Control with a gap of 0.183.

If efforts are focused on these three areas, then a potential score of 3.459 can be reached to be close to excellence with a rating of 4.

Hence, the efforts will be focused on the following processes:

a) Receipt, Preparation of the Shipping Order & Inventory Control. It interrupts the shipment's preparation flow due to finished goods and raw material shortage in the main warehouse, which causes reschedule shipments to the distribution centers based on the promised delivery date of production and suppliers. Evaluate the option to implement a pre-shipment activity and implement a performance metric of hours of preparation and loading of the shipment.
b) Dispatching requires a shipment schedule to increase flexibility associated with sales to customers based on available resources as workforce, space, and equipment.
c) One of the audit opportunities is the lack of KPI's to monitor CD's performance. Customer's delivery time, one of the top priorities identified by the commercial area, currently is 5 hours, being the goal of an average time of 90 minutes.
d) In the stock control process, the opportunity is detected to implement an inventory control to guarantee the inventory's availability due to the constant interruption in the picking process due to finished goods inventory shortage compromising the shipment program and impacting the customer service level.

Associated with the delivery of the finished product to the customer, transformation processes were evaluated to understand the flow of the value chain of the Company, from suppliers, production processes, distribution and final delivery of customer, as well as the activities that provide service to primary processes, which involves the creation of the customer order, the administration area authorizes the order, the ordered product inventory is booked, delivery of documents to logistics, processing the loading order, delivery of documents to the loading area, loading of order and delivery of documents to the customer. Value stream mapping is a tool to achieve competitiveness in any company's value chain, highlighting the inefficiencies of processes such as areas for improvement [22].

The value stream map of the current state indicates:

- The model is pushed and does not have a sales estimate from the commercial area.
- Orders flow into the commercial area, and there is no supermarket of the finished product to shorten the delivery time to the customer and act as a buffer to absorb changes in demand.
- The production plan is made based on the demonstrated demand, and a production signal is sent to production based on a weekly production plan that applies to each of the work centers, which takes into account proven capacity and labor requirements.
- Material planning is done based on an MRP that explodes customer requirements to validate inventory position and release the production orders.
- The suppliers do not estimate the future requirement of the raw materials, so they react based on the monthly supply signal sent by the purchasing department.

ABC analysis was performed, Table 2, which considered the sale of each Distribution Center (DC) for each of the finished products by book value, based on manufacturing and sorted from highest to lowest. The Company's approach resides in 52 catalogs to implement the inventory control policy of maximums and minimums in Central Distribution Center, 18 catalogs in Distribution Center in City one, and 21 catalogs in City two Distribution Center, as it is shown below:

Table 2: ABC Classification

DC Central

ABC	Total Catalogs	%Sales Value
A	52	80
B	68	15
C	207	5

DC City 1

ABC	Total Catalogs	%Sales Value

A	18	80
B	31	15
C	58	5

DC City 2

ABC	Total Catalogs	%Sales Value
A	21	80
B	26	15
C	68	5

When reviewing the inventory's coverage in the days of the finished goods classified as type A, there is an average inventory coverage of 27 days, compared against the Company's goal of 14 inventory days. The following Table 3 shows a summary of coverage in the inventory days before the implementation of the improvement:

Table 3: Coverage in days and distribution as a percentage of the value of inventory catalogs classified as Type A

DC Central

DOS	Number of Finished Goods	% of Inventory Value
0 Days	7	3%
From 1 to 14 days	12	7%
From 15 to 22 days	13	53%
From 23 to 30 days	6	17%
Greater than 30 days	12	20%
Total	50	100%

DC City 1

DOS	Number of Finished Goods	% of Inventory Value
0 Days	1	0%
From 1 to 14 days	8	57%
From 15 to 22 days	2	18%
From 23 to 30 days	3	9%
Greater than 30 days	4	16%
Total	18	100%

DC City 2

DOS	Number of Finished Goods	% of Inventory Value
0 Days	2	0%
From 1 to 14 days	10	43%
From 15 to 22 days	4	19%
From 23 to 30 days	3	33%
Greater than 30 days	2	5%
Total	21	100%

Next, the state of statistical process control regarding the original process is shown and the capability index, Z level, and associated sigma [23].

A prerequisite to determine the status of control of a process is to validate that the data analyzed corresponds to a standard distribution. Figure 2, which can be validated and with the value p=0.191, is higher than the level of type I error or α=0.05.

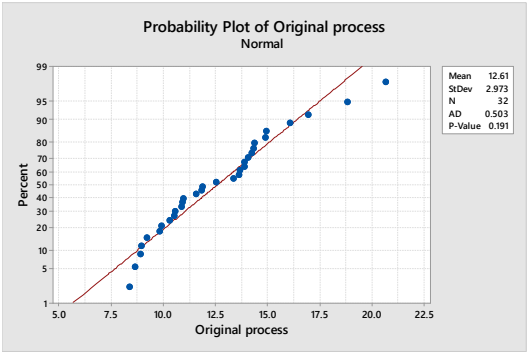


Figure 2: Normal Probability Graph

In Figure 3, it can be noticed that the process is out of statistical control due to several measuring points, which are outside control limits, which are associated with sources of variation attributable to the conditions of the current process [23].

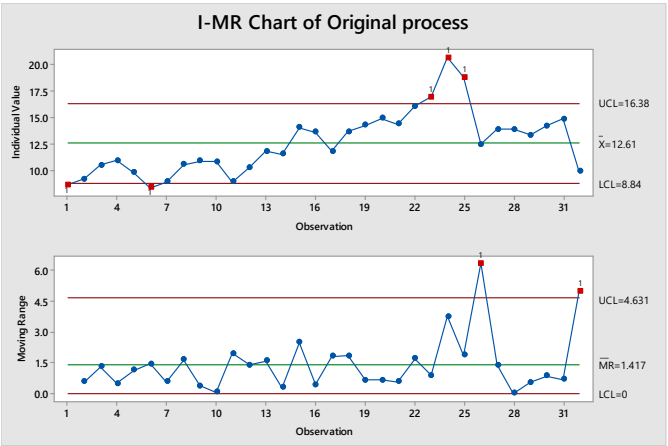


Figure 3: Control Chart for Individual Values and Mobile Range of the original process

Figure 4 shows the capability process analysis report for the original process. The cpk index establishes the actual process capability to meet specifications to consider that a process is adequate; the value must be higher than 1.25 [23]. the current cpk value of the process is 0.37, so the process under current conditions cannot meet specifications and requires serious intervention to restructure its activities and operation controls. The Z level of the process is 1.11, which is reflected in the sigma level = 2.61, the sigma level of a process is determined by the value of the Z + 1.5 level associated with the general standard deviation of the process. It should be noted that the lower specification limit is considered suitable; hence efforts will be concentrated on the upper specification limit [23].

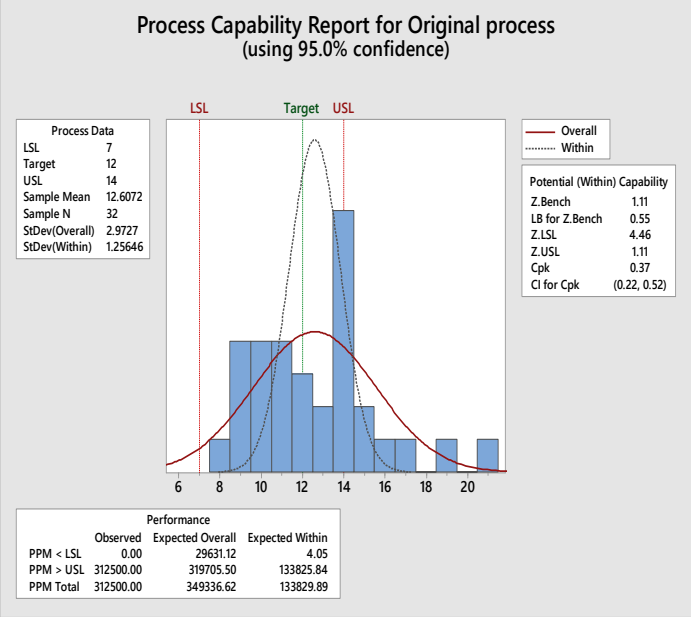


Figure 4: Process Capability report for the original process

The graph of the performance of the process shown in Figure 5 allows observing the fraction of the process that fails to comply with the established specifications. It can be seen in the area under the curve that exceeds the upper control limit.

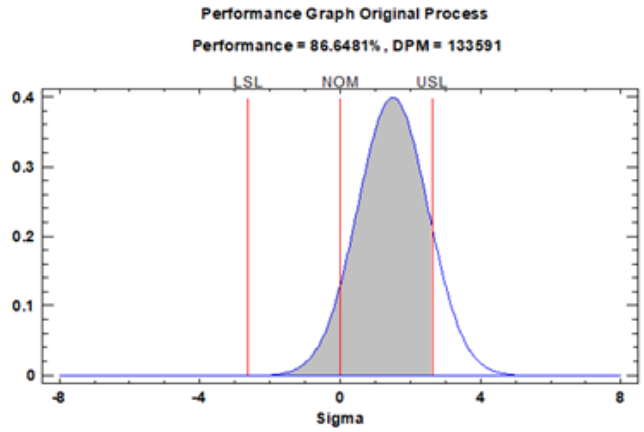


Figure 5: Performance Graph for the original process.

In the six sigma calculator associated with the process Table 4, the values obtained are the Z value of 1.11, Defects Per Million (DPM) of 133,519, Percentage of Defects 13.3, Yield of 86.64% with a process capability (Cpk) of 0.37, and a sigma quality level (SQL) of 2.6 [23].

Table 4: Six Sigma Calculator
Entry: Value-Z = 1.11
Change in sigma = 1.5
Equivalent values

Table of Contents	Value
Value-Z	1.11
DPM	133519
Defects	13.3519
Performance	86.6481
Cpk	0.37
SQL	2.61

5. Issues, Controversies and Problems

Based on the audit, other opportunities were identified and then raised to the Company:

1) Improve the receiving flow of raw materials and finished product, since the schedule for the reception of raw material as complementary products for sale is interspersed with the hours of shipment preparation, which generates a conflict of space in the yard and rescheduling of shipments of the finished product.
2) Review the flow of finished products from other manufacturing plants to ensure the availability of inventory in the central distribution center, positioning a supermarket to facilitate the consolidation of shipments.
3) Streamline the work used in the Operation of the process put away (pre-stowage to the final location of the finished product) combined with the efforts of picking Operation for the shipment of the product and create a standardized work for the truck loading process.

6. Research Methodology

In the implementation of the improvement of the customer service level to guarantee the availability of the finished product and improve the delivery time in the distribution center, the following results were obtained:

1) **Audit in the central distribution center** to detect improvement opportunities, where the detail of such is presented in the analysis of Table 1.
2) **The ABC classification** was performed to determine the items that represent 80% of the sales valued, to focus on the control of inventory presented in Table 2.
3) **The analysis of the coverage of inventory of finished product catalogs**, Table 3, classified as type A, indicates that seven catalogs of finished product do not have inventory for sale, only twelve catalogs are within the target set by the Company of fourteen days of inventory, and the rest of the twenty-nine catalogs are above the goal of fourteen days of inventory. Only 7% of the total monetary value of the stock is located within the established goal of inventory coverage.
4) **Lead Time Component Analysis**
   a) Information Lead Time: Time from the moment the manufacturing facility receives the request for inventory replenishment from the distribution center supermarket.
   b) Manufacturing Lead Time: Considers the days that require the manufacturing facility to produce the finished product, plus the quarantine time.
   c) In Transit Lead Time: Time in transit from the warehouse of the manufacturing facility to the distribution centers.
   d) In Transit Lead Time Variation: Time that can delay the shipment from the warehouse manufacturing facility to the distribution center due to external factors such as climate, holidays, transport breakdown, among others.
   e) Frequency of Shipments: Considers the period in days between each shipment.
   f) Variation of Demand: Days of daily inventory coverage regarding the daily sales average to cover any variation in demand required by the final customer.

g) Variation of Supply: Daily coverage based on average daily production to cover any variation in the supply as quality defects, equipment stoppage, lack of personnel, shortage of raw material.

The evaluation of lead time's components for the manufacturing facility is considered in the following Table 5:

Table 5: Manufacturing Lead Time Components

Manufacturing Lead Time Components	Days
Information Lead Time	2
Manufacturing Lead Time	1
Quarantine Lead Time	7
Supply Variation	10

For the case of a distribution center, the lead time components are shown in the following Table 6:

Table 6: Lead Time Components for the Distribution Centers

DC Central Lead Time Components	Days
Information Lead Time	2
Manufacturing Lead Time	7
In Transit Lead Time	0
In Transit Variation	0
Shipment Frequency	1
Demand Variation	7

DC City 1 Lead Time Components	Days
Information Lead Time	2
Manufacturing Lead Time	7
In Transit Lead Time	1
In Transit Variation	0
Shipment Frequency	3
Demand Variation	3

DC City 2 Lead Time Components	Days
Information Lead Time	2
Manufacturing Lead Time	7
In Transit Lead Time	1
In Transit Variation	0
Shipment Frequency	3
Demand Variation	2

DC Central Lead Time Components	Days
Information Lead Time	2
Manufacturing Lead Time	7
In Transit Lead Time	0
In Transit Variation	0
Shipment Frequency	1
Demand Variation	7

5) Inventories Planning for the distribution centers and manufacturing facility

The sales estimate horizon is three months (current month plus two months based on the historical demand), the demand is analyzed based on the current sales and sales estimate provided by the commercial area Table 7 for calculations of supermarkets to position each of them in the distribution centers.

Table 7: Demand for each Distribution Center

Central Distribution Center

Catalog	Aug 19	Sep 19	Oct 19	Nov 19	Dec 19	
8750	286	92	138	272	227	
6890	271	219	210	169	150	
5151	78	170	61	162	94	

Central Distribution Center

Catalog	Jan 20	Feb 20	Mar 20	Apr 20	May 20	Jun 20
8750	307	307	314	215	205	205
6890	260	313	282	215	200	200
5151	18	62	59	60	50	50

City 1 Distribution Center

Catalog	Aug 19	Sep 19	Oct 19	Nov 19	Dec 19	
8750	2	8	23	6	7	
6890	19	139	119	197	140	
5151	58	28	155	140	100	

City 1 Distribution Center

Catalog	Jan 20	Feb 20	Mar 20	Apr 20	May 20	Jun 20
8750	2	1	0	13	13	9
6890	230	211	293	171	140	128
5151	177	88	126	92	43	43

City 2 Distribution Center

Catalog	Aug 19	Sep 19	Oct 19	Nov 19	Dec 19	Catalog
8750	53	25	6	39	52	8750
6890	8	60	24	21	6	6890
5151	26	24	8	34	19	5151

City 2 Distribution Center

Catalog	Jan 20	Feb 20	Mar 20	Apr 20	May 20	Jun 20
8750	64	35	61	63	67	63
6890	2	6	6	12	12	12
5151	16	18	16	25	29	25

The daily requirement sale is calculated as well the reorder point as the maximum, optimum, and minimum of inventory of finished product Table 8, for each of the distribution centers of the Company, which consider the lead time components.

Table 8: Supermarket Positioning Distribution Center

DC Central

Catalog	ROP	Inventory	MIN	IDEAL	MAX
8750	171	247	96	139	182
6890	166	231	93	135	177
5151	16	49	9	14	19

DC City 1

Catalog	ROP	Inventory	MIN	IDEAL	MAX
8750	17	4	6	13	20
6890	147	76	52	113	173
5151	51	35	24	43	63

DC City 2

Catalog	ROP	Inventory	MIN	IDEAL	MAX
8750	35	18	12	26	41
6890	17	27	6	13	20
5151	9	25	5	8	11

In the manufacturing facility, the production plan Table 9 is analyzed to implement the supermarket and act as a buffer before any variation of supply to ensure the availability of the product. The production plan considers the distribution centers' net requirement, less the positioned inventory on the facility to calculate the requirement to produce.

Table 9: Production Plan - Manufacturing Facility

Catalog	Aug 19	Sep 19	Oct 19	Nov 19	Dec 19	
8750	456	332	338	100	325	
6890	549	542	467	320	360	
5151	38	286	65	241	229	

Catalog	Jan 20	Feb 20	Mar 20	Apr 20	May 20	Jun 20
8750	430	323	450	449	408	400
6890	580	493	584	599	469	457
5151	128	59	191	235	65	65

Determines the supermarket Table 10 for the production area that considers the maximum and minimum inventory policy:

Table 10: Supermarket Manufacturing Facility.

Catalog	ROP	Inventory	MIN	IDEAL	MAX
8750	154	130	154	170	185
6890	212	230	170	212	255
5151	19	25	10	22	34

6) Visual Management of the Inventory of Distribution Centers

The current inventory of finished products is compared against the maximum, optimum, and minimum inventory to maintain visual management. Green indicates that it is within the maximum level and optimum inventory; this is the ideal position. Yellow between the optimum and minimum levels requires planning to place a manufacturing order to replenish the inventory to its position. Red requires expediting the production order, given that the current inventory is below the minimum level with a possible negative impact on the service level. When the current position is blue, it is above the maximum allowed level, and action is required to balance the inventory levels to those approved. It is essential to run this report daily since it is the tool that manages the release, issuance, and establishment of priorities of the manufacturing orders to production and the valuation of the position of inventory in units. This report is required per the manufacturing facility and distribution center.

7) Planning Horizons

Once the supermarket was positioned in the Distribution Center and manufacturing facility, the following was agreed with the Commercial area:

The sales estimate horizon is three months (current month plus two months based on the historical demand), delivery takes place between 3 - 4 days business days of the month, and formed on the following and the basis of the planning zones:

- *Fixed Zone*: Consists of the time required upon releasing the production order and until the product is available for sale. The horizon is of 14 days with zero percentage of variation. This planning zone is conformed of labor, materials, and burden.
- *Firm Zone*: This zone considers, from the moment the purchase order is placed to the suppliers, considers the longest lead time of the resource, either materials, tools, and/or equipment, as labor in compromising the production requirements. The time horizon is 20 days; the variation percentage of the sale estimated is +/- 20%. In this horizon, planning involves materials and any other resource required to be contracted as labor and/or equipment.
- *Planning Zone*: Consists of the horizon when no resource for production is committed; therefore, the percentage of variation is open.

8) Delivery Time to Customer Service

The delivery time to the customer is essential as part of the flexibility offered by the Company, to meet the immediate need for what was agreed with the Commercial area, the following delivery times according to the ABC classification of the finished product:

- *Family Products A*: Delivery time is the time upon the receipt of the order, preparation, shipment, and invoicing; this delivery time is of 90 minutes for the external customer; in the case of an internal customer as the Distribution Centers located in other cities, it must be considered two days upon the receipt of the inventory replenishment requirement, preparation of shipment and billing plus the in-transit time. If demand variation consumes the inventory available in the supermarket, the planning area should coordinate the managing of materials and resource capacity with Production and Purchasing; consideration should be given to the delivery to customers and communicate to the commercial area the best delivery commitment to customer's order. In coordination with planning, the Commercial area shall manage the priorities of the inventory allocation until the inventory is regularized in the supermarket according to the maximum and minimum levels of inventories approved. Production should communicate the plan to recover inventory to approved levels.

- *Family Products B*: These finished goods products are not supplied within the calculation of the maximum and minimum inventory. It is a product that is managed to make an order basis. Delivery Time is seven days. The way to schedule this finished product to production is FIFO (Customer's First orders entered are the first to be scheduled for production).
- *Family Products C*: This is a new product with a particular design in which the delivery time is variable according to the project's timeline.

9) Production Balance

Based on the month's production requirement, the production requirement is planned in Table 11, 30% in week 1, 30% in week 2, 20% in week 3, and 20% in Week 4; this production balance enables demand confrontation since 60% of the sale occurs in the last two weeks of the month, which makes the labor requirement planning Table 12 as the materials requirements more predictable, for inventory positioning of raw materials in supermarkets, this enables better visibility of the requirement to the consumption of raw materials to suppliers and increase the flexibility of the productive capacity, especially when they are manufacturing make to order products for special projects for which demand is random.

Table 11: Monthly Balanced Production Plan

Catalog	Monthly Prod. Req"	Prod. Req Wk 1	Prod. Req Wk 2	Prod. Req Wk 3	Prod. Req Wk 4	Hr. Req Process 1 Wk1	Hr. Req Process 2 Wk1	Hr. Req Process 1 Wk2	Hr. Req Process 2 Wk2	Hr. Req Process 1 Wk3	Hr. Req Process 2 Wk3	Hr. Req Process 1 Wk4	Hr. Req Process 2 Wk4
Product 1	493	148	148	99	99	166	243	166	243	111	162	111	162
Product 2	323	97	97	65	65	109	159	109	159	73	106	73	106
Product 3	59	18	18	12	12	8	16	8	16	5	11	5	11
Product 4	0	0	0	0	0	0	0	0	0	0	0	0	0
Product 5	232	70	70	46	46	63	63	63	63	42	42	42	42
Product 6	0	0	0	0	0	0	0	0	0	0	0	0	0
Product 7	145	44	44	29	29	49	71	49	71	33	48	33	48
Product 8	430	129	129	86	86	145	212	145	212	97	141	97	141
Product 9	1	0	0	0	0	1	3	1	3	1	2	1	2
Product 10	57	17	17	11	11	19	51	19	51	13	34	13	34
Product 11	40	12	12	8	8	14	36	14	36	9	24	9	24
Product 12	151	45	45	30	30	27	20	27	20	18	14	18	14
Product 13	241	72	72	48	48	43	33	43	33	29	22	29	22
Product 14	78	23	23	16	16	14	11	14	11	9	7	9	7
Product 15	372	112	112	74	74	67	50	67	50	45	33	45	33
Product 16	300	90	90	60	60	54	41	54	41	36	27	36	27
Product 17	109	33	33	22	22	20	15	20	15	13	10	13	10
Product 18	92	28	28	18	18	25	83	25	83	17	55	17	55
Product 19	165	50	50	33	33	22	22	22	22	15	15	15	15
Product 20	218	65	65	44	44	0	75	0	75	0	50	0	50
Product 21	1089	327	327	218	218	0	399	0	399	0	266	0	266
Wk Req. hrs.						540.1	762.9	540.1	762.9	360.0	508.6	360.0	508.6
Daily Req. hrs.						90.0	127.2	90.0	127.2	60.0	84.8	60.0	84.8

Table 12: Labor Weekly Requirement

Labor Weekly Requirement								
	Week 1		Week 2		Week 3		Week 4	
	Process 1	Process 2	Process 1	Process 2	Process 1	Process 2	Process 1	Process 2
Product Family 1	9.7	14.1	9.7	14.1	6.5	9.4	6.5	9.4
Product Family 2	7.8	11.2	7.8	11.2	5.2	7.4	5.2	7.4
Product Family 3	1.7	1.8	1.7	1.8	1.1	1.2	1.1	1.2
Product Family 4	0.9	1.8	0.9	1.8	0.6	1.2	0.6	1.2

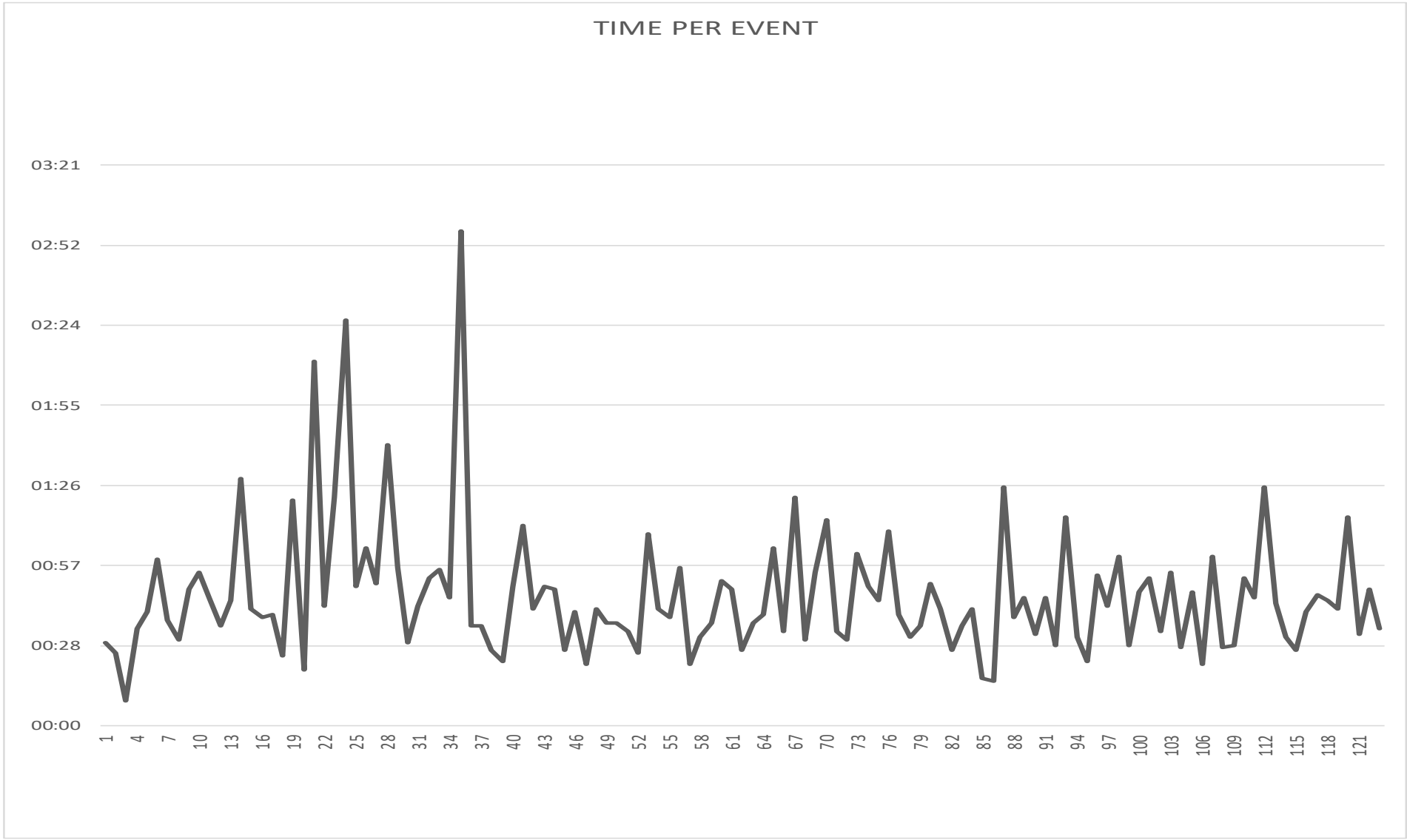


Figure 6: Customer delivery Time per event.

10) Customer delivery Time

The customer delivery time Figure 6 is measured from the moment the customer order is created, the administration area authorizes the order, the ordered product inventory is booked, delivery of documents to logistics, processing the loading order, delivery of documents to the loading area, loading of order and delivery of documents to the customer.

Registers an average customer delivery time of 46 minutes below the goal of 90 minutes, which encourages the customer to be recurrent customers considering the availability of inventory to cover their requirements with a short delivery time.

11) Process Capability and Six Sigma

Through the implementation of the proposed improvements, the process is in statistical control, as well as in the assessment of capability indexes, the level of cpk increased from 0.37 to 0.86, the associated z level moved from 1.11 to 2.59, and the sigma level moved from 2.61 to 4.09 with a decrease of defects per million of 133,519 to 4,798.83. It should be noted that an optimum level is not yet achieved. However, this represents a strategic impact and the opportunity to deploy additional improvements to arrive at an ideal level of six sigma, that is to say, with only 3.4 ppm of failure. It can be observed in Figure 7 that the associated data with the subsequent measurements to improvements come from a normal distribution with a p value=0.407, which is higher than the value α=0.05 [23].

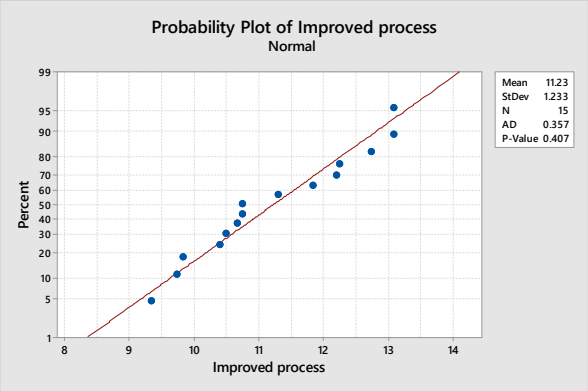


Figure 7: Probability Plot of Improved Process

Figure 8 shows the process under statistical control without points outside the control limits without any anomaly trends.

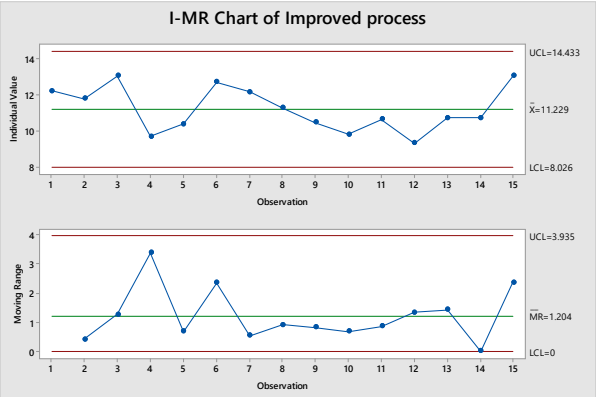


Figure 8: Graph of Individual Values and Mobile Range for the Improved Process

Figure 9 shows the process capability analysis report in which it is possible can see the improvements obtained with a Cpk index of 0.86; although it is not the optimal state of the process represents a significant improvement.

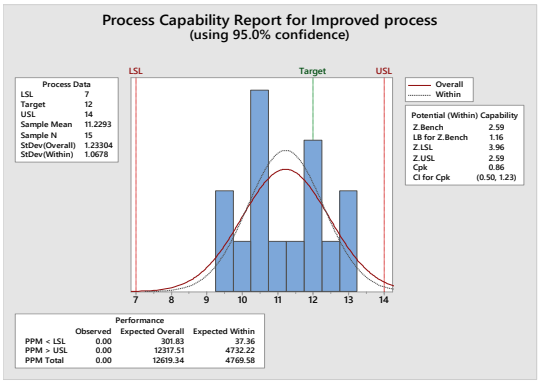


Figure 9: Process Capability Report for Improved process

Figure 10 shows the improvement in performance in which there is a significant reduction in the fraction of the probability distribution outside the upper limit of the specification.

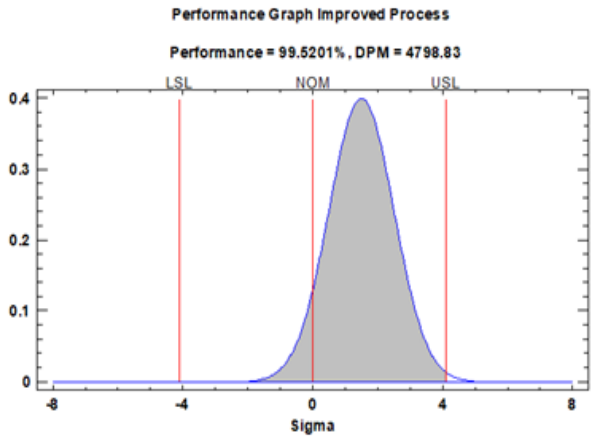


Figure 10: Performance Graph Improved Process

Table 13 shows the values obtained in the six-sigma calculator associated with the improved process the Z value of 2.59, Defects Per Million (DPM) of 4,798.33, Percentage of Defects 0.47, Yield of 99.52% with a process capability (Cpk) of 0.86, and a sigma quality level of (SQL) 4.09.

Table 13: Six Sigma Calculator
Entry: Value-Z = 2.59, Change in sigma = 1.5
Equivalent values

Table of Contents	Value
Value-Z	2.59
DPM	4798.83
Defects	0.479883
Performance	99.5201
Cpk	0.863333
SQL	4.09

12) Pull System

The Commercial area estimates the monthly sales to determine the maximum and minimum level of finished product inventory. A pull system is implemented, the customer's orders are supplied directly from the CD's finished good product supermarket. The

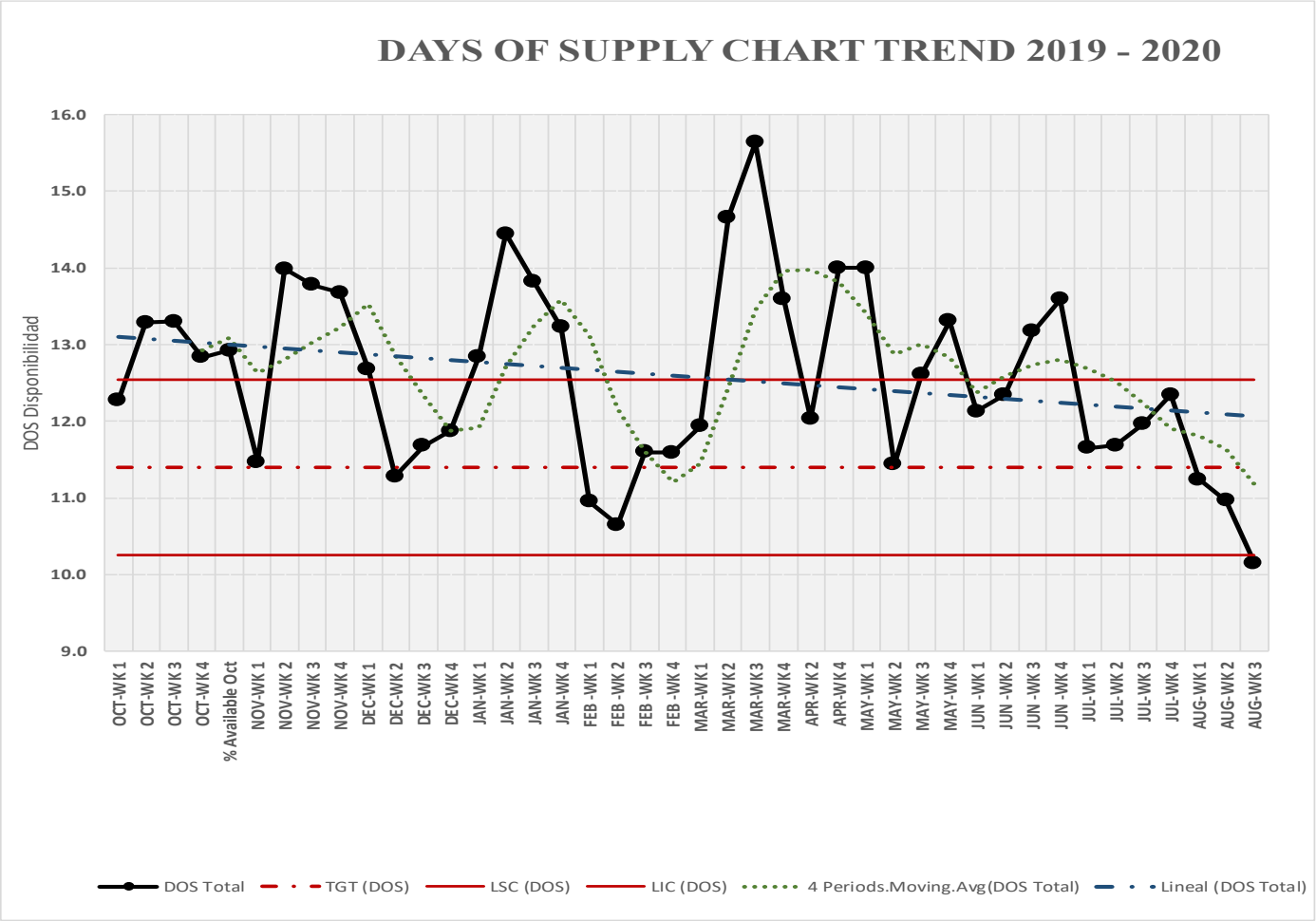


Figure 11: Days of Supply Finished Goods Inventory

consumption of the supermarket inventory generates a manufacturing order for production to replenish the finished product supermarket. As the supermarket's raw materials are consumed when producing the manufacturing orders, the purchasing requisitions of raw material are released to the purchasing department to generate the purchase orders for the suppliers and thus guarantee the replenishment in the raw's materials supermarket.

13) Days of Supply Finished Goods Inventory

The trend chart of days of supply Figure 11 shows a favorable trend in days of coverage, having a supply of twelve days of inventory against a goal of 14 days.

7. Results and Conclusions

During the COVID 19 pandemic, the Company has been operated with a Pull system, in which supermarkets were positioned in Central, City 1, and City 2 distribution centers for finished products classified A; The inventory is managed with a maximum, and minimum policy, based on customer consumption, the supply signal to the manufacturing plant is triggered to replenish and position the inventory in the distribution centers. Based on the inventory consumption for the supermarket of the manufacturing plant, the manufacturing orders are released to replenish the inventory in the supermarket of the plant, which has generated a reduction of the inventory to a coverage of 12 days before starting the improvement in the process, the catalogs classified type A had a 27-day average coverage, against the Company's goal of 14 days of inventory. The service level, the baseline of performance 89% guaranteeing inventory availability, exceeds expectations with a 98% service level. An average customer delivery time of 46 minutes is recorded below the 90-minute goal, which encourages the customer to be a recurring customer considering inventory availability to meet their requirements with lead time delivery. The savings in inventory optimization on an annual basis represented USD 530,785. (Five hundred thirty thousand seven hundred eighty-five dollars) helping the Company to gain liquidity and to create the conditions for the recovery of economic growth, and helping to preserve employment as much as possible in a constrained market. Through the implementation of the proposed improvements, the process is under statistical control, the process capability (Cpk) is increased from 0.37 to 0.86, and the level six sigma (SQL) is increased from 2.61 to 4.09, reducing the percentage of defects to

13.3 to 0.47. An additional benefit is that the Company implemented the methodology nationwide in its other Manufacturing Plants and Distribution Centers. This methodology supports the acceptable practices that the Company is implementing in search of continuous improvement as part of the strategic framework to fulfill the strategic objectives based on the business vision.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The corresponding author wants to acknowledge the Popular Autonomous University of the State of Puebla (UPAEP University).

References

[1] B. Keyte, D.A. Locher, The Complete Lean Enterprise: Value Stream Mapping for Administrative and Office Processes, CRC Press, 2004.
[2] R. Martichenko, K. Von Grabe, Building a Lean Fulfillment Stream, Lean Enterprise Institute, 2010.
[3] M. Imai, Kaizen: La clave de la ventaja competitiva japonesa, Compañía Editorial Continental, 2000.
[4] J.P. Womack, D.T. Jones, Lean Thinking: Banish Waste and Create Wealth in Your Corporation, Simon & Schuster, 1996.
[5] J.K. Liker, The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer, McGraw-Hill, 2004.
[6] S. Emmett, Excellence in warehouse management, Wiley, 2005.
[7] M. Malindzakova, D. Zimon, "Design Supply Cycle for Inventory Management," TEM Journal, **8**(3), 894-899, 2019, doi: 10.18421/TEM83-29.
[8] S. Kumar, D. Choe, S. Venkataramani, "Achieving customer service excellence using Lean Pull Replenishment," International Journal of Productivity and Performance Management, **62**(1), 85-109, 2012, doi: 10.1108/17410401311285318.
[9] L. Cuatrecasas-Arbós, J. Fortuny-Santos, P. Ruiz-de-Arbulo-López, C. Vintró-Sanchez, "Monitoring processes through inventory and manufacturing lead time," Industrial Management & Data Systems, **115**(5), 951-970, 2015, doi: 10.1108/IMDS-12-2014-0375.
[10] R. Rodríguez-Méndez, D. Sánchez-Partida, J.L. Martínez-Flores, E. Arvizu-Barrón, "A case study: SMED & JIT methodologies to develop continuous flow of stamped parts into AC disconnect assembly line in Schneider Electric Tlaxcala Plant," IFAC-Papersonline, **48**(3), 1399-1404, 2015, doi: https://doi.org/10.1016/j.ifacol.2015.06.282.
[11] D. Sánchez-Partida, R. Rodríguez-Méndez, J.L. Martínez-Flores, S.O. Caballero-Morales, "Implementation of Continuous Flow in the Cabinet Process at the Schneider Electric Plant in Tlaxcala, Mexico," Interfaces, **48**(6), 566-577, 2018, doi: 10.1287/inte.2018.0966.
[12] Y. Ge, Y. Ye, "A bibliometric analysis of inventory management research based on knowledge mapping," The Electronic Libray, **37**(1), 127-154, 2018, doi: 10.1108/EL-11-2017-0241.
[13] Y. Lin, B. Liang, X. Zhu, "The effect of inventory performance on product quality: The mediating effect of financial performance," International Journal of Quality and Reliability Management, 35(10), 2227-2247, 2018, doi: 10.1108/IJQRM-08-2017-0162.
[14] N.A. Karim, A. Nawawi, A.S. Salin, "Inventory control weaknesses – a case study of lubricant manufacturing Company," Journal of Financial Crime, **25**(2), 436-449, 2018, doi: 10.1108/JFC-11-2016-0077.
[15] R.H. Teunter, M.Z. Babai, A.A. Syntetos, "ABC Classification: Service Levels and Inventory Costs," Production and Operations Management, **19**(3), 343-352, 2009, doi: 10.1111/j.1937-5956.2009.01098.x.
[16] J. Shook, T. Narusawa, Kaizen express: Fundamentals for your Lean Journey, The Lean Enterprise Institute, 2009.
[17] D. Kumar, D. Kumar, "Managing the essential medicines stock at rural healthcare systems in India," International Journal of Health Care Quality Assurance, **31**(8), 950-965, 2018, doi: 10.1108/IJHCQA-12-2016-0186.
[18] E.M. Goldratt, J. Cox, The Goal: A Process of Ongoing Improvement, North River Press, 1986.
[19] E.M. Goldratt, J. Cox, The Race, North River Press, 1986.
[20] G. Richards, S. Grinsted, The logistics and supply chain toolkit, Kogan Page Limited, 2016.
[21] K.B. Ackerman, Auditing Warehouse Performance, Ackerman Publications, 2003.
[22] H. Singh, A. Singh, "Application of lean manufacturing using value stream mapping in an auto-parts manufacturing unit," Journal of Advances in Management Research, **10**(1), 72-84, 2018, doi: 10.1108/09727981311327776.
[23] T.M. Kubiak, D.W. Benbow, The certified Six Sigma Black Belt Handbook, American Society for Quality, 2009.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 627-635 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Chatbot Developments in The Business World

Azani Cempaka Sari*, Natashia Virnilia, Jasmine Tanti Susanto, Kent Anderson Phiedono, Thea Kevin Hartono

Computer Science Department, School of Computer Science, Bina Nusantara University Kav 21, Alam Sutera Jalan Jalur Sutera Barat Panunggangan Timur, RT.001/RW.004, Panunggangan, Kec. Pinang, Kota Tangerang, Banten, 15143, Indonesia

ARTICLE INFO

Article history:
Received: 23 June, 2020
Accepted: 12 October, 2020
Online: 24 November, 2020

Keywords:
Chatbot
Artificial Intelligence
Natural Language Processing (NLP)
Business
AIML

ABSTRACT

Chatbot is an application that is designed to help humans communicate with machines. NLP and API.AI are some of the components that are crucial in making chatbot as they help chatbot in performing chatting operations. Chatbot could be implemented in many aspects of life such as education, social media platforms, games, even in business. Business has always been an important part of human life and it is irreplaceable. As the era evolves, the business industry is also getting modernized and automated which resulted in implementing the latest technology into the industry like chatbot. In the context of the business industry, chatbot can be used as an assistant that will help dealing with the customer service section as well as giving the latest information about the business. This study aims to determine how the chatbot gives impact to the business as well as the owners, what kind of features that the chatbot needs, and how the chatbot should be developed so that it will be compatible in the business environment. The research was conducted using two methods which are literature review and survey: questionnaire. The result of the research conducted are the impacts that the owner gets by using chatbot in their business, a set of features that are relevant in developing chatbot and recommendation for further implementation.

1. Introduction

Chatbot is an application or service that can interact with users by communicating using text or audio. The first chatbot produced was ELIZA [1] in which this chatbot could communicate using simple responses with the user. The usage of chatbot nowadays is inclining significantly that it is easy to find chatbot in web sites with each different functionality. Chatbot both on applications and web sites usually give services like giving important information, helping users in operating data, and supporting business processes. Surprisingly, chatbots can fit in every kind of website or program, even gaming sites. Thus, it is easy to say that chatbot is also gaining popularity in the business environment since chatbot could confirm that messages or complaints from customers can be handled fast and accurately. In addition, companies that use chatbot could cut down on their operational fees [1] and also their business could be active for 24 hours every day. In technology development nowadays, there are lots of research that uses chatbot as virtual assistant or shop assistant. Chatbots are programmed so that they can interact with users using natural human language. These days, there are many chatbots that are based on AIML or Artificial Intelligence Markup Language like virtual assistants such as Voice Actions, Skyvi, Iris, and Call Mom (Pandorabots.com). AIML is often used for programming the artificial intelligent software of chatbot or also known as knowledge base [2, 3]. Technology these days is developing so fast that it is now reaching to all kinds of people and covering different aspects in life like health, business, education, and many more. This is because information plays an important role in this globalization era. Everything including daily activities needs information before doing it and every action that was taken is demanded to produce another information that is helpful for many people. Thus, as technology is something that makes information reachable, technology including chatbots are developing and expanding fast and the demand of technology is for sure high. In Indonesia, chatbots are becoming popular as they can communicate to customers as if they are literally talking naturally and they can answer different questions faster than humans. It is usually implemented to support the company's service as chatbots will be likely to increase the service rate. Moreover, chatbot will not forget any information unlike humans who will be most likely to forget the information that is needed to be delivered. Thus, in this research we are planning to know

* Corresponding Author: Azani Cempaka Sari, kent.phiedono@binus.ac.id

whether chatbots are actually important in running businesses, if it is needed in the business, and what kind of chatbot model that is suitable for supporting a business.

2. Literature Review

The basic mechanism of chatbot starts with messages sent by the user. The message is then processed by NLP (Natural Language Processing), and chatbot responds by replying to the message according to the existing database [4].

2.1. Chatbot

Chatbot is a character that communicates with its user or people using natural human language in chatting forums such as messenger, instant web, email, usenet, forum, web, and even using speech like telephone. Chatbot has so many aliases and some of them are chat robot, bot, chatter bot, bot chatting, chatterbox, V-Host, V-People, agent, and virtual human. Chatbot can also be connected to an avatar or animation that includes speech synthesis so that chatbot could be seen as something livelier as we, human beings, could see what it is and what it is speaking. To improve the service given, chatbots are programmed so that it could analyze words, phrases, and the construction of sentences of inputs so then it could predict the users' personality, preferences of products, and give responses according to what it had analyzed. While doing the conversations, chatbots usually are collecting previous responses and inputs so their future responses could vary and it will also improve the quality of its future responses. Giving personal recommendations is an example of the result in collecting data from previous conversations.

In the context of e-commerce, chatbots can handle conversations with customers and help them in the process of selling which are expected to give positive effects toward online businesses [5]. It is believed that experienced users feel more satisfied when communicating with virtual communicating agents rather than communicating with inexperienced users in all kinds of aspects that were evaluated including impression, control, effectiveness, navigation, learning, and many more [5].

There are 3 things that form a combination and this combination is the main pillar of chatbot. The 3 things are:

2.1.1. User Interface

User interface referencing to the display interface of a chatbot that bridges between user and chatbot to interact with each other. User interface has to give the best experience to the users when interacting with the user. Without it, chatbot will be useless as it is not convenient to use. The interface includes the text's font, background color or theme, animation, images, and navigation tools.

2.1.2. Artificial Intelligence

AI or Artificial Intelligence is like the core of the whole chatbot application because it is the main brain where everything is processed [6]. The chatbot application could understand inputs when interacting with users because of this. Chatbot is the expansion of the artificial intelligence field so that is why artificial intelligence is very crucial for the whole development of chatbot.

2.1.3. Integration

Integration with other applications or systems can enrich the chatbots' features. Chatbot could get additional or supporting materials and information by integrating it with another system. Thus, integration could improve the user experience and satisfaction. Although it is not as important as artificial intelligence, integration plays a major role too in promoting the usefulness of chatbot.

In a business environment, chatbot can help solve problems regarding communicating with users or customers especially in terms of service and experience when interacting. Thus, chatbot is very effective particularly in problems that are fixated, specific, and predictable. In the rapid business development environment, chatbot becomes an alternative solution that can help customers to reach and interact with business.

2.2. Platform API.AI

API.AI is a platform that provides NLP (Natural Language Processing) and NLU (Natural Language Understanding) services. These services are used to make chatbot smarter and can understand the purpose of what is asked by the user. Natural language processing is one of the disciplines of Artificial Intelligence that focuses on human and computer interactions through natural language that humans use. In NLP the goal to be achieved is the ability of an NLP system to have natural language knowledge both from the sentence structure, the meaning of the word and the purpose of a sentence. While NLU itself is a sub-field of NLP, where the focus of the purpose of NLU itself is to make an understanding of a sentence and conduct semantic analysis. API.AI itself does not support the Indonesian knowledge domain but has a lot of knowledge domains in English. Domain is a collection of knowledge and data structures. Within API.AI itself many knowledge bases have been embedded in this API.AI service system. The collection of knowledge contained in this service includes musician entities, airports, etc. However, the absence of Indonesian language support in this service does not mean this service cannot be used by researchers who speak Indonesian language [7].

2.3. Chatbot Type (Based on its functionality

2.3.1. Business Management Chatbot

This chatbot type is usually found in business environments. Aside from communicating with users, there are some other abilities which are processing business process data and giving information related to certain business. Usually the chatbot is implemented in applications or websites that are related to business and it acts as a supporting application. Chatbots that can be classified as business management chatbot are CardBot, Naver TalkTalk, SuperAgent, and many more [8, 9, 10].

2.3.2. Game Chatbot

This kind of chatbot is the most popular chatbot type among youngsters, especially gamers. It is usually used for entertainment as it is used when users are playing games. The chatbot is usually used when playing multiplayer games but it is compatible with single player games too. One of the examples is SimSimi in which we can use it in mobile and web platforms.

2.3.3. Music Chatbot

Just like its name, this chatbot type is used for playing music. The chatbot is actually pretty easy to use as the chatbot will ask the user to input their choice of songs and the chatbot will automatically search the song in platform YouTube or any accessible platform then it will play the song. This chatbot is usually found in chatting applications that have voice related features in it like Discord application. The chatbot examples are Groovy, FredBoat, and Rythm.

2.3.4. Assistant Chatbot

Assistant Chatbot just like its name means that it helps users in doing daily activities just like how secretaries in real life do [11]. Telling weather forecasts, planning plans, setting important alarms, and asking to call someone are some of the features that this type of chatbot has. The examples of this kind of chatbot are Google Assistant, Siri from Apple, Alexa from Amazon, and Cortana from Microsoft [11].

2.3.5. Education Chatbot

Education chatbot is a chatbot that is used to help users in learning. This type of chatbot is usually implemented in the education field. The benefits of using this chatbot is that it can cut down operational fee and save time when learning. The examples of this kind of chatbot are Duolingo, Gengobot [4].

2.4. Chatbot Examples

2.4.1. Telegram Bot

Telegram Bot is a social media bot that many people installed. This bot gives user feature API (Application Programming Interface) this feature which allow chatbot exist. This feature used for direct Chatting with using chatbot that made user can make chatbot suitable as user wish. Telegram bot started from registration about bot used for chatbot [7].

2.4.2. Pandorabots

Pandorabots is an online service which allows developers to host and deploy Chatbot. This chatbot has 225.000 Developers registered and tools already help 300.000 chatbot. Pandorabots using AIML(Artificial Intelligence Markup Language) which is a standard for writing code in chatbot. This chatbot offers many tools including 'Artificial Intelligence as a Service' these tools provide access API to platform bot hosting and pandorabots playground : a development environment used for made chatbot. Pandorabots are working on constant and have very large active communication. Pandorabots obey the standard for making chatbot. Up until now, open source software has been used for building pandorabots. Pandorabots are also known to be using AIML as knowledge content markup language [12].

2.4.3. CardBot

CardBot is a chatbot which is a result of a research by Meng-Chieh Ko and Zih- Hong Lin in his research entitled CardBot: A Chatbot for Business Card Management with the additional function of being able to read data from business cards with OCR techniques collaborated with cameras and also has a function to organize the data it has acquired into a knowledge base [10]. The data that has been obtained will be managed, combined and stored so that when a user requests data from that already obtained chatbot can immediately retrieve the data to the user. This chatbot uses NLP development technique in its development [10] and includes a type of chatbot for business management.

2.4.4. SuperAgent Chatbot

SuperAgent Chabot is a chatbot with features that can read data from web pages (about products) and can answer questions from users based on data available on these web pages [9]. In addition to these features, chatbot can carry out regular conversations (chit-chat) with users and the knowledge base of this chatbot can be added by the developer as well. This chatbot was made with machine learning techniques, NLP, and using the DSSM model for attribute matching and the first candidate will be the result of the response of this chatbot [9]. This chatbot is used as an extension on the web and can be used on any website (not just linked to a web or application) [9].

2.5. Natural Language Processing (NLP)

This subfield of artificial intelligence is a technique that helps computers understand human language. Using NLP, machines can make sense of unstructured online data so that we can gain valuable insights. NLP primarily comprises natural language understanding (human to machine) and natural language generation (machine to human). This article will mainly deal with natural language understanding (NLU). In recent years there has been a surge in unstructured data in the form of text, videos, audio and photos. NLU aids in extracting valuable information from text such as social media data, customer surveys, and complaints.

2.5.1. Common Techniques used for extracting information

Named Entity Recognition

This technique is one of NLP techniques that is used to extract texts, name and categorize them into a certain topic or context. Named entity recognition or NER is serving as important targets for most language processing systems because they usually carry key information in a sentence [13]. Accurate named entity recognition can be used as a useful information source for different NLP applications. There are lots of topics that NER can identify but there are 3 classes that NER systems use the most: person (PER), location (LOC), and loosely defined miscellaneous (MIS) [13].

Sentiment Analysis

Sentiment analysis or also known as opinion mining is a technique where it aims to determine the attitude of a writer with respect to some topic or the overall contextual polarity of a document [14]. This technique differentiates texts or documents into two main parts: objective (facts) and subjective (opinions). Here are text examples on both objective and subjective:

Objective: I bought a book.

Subjective: The paper was nice.

Sentiment could be identified through three methods which are manual opinion words generation, dictionary-based opinion

words generation and corpus-based opinion words generation. Manual opinion words generation is a method in which we collect the opinion words manually and it is less effective than the other two. The dictionary-based opinion words generation is a method in which we use a seed list and grow the list using words from the dictionary until there are no new opinion words. The last method is a method in which it depends on syntactic or co-occurrence patterns in large text corpora.

In sentiment analysis, opinion words are counted and grouped into 3 categories: positive, negative, and neutral. Opinion words such as "good", "amazing", "great" are considered positive meanwhile "bad", "suck", "terrible" are considered negative. In sentences, every word that is considered positive will be graded as (+1) while negative opinion words will be graded as (-1). If the end result is 0 then the text will be considered neutral, if the result is above 0 then the text will be considered as positive text, and if the result is below 0 then the text will be considered as negative text [14].

Text Summarization

Just like the topic's name, this technique aims to create a summary of a certain document that contains the most important information of the original one [15]. The technique mainly involves large chunks of text such as news articles.

There are two approaches to text summarization: abstraction and extraction. The extraction method works by identifying important sections of the text and generating them verbatim; thus, they depend only on extraction of sentences from the original text [16]. Meanwhile, the abstraction method aims to produce important material in a new way. Since existing abstractive summarizers often rely on an extractive preprocessing component to produce the abstract of the text [16], the paper will be focused on discussing the extractive summarization method. There are three main tasks of extractive text summarization which are to construct an intermediate representation, give sentence score, and summary sentence selection.

Aspect Mining

Aspect mining is a technique that is similar to sentiment analysis but focuses mainly on identifying different aspects in text. There are two major tasks in aspect mining which are aspect extraction and aspect sentiment classification. Process of identifying the opinion words from the given sentence is called aspect extraction and categorizing the extracted opinion words into one of the polarity scales is called aspect sentiment classification [17]. The polarity scales in aspect mining are broad but we can narrow them into 3 major scales: positive, negative, and neutral.

Topic Modeling

Topic modeling is a technique where it aims to give the reader better understanding by identifying natural topics in the text. There are four most popular methods in topic modeling which are Latent Semantic Analysis (LSA), Non-Negative Matrix Factorization (NNMF), Probabilistic Latent Semantic Analysis (PLSA), and Latent Dirichlet Allocation (LDA) [18].

LSA is an algebraic method that presents the semantic space of documents. The more semantic the relationship, the contextual usage will come closer. LSA is based on single value decomposition (SVD). NNMF on the other hand is another reduction technique where there are negative numbers in the dataset which is a problem and are addressed by placing non-negativity constraints on the data model. PLSA is another technique of reduction for detecting semantic co-occurrence of terms in text mining based on bag of words (BOW). PLSA uses a probabilistic framework in a corpus. Finally, LDA is an approach to capture significant inter as well as intra document statistical structure by mixing distributional assumptions that document could arise from multiple topics. The topic is defined as distribution over a vocabulary. This approach is based on the Definneti theorem.

2.5.2. Advantages of NLP

Natural Language Processing can perform a large-scale analysis which can help machines perform language-based tasks such as reading text, identifying topics or what is important, or even extracting sentiment in a text. Moreover, NLP can perform these tasks in such a short time while still applying consistent and unbiased criteria. Thus, it can help people in doing these kinds of tasks faster and more accurately.

Other than large-scale analysis, Natural Language Processing has increased the level of sophistication in engineering because generic engines now can deliver semantic representations for sentences or sentences for representations [19]. Natural Language Processing also makes engines to be more user-friendly since it is highly expressive, highly flexible, and highly representative of reality [20]. On the other hand, NLP helps to bridge the gap between machines and humans because it can decipher information to computers and interpret the way people talk.

3. Research Method

3.1. Literature Review

Literature review is a process in which researchers conduct surveys from scientific sources on a specific topic. In this study, the collection of scientific sources based on international journals, international conferences, and the results of research on chatbots in the business world. All scientific sources are obtained through the internet namely, Google Scholar, ScienceDirect, Academia.edu, and the IEEE website. The sources that had been collected will be analyzed and summarized into important points.

3.2. Survey: Questionnaire

The survey is classified as qualitative data collection. A qualitative approach with case studies is applied in this study where qualitative view of field conditions by collecting data from informants. The survey was done online using Google Forms and each of the results were checked and analyzed thoroughly. After the checking, the results were then summarized as points.

The questions mainly ask about people's perspective on chatbot, its effectiveness, and its features to be added on the chatbot in a business-related environment. This survey method is reliable as the questions are based on this high technology era situation and respondents from all over the world might have the

same responses towards the same question. Thus, this method may produce results that are consistent and reproducible. For the validity of the method, the results would correspond well to the questions and would produce accurate results as the questions were made and distributed to people who are exposed to technology and business at the same era of time.

4. Results

4.1. Questionnaire and Literature Review Result

Based on that questionnaire results, we get that most of the respondents feel that chatbot will be helpful for online shop owners or sellers when handling customers (91 out of 100 respondents). Respondents also feel that an up-to-date answer from the seller is important thus, we can conclude that chatbot will be useful since it responds faster than humans (98 out of 100 respondents). Although most of the respondents feel that chatbot may be able to handle an online shop alone (39 out of 100 respondents) or with human's help (38 out of 100), there are still some respondents who feel that it is better for an online shop to be handled by the seller himself (27 out of 100). From the results shown above, we can conclude that the chatbot which is suitable in the business environment is a fast-responding chatbot with up-to-date data regarding the products or services. Although the idea of having a chatbot handling the online business alone is wonderful, we still believe that the seller plays a major role in the online business and it is better to have both chatbot and the seller handling the online business. In conclusion, chatbots could be accepted by people in business and should be helping the seller instead of handling the business alone. These questionnaire results are from our online questionnaire that we shared on online messaging platforms like LINE application and WhatsApp. The respondents vary from university students (Binus University Alam Sutera) to online shop owners or sellers in Indonesia.

On the other hand, there are some suitable features of a chatbot for business environments based on our literature review. We found out that the SuperAgent chatbot researched by Lecu, Shaohanh, Fuwei, V-Chutan, V-Chadu, and Mingzhou can identify texts whether it was about fact Q&A (information detail about products or services), text Q&A (opinion about the products or services), or chatting [9]. From another journal titled Chatbot Using A Knowledge in Database which was researched by Bayu Setiaji and Ferry Wahyu Wibowo has great features for chatbot especially for online business that need to be pre-trained by a knowledge base. Moreover, it can be used in any language but with some boundaries so chatbots can be more flexible [21]. Lastly, Intellibot which was researched by Mohammad Nuruzzaman and Omar Khadeer Hussain can set a great example for chatbots because it uses a movie as its knowledge base for answering chit-chats and also another personalized knowledge base for questions about products or services [22]. There are some other chatbots that we considered as suitable for the business environment with each feature that will be described in Table 1: Chatbot Description Table.

Table 1: Chatbot Description Table

Chatbot	Description
Chatbot [21]	Chatbot created using database as its knowledge base (RDBMS MySQL), pre-trained chatbot, using bigram method for processing text and identifying response, can be used in any language but with some boundaries, can understand similar words, typo, or even different words and similar meaning (responses are based on the data inside the database). Refer to *Fig. 1* for the example of its usage. The text is in Indonesian but it mainly talks about daily conversation.
Naver TalkTalk [8]	Naver TalkTalk chatbot is a chatbot with cardbot-based interface and using RPA (Robotic Process Automation) as its automation. The bot is pre-trained because the bot will proactively ask questions that is based on its data or knowledge base that previously must be trained or if not the chatbot will not be able to ask related questions. The interface is using cardbot interface and it is simpler to use and to understand what the bot is trying to say. This chatbot is much more compatible with business environment but sadly no chit-chatting can be held in this type of chatbot. Refer to *Fig. 2* for its example of chatbot usage and its cardbot-based interface.
SuperAgent [9]	SuperAgent chatbot is a chatbot that implements machine learning and NLP (Natural Language Processing). The chatbot is connected to business environment and act as a marketing agent. The bot is not pre-trained because it takes what was written on the web page as its knowledge base and will change when the web page is changed too. It is flexible and easy to use as it is used as an extension of an e-commerce web sites. Although it is pretty good for business chatbot, datasets from external can not be implemented as it is not pre-trained. Refer to *Fig. 3* to differentiate between fact QA, text QA, and chit-chat in SuperAgent chatbot.
Intellibot [22]	Intellibot for insurance needs, it is a pre-trained chatbot that uses 2 datasets (Cornell movie and insurance custom dataset) and it is dialogue-based with NLP techniques implementation. The result of this paper's research turns out that the chatbot is more superior than the other bots tested. Refer to *Fig. 4* as the questions given by the user and answers from the chatbots with its strategy selection level ranges from 1-5 as numbers of accuracy and relevancy with 5 as the highest level.
Chatbot with Business Intelligence with Big Data Integration (Enquiry Bot) [23]	The chatbot uses big data analytics for the process of collecting, organizing, and analyzing large data sets to discover hidden patterns or unknown information. The chatbot uses AIML for its knowledge base and it is a pre-trained chatbot as its knowledge base is from the large data set mentioned earlier. Refer to *Fig. 5* on how Enquiry Bot answers questions given however the answers given are still not human-like.
CardBot [10]	CardBot is a chatbot that implements OCR techniques with a goal to help people manage business cards cleverly. It uses NLP for processing the text and OCR to process the text from physical to digital. The knowledge base is based on what was taken by OCR and NLP thus, it is not a pre-trained chatbot. Refer to *Fig. 6* on how CardBot read information with OCR Technique and how the user can manage business cards efficiently.

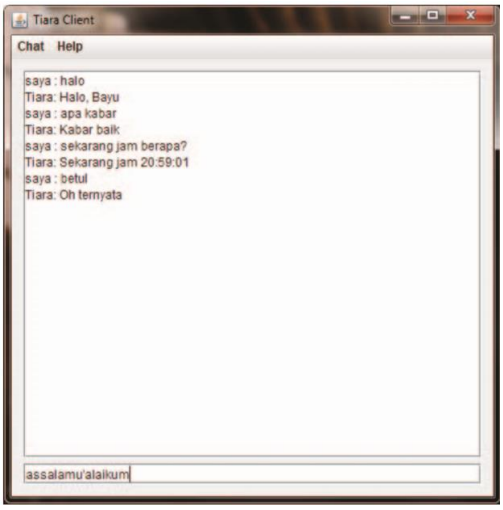


Figure 1: Chatbot

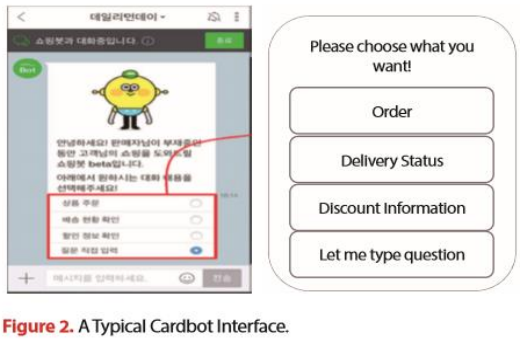


Figure 2: Naver TalkTalk Chatbot

4.2. Natural Language Processing (NLP) Techniques Corresponding to The Result

From the result of both questionnaire and literature review, we can conclude that the ideal chatbot for a business environment must be fast and accurate when answering questions with up-to-date information. From what we had researched, NLP is commonly used in chatbots for text processing. Moreover, chatbots with NLP are fast and accurate when answering questions and their answers are mostly in context so NLP is the most compatible text processing method for developing business related chatbots [13]. There are a lot of techniques in NLP but Named Entity Recognition and Sentiment Analysis will be the most frequently used techniques in chatbots.

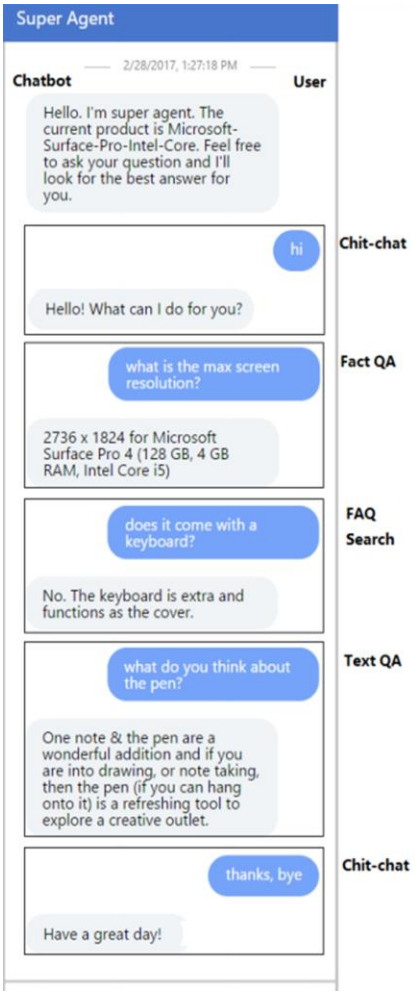


Figure 3: SuperAgent Chatbot

Greetings

	User Input	RootyAI	ChatterBot	DeepQA	IntelliBot	Strategy selection
	Bot Initialization:	Hello, I'm AI assistance. How may I help you?	Hi there, How may I assist you?	Hello, I'm Papaya. I'm your assistant.	Hello, I'm IntelliBot. Your 24/7 h assistance. How may I assist you?	1
1	Hello	Hello	Hello	Hello	Hello	1
2	Good morning.	Good morning.	Good morning.	Good morning.	Good morning.	1
3	My name is Nur.	Nice to meet you, Nur	Ok, Nur	Ok. How can I help you, Nur?	Nice to meet you, Nur. How may I help you?	1
4	What is your name?	I don't know.	I don't have it.	I'm Papayay.	My name is IntelliBot. What yours?	1

Figure 4: Intellibot

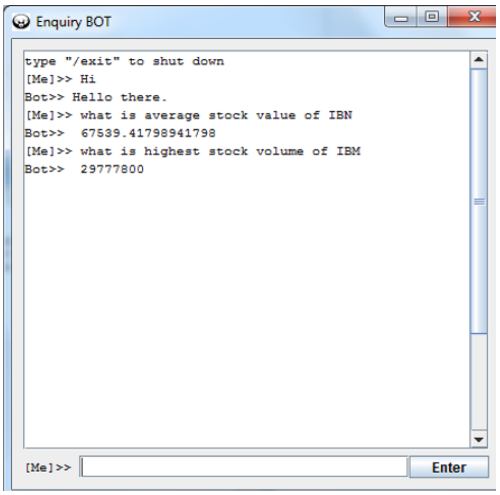


Figure 5: Enquirybot

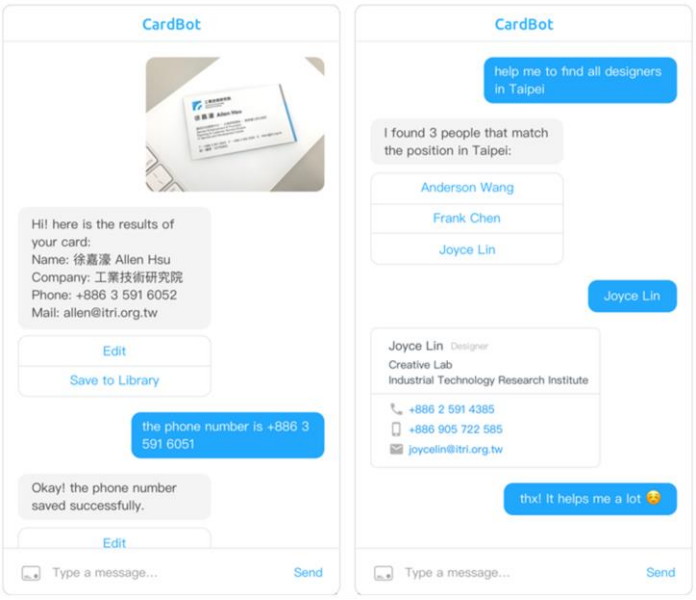


Figure 6: Cardbot

By using NLP named entity recognition technique, it can identify the key words such as, 'how much' and 'red chair'. After extracting the key words, those words will then be processed and will be searched thoroughly in the database as the chatbot's knowledge base will be primarily from the database. After understanding and processing the text, the chatbot can give responses. On the other hand, using NLP sentiment analysis technique can help the chatbot identify whether the interlocutor is feeling great or not about the products or services. This technique will be helpful especially when chit-chatting so the chatbot's attitude of replying can be more accurate, appropriate, and human-like.

Lastly, the training type of the chatbot will be better if it is pre-trained with datasets because it will be more prepared that way. If the chatbot relies on what was written on the website just like SuperAgent chatbot then the website has to be opened beforehand so that the chatbot could read what was written and it is not that efficient. However, OCR technique can be implemented for business chat bots if needed. The datasets will consist of business related datasets (about products or services) and daily conversation datasets so the chatbot could handle both questions related to the business side and daily conversations. These datasets then will be processed by NLP techniques mentioned in the second paragraph and resulted in contextual replies.

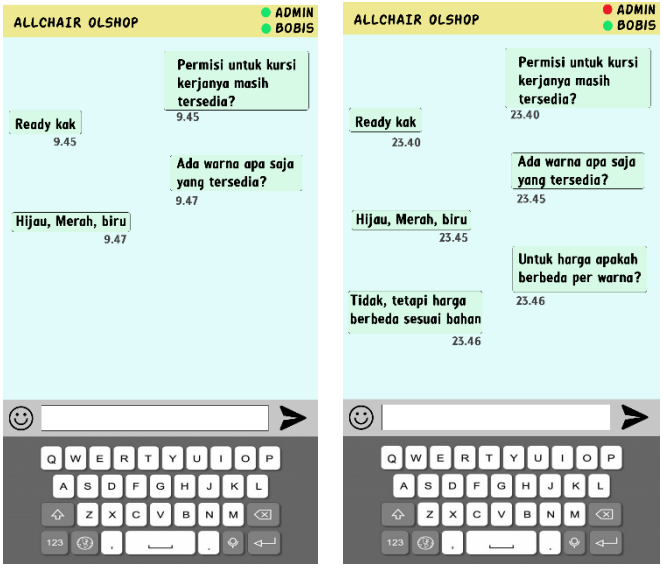


Figure 7: Prototype Chat Example

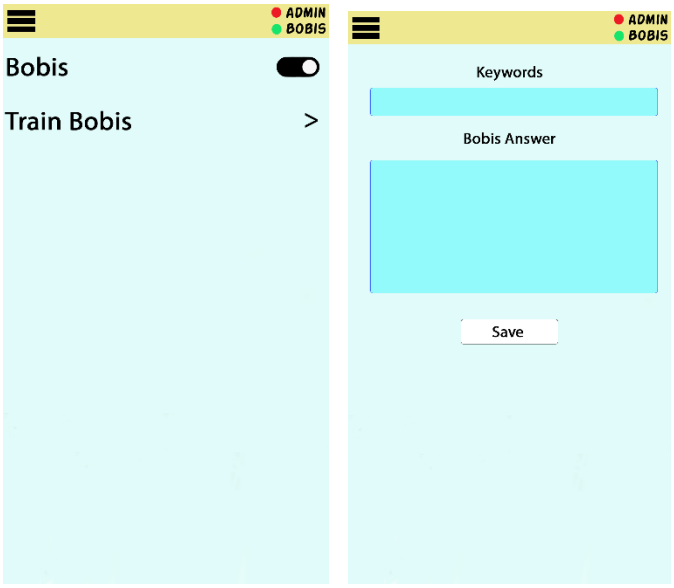


Figure 8: Prototype Setting and Training Menu Example

4.3. Prototype Corresponding to The Result

Based on the results our team got from our research, the chatbot should be able to be online 24 hours a day. The prototype that corresponds to this feature is depicted in Figure 7 where the chatting process could take place at any time and the chatbot would kindly reply. The other feature is that the chatbot should be able to work together with the seller just like an assistant would do, thus the chatbot can help replying to customers' chats when the seller is online or offline. For example, when there are a lot of chats coming through and the seller is online, but the seller only

can reply one chat at a time. In that situation, chatbot could help replying to the other chats that the seller could not cover.

The seller also can train the chatbot outside the datasets they were trained before if needed. The feature is depicted in Figure 8 where the seller could give the keywords and what to answer towards that given keywords. This feature is just an additional feature so that the chatbot could study other keywords other than what it had been trained before. The home menu page depicted in Figure 9 is how the seller would see when it is full of customers' chats.

Figure 9: Prototype Home Menu Example

5. Conclusion

From the research that we did, we can conclude that there are some features that chatbots in a business environment need to have. First, the chatbot needs to be fast and accurate when answering questions and the information in the knowledge base need to be up to date. Then, the chatbot needs to have a text processing technique to help it process texts and thus we can use NLP named entity recognition technique and NLP sentiment analysis technique. The chatbot also needs to be ready 24 hours a day and should act like an assistant helping the seller rather than handling the online business alone. The datasets used should consist of daily conversation dataset and business-related dataset (about products or services). The datasets can be expanded if there are much more datasets needed in the business. The chatbot should be flexible when using, thus the self-training feature by the seller can be used as an additional feature inside the chatbot. However, the self-training feature by the seller can also be removed if the seller does not need it. Lastly, the chatbot interface should be simple as its main aim is to help sellers to enhance their online business. We also can conclude that there is a high chance of acceptability for the chatbot technology especially from younger people [24, 25].

For the future implementation of the chatbot, we recommend that the chatbot should become a part of the application or website the owner has. Since the chatbot should be able to work like an assistant in the business thus the chatbot should be placed in places that are easy to find. If the chatbot is implemented in e-commerce business, it would be convenient to have a chatting menu on the navigation bar and in each page of products respectively. If the chatbot is implemented in company websites, it would be convenient to have the chatting page relatively small to medium size and keep it on a fixed position on the page like a pop-up. Make sure to make it not too eye-catching nor too boring or else users may lose their attention to the chatting pop-up or even not knowing that the chatting pop-up is there. The self-training feature of the chatbot should be implemented in another platform (not a part of the application or website) because we would not want any people controlling the data set without owners' consent.

Conflict of Interest

This paper is created due to a project in our research class and this is our group's research idea about chatbot in the business environment.

Acknowledgment

Special thanks to our lecturer Mrs. Azani that is willing to help us in writing this paper and our respondents that participated in giving their honest opinion for this research.

References

[1] R. Sutoyo, A. Chowanda, A. Kurniati, R. Wongso"Designing an emotionally realistic chatbot framework to enhance its believability with AIML and information states," Procedia Computer Science, **157**, 621–628, 2019, doi:10.1016/j.procs.2019.08.226.

[2] G.D. Souza, "Chatbot for Organizational FAQs," (May), 5591–5594, 2019.

[3] S.A. Abdul-Kader, J.C. Woods, "Survey on Chatbot Design Techniques in Speech Conversation Systems," International Journal of Advanced Computer Science and Applications, **6**(7), 72–80, 2015, doi:10.14569/ijacsa.2015.060712.

[4] N. Haristiani, "Artificial Intelligence (AI) Chatbot as Language Learning Medium: An inquiry," Journal of Physics: Conference Series, **1387**(1), 2019, doi:10.1088/1742-6596/1387/1/012020.

[5] Q.N. Nguyen, A. Sidorova, "Understanding user interactions with a chatbot: A self-determination theory approach," 2018.

[6] E. Bittner, O. Shoury, "Designing Automated Facilitation for Design Thinking: A Chatbot for Supporting Teams in the Empathy Map Method," Proceedings of the 52nd Hawaii International Conference on System Sciences, 227–236, 2019, doi:10.24251/hicss.2019.029.

[7] E. Yuniar, H. Purnomo, "Implementasi Chatbot 'Alitta' Asisten Virtual Dari Balittas Sebagai Pusat Informasi Di Balittas," Antivirus : Jurnal Ilmiah Teknik Informatika, **13**(1), 24–35, 2019, doi:10.35457/antivirus.v13i1.714.

[8] M. Heo, K.J. Lee, "Chatbot as a New Business Communication Tool: The Case of Naver TalkTalk," Business Communication Research and Practice, **1**(1), 41–45, 2018, doi:10.22682/bcrp.2018.1.1.41.

[9] L. Cui, S. Huang, F. Wei, C. Tan, C. Duan, M. Zhou, "Superagent: A customer service chatbot for E-commerce websites," ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of System Demonstrations, 97–102, 2017, doi:10.18653/v1/P17-4017.

[10] M.C. Ko, Z.H. Lin, "Cardbot: A chatbot for business card management," International Conference on Intelligent User Interfaces, Proceedings IUI, 1–2, 2018, doi:10.1145/3180308.3180313.

[11] P. Smutny, P. Schreiberova, "Chatbots for learning: A review of educational chatbots for the Facebook Messenger," Computers and Education, **151**(June 2019), 103862, 2020, doi:10.1016/j.compedu.2020.103862.

[12] N.S. Raj, B.R. Suteja, "Implementasi AIML pada Pandorabot untuk studi kasus Fakultas Teknologi Informasi," **1**, 13–22, 2019.

[13] B. Mohit, Named entity recognition, Springer: 221--245, 2014, doi:10.1007/978-3-642-45358-8_7.

[14] T. Luo, S. Chen, G. Xu, J. Zhou, Sentiment Analysis, Springer: 53--68, 2013, doi:10.1007/978-1-4614-7202-5_4.

[15] L. Abualigah, M.Q. Bashabsheh, H. Alabool, M. Shehab, "Text Summarization: A Brief Review," Studies in Computational Intelligence, **874**(December 2019), 1–15, 2020, doi:10.1007/978-3-030-34614-0_1.

[16] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D., J. B., K. Kochut, "Text Summarization Techniques: A Brief Survey," International Journal of Advanced Computer Science and Applications, **8**(10), 2017, doi:10.14569/ijacsa.2017.081052.

[17] K. Vivekanandan, J.S. Araviandan, "Aspect-based Opinion Mining: A Survey," International Journal of Computer Applications, **106**(3), 21–26, 2014.

[18] P. Kherwa, P. Bansal, "Topic Modeling: A Comprehensive Review," ICST Transactions on Scalable Information Systems, **0**(0), 159623, 2018, doi:10.4108/eai.13-7-2018.159623.

[19] S.R. Joseph, H. Hloman, K. Letsholo, K. Sedimo, "Natural Language Processing: A Review," International Journal of Research in Engineering and Applied Sciences, **6**(3), 1–8, 2016.

[20] A. Kawtrakul, "An Overview of a Role of Natural Language Processing in An Intelligent Information Retrieval System Conventional IR V . S . Intelligent IR."

[21] B. Setiaji, F.W. Wibowo, "Chatbot Using a Knowledge in Database: Human-to-Machine Conversation Modeling," Proceedings - International Conference on Intelligent Systems, Modelling and Simulation, ISMS, **0**, 72–77, 2016, doi:10.1109/ISMS.2016.53.

[22] M. Nuruzzaman, O.K. Hussain, "IntelliBot: A Dialogue-based chatbot for the insurance industry," Knowledge-Based Systems, **196**(xxxx), 105810, 2020, doi:10.1016/j.knosys.2020.105810.

[23] R. Sankar, "Empowering Chatbots With Business Intelligence By Big Data Integration," International Journal of Advanced Research in Computer Science, **9**(1), 627–631, 2018, doi:10.26483/ijarcs.v9i1.5398.

[24] M. Kottorp, F. Jäderberg, "Chatbot as a potential tool for businesses A study on chatbots made in collaboration with Bisnode," 2017.

[25] M. Van Eeuwen, "Mobile conversational commerce: messenger chatbots as the next interface between businesses and consumers," University of Twente, 15, 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 636-641 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Burnout Among Primary School Teachers in the Wazzane Region in Morocco: Prevalence and Risk Factors

Abdeslam Amri[1], Zakaria Abidli[2,*], Mounir Bouzaaboul[1], Rabea Ziri[1], Ahmed Omar Touhami Ahami[1]

[1]*Cognitive Behavioral Neuroscience and Applied Nutrition Team, Faculty of Sciences, University of Ibn Tofail, Kenitra, 1400, Morocco*

[2]*Genetics and Biometrics Laboratory, Faculty of Sciences, University of Ibn Tofail, Kenitra, 1400, Morocco*

ARTICLE INFO

Article history:
Received: 09 August, 2020
Accepted: 16 September, 2020
Online: 24 November, 2020

Keywords:
Burnout
Prevalence
Risk factors
Teachers
Primary
Morocco

ABSTRACT

Introduction: Burnout is a real malaise that affects the mental health of teachers. The objective is to determine the prevalence of burnout among primary school teachers in the Wazzane region and to look for associated risk factors. Methods: Descriptive cross-sectional study conducted in 2017- 2018 with a sample of 330 teachers. Socio-demographic and work-related data were assessed by a self-administered questionnaire and the Karasek Job Content Questionnaire (JCQ) the Karasek, the burnout by the Maslach Burnout Inventory (MBI). Results: 330 teachers participated in the study with a male predominance of 62.4% and an average age of 38.4 (SD = 8.9). The average scores for emotional exhaustion, depersonalization and personal accomplishment were (21.8 ± 9.2); (14.1 ± 9.9) and (15.1 ± 8.4) respectively. Pathologically, emotional exhaustion was high in 23.9%, depersonalization was high in 24.2% and personal accomplishment was low in 26.1% of participants. 46.6% were victims of burnout, of which 23.9% had a low level; 14.2% had a moderate level and 8.5% had severe burnout. Logistic regression analysis showed that the predictors of burnout were female gender ($p<0.01$), Problems of equipment didactic ($p<0.01$), student demotivation ($p<0.05$), conflict with colleagues ($p<0.05$), work overload ($p<0.01$) and low decision-making autonomy ($p<0.01$).Conclusion: These results justify the interest of a prevention program and intervention at different levels to protect teachers.

1. Introduction

Burnout is a negative psychic experience experienced by an individual; it is related to chronic emotional stress caused by working to help people [1]. Maslach and Jackson [2] identified three characteristics of the burnout syndrome:

Emotional exhaustion: It refers to the feeling of being drained of one's emotional and physical resources and leads to difficulties in regarding to the emotions of the other. It represents the basic response to stress.

Depersonalization: It represents a negative or excessively detached response towards other people who are considered as objects.

Loss of personal accomplishment: It refers to feelings of incompetence and inefficiency, lack of success and productivity at work.

Teachers are part of at-isk populations of workers who may be exposed to the burnout syndrome [3]. Indeed, Maslach and Travers have identified several socio-demographic and occupational risk factors for teacher burnout [4, 5]. In our context, no study has been carried out among primary school teachers. For this reason, we aim to determine the prevalence and risk factors of burnout among a population of primary school teachers in the Wazzane district (Morocco).

2. Methods

This cross-sectional study lasted six months, from December 2017 to June 2018. It concerned teachers working in public primary schools in the region of Wazzane in North-West of Morocco. Socio-demographic and work-related data were evaluated by a self-administered questionnaire and by the 29-item Karasek Job Content Questionnaire (JCQ) [6], which measures three different scales: Work overload with 9 items: 10, 11, 12, 13, 14, 15, 16, 17 and 18. Decision latitude with 9 items, subdivided into two sub-scales : the use of skills or qualifications with 6 items:

*Corresponding Author: Zakaria ABIDLI, zakaria.abidli@uit.ac.ma

1, 2, 3, 4, 5, 6 and decision-making autonomy with 3 items: 7, 8, 9. Social support with 11 items, 6 items for support from colleagues : 19, 20, 21, 22, 23, 24 and 5 items for hierarchical support: 25, 26, 27, 28, and 29. The Maslach Burnout Inventory MBI to 22 items that explore three scales has evaluated burnout [7]: Emotional exhaustion assessed by items 1,2,3,6,8,13,14,16,20; Depersonalization assessed by items 5,10,11,15,22; and loss of personal accomplishment assessed by items 4,7,9,12,17,18,19 and 21.These two questionnaires were validated and adapted to the context of Moroccan teachers.

For burnout, a score was calculated and a severity rating (low, moderate, and high) was assigned for each dimension. Emotional exhaustion was considered pathological for a score above 29, depersonalization for a score above 22, and personal accomplishment was considered low for a score below 11. The number of dimensions affected characterizes the degree of severity of burnout, so a single affected dimension determines low burnout, two affected dimensions determine moderate burnout, and three affected dimensions determine severe burnout (high emotional exhaustion and depersonalization scores, low personal accomplishment score). Statistical analysis was realized using SPSS version 21. Frequencies and percentages were calculated for the categorical variables, and the chi-square test (X^2) was used to evaluate the differences between these variables. Continuous numerical data were summarized as mean and standard deviation. Logistic regression was used to determine the predictors of burnout; a p-value<was considered statistically significant.

3. Results

For the 330 participants who completed the data sheet, the average age was 38.4±8.9 years, with a minimum age of 24 and a maximum age of 59. The sex ratio was 1.7 in favor of males. The average years of service was 16.3±8.9 years. 46.7% are between the ages of 35 and 49; 62.4% are male; 52.4% are married and 48.5% have years of service between 11 and 24.

In terms of work-related characteristics, 40% managed multigrade classes; 61.5% had problems due to student demotivation; 54.5% had problems due to equipment didactic; 19.1% reported conflicts with colleagues; 12.4% with administration; workload was high for 57.5%; decision-making autonomy was low for 42.7%, and social support was low for 45.2% of participants (Table 1).

Table 1: Sociodemographic and work-related characteristics of primary school teachers (n = 330) in Wazzane, 2017/2018.

Variable	Number	Percentage
Age		
24-34	125	37.9%
35-49	154	46.7%
≥ 50	51	15.4%
Sex		
Male	206	62.4%
female	124	37.6%
Marital status		
Married	173	52.4%
Unmarried	157	47.6%
Years of service		
1-10	107	32.4%
11-24	162	49.1%
≥25	61	18.5%
Multigrade classes		
Yes	132	40%
No	198	60%
student demotivation		
Yes	163	49.4%
No	167	50.6%
Problems of equipment didactic		
Yes	180	54.5%
No	150	45.5%
Conflicts with colleagues		
Yes	63	19.1%
No	267	80.9%
Conflicts with administration		
Yes	41	12.4%
No	289	87.6%
Workload		
High	187	57.5%
Low	138	42.5%
Decision-making autonomy		
Low	141	42.7%
High	189	57.3%
Social support		
Low	149	45.2%
High	181	57.3%

For the three dimensions of the burnout syndrome (emotional exhaustion, depersonalization and personal accomplishment), the average scores were respectively 21.8±9.1; 14.1±9.9 and 15.1±8.4.

Pathologically, emotional exhaustion was high for 23.9%; depersonalization was high for 24.2% and personal accomplishment was low for 27.6% of participants (table 2).

Similarly, the Maslach Burnout Inventory revealed that 154 teachers (46.6%) were victims of burnout, which 79 (23.9%) had a low level; 47 (14.2%) had a moderate level and 28 (8.5%) had a severe burnout (Figure 1).

Regarding sociodemographic risk factors, the chi-square test showed that age and years of service were not significantly associated with burnout, while gender was found to be significantly associated with burnout (p<0.01). Marital status was also significantly associated with burnout: married were more burned out than unmarried (p<0.001) (Table 3).

Table 2: Means, standard deviations and prevalence of the three dimensions of Burnout among primary school teachers (n=330) in Wazzane, 2017/2018.

Variable	Mean	Standard- Deviation	Pathological cases	Prevalence
Emotional exhaustion	21.8	9.1	79	23.9%
Depersonalization	14.1	9.9	80	24.2%
Personal accomplishment	15.1	8.4	91	27.6%

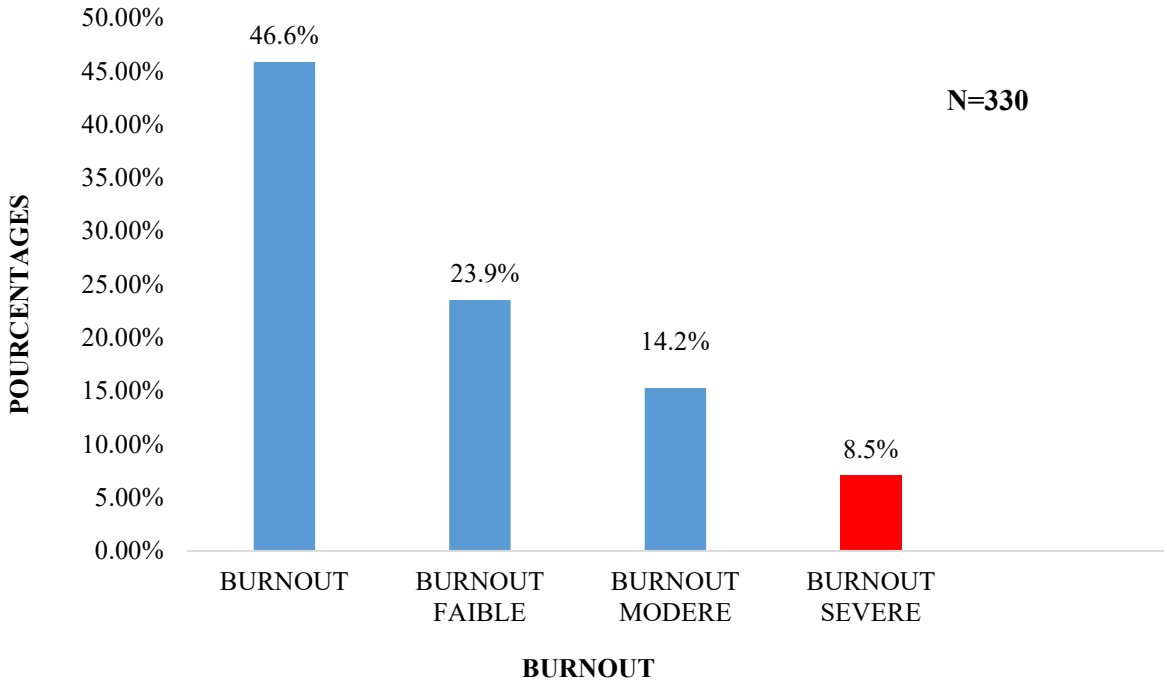


Figure 1: Profile of burnout syndrome among primary school teachers (n = 330) in Wazzane, 2017/2018.

Table 3: Association of burnout with sociodemographic risk factors among primary school teachers (n = 330) in Wazzane, 2017/2018.

Variable	Burnout		No burnout		Chi-square	p-value
	Number	%	Number	%		
Age						
24-34	57	37.0%	68	38.6%		
35-49	75	48.7%	79	44.9%	0.56	**n .s**
≥ 50	22	14.3%	29	16.5%		
Sex						
Male	78	50.6%	128	72.7%	17.06	**<0.001**
female	76	49.4%	48	27.3%		
Marital status						
Married	101	65.6%	72	40.9%	20.05	**<0.001**
Unmarried	53	34.4%	104	59.1%		
Years of service						
1-10	51	33.1%	56	31.8%		
11-24	73	47.4%	89	50.6%	0.36	**n .s**
≥25	30	19.5%	31	17.6%		

For work-related risk factors, the chi square test showed that burnout was significantly related to the management multigrade classes ($p<0.001$), the student demotivation ($p<0.001$), the problems of equipment didactic ($p<0.01$), the conflicts with colleagues ($p<0.001$), the conflicts with administration ($p<0.01$), the work overload ($p<0.01$), the low decision-making autonomy ($p<0.001$) and the low social support ($p<0.01$) (Table 4).

Logistic regression analysis showed that gender was the only sociodemographic variable that predicted burnout; while the work-related variables that predicted it were:

Problems of equipment didactic, student demotivation, conflicts with colleagues, work overload and Low Decision-making autonomy (Table 5).

These results show the importance of knowing certain stress factors specific to the teaching profession. Thus, problems of didactic equipment, student demotivation, work overload and low decision-making autonomy are risk factors that threaten the psychological health of the Moroccan teacher.

Table 4: Association of burnout with work-related risk factors among primary school teachers (n = 330) in Wazzane, 2017/2018**.**

Variable	Burnout		No burnout		Chi-square	p-value
	Number	%	Number	%		
multigrade classes						
Yes	80	51.9%	52	29.5%		
No	74	48.1%	124	70.5%	17.18	<**0.001**
student demotivation						
Yes	95	61.7%	68	38.6%	17.46	<**0.001**
No	59	38.3%	108	61.4%		
Problems of equipment didactic						
Yes	106	68.8%	74	42.0%	23.76	<**0.001**
No	48	31.2%	102	58.2%		
conflicts with colleagues						
Yes	44	28.6%	19	10.8%		
No	110	71.4%	157	89.2%	16.80	<**0.001**
Conflicts with administration						
Yes	28	18.2%	13	7.4%		
No	126	81.8%	163	92.6%	8.79	<**0.01**
workload						
High	103	66.9%	89	50.6%		
Low	51	33.1%	87	49.4%	8.99	<**0.01**
Decision-making autonomy						
Low	82	53.2%	59	33.5%		
High	72	46.8%	117	66.5%	13.05	<**0.001**
Social support						
Low	83	53.9%	66	37.5%		
High	71	46.1%	110	62.5%	8.91	<**0.01**

Table 5: Logistic regression analysis of independent predictors of burnout among primary school teachers (n=330) in Wazzane, 2017/2018

Variable	B	p-value	OR	95%CI	
Sex (female)	-0,743	0,004	0,476	0,286	0,791
Problems of equipment didactic	0,774	0,003	2,169	1,302	3,615
Student demotivation	-0,59	0,032	0,554	0,324	0,95

Conflicts with colleagues	0,733	0,029	2,081	1,079	4,015
Work overload	0,718	0,005	2,05	1,24	3,389
Low decision-making autonomy	-0,787	0,003	0,455	0,272	0,764

OR: 0dds-Ratio; CI: Confidence Interval

These results show the importance of knowing certain stress factors specific to the teaching profession. Thus, problems of didactic equipment, student demotivation, work overload and low decision-making autonomy are risk factors that threaten the psychological health of the Moroccan teacher.

4. Discussion

The objective of our study was to evaluate burnout and determine its prevalence and risk factors among a sample of Moroccan primary school teachers.

In Morocco, no study had yet been carried out among primary school teachers. Most of the studies on burnout concerned the health care professions. However, the teacher may also be exposed to this pathology, as confirmed by many of foreign studies [8, 9, 10]. The prevalence of burnout found in our study is 46.6%, which is higher than that found in the studies of some Arab countries, for example 21% in Tunisia [11], 24.5% in Iraq [12]. However, it remains lower than that found in a Moroccan study conducted among university teachers where it exceeds 60% [13] and that found in some Asian countries where it can reach 50 to 70% [5]. The difference in the prevalence of burnout between studies can be explained by variations in assessment tools and methods, socio-cultural factors, the timing of data collection and teachers' coping strategies to deal with stress. In this study, age and years of service were not associated with burnout, which is consistent with the study by Van Der Doef and Maes [14]. However, in a meta-analysis involving 5 to 35 studies and 1,092 to 10,818 subjects, the association between burnout and age, and burnout and experience was demonstrated, the older and more experienced subjects had lower burnout levels [15].

For gender, it was significantly related to burnout, and women were more exhausted than men, which is consistent with other studies that have reported that female teachers experienced burnout more than teachers [16,17], and a meta-analysis that looked at the effect of gender on burnout reported significantly higher overall burnout scores for women [18].

With regard to teacher burnout, the literature has clearly objectified the link between work-related risk factors and the development of burnout [5]. The main stressors described in the international literature include poor material conditions, work overload, organizational factors, but also administrative difficulties and role conflicts [9, 19, 20]. In our study, logistic regression revealed that the work-related risk factors that predict burnout were work overload ($p<0.01$); low decision-making autonomy ($p<0.01$); the Problems of equipment didactic ($p<0, 01$); work overload ($p<0.01$); student demotivation ($p<0.05$) and conflicts with colleagues ($p<0.05$).

For work overload and low decision-making autonomy, the results found are consistent with a study that looked at associations between work demands (workload) and resources (autonomy, degree of control) and burnout. Positive correlations were found for work demands and negative correlations were found for perceived resources [21]. In the same vein, a meta-analysis of one hundred and seventy-nine articles, representing 186,440 topics treating the associations between burnout and perceived work demands and resources, showed that there are positive correlations with cognitive workload at work and negative correlations with the "Resources" category: autonomy, support from colleagues, knowledge, management style, and feeling of security [22]. For other work-related factors, our work revealed a significant link between burnout and the Problems of equipment didactic, which is consistent with several studies that have found that the material conditions of the school environment play a significant role in teacher distress and dissatisfaction [8, 23]. Addition, student demotivation also represents a risk factor for the development of burnout in this study, and the literature has widely reported the association of burnout with relationship problems with students. Thus, a study, which involved 2558 teachers in France, showed a significant association between difficulties with students and the three pathological dimensions of burnout [24].

5. Conclusion

The Moroccan educational context is affected by a plethora of risk factors that can affect the psychological health of teachers. It is necessary to carry out serious studies on the quality of life at work, teacher satisfaction and psychological health in order to establish a prevention program at different levels and protect this true wealth of Morocco.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] H.J Freudenberger. New psychotherapy approaches with teenagers in a new world. Psychotherapy, 1971, **8**(1). DOI: https://doi.org/10.1111/j.1540-4560.1974.tb00706.x

[2] C. Maslach, MP. Leiter, Early predictors of job burnout and engagement. J Appl Psychol, 93,498-512, 2008. DOI: 10.1037 / 0021-9010.93.3.498.

[3] Direction de l'Animation de la Recherche, des Études et des Statistiques. Les facteurs psychosociaux au travail, les indicateurs disponibles. Consulted on 21/09/2020. Available on: https://travail-emploi.gouv.fr/IMG/pdf/2014-031.pdf

[4] C. J Travers, C. L Cooper. Stress in teaching. In D. Shorrocks-Taylor (Ed.), Directions in educational psychology. London: Whurr publishers, 365-390, 2018.

[5] C. Maslach, WB. Schaufeli, MP. Leiter, Job burnout. Annu Rev Psychol, 52:397–422, 2001. DOI:https://doi.org/10.1146/annurev.psych.52.1.397

[6] R. Bourbonnais, M. Comeau, M. Vézina, G. Dion. Job strain, psychological distress, and burnout in nurses. Am J Ind Med. Jul; **34**(1):20-8. 1998. DOI:10.1002/(sici)10970274(199807)

[7] C. Maslach, S.E Jackson. The measurement of experienced burnout. Journal

of Occupational Behavior,2,99-113,1981. DOI:https://doi.org/10.1002/job.403002020

[8] G. Jaoul, V. Kovess. FSP-MGEN. Le burnout dans la profession enseignante. Ann Medpsychol, **162**(1):26—35,2004.

[9] PJ. Mazur, MD. Lynch, Differential impact of administrative, organizational and personality factors on teacher burnout. Teach Teach Educ, **5**(4):337-53, 1989.

[10] V. Tsenova, T. Tat'ozov, TS. Antonova, et al. The prevalence of the burnout syndrome in the personnel of children's institutions. Probl Khig, 20:90-100,1995.

[11] L. Channoufi, F. Ellouze, W. Cherif, M. Mersni, M.F, M'rad. Stress et épuisement professionnel des enseignants tunisiens. L'Encéphale, **38**, 480-487,2012.DOI: 10.1016/J.ENCEP.2011.12.012

[12] J. Al-Asadi, S. Shukrya Khalaf, A. Al-Waaly, A. Abed, S. Shami. Burnout among primary school teachers in Iraq: prevalence and risk factors EMHJ, **24**, 3, 2018. Consulted on 21/09/2020. Availableon :https://applications.emro.who.int/emhj/v24/03/EMHJ_2018_24_3_262_268.pdf?ua=1&ua=1

[13] M.S., Elmossati, A.O.T. Ahami, H. Oudda, Y. Elkettani. Évaluation de l'épuisement professionnel chez les enseignants universitaires au Maroc. International Journal of Innovation and Applied Studies. IJIAS, **16**, 914-921, 2016.

[14] M. Van der Doef, S. Maes. Teacher-specific quality of work versus general quality of work assessment: A comparison of their validity regarding burnout, psychosomatic well-being and job satisfaction. Anxiety, Stress and Coping. **15**, 327-344, 2002.

[15] EW. Brewer, L. Shapard. Employee burnout: A meta-analysis of the relationship between age or years of experience. Human Resource Development Review, **3**(2),102-23,2004. DOI: https://doi.org/10.1177/1534484304263335

[16] PSY Lau, MT Yuen, RMC Chan. Do demographic characteristics make a difference to burnout among Hong Kong secondary school teachers? Soc Indic Res, **71**, 491–516, 2005. DOI: https://doi.org/10.1007/s11205-004-8033-z.

[17] L. Zhang, J. Zhao, H. Xiao , H. Zheng , Y. Xiao, M. Chen, et al. Mental health and burnout in primary and secondary school teachers in the remote mountain areas of Guangdong Province in the People's Republic of China. Neuropsychiatr Dis Treat, 10:123–30; 2014. DOI: 10.2147 / NDT.S56020.

[18] RK. Purvanova, JP. Muros. Gender differences in burnout: A meta-analysis. Journal of Vocational Behavior, **77**(2) :168-85;2010. DOI: 10.1016/j.jvb.2010.04.006

[19]. Genoud PA, Brodard F, Reicherts M. Facteurs de stress et burnout chez les enseignants de l'école primaire. Rev Eur Psychol Appl, **59**(1):37-45;2009. DOI:10.1016/j.erap.2007.03.001 A

[20]. Rudow B. Stress and Burnout among teaching profession: european studies, issues and research perspectives. In: Vandenberghe R, Huberman AM, editors. Understanding and preventing teacher burnout: a sourcebook of international research and practice. Cambridge: Cambridge University Press; 38-58,1999. DOI:10.1017/CBO9780511527784.004

[21] GM. Alarcon. A meta-analysis of burnout with job demands, resources, and attitudes. Journal of Vocational Behavior, **79**(2):549-62, 2011. DOI: 10.1016/j.jvb.2011.03.007

[22]. Nahrgang JD, Morgeson FP, Hofmann DA. Safety at work: a meta-analytic investigation of the link between job demands, job resources, burnout, engagement, and safety outcomes. The Journal of applied psychology, **96**(1):71- 94, 2011. DOI: 10.1037 / a0021484.

[23] A. Khebbeb. Stress et satisfaction au travail dans le métier d'enseignant universitaire. Rev Sci Hum, 25:5-18,2006.

[24] MN. Vercambre, P. Brosselin, F. Gilbert, et al. Individual and contextual covariates of burnout: a cross-sectional nationwide study of French teachers. BMC Public Health, **9**(333):1-12, 2009. DOI: 10.1186 / 1471-2458-9-333.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 642-651 (2020)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Performance Analysis and Enhancement of Spline Adaptive Filtering based on Adaptive Step-size Variable Leaky Least Mean Square Algorithm

Sethakarn Prongnuch[*,1], Suchada Sitjongsataporn[2]

[1]*Department of Computer Engineering, Faculty of Industrial Technology, Suan Sunandha Rajabhat University Bangkok, 10300, Thailand*

[2]*Department of Electronic Engineering, Mahanakorn Institute of Innovation (MII), Faculty of Engineering, Mahanakorn University of Technology, Bangkok, 10530, Thailand*

ARTICLE INFO

Article history:
Received: 28 September, 2020
Accepted: 11 November, 2020
Online: 24 November, 2020

Keywords:
Spline adaptive filter
Variable leaky
Nonlinear filter

ABSTRACT

This paper presents an adaptive step-size and variable leaky least mean square algorithm based on nonlinear adaptive filter with the adaptive lookup table using spline interpolation. An adaptive step-size approach is proposed with the energy of squared previous and present errors to boost up the convergence rate. A modified variable leaky mechanism is proposed with the optimal leaky parameter by using the recursion form. The proposed algorithm merges an adaptive step-size and a modified variable leaky method with least mean square algorithm for linear and nonlinear network part of spline adaptive filtering in term of fast convergence enhancement. Experimental results demonstrate that proposed algorithm can notably achieve a competitive performance on the convergence rate in comparison with the conventional least mean square algorithm for spline adaptive filter. Simulation results suggest that mean square error performance of proposed algorithm can be partially assessed using adaptive step-size with the variable leaky parameters indicating better than the conventional least mean square algorithm by 16.76%.

1 Introduction

This paper is an extension of work originally presented in the International Electrical Engineering Congress [1], which has been proposed the variable leaky mechanism based on the least mean square (LMS) algorithm for spline adaptive filter. The combination of a linear finite impulse response (FIR) filter and a nonlinear adaptive lookup table (LUT) based on the spline interpolation is called the spline adaptive filtering (SAF) with the adaptation process [2]. It is a class of nonlinear adaptive filtering with a spline function [1]-[12].

According to the practical models, the linear adaptive filter should be inadequate [3]. The models of nonlinear dynamic systems have demonstrated with more robust performance. Linear adaptive filtering should be insufficient in real-world models. Moreover, many dynamic systems using model of nonlinear structure have been extended to the operating model.

Based on the SAF, many research works have achieved efficiently in the practical system identification. In [2], the authors have presented the sign approach with the normalized version of least mean square (NLMS) based on Wiener spline adaptive filter in order to fight against the impulsive noise environment. In [4], the authors have analyzed the convergence and stability analysis of SAF based on LMS algorithm. A steady-state performance of SAF has been examined in [5].

In order to model the nonlinear system identification, the SAF is more attention for the practical use [6],[7]. The linear time-invariant model with the cubic spline function [8] can obtain the good performance working with the adaptive lookup table (LUT) [9] and the spline basis matrix on the adaptive control points coefficient vector.

Applications of SAF architecture have been applied in the infinite impulse response [10] and the system environment with impulsive noise [11]. A set-membership mechanism with the Wiener spline adaptive filtering and normalized least M-estimate algorithm have been obtained significantly the fast convergence in the environment of impulsive noise [11]. In [12], the authors have proposed the Hammerstein function with SAF based on LMS algorithm for improving the convergence rate. Further, a SAF based on the maximum correntropy criterion [13], [14] has been discussed that correntropy is robustness to non-Gaussian noise.

Least mean square (LMS) algorithm which is simple and low computation has been widely used [15]. Most researchers have modified accurately on LMS and NLMS algorithms [16] and [17], the variable step-size LMS [18], the sign mechanism with NLMS [2] and so on.

[*]Corresponding Author: Sethakarn Prongnuch, Suan Sunandha Rajabhat University, Bangkok, Thailand, sethakarn.pr@ssru.ac.th

As stated in a steady-state of convergence analysis, a variable leaky mechanism has been depicted in use of tracking the systems. In [19], a variable leaky LMS algorithm based on the greedy heuristic algorithm has been leveraged to enhance with the high eigenvalue input. In [20], the authors have implemented the variable leaky LMS in the field programmable gate array implementation. The results have been shown that the algorithm can explore the various digital filter architecture.

Based on the variable leaky LMS algorithm which has been orchestrated against and attenuating the drifting [21]. In the field of adaptive signal processing, a variable leaky based on the orthogonal gradient adaptive algorithm with the minimized cost function has been applied for the wireless communications [22].

The novelty of this paper is to merge an adaptive step-size and a modified variable leaky method with least mean square algorithm for linear and nonlinear network part of spline adaptive filtering in term of fast convergence enhancement. The analysis of the convergence performance of the proposed algorithm is derived in forms of the properties and mean square error performance.

In this paper, we organize this paper as follows. Section 2 explains shortly in the structure of SAF. Next, Section 3 introduces the modified adaptive step-size mechanism and the modified variable leaky criterion on the minimization cost function based on the LMS algorithm. Section 4 describes the performance analysis of the proposed adaptive step-size and variable leaky LMS algorithm in terms of the properties and the mean square error performance. Further, the experiment results and discussion show in Section 5 and Section 6, respectively. Finally, Section 7 summarizes the proposed algorithm.

2 Spline Adaptive Filter

Spline adaptive filter (SAF) is namely a combination of linear and nonlinear structures shown in Fig. 1, where $\mathbf{x}_k$ is the input of SAF structure and y_k is the output of system.

The objective is that adaptive lookup table in the nonlinear structure generates an output of SAF y_k nearly to a desired sequence d_k as

$$d_k = y_k + e_k \,, \tag{1}$$

where the error of system e_k should be small. Thus, the adaptive FIR filter brings an output ϕ_k in the linear structure, while the input signal is a sequence of x_k at the linear structure.

An adaptive FIR filter output ϕ_k is given by [3]

$$\phi_k = \mathbf{w}_k^T \, \mathbf{x}_k \,, \tag{2}$$

where $\mathbf{w}_k$ is the update FIR coefficient vector.

Following [3], the SAF output y_k can be examined by following

$$y_k = \mathbf{u}_k^T \, \mathbf{C}_b \, \mathbf{g}_{m,k} \,, \tag{3}$$

$$\mathbf{v}_k = [v_k^3, \; v_k^2, \; v_k, \; 1]^T \,, \tag{4}$$

where $\mathbf{g}_{m_k}$ is the coefficient of control point vector and $\mathbf{C}_b$ is a spline basis matrix.

The local parameter v_k and index m are usually indicated as [5]

$$v_k = \frac{\phi_k}{\Delta x} - \left\lfloor \frac{\phi_k}{\Delta x} \right\rfloor \,, \tag{5}$$

$$m \; = \left\lfloor \frac{\phi_k}{\Delta x} \right\rfloor + \frac{P-1}{2} \,, \tag{6}$$

where Δx is the uniform space between two adjacent-coefficients of control points [6], P is the size of control point coefficient, and the floor operator $\lfloor \cdot \rfloor$ is applied.

By using the minimized objective function based on least mean square (LMS) [5], it becomes

$$J_{\mathbf{w},\mathbf{g}}^{\mathrm{LMS}}(k) = \min_{\mathbf{w},\mathbf{g}} \left\{ \frac{1}{2} \mid e_k \mid^2 \right\} , \tag{7}$$

where *a priori* error e_k is defined as

$$e_k = d_k - y_k = d_k - \mathbf{v}_k^T \, \mathbf{C}_b \, \mathbf{g}_{m,k} \,. \tag{8}$$

The update FIR coefficient vector $\mathbf{w}_k$ can compute by the gradient vector of (7) with respect to the coefficients $\mathbf{w}_k$ in terms of the recursion form as [3]

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu_w \, \nabla J_{\mathbf{w}}^{\mathrm{LMS}} \,, \tag{9}$$

where μ_w is a step-size and $\nabla J_{\mathbf{w}}^{\mathrm{LMS}}(k)$ is the gradient vector for $\mathbf{w}_k$.

Similarly, the update control points vector $\mathbf{g}_{m,k}$ can be expressed by the gradient vector of (7) with respect to the coefficients $\mathbf{g}_{m,k}$ [3]

$$\mathbf{g}_{m,k+1} = \mathbf{g}_{m,k} - \mu_g \, \nabla J_{\mathbf{g}}^{\mathrm{LMS}}(k) \,, \tag{10}$$

where μ_g is a step-size and $\nabla J_{\mathbf{g}}^{\mathrm{LMS}}(k)$ is the gradient vector for $\mathbf{g}_{m,k}$. Thus, the FIR weight vector $\mathbf{w}_k$ and the control points update coefficient vector $\mathbf{g}_{m,k}$ are the particularly simple forms as [5]

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_w \, \mathbf{v}_k' \, \mathbf{C}_b \, \mathbf{g}_{m,k}^T \, \mathbf{x}_k \, e_k \,, \tag{11}$$

$$\mathbf{g}_{m,k+1} = \mathbf{g}_{m,k} + \mu_g \, \mathbf{C}_b^T \, \mathbf{v}_k \, e_k \,, \tag{12}$$

where $\mathbf{v}_k'$ is given by

$$\mathbf{v}_k' = [3 \, v_k^2, \; 2 \, v_k, \; 1, \; 0]^T \,. \tag{13}$$

and the local parameter $\mathbf{v}_k$ is given in (5).

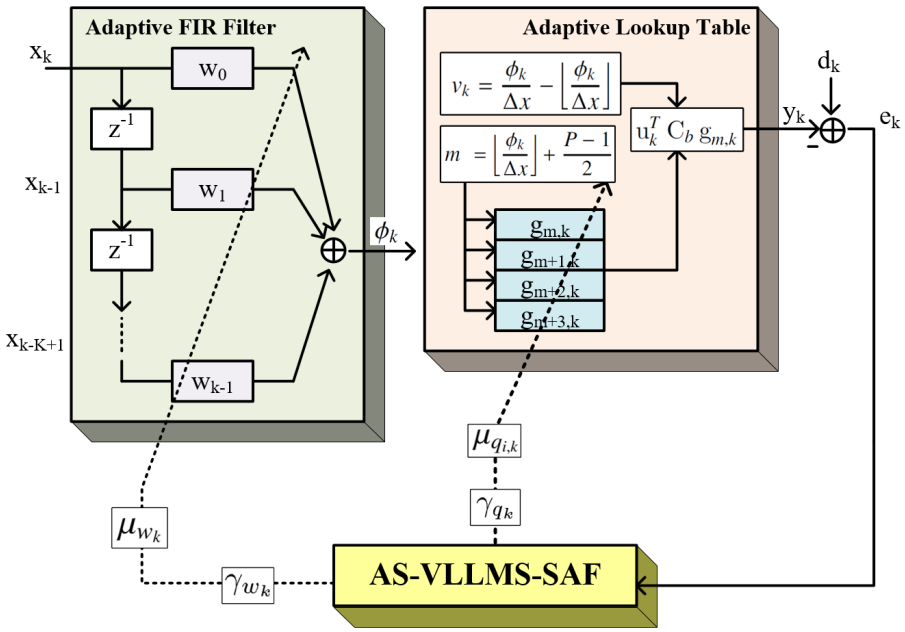


Figure 1: Linear-Nonlinear structure of proposed adaptive step-size variable leaky LMS algorithm for spline adaptive filter (AS-VLLMS-SAF).

3 Proposed Adaptive Step-size and Variable Leaky Least Mean Square based on Spline Adaptive Filter

Based on the least mean square (LMS) algorithm, the advantage of the leaky LMS algorithm is that it avoids the drift of weights [23]. Meanwhile, the step-size parameter is an efficient approach to improve the convergence rate [16].

Following [1] and [11], the cost function using adaptive step-size and variable leaky criterion for SAF based on least mean square (AS-VLLMS-SAF) algorithm can be minimized as

$$J(\mathbf{w}_k, \mathbf{g}_{m,k}) = \min_{\mathbf{w}_k, \mathbf{g}_{m,k}} \left\{ \frac{1}{2} \mid e_k \mid^2 + \gamma_w \|\mathbf{w}_k\|^2 + \gamma_g \|\mathbf{g}_k\|^2 \right\}, \tag{14}$$

where γ_w and γ_g are the leaky paramters for the linear FIR coefficient vector $\mathbf{w}_k$ and the control points weight vector $\mathbf{g}_{m,k}$, respectively. The *a priori* error of system e_k is given as

$$e_k = d_k - \mathbf{v}_k^T \, \mathbf{C}_b \, \mathbf{g}_{m,k} \, . \tag{15}$$

Considering the chain rule on the cost function in (14) with respect to (w.r.t) $\mathbf{w}_k$, we get

$$\begin{aligned} \frac{\partial J(\mathbf{w}_k, \mathbf{g}_{m,k})}{\partial \mathbf{w}_k} &= \left\{ - e_k \frac{\partial y_k}{\partial \mathbf{v}_k} \frac{\partial \mathbf{v}_k}{\partial \phi_k} \frac{\partial \phi_k}{\partial \mathbf{w}_k} + \gamma_w \mathbf{w}_k \right\} \\ &= \left\{ - \frac{e_k}{\Delta x} \mathbf{v}'_k \, \mathbf{C} \, \mathbf{g}_{m,k} \, \mathbf{x}_k + \gamma_w \mathbf{w}_k \right\}, \end{aligned} \tag{16}$$

where $\mathbf{v}'_k$ is defined as

$$\mathbf{v}'_k = [3v_k^2 \;\; 2v_k \;\; 1 \;\; 0] \, . \tag{17}$$

The gradient of cost function in (14) w.r.t $\mathbf{g}_{m,k}$ can be expressed by using the chain rule in the form of vector below

$$\begin{aligned} \frac{\partial J(\mathbf{w}_k, \mathbf{g}_{m,k})}{\partial \mathbf{g}_{m,k}} &= \left\{ - e_k \frac{\partial y_k}{\partial \mathbf{v}_k} \frac{\partial \mathbf{v}_k}{\partial \phi_k} \frac{\partial \phi_k}{\partial \mathbf{g}_{m,k}} + \gamma_g \mathbf{g}_{m,k} \right\} \\ &= \{ - e_k \, \mathbf{C}^T \, \mathbf{v}_k + \gamma_g \mathbf{g}_{m,k} \} \, . \end{aligned} \tag{18}$$

Hence, the proposed update linear FIR coefficient vector $\mathbf{w}_k$ of AS-VLLMS-SAF algorithm is the stochastic adaptation formula as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu_{w_k} \frac{\partial J(\mathbf{w}_k, \mathbf{g}_{m,k})}{\partial \mathbf{w}_k} \, , \tag{19}$$

where μ_{w_k} is a adaptive step-size parameter at symbol k.

By substituting (16) into (19), the proposed update tap-weight vector $\mathbf{w}_k$ can be expressed as

$$\therefore \mathbf{w}_{k+1} = (\mathbf{1} - \mu_{w_k} \gamma_{w_k}) \, \mathbf{w}_k + \frac{\mu_{w_k}}{\Delta x} \, \mathbf{v}'_k \, \mathbf{C}_b \, \mathbf{g}_{m,k} \, \mathbf{x}_k \, e_k \, , \tag{20}$$

where γ_{w_k} is a variable leaky parameter for $\mathbf{w}_k$. It is noticed that $(1 - \mu_{w_k} \gamma_{w_k})$ is also defined as the leakage factor [23] for updated weight $\mathbf{w}_k$ in which its value is generally close to 1.

The adaptive control points coefficient vector $\mathbf{g}_{m,k}$ of AS-VLLMS-SAF becomes

$$\mathbf{g}_{m,k+1} = \mathbf{g}_{m,k} - \mu_{\mathbf{g}_k} \frac{\partial J(\mathbf{w}_k, \mathbf{g}_{m,k})}{\partial \mathbf{g}_{m,k}} \, , \tag{21}$$

By substituting (18) into (21), the proposed control points vector $\mathbf{g}_{m,k}$ is given as

$$\therefore \mathbf{g}_{m,k+1} = (\mathbf{1} - \mu_g \gamma_{g_k})\, \mathbf{g}_{m,k} + \mu_q\, \mathbf{C}_b^T\, \mathbf{v}_k\, e_k\ , \tag{22}$$

where μ_g is a step-size parameter and γ_{g_k} is a variable leaky parameter for $\mathbf{g}_{m,k+1}$. It is noted that $(1 - \mu_g \gamma_{g_k})$ is also defined as the leakage factor [23] for updated weight $\mathbf{g}_{m,k+1}$ in which its value is generally close to 1.

3.1 Modified Variable Leaky mechanism

Following [24], we introduce the modified variable leaky algorithm with the optimal leaky parameter for weights $\mathbf{w}_k$ and $\mathbf{g}_{m,k}$ in the recursion form as

$$\gamma_{w_k} = \gamma_{w_{k-1}} + \rho_w\, \gamma_w^{\text{opt}}\ , \tag{23}$$

$$\gamma_{g_k} = \gamma_{g_{k-1}} + \rho_g\, \gamma_g^{\text{opt}}\ , \tag{24}$$

where ρ_w, ρ_g and γ_w^{opt}, γ_g^{opt} are the adaptation rate and the optimal leaky parameters of weights $\mathbf{w}_k$ and $\mathbf{g}_{m,k}$.

We rewrite (20) as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu_{w_k} \gamma_{w_k} \mathbf{w}_k + \frac{\mu_{w_k}}{\Delta x}\, \mathbf{v}_k'\, \mathbf{C}_b\, \mathbf{g}_{m,k}\, \mathbf{x}_k\, e_k\ , \tag{25}$$

Assumption 1: *We consider the steady-state value of* $E\{\boldsymbol{w}_{k+1}\}$ *for* $k \to \infty$ *by*

$$E\{\mathbf{w}_{k+1}\} \simeq E\{\mathbf{w}_k\}.$$

We determine the optimal leaky parameter for $\mathbf{w}_k$ by using this assumption above in (25), we arrive at

$$\gamma_w^{\text{opt}} \simeq \frac{\mathbf{v}_k'\, \mathbf{C}_b^T\, \mathbf{g}_{m,k}\, \mathbf{w}_k^{-1}\, \mathbf{x}_k\, e_k}{\Delta x}\ . \tag{26}$$

Therefore, the modified variable leaky algorithm in the recursion form can be expressed as

$$\gamma_{w_k} = \gamma_{w_{k-1}} - \rho_w\, \frac{\mathbf{v}_k'\, \mathbf{C}_b^T\, \mathbf{g}_{m,k}\, \mathbf{w}_k^{-1}\, \mathbf{x}_k\, e_k}{\Delta x}\ , \tag{27}$$

where ρ_w is an adaptation rate for $\mathbf{w}_k$.

Consequently, the *a posterior* AS-VLLMS error [1] is rewritten as

$$e_{p_k}^{\text{VL-LMS}} = (1 - \mu_{g_k}\, \mathbf{v}_k\, \mathbf{C}_b^T\, \mathbf{v}_k\, \mathbf{C}_b)\, e_k + \mu_{g_k}\, \gamma_{g_k}\, \mathbf{v}_k\, \mathbf{C}_b^T\, \mathbf{g}_{m,k}\ . \tag{28}$$

Assumption 2: *We consider the steady-state value of proposed algorithm is stable for* $k \to \infty$ *by*

$$E\{e_{p_k}^{\text{VL-LMS}}\} \simeq E\{e_k\}.$$

By using Assumption 2 in (28), the optimal leaky parameter can be defined as

$$\gamma_g^{\text{opt}} \simeq \frac{\Omega_k^T\, \Omega_k\, e_k}{\Omega_k^T\, \mathbf{g}_{m,k}}\ , \tag{29}$$

where $\Omega_k = \mathbf{v}_k^T\, \mathbf{C}$.

Table 1: Proposed Variable Leaky Least Mean Square based on Spline Adaptive Filtering with the adaptive step-size (AS-VLLMS-SAF) algorithm.

Initialize : $\mathbf{w}(0) = \delta\, \mathbf{I}$ and $\mathbf{I}$ is the identity matrix.

$$\mathbf{w}_k = [\ w_0\ w_1\ \ldots\ w_{K-1}\ ]$$
$$\mathbf{x}_k = [\ x_k\ x_{k-1}\ \ldots\ x_{k-K+1}\ ]$$
$$\mathbf{g}_{m,k} = [\ g_{m,k}\ \ g_{m+1,k}\ \ g_{m+2,k}\ \ g_{i+3,k}\ ]$$

for $k = 0, 1, 2, \ldots, K-1$.

$$\phi_k = \mathbf{w}_k^T\, \mathbf{x}_k$$
$$v_k = \frac{\phi_k}{\Delta x} - \left\lfloor \frac{\phi_k}{\Delta x} \right\rfloor$$
$$m\ = \left\lfloor \frac{\phi_k}{\Delta x} \right\rfloor + \frac{P-1}{2}$$
$$\mathbf{v}_k = [v_k^3,\ v_k^2,\ v_k,\ 1]^T$$
$$\mathbf{v}_k' = [3v_k^2,\ 2v_k,\ 1,\ 0]^T$$
$$e_k = d_k - \mathbf{v}_k^T\, \mathbf{C}_b\, \mathbf{g}_{m,k}$$
$$\gamma_{w_k} = \gamma_{w_{k-1}} + \rho_w\, \frac{\mathbf{v}_k'\, \mathbf{C}_b^T\, \mathbf{g}_{m,k}\, \mathbf{w}_k^{-1}\, \mathbf{x}_k\, e_k}{\Delta x}$$
$$\gamma_{q_k} = \gamma_{q_{k-1}} + \rho_q\, \frac{\Omega_k^T\, \Omega_k\, e_k}{\Omega_k^T\, \mathbf{g}_{m,k} + \epsilon}$$
$$\Omega_k = \mathbf{v}_k^T\, \mathbf{C}_b$$
$$\mu_{w_k} = \alpha_w\, \mu_{w_{k-1}} + \beta_w\, \xi_k^2$$
$$\xi_k = \lambda\, \xi_{k-1} + (1-\lambda)\, e_{k-1}^*\, e_k$$
$$\mu_{q_k} = \alpha_q\, \mu_{q_{k-1}} + \beta_q\, e_{k-1}^*\, e_k$$
$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu_{w_k} \gamma_{w_k} \mathbf{w}_k + \frac{\mu_{w_k}}{\Delta x}\, \mathbf{v}_k'\, \mathbf{C}_b^T\, \mathbf{g}_{m,k}\, \mathbf{x}_k\, e_k$$
$$\mathbf{q}_{i,k+1} = \mathbf{g}_{m,k} - \mu_{q_k} \gamma_{q_k} \mathbf{g}_{m,k} + \mu_{q_{m,k}}\, \mathbf{C}_b^T\, \mathbf{v}_k\, e_k$$

end

Therefore, the modified variable leaky algorithm can be expressed in the recursion method as

$$\gamma_{g_k} = \gamma_{g_{k-1}} - \rho_q\, \frac{\Omega_k^T\, \Omega_k\, e_k}{\Omega_k^T\, \mathbf{g}_{m,k} + \epsilon}\ , \tag{30}$$

where ρ_g is an adaptation rate for $\mathbf{g}_{m,k}$ and ϵ is a regularization parameter with a small constant.

3.2 Modified Adaptive Step-size Approach

According to the better convergence of proposed AS-VLLMS-SAF algorithm, the step-size parameter should be adapted recursively.

The behavior of convergence rate is which the algorithm starts converging, the value of step-size parameter should be large in order to boost up the learning rate of convergence. At closely to the steady-state, the step-size parameter might be decreased adequately to get the lower adjustment. This leads to adjust accordingly the step-size parameter.

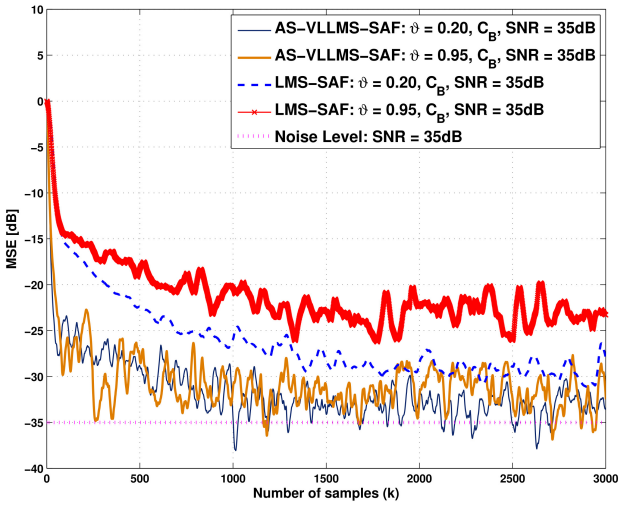


Figure 2: MSE of proposed AS-VLLMS-SAF and LMS-SAF [3] on the *B-spline* matrix $\mathbf{C}_B$ with the different $\vartheta = 0.20, 0.95$ and SNR = 35dB.

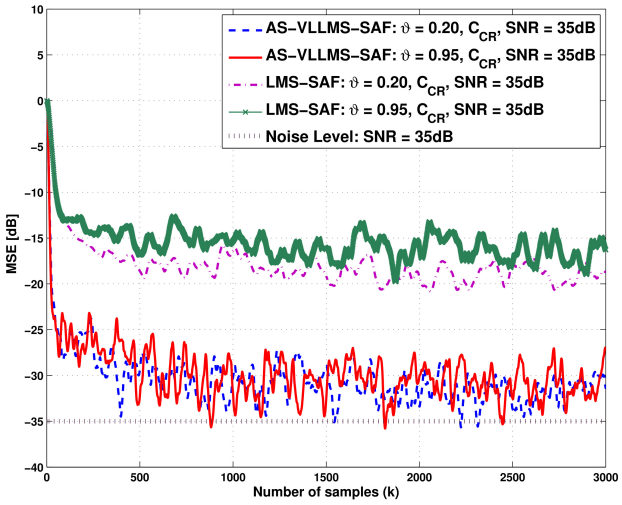


Figure 3: MSE of proposed AS-VLLMS-SAF and LMS-SAF [3] on the *Catmul-Rom* matrix $\mathbf{C}_{CR}$ with $\vartheta = 0.20, 0.95$ and SNR=35dB.

We assume that the step-size parameters μ_{w_k} and μ_{g_k} are regulated by squared estimated error. Hence, the proposed adaptive step-size μ_{w_k} for weight $\mathbf{w}_k$ based on AS-VLLMS algorithm is introduced for spline adaptive filtering as [4]

$$\mu_{w_k} = \alpha_w \mu_{w_{k-1}} + \beta_w \xi_k^2 , \tag{31}$$

$$\xi_k = \lambda \xi_{k-1} + (1-\lambda) e_{k-1}^* e_k , \tag{32}$$

and the proposed adaptive step-size μ_{g_k} for weight $\mathbf{g}_{m,k}$ can be expressed as

$$\mu_{g_k} = \alpha_g \mu_{g_{k-1}} + \beta_g e_{k-1}^* e_k , \tag{33}$$

A summary of proposed variable leaky least mean square for spline adaptive filtering based on adaptive step-size algorithm (AS-VLLMS) is depicted in Table 1.

4 Performance Analysis

In this section, we analyze the convergence performance of the proposed variable leaky LMS based on SAF. Compared to the LMS-SAF algorithm, the proposed AS-VLLMS algorithm is presented by introducing an additional concept.

Variable leaky algorithm is used in the iterative of weights $\mathbf{w}_k$ and $\mathbf{g}_{m,k}$, respectively. So, the computational complexity of AS-VLLMS-SAF is more complex than that of the LMS-SAF algorithm.

4.1 Properties of AS-VLLMS Algorithm

The properties of adaptive step-size and variable leaky LMS algorithm may be derived by examining the behaviour of $E\{\mathbf{w}_k\}$ and $E\{\mathbf{g}_{m,k}\}$ of SAF.

Taking expectation value and using the independence assumption [25], we then have

$$E\{\mathbf{w}_{k+1}\} = (1 - \mu_{w_k}\gamma_{w_k})\, E\{\mathbf{w}_k\} + \frac{\mu_w}{\Delta x} E\{\mathbf{v}_k' \mathbf{C}_b \mathbf{g}_{m,k} \mathbf{x}_k e_k\} . \tag{34}$$

Specially, we note that if $\mathbf{w}_k$ converges to the steady-state solution, we get

$$\lim_{k\to\infty} E\{\mathbf{w}_k\} = \frac{\mathbf{v}_k' \mathbf{C}_b \mathbf{g}_{m,k}^T \mathbf{x}_k e_k}{\gamma_{g_k} \Delta x} . \tag{35}$$

In a similar fashion, we substitute *a priori* error e_k in (15) into (22), then we have

$$\mathbf{g}_{m,k+1} = \left[\mathbf{I} - \mu_{g_k} \{\gamma_{g_k} \mathbf{I} + \mho_k^T \mho_k\}\right] \mathbf{g}_{m,k} + \mu_{g_k} \mho_k d_k , \tag{36}$$

where $\mho_k$ is given by

$$\mho_k = \mathbf{C}_b^T \mathbf{v}_k . \tag{37}$$

Thus, we take the expectation value and using the independence assumption [25], it becomes

$$E\{\mathbf{g}_{m,k+1}\} = \left[\mathbf{I} - \mu_{g_k}\left\{\mathbb{R}_\mho + \gamma_{g_k} \mathbf{I}\right\}\right] E\{\mathbf{g}_{m,k}\} + \mu_{g_k} \mathbf{r}_{d\mho} , \tag{38}$$

where $\mathbb{R}_\mho$ and $\mathbf{r}_{d\mho}$ are defined by

$$\mathbb{R}_\mho = \mho_k^T \mho_k , \tag{39}$$

$$\mathbf{r}_{d\mho} = d_k \mho_k . \tag{40}$$

Clearly, we note that if $\mathbf{g}_{m,k}$ converges to the steady-state solution, we have

$$\lim_{k\to\infty} E\{\mathbf{g}_{m,k}\} = \left(\mathbb{R}_\mho + \gamma_{g_k} \mathbf{I}\right)^{-1} \mathbf{r}_{d\mho} . \tag{41}$$

It is noticed that the leakage coefficient introduces a bias into the steady-state solution.

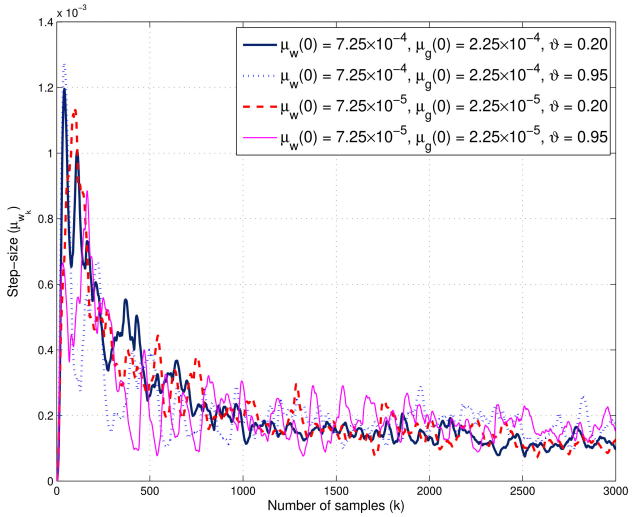


Figure 4: Learning curves of adaptive step-size μ_{w_k} of coefficient vector $\mathbf{w}_k$ of proposed AS-VLLMS-SAF on the *B-spline* matrix $\mathbf{C}_B$ with $\alpha = 0.20, 0.95$ and SNR = 35dB.

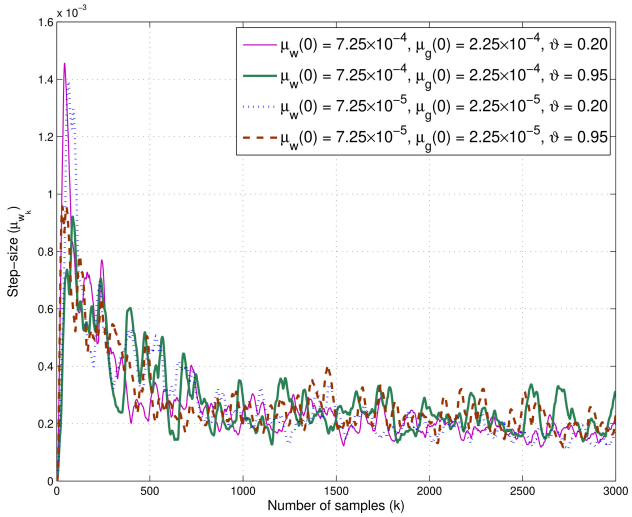


Figure 6: Learning curves of adaptive step-size μ_{w_k} of coefficient vector $\mathbf{w}_k$ of proposed AS-VLLMS-SAF on the *Catmul-Rom* matrix $\mathbf{C}_{CR}$ with $\alpha = 0.20, 0.95$ and SNR = 35dB.

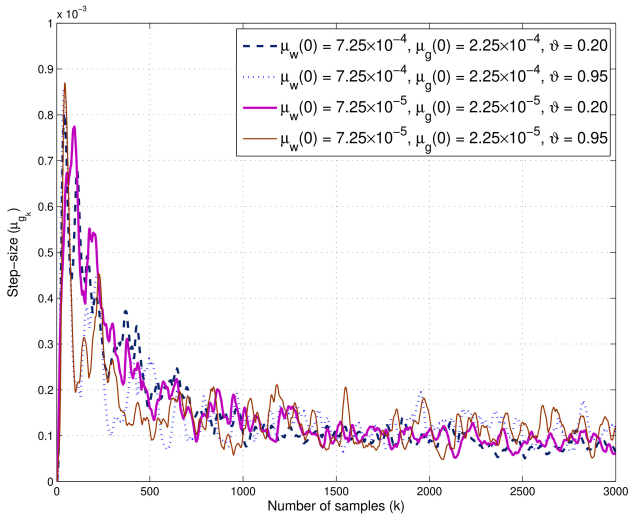


Figure 5: Learning curves of adaptive step-size μ_{g_k} of coefficient vector $\mathbf{g}_{m,k}$ of proposed AS-VLLMS-SAF on the *B-spline* matrix $\mathbf{C}_B$ with $\alpha = 0.20, 0.95$ and SNR = 35dB.

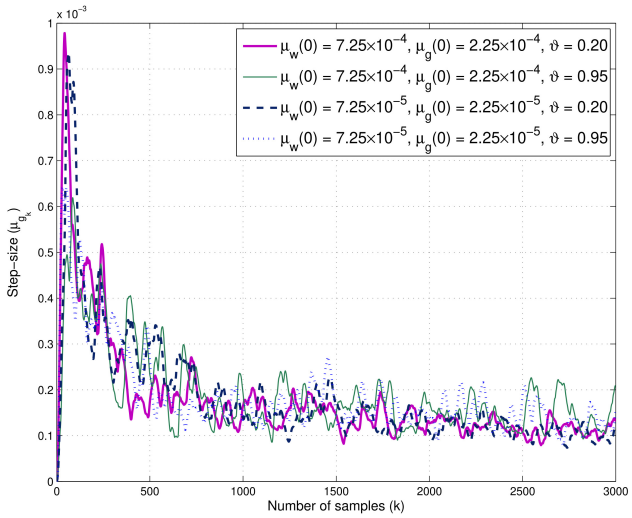


Figure 7: Learning curves of adaptive step-size μ_{g_k} of coefficient vector $\mathbf{g}_{m,k}$ of proposed AS-VLLMS-SAF on the *Catmul-Rom* spline matrix $\mathbf{C}_{CR}$ with $\alpha = 0.20, 0.95$ and SNR = 35dB.

4.2 *The Leaky Adjustment*

In this section, the leaky adjustment is investigated in form of the *a posterior* error compared with *a priori* error in the spline adaptive filtering. We determine the *a posteriori* LMS error as [1]

$$\tilde{e}_k^{LMS} \triangleq d_k - \mathbf{v}_k^T \, \mathbf{C}_b \, \mathbf{g}_{m,k+1} = e_k \left(1 - \mu_g \, \mho_k^T \, \mho_k \right), \qquad (42)$$

where $\mho_k$ is described in (37).

Correspondingly, we examine the *a posteriori* variable leaky least mean square (VL-LMS) error as

$$\begin{aligned} \tilde{e}_k^{VLLMS} &\triangleq d_k - \mathbf{v}_k^T \, \mathbf{C}_b \, \mathbf{g}_{m,k+1} \\ &= e_k(1 - \mu_{g_k} \mho_k^T \, \mho_k) + \mu_{g_k} \gamma_{g_k} \mho_k^T \mathbf{g}_{m,k} \, . \end{aligned}$$

It is noticed that the leaky adjustment will be achieve based on a greedy heuristic algorithm with each iteration, which can get the appropriate leaky value, if $|\tilde{e}_k^{VLLMS}| < |\tilde{e}_k^{LMS}|$. That means the proposed VL-LMS algorithm would allow to get the efficient LMS algorithm. Otherwise, the leak parameter should be diminished.

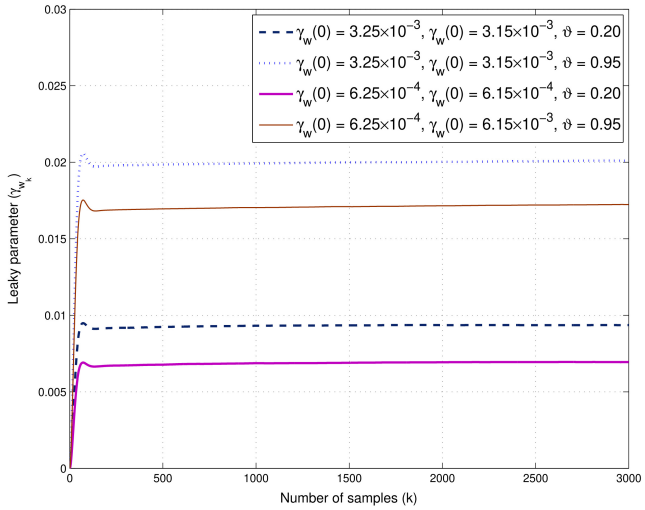


Figure 8: Learning curves of variable leaky γ_{w_k} of coefficient vector $\mathbf{w}_k$ of proposed AS-VLLMS-SAF on $\mathbf{C}_B$ with $\vartheta = 0.20, 0.95$ and SNR = 35dB.

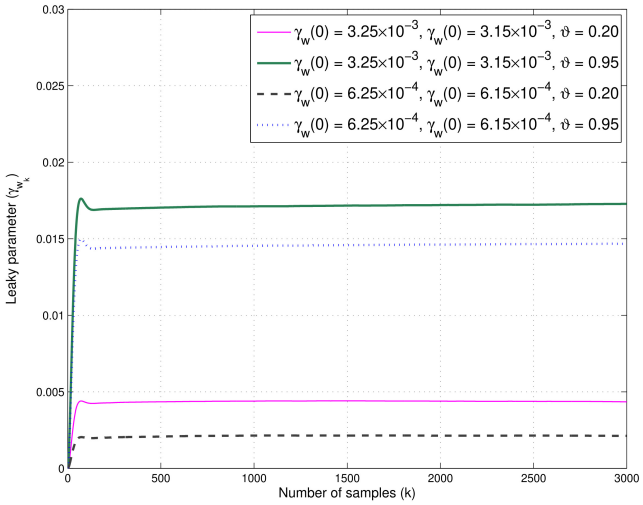


Figure 10: Learning curves of variable leaky γ_{w_k} of coefficient vector $\mathbf{w}_k$ of proposed AS-VLLMS-SAF on $\mathbf{C}_{CR}$ with $\vartheta = 0.20, 0.95$ and SNR = 35dB.

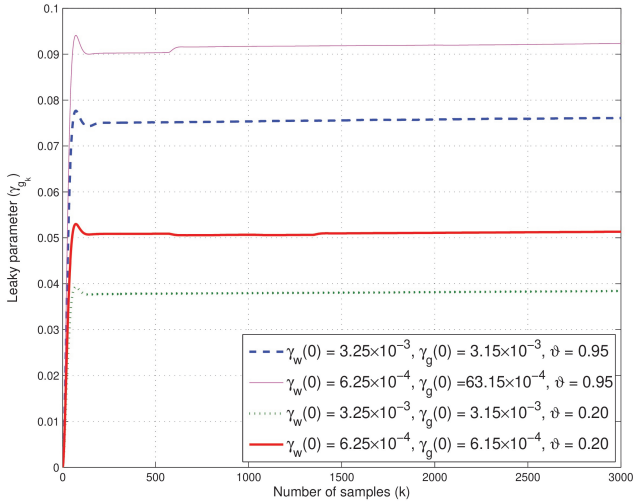


Figure 9: Learning curves of variable leaky γ_{g_k} of coefficient vector $\mathbf{g}_{m,k}$ of proposed AS-VLLMS-SAF on $\mathbf{C}_B$ with $\alpha = 0.20, 0.95$ and SNR = 35dB.

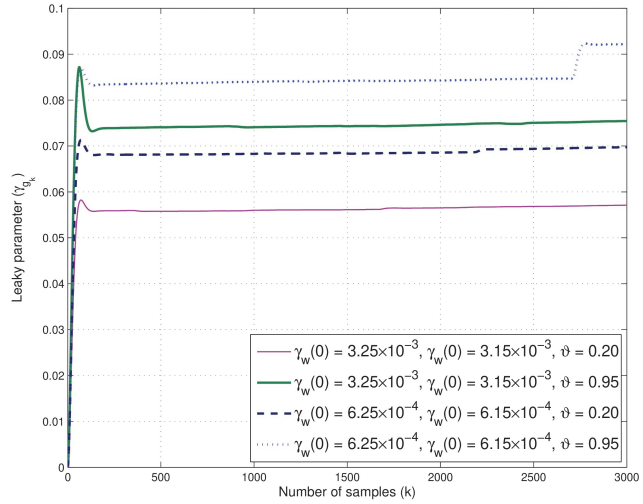


Figure 11: Learning curves of variable leaky γ_{w_k} of coefficient vector $\mathbf{w}_k$ of proposed AS-VLLMS-SAF on $\mathbf{C}_{CR}$ with $\alpha = 0.20, 0.95$ and SNR = 35dB.

4.3 Mean Square Error Performance

We investigate the mean square error (MSE) performance of the proposed AS-VLLMS-SAF algorithm at the steady-state.

Assumption 3: *We assume that the a priori and a posteriori optimal errors are identical, we have*

$$E\{\epsilon_g^{\text{opt}}\} \approx e_{\text{opt},g} \, .$$

Assumption 4: *We consider the convergence condition for $k \to \infty$, that is of*

$$E\{\epsilon_g^{\text{opt}}\} \to 0 \, , \text{ as } \; k \to \infty$$
$$E\{\mathbf{g}_{m,k}\} \to \mathbf{g}_{m,k}^{\text{opt}} \, , \text{ as } \; k \to \infty$$

Following [12], we decompose the MSE under these assumptions above as follows.

$$\mathbb{J}_{MSE}^{P} = \mathbb{J}_{MMSE}^{P} + \mathbb{J}_{EX}^{P} = E\{\|\epsilon_g\|^2\} \, , \tag{43}$$

where $\mathbb{J}_{MMSE}^{P}$ is the minimum mean square error (MMSE) as

$$\mathbb{J}_{MMSE}^{P} = E\{\|\epsilon_g^{\text{opt}}\|^2\} \, , \tag{44}$$
$$\epsilon_g^{\text{opt}} = d_k - \mathbf{v}_k^T \, \mathbf{C}_b \, \mathbf{g}_{\text{opt}} \, , \tag{45}$$

where ϵ_g^{opt} is the *a posteriori* optimal error of $\mathbf{g}_{\text{opt}}$.

Certainly, $\mathbb{J}_{EX}^{P}$ is the *a posteriori* excess mean square error (EMSE) given by

$$\mathbb{J}_{EX}^{P} = \mathbb{J}_{MSE}^{P} - \mathbb{J}_{MMSE}^{P} = E\{\|\epsilon_g\|^2\} - E\{\|\epsilon_g^{\text{opt}}\|^2\} \, . \tag{46}$$

4.4 Experimental conditions on parameters setting

In this section, the theoretical experiments are simulated in system identification over the random process and under the Gaussian noise scenario. We evaluate the performance of proposed adaptive step-size variable leaky least mean square algorithm based on spline adaptive filter (AS-VLLMS-SAF) as compared to the conventional least mean square algorithm based on spline adaptive filter (LMS-SAF) [3].

The coloured input signal generates for all experiments over 100 Monte Carlo trials and 5,000 samples used that is generated by [2]

$$x_k = \vartheta \cdot x_{k-1} + \sqrt{1-\vartheta^2}\zeta_k \,, \tag{47}$$

where ζ_k is a unitary variance type of zero mean white Gaussian noise, $\vartheta = [0, 0.95]$ and an interval sampling Δx is used at 0.2 [11].

An unknown Wiener system is composed by a linear component as [3]

$$\mathbf{w}_0 = [0.6,\ -0.4,\ 0.25,\ -0.15,\ 0.1]^T \,,$$

and a 23-point length lookup table (LUT) $\mathbf{g}_0$ is a nonlinear memoryless target function applied by

$$\mathbf{g}_0 = [\ -2.2,\ \ldots\ , -0.8,\ -0.91,\ -0.4,\ -0.2,\ 0.05,\ 0,\ -0.4,\ 1.0,\ 1.0,\ 1.2,\ \ldots\ , 2.2\ ] \,.$$

The linear filter $\mathbf{w}_0$ is initialzed as $\mathbf{w}_0 = [1,\ 0, \ldots,\ 0]$ and $\delta = 0.001$. Other parameters are fixed at the length of coefficient vector $T = 5$, a signal to noise ratio $SNR = 35$dB . The spline basis matrices are used as the *B-spline* matrix $\mathbf{C}_B$ and the *Catmul-Rom spline* matrix $\mathbf{C}_{CR}$ in [8] which are selected for simulation experiments as follows:

$$\mathbf{C}_B = \frac{1}{6}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -3 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix}, \tag{48}$$

$$\mathbf{C}_{CR} = \frac{1}{2}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 2 & -5 & 4 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix}. \tag{49}$$

The fixed parameters of proposed AS-VLLMS-SAF algorithm are as follows: $\alpha = 0.20, 0.95$, $\alpha_w = 7.55 \times 10^{-3}$, $\beta_w = 2.75 \times 10^{-3}$, $\alpha_g = 6.55 \times 10^{-3}$, $\beta_g = 1.85 \times 10^{-3}$, $\lambda = 0.97$, $\epsilon = 1 \times 10^{-6}$ and $\rho_w = 1.5 \times 10^{-6}$, $\rho_g = 1.125 \times 10^{-6}$. And the initial parameters for proposed AS-VLLMS-SAF are $\gamma_w(0) = 3.25 \times 10^{-2}$, $\gamma_g(0) = 3.25 \times 10^{-2}$, $\mu_w = 7.75 \times 10^{-4}$, $\mu_g = 2.25 \times 10^{-4}$. Other fixed parameter of SAF-LMS [3] are as: $\mu_w = 0.05$ and $\mu_g = 0.05$.

Table 2: Summary of MSE of proposed AS-VLLMS-SAF with the initial parameters: $\mu_w(0) = 7.25 \times 10^{-5}$, $\mu_g(0) = 2.25 \times 10^{-5}$, $\gamma_w(0) = 6.25 \times 10^{-4}$, $\gamma_g(0) = 6.15 \times 10^{-4}$, of VL-LMS in [1] with $\gamma_w(0) = 3.15 \times 10^{-4}$, $\gamma_g(0) = 3.15 \times 10^{-4}$ and of LMS-SAF in [3] with $\mu_w = \mu_g = 0.05$ over 100 Monte Carlo trials and 5,000 samples used at SNR = 35dB

Algorithm	ϑ in (47)	Spline matrix	MSE at steady-state condition	
			MSE	dB
AS-VLLMS	$\vartheta = 0.20$	$\mathbf{C}_B$	1.268×10^{-3}	-28.969
		$\mathbf{C}_{CR}$	1.448×10^{-3}	-28.392
	$\vartheta = 0.95$	$\mathbf{C}_B$	1.768×10^{-3}	-29.293
		$\mathbf{C}_{CR}$	2.342×10^{-3}	-26.304
VL-LMS in[1]	$\vartheta = 0.20$	$\mathbf{C}_B$	1.815×10^{-3}	-27.409
		$\mathbf{C}_{CR}$	1.931×10^{-3}	-27.144
	$\vartheta = 0.95$	$\mathbf{C}_B$	2.175×10^{-3}	-26.625
		$\mathbf{C}_{CR}$	2.592×10^{-3}	-25.862
LMS in[3]	$\vartheta = 0.20$	$\mathbf{C}_B$	3.303×10^{-3}	-24.811
		$\mathbf{C}_{CR}$	1.352×10^{-2}	-18.688
	$\vartheta = 0.95$	$\mathbf{C}_B$	8.345×10^{-3}	-20.786
		$\mathbf{C}_{CR}$	2.858×10^{-2}	-15.439

5 Simulation Results

For the experiment results, the mean square error (MSE) is simulated at $\vartheta = 0.20, 0.95$. Fig. 2 and Fig. 3 show the MSE convergence curves of proposed AS-VLLMS-SAF compared with the original LMS-SAF [5] based on the *B-spline* matrix ($\mathbf{C}_B$) matrix and the *Catmul-Rom* spline matrix ($\mathbf{C}_{CR}$) with the parameters $\vartheta = 0.2, 0.95$ in (47) shown in dB and SNR=35dB, respectively. We notice that the curves of MSE of proposed AS-VLLMS-SAF based on both spline basis matrices outperform when comparable to that of the LMS-SAF algorithm.

Furthermore, Fig. 4 and Fig. 5 based on the *B-spline* matrix and Fig. 6 and Fig. 7 based on the *Catmul-Rom* spline matrix demonstrate the curves of learning rate of μ_{w_k} of coefficient vector $\mathbf{w}_k$ and μ_{g_k} of control points $\mathbf{g}_{m,k}$ for the proposed AS-VLLMS-SAF at $\vartheta = 0.20, 0.95$ and SNR=35dB, respectively. Their learning curves are depicted to converge to their equilibrium at the steady-state, even the initial of μ_{w_k} and μ_{g_k} are varied.

Finally, Fig. 8 and Fig. 9 based on the *B-spline* matrix and Fig. 10 and Fig. 11 based on the *Catmul-Rom* spline matrix present the curves of learning parameters of γ_{w_k} of coefficient vector $\mathbf{w}_k$ and γ_{g_k} of adaptive control points vector $\mathbf{g}_{m,k}$ for the proposed AS-VLLMS-SAF at $\vartheta = 0.20, 0.95$ and SNR=35dB, respectively. It can seen clearly that their learning rates are shown to converge for the tracking ability at the steady-state with the different initial parameters.

Summary of MSE of proposed AS-VLLMS-SAF with the initial parameters as $\mu_w(0) = 7.25 \times 10^{-5}$, $\mu_g(0) = 2.25 \times 10^{-5}$, $\gamma_w(0) = 6.25 \times 10^{-4}$, $\gamma_g(0) = 6.15 \times 10^{-4}$ and of LMS-SAF in [3] with the fixed parameter as $\mu_w = \mu_g = 0.05$ over 100 Monte Carlo trials and 5,000 samples used at SNR = 35dB and $\vartheta = 0.20, 0.95$ is presented in Table. 2. Simulation results suggest that mean square error performance of proposed algorithm can be partially assessed using adaptive step-size with the variable leaky parameters indicating better than the conventional least mean square algorithm by 16.76% with the *B-spline* matrix.

6 Discussion

The comparison of MSE averaged over 100 trials are shown the robustness and superiority of proposed AS-VLLMS algorithm over the conventional LMS algorithm. We have plotted the learning curves of adaptive step-size μ_{w_k} of coefficient vector $\mathbf{w}_k$, μ_{g_k} of control points $\mathbf{g}_{m,k}$ for the proposed AS-VLLMS algorithm and of variable leaky γ_{w_k} of coefficient vector $\mathbf{w}_k$ and γ_{g_k} of adaptive control points vector $\mathbf{g}_{m,k}$ after 5,000 iterations, which are seen that their learning curves converges to their equilibrium, even the initial values are assigned to be varied.

7 Conclusion

We have orchestrated an adaptive step-size and a variable leaky approach based on least mean square algorithm for spline adaptive filtering. The proposed AS-VLLMS-SAF algorithm has been explained how to derive using the variable leaky mechanism. We have designed an adaptive step-size algorithm with the methods of an energy of squared previous and present estimated error. We have designed a modified leaky algorithm with the methods of an optimal leaky parameter. Simulation experiments have shown that the proposed AS-VLLMS-SAF algorithm can perform well with the corresponding LMS-SAF algorithm.

In general, spline adaptive filtering structure is already fascinatingly applied in many applications such as signal processing for communications in nonlinear channel estimation and acoustic processing in bio-acoustic signal.

Conflict of Interest The authors declare no conflict of interest.

References

[1] A. Saengmuang, S. Sitjongsataporn, "Spline Adaptive Filtering based on Variable Leaky Least Mean Square Algorithm," in Proc. IEEE International Electrical Engineering Congress (iEECON), Chiang Mai, Thailand, 2020.

[2] C. Liu, Z. Zhang, X. Tang, "Sign Normalised Spline Adaptive Filtering Algorithms Against Impulsive Noise", Signal Processing, **148**, 234-240, 2018, doi:10.1016/j.sigpro.2018.02.022.

[3] M. Scarpiniti, D. Comminiello, R. Parisi, A. Uncini, "Nonlinear spline adaptive filtering," Signal Processing, **93**(4), 772-783, 2013, doi:10.1016/j.sigpro.2012.09.021.

[4] A. Saengmuang, S. Sitjongsataporn, "Convergence and Stability Analysis of Spline Adaptive Filtering based on Adaptive Averaging Step-size Normalized Least Mean Square Algorithm ", International Journal of Intelligent Engineering and Systems, **13**(2), 267-277, 2020, doi:10.22266/ijies2020.0430.26.

[5] M. Scarpiniti, D. Comminiello, G. Scarano, R. Parisi, A. Uncini, "Steady-State Performance of Spline Adaptive Filters", IEEE Transactions on Signal Processing, **64**(4), 816-828, 2015, doi:10.1109/TSP.2015.2493986.

[6] M. Scarpiniti, D. Comminiello, R. Parisi, A. Uncini, "Novel cascade spline architectures for the identification of nonlinear systems," IEEE Transactions on Circuits and Systems I: Regular Papers, **62**(7), 1825-1835, 2015, doi: 10.1109/TCSI.2015.2423791.

[7] M. Scarpiniti, D. Comminiello, R. Parisi, A. Uncini, "Hammerstein uniform cubic spline adaptive filtering : learning and convergence properties," Signal Processing, **100**, 112-123, 2014, doi:10.1016/j.sigpro.2014.01.019.

[8] S. Guan, Z. Li, "Normalised Spline Adaptive Filtering Algorithm for Nonlinear System Identification," Neural Processing Letter, **46**(2), 595-607, 2017, doi:10.1007/s11063-017-9606-6.

[9] S. Prongnuch, R.E. Valmoria, "Applied of Co-design in Reconfigurable System for Remote Image Noise Filtering via Ethernet Technology", in Proc. International Science, Social Science, Engineering and Energy Conference, 92-98, 2013.

[10] M. Scarpiniti, D. Commiiniello, R. Parisi, A. Uncini, "Nonlinear system identification using IIR spline adatpive filers," Signal Processing, **108**, 30-35, 2015, doi:10.1016/j.sigpro.2014.08.045.

[11] C. Liu, Z. Zhang, "Set-membership normalised least M-estimate spline adaptive filtering algorithm in impulsive noise," Electronics Letters, **54**(6), 393-395, 2018, doi:10.1049/el.2017.4434.

[12] S. Prongnuch, S. Sitjongsataporn, "Stability and Steady-State Performance of Hammerstein Spline Adaptive Filter Based on Stochastic Gradient Algorithm ", International Journal of Intelligent Engineering and Systems, **13**(3), 112-123, 2020, doi:10.22266/ijies2020.0630.11.

[13] W. Wang, H. Zhao, X. Zeng, K. Dogancay, "Steady-State Performance Analysis of Nonlinear Spline Adaptive Filter Under Maximum Correntropy Criterion", IEEE Transactions on Circuits and Systems II: Express Briefs, vol. **67**(6), 1154-1158, 2020, doi: 10.1109/TCSII.2019.2929536.

[14] W. Shi, Y. Li, Y. Wang, "Noise-Free Maximum Correntropy Criterion Algorithm in Non-Gaussian Environment", IEEE Transactions on Circuits and Systems II: Express Briefs, **67**(10), 2224-2228, 2020, doi:10.1109/TCSII.2019.2914511.

[15] D. Bismor, M. Pawelczyk, "Stability Conditions for the Leaky LMS Algorithm based on Control Theory Analysis", Archives of Acoustics, **41**(4), 731–739, 2016, doi:10.1515/aoa-2016-0070.

[16] S. Sitjongsataporn, W. Chimpat, "Adaptive Step-size Normalised Least Mean Square Algorithm for Spline Adaptive Filtering", in 2019 IEEE International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), 544-547, 2019, doi:10.1109/ITC-CSCC.2019.8793383.

[17] L. Deng, Y. M. Huang, Q. Chen, Y. He, X. Sui, "Collaborative Blind Equalization for Time-Varying OFDM", IEEE Access, **8**, 103073-103087, 2020, doi: 0.1109/ACCESS.2020.2999387.

[18] K. Xiong, S. Wang, B. Chen, "Robust Normalized Least Mean Absolute Third Algorithm", IEEE Access, **7**, 10318-10330, 2019, doi:10.1109/ACCESS.2019.2891549.

[19] M.Kamenetsky, B.Widrow, "A Variable Leaky LMS Adaptive Algorithm," in *2004 IEEE Asilomar Conference on Signals, Systems and Compute (ASILOMAR)*, 125-128, 2004, doi:10.1109/ACSSC.2004.1399103.

[20] P. Kalkar, J.S.R.Alex, "Field Programmable Gate Array implementation of a Variable Leaky Least Mean Square Adaptive Algorithm," Asian Journal of Pharmaceutical and Clinical Research, 2017, doi:10.22159/ajpcr.2017.v10s1.19566.

[21] T.R.Gwadabe, M.S.Salman, H.Abuhilal, "A Modified Leaky-LMS Algorithm," International Journal of Computer and Electrical Engineering, **6**(3), 222-225, 2014, doi:10.1109/4234.1001664.

[22] S. Sitjongsataporn, P. Nurarak, "Low Complexity Variable Leaky Normalised Orthogonal Gradient Adaptive Algorithm", in 2017 IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA), 1-4, 2017, doi:10.1109/WAINA.2017.110.

[23] H. Kaur, K. Kaur, R. Mehra, "Adaptive Filter For Audio Signal Enhancement Using Leaky and non leaky algorithms," International Journal of Electrical and Electronic Engineering and Telecommunications, **3**(2), 114-122, 2014.

[24] S. Sitjongsataporn, "Modified adaptive forgetting-factor recursive least squares-based frequency-domain equalisation," in 2013 IEEE International Electrical Engineering Congress (iEECON), 1-4, 2013. doi:10.1109/iEECON.2013.110.

[25] M. H. Hayes, Statistical Digital Signal Processing and Modeling, John Wiley&Sons, 1996.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 652-663 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Automatic Stochastic Dithering Techniques on GPU: Image Quality and Processing Time Improved

Giorgia Franchini*, Roberto Cavicchioli, Jia Cheng Hu

University of Modena and Reggio Emilia, Department of Physics, Informatics and Mathematics, 41125, Modena, Italy

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 01 November, 2020
Online: 24 November, 2020

Keywords:
Dithering
Half-tone
Floyd-Steinberg
Stochastic Algorithm
Parallel Computing

ABSTRACT

Dithering or error diffusion is a technique used to obtain a binary image, suitable for printing, from a grayscale one. At each step, the algorithm computes an allowed value of a pixel from a grayscale one, applying a threshold and, therefore, causing a conversion error. To obtain the optical illusion of a continuous tone, the obtained error is distributed to adjacent pixels. In literature there are many algorithms of this type, to cite some Jarvis, Judice and Ninke (JJN), Stucki, Atkinson, Burkes, Sierra but the most known and used is the Floyd-Steinberg. We compared various types of dithering, which differ from each other for the weights and number of pixels involved in the error diffusion scheme. All these algorithms suffer from two problems: artifacts and slowness. First, we address the artifacts, which are undesired texture patterns generated by the dithering algorithm, leading to a less appealing visual results. To address this problem, we developed a stochastic version of Floyd-Steinberg's algorithm. The Weighted Signal to Noise Ratio (WSNR) is adopted to evaluate the outcome of the procedure, an error measure based on human visual perception that also takes into account artifacts. This measure behaves similarly to a low-pass filter and, in particular, exploits a contrast sensitivity function to compare the algorithm's result and the original image in terms of similarity. We will show that the new stochastic algorithm is better in terms of both WSNR measurement and visual analysis. Secondly, we address the method's inherent computational slowness: we implemented a parallel version of the Floyd-Steinberg algorithm that takes advantage of GPGPU (General Purtose Graphics Processing Unit) computing, drastically reducing the execution time. Specifically, we observed a quadratic time complexity with respect to the input size for the serial case, whereas the computational time required for our parallel implementation increased linearly. We then evaluated both image quality and the performance of the parallel algorithm on a exhaustive image database. Finally, to make the method fully automatic, an empirical technique is presented to choose the best degree of stochasticity.

1 Introduction

Most printing devices are constrained by a limited number of color intensities (typically 2/4/8/16 values), as consequence, before printing an image they must first convert the original color domain into a new set of values that can be represented with the color intensities at their disposal. A process we call quantization.

An intuitive and trivial quantization method is to map each input pixel to the nearest available color. However, most of the time this leads to a visually unpleasant and low quality output image due to a large number of artifacts. Here, dithering comes to help and it has been proved to be able of achieving a higher quality results. In particular, the best known and popular dithering algorithms is the Floyd-Steinberg (FS) [1] which is based on errors diffusion. Besides the FS, there are other algorithms that differ from this for the weight distribution matrix. The error scattering technique is straightforwardly described in Algorithm 1.

Algorithm 1 Dithering algorithm

```
for rows ∈ image do
    for pixels p ∈ row do
        find closest color τ to p
        err = p − τ
        diffuse err to neighbouring pixels according to the error diffusion scheme
    end for
end for
```

*Corresponding Author: Giorgia Franchini, University of Modena and Reggio Emilia, Email: giorgia.franchini@unimore.it

In case of binary dithering applied to a monotone image (e.g. grayscale to black and white), Algorithm 1 is just a simple threshold operation. Extending it to a larger set of colors (or tones) is a trivial change in *finding the closest color*. It was one of the most revolutionary methods and is widely used in digital printers, which prints small and isolated dots.

The other algorithms against which we will compare for the numerical experiments are: Atkinson, Burkes, Fan, Sieraa, Filter Lite, Jarvis, Judice and Ninke (JJN), Jarvis, Stucki and Shio Fan [2].

There are several error distribution scheme and different ways to scan the image, such as the traditional left-to-right direction and the top-to-bottom raster. Considering all the matrices of weights related to the various methods would be redundant, therefore we show only a few in Figure 1 to clarify the concept.

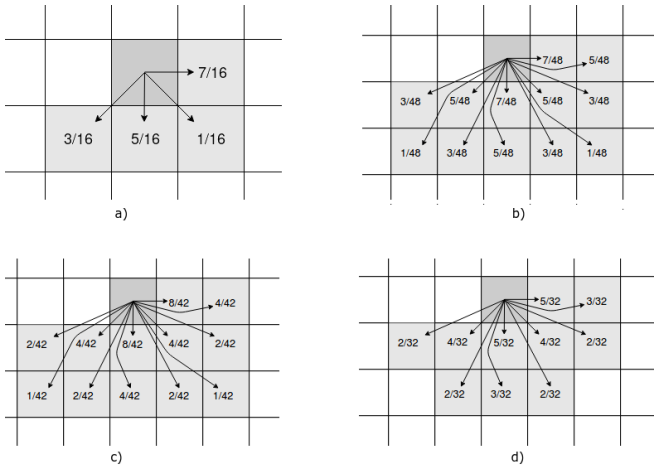


Figure 1: Error diffusion scheme for different dithering algorithms: Floyd-Steinberg (a); Jarvis, Judice and Ninke (b); Stucki (c); Sierra (d). The flux of the whole image remains the same.

We also took care of the borders, in particular, we ensured that the image flux or average is maintained through the entire image, Figure 2 shows the modified FS error distribution *weights* in all particular cases.

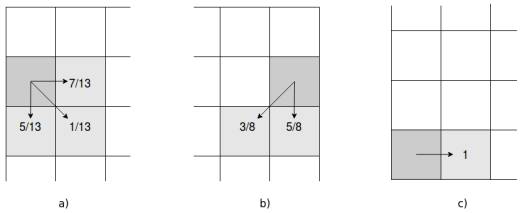


Figure 2: Error diffusion in the FS dithering in case of boundaries. a) first column, b) last column and c) last row.

The major shortcomings of the FS dithering algorithm are the generation of disturbing hysteresis artifacts or worm patterns at extreme gray-levels[3, 4] and the slowness [5].

2 Background and preliminaries

The purpose of image dithering, or halftoning, is the procedure that generates a pattern of quantized pixels able to create a continuous-tone image illusion. A comprehensive analysis about halftoning techniques can be found in [2]. Dithering is necessary for showing gray scale images on printed surfaces when the only available tones are black and white. Dithering algorithms will produce different results with respect to quality and image characteristics. To allow a comparison between different algorithms, we must therefore quantify the performance. Doing so by means of psychovisual tests is hard due to the need of controlled conditions and long duration of the tests. A possible solution to this problem is developing quality measures that can numerically express the perceived visual difference between the dithered image and the original one. The ideal target would be to include several objective aspects of image quality within one single, robust and reliable measure, so to evaluate and improve dithering algorithms.

We cans also describe digital halftoning as an artifact minimization problem, because we search for the dithered image that minimizes the measured visibility of artifacts. We therefore need a computational model that allows us to compute the visible error, so to automatically rank images in an optimization algorithm. Nevertheless, measuring in a quantitative way the visual quality of a dithered image is still a challenging task.

Objective quality measures can be divided in two classes: mathematical/statistical ones and human visual system (HVS) based.

The first ones could be used to evaluate dithered patterns of constant grey values. These metrics provide a great understanding of the possible relationships for a given point distribution. These kinds of measures are usually independent of viewing conditions or observer, easy to compute thanks to the low computational complexity. It is possible to find in literature different examples of this kind of measures, such as: Structural Similarity (SSIM) [6], Peak Signal-to-Noise Ratio (PSNR) [7] and Mean Squared Error (MSE) [8]; nevertheless, none of these measures consider the presence of artifacts as explained in [9, 10].

The second one tries to model the perceived visual quality and predict artifacts generated by the dithering procedure. The ideal halftone algorithm is able to minimize this visibility. Since the eye detects more distortions at certain spatial frequencies, the development of a halftone visibility metric is based on a model of the HVS.

HVS models are relatively simple and have proven to be quite successful when applied to algorithms that seek the best possible distribution scheme. All digital halftone techniques, whether they are based on screening algorithms, error scattering or iterative halftone, all adopt these models, either explicitly or implicitly. In fact, even those methods that are categorized as model-free because they do not explicitly include an HVS block in their block diagram, still replace it with a low pass filter. Moreover, not only is an HVS model fundamental to the design of most halftone techniques, but additionally, the shape of the HVS model can be modified to achieve increased texture quality in the dithering result.

Therefore, the performance of a certain halftone algorithm can be maximized by adequately designing improved HVS models.

Digital halftoning usually refers to methods and algorithms able

to convert continuous-tone images to binary ones, for displaying them in bilevel devices such as printers, both ink jet and laser ones. Within the computer industry, the market demand for fast, high resolution printing devices has recently escalated. Halftoning's contribution to this industry has been essential to its success. For this reason investigating image quality, and in particular finding an objective measure to evaluate the results is of fundamental importance.

2.1 A Quantitative Measure

As discussed, HVS models use human visual selectivity and sensitivity to model and improve the perceived image quality. HVS is based on the psycho-physical process that relates psychological phenomena (contrast, brightness, etc.) with physical phenomena (light intensity, spatial frequency, wavelength, etc.). It determines which physical configurations give rise to a particular psychological sensation (perceptual, in this case).

The two dimensional Discrete Fourier Transform (DFT) is applied on the image and then multiplied using Hadamard product with a Contrast Sensitivity Function (CSF), so that a component of the image at a particular angular frequency is weighted by the CSF value at that frequency. The result of this computation is the two dimensional DFT of an image that would lead to the same psychological response when viewed from a visual system with a flat CSF to which the original image leads when viewed by HVS [11]. The weighted signal-to-noise ratio (WSNR) of an $M \times N$ pixel binary image (y), given the original (x) of the same size is computed as:

$$WSNR(dB) = 101 \log_{10}\left(\frac{\sum_{u,v} |X(u,v)C(u,v)|^2}{\sum_{u,v} |(X(u,v) - Y(u,v))\, C(u,v)|^2}\right) \qquad (1)$$

where $X(u,v)$, $Y(u,v)$ and $C(u,v)$ are correspondingly the DFT of the input image, of the output image and the CSF.

2.2 Stochastic version

As already described in the previous sections, halftoning converts the input image into a black and white version of it to be reproduced on a binary output device, such as an ink-jet printer, which can only choose whether to print a dot or not in each position. The human eye, behaving as a low-pass filter, blurs the dots and spaces together and creates the illusion of many continuous shades of gray tones. Depending on the specific way the dots are distributed, a display device can produce different degrees of image fidelity with more or less granularity. According to the HVS, isolated and randomly arranged dots, if distributed in a proper manner, should produce images with the highest quality, maintaining fine details and sharp edges. Nevertheless, some displays and printing devices are unable to reproduce isolated dots consistently in their entirety and, as a result, introduce printing artifacts that greatly degrade the aforementioned details that the computed dot distribution is designed to preserve. For this reason, many printing devices create periodic patterns of grouped dots, which are easier to produce on the printed page in a consistent manner. For this reason, in our halftone study, the main objective is to determine the optimal dot distribution for that HVS, and then produce these models computationally efficiently.

Since the first article in which Floyd and Steinberg's algorithm was mentioned, many changes to the original error spread algorithm have been introduced to avoid the undesirable artifacts present in the original algorithm. But while these modifications eliminate disturbing artifacts at certain levels of gray, many do so therefore at the expense of other levels. Jarvis [12] and Stucki [13], for example, introduced a different mask model and weights, called 12-element error filter, but the artifacts remain during the algorithm because of the fixed behavior manifested by these algorithms. Hence we realize that the problem is not related to the number of elements involved but rather to the determinism of the method. Other approaches suggested to change the scanning path in which pixels are processed, including the serpentine as the most trivial up to the space-filling curves like Hilbert's curves. Unfortunately, these methods still create strong periodic patterns. To mitigate the problem of artifact our idea is to transform the FS algorithm, a deterministic algorithm, into a Stochastic (SFS) version inspired by [14]. A crucial point of our new approach is the preservation of the average image. As in all halftone methods, where everything and only the error is spread. Even in our method, while modifying the weight matrix at each pixel, the average of the values of the original image is not changed.

In particular we choose a real number p and, for each pixel, we generate 2 random number:

$$r_1 \in [-5/16, 5/16] \text{ and } r_2 \in [-1/16, 1/16]$$

and we compute

$$\frac{7}{16} + p \cdot r_1, \frac{5}{16} - p \cdot r_1, \frac{3}{16} + p \cdot r_2 \text{ and } \frac{1}{16} - p \cdot r_2 \qquad (2)$$

as error diffusion coefficients. This method, as a modification of Ulichney's one[15], is a good approximation of a blue-noise process. Blue-noise halftoning is characterized by a distribution of binary pixels where the minority pixels are spread as homogeneously as possible. This method of pixel distribution creates a pattern that is isotropic (or radially symmetric), aperiodic and does not contain any low-frequency spectral components. From the considerations made about HVS, it's not surprising that blue-noise creates the visually optimal arrangement of dots. Unfortunately, blue-noise is not trivial or fast to generate [16], but whit our method this can be approximated.

2.3 Optimal choice of p

By introducing our new method we have underlined the need to choose a p parameter.

In our previous work the choice of p was totally empirical and not automated, that is, we tried different values of p in a predetermined range and picked the best according to the image quality measured by WSNR. Here we introduce a method for the automatic setting of p. As pointed out above, these algorithms are often used for printers in different areas, but often as an integral part of industrial processes. For this reason and for the sake of completeness, finding a fully automatic method would make the method more effective and usable on a large scale. There are two critical issues to be addressed: the choice of an appropriate interval and the need for an intervention by the operator for the choice. First of all, let's discuss the choice of the interval. As we have defined p there are no

limitations, it simply has to be a real value. Since it is not a probability, it can certainly assume both values greater than 1 and negative values. Conceptually there is a symmetry, at least on average, between negative and positive values so we can only take into account positive values without losing in the generality. On the other hand, choosing values much greater than 1 means to substantially change the coefficients of the matrix used for the diffusion of the errors, for example choosing $p = 2$ the values, on average, are doubled. In addition to these preliminary considerations we conducted several experiments aimed at understanding where the maximum value of WSNR was in relation to the p choice. As it can be seen in Figure 3 , the maximum is usually included in the $[0, 2]$ interval.

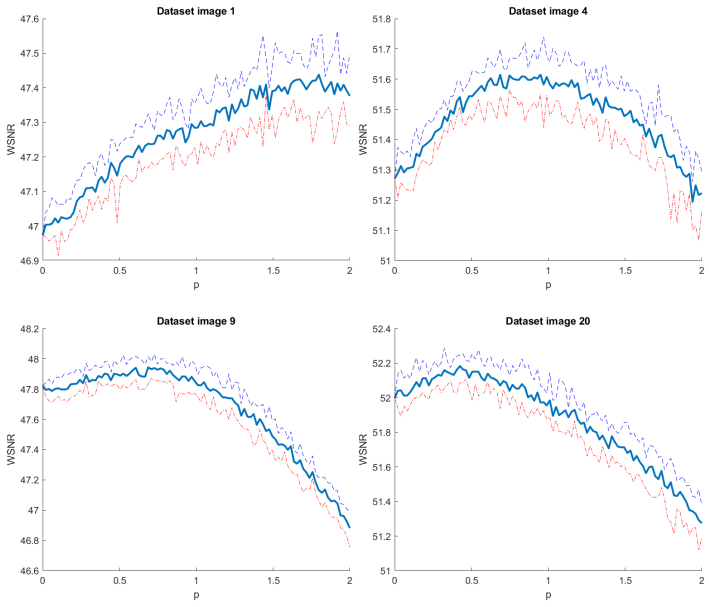


Figure 3: Sample of the algorithm behaviour for p in the interval $[0, 2]$ for some images of the dataset available in [2]. Due to the stochastic nature of the algorithm, we repeat the algorithm 10 times with different random seeds for each value of p, plotting minimum, average and maximum value for the resulting WSNR.

Once the range is fixed, we address the second issue by means of an optimization algorithm that provides the p value that maximizes the WSNR. The first idea, given the amount of references in the literature, was to use a first-order technique as a gradient method [17]. But we immediately noticed that, due to the lack of convexity of the function to be maximized, such methods are often not effective. There are certainly techniques to use gradient methods even in non-convex areas but they require at least the differentiability of the function. Since we have to maximize a stochastic process we see both the basic hypotheses of the gradient methods fall: convexity and differentiability. Fortunately, our implementation of the parallel algorithm on GPUs makes execution time particularly low, even for large images. This allows us to approach efficiently the problem in an empirical way without increasing the time excessively. Considering all these premises, our final idea is to calculate, first, the WSNR for different values of $p \in S_n$ where S_n is an equally spaced partition of the interval $[0, 2]$ with n elements. Due to the stochasticity of the method, for each $p \in S_n$ we consider the average over m runs.

Finally, choose the p with the highest average WSNR. We summarize our method in the equation 3 where m is the number of runs, I and $SFS(p)$ represent respectively the input image and the output image obtained from a single run of our stochastic dithering algorithm with parameter p.

$$p* = \arg\max_{p \in S_n} \frac{1}{m} \sum^{m} WSNR(I, SFS(p)) \tag{3}$$

We obtained the best results with the following configuration:

- Size of partition S_n, n = 100
- Number of repetitions for each p, m = 10

In this way, given any image in our experiments, the method lead to satisfying results in terms of similarity to the initial image.

3 GPU implementation

Dithering algorithm are infamously known for being "embarassingly" unparallel due to the inherently sequential nature of the dithering operation: the every pixel output depends on all the previously computed outputs. Achieving good speed ups in such problems is a challenging task and several work from the research community tried to overcome this issue, either proposing a "friendlier" dithering algorithm or a more efficient implementation. For example in [18]–[20] they introduce a block-based error diffusion algorithm which allows to simultaneously process blocks of pixels and produced good visual results, however is not clear which block dimension should be taken in order to get the best image quality; whereas in the work [21] the authors propose to parallelize over multiple pages in order to saturate the GPU utilization, in this work we focus on a single image case; in [22] they propose an efficient implementation that consists of a hybrid approach where CPU and GPU are both exploited but the CPU is used in the regions where GPU is expected to perform worse, in particular, towards image corners; finally [5] provided an optimal parallelization algorithm from a theoretical point of view, unfortunately, assigning each thread a diagonal portion of the image is currently not suited for general purpose devices as it causes more time spent for additional computational overhead than what parallel devices can manage to reduce.

In this work, we adopt the classical Floyd-Steinberg algorithm as it's still the dithering algorithm that yields the best result and propose an implementation that rely solely on GPUs. Graphics Processing Units (GPUs), in particular General Purpose GPUs, are massively parallel architectures that comprises thousands of cores designed to compute in an extremely efficient way repetitive and simple operations such as those involved in the rendering pipeline in Computer Graphics, hence "Graphics" in the name. Each core are less performing and simpler than state of art processors but the computational power in such devices lies in the number of cores rather than the capability of each single core. Nowadays GPUs are widely adopted for many tasks in data mining, physics simulation, medical imaging and machine learning, to name a few, hence the "general purpose" in the name. We propose an implementation [1] that

[1] https://git.hipert.unimore.it/hjiacheng/stochastic_floydsteinberg

is optimized for the NVIDIA GPUs, from which we will also get the results reported in section 4.1, and we program with the architecture's standard Application Programming Interface (API), called CUDA [23] developed by NVIDIA itself. Nonetheless, our implementation takes into account obstacles and design principles valid for almost all GPUs so it's not limited to NVIDIA architectures.

Our implementation consists of two parallelization levels we call stream level and pixel level. First of all we, describe the pixel level parallelism consists of a function, named Parallel-FS, that implements the Floyd-Steinberg on GPU and assumes the image being already stored in the GPU memory, this function will be then invoked by the stream level function.

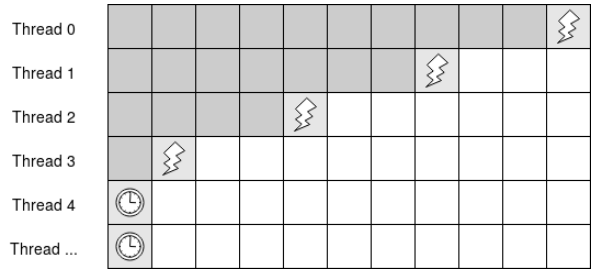


Figure 4: Each thread is assigned a row and waits for the previous one to be three pixels ahead instead of two. The thunder symbol on the diagonal represents the pixels processed simultaneously, whereas the clock indicates a waiting thread

In order to achieve good results one must take into account the of hardware on which the algorithm is executed. GPUs are SIMT (Single Instruction Multiple Thread) architectures partitioned in several Streaming Multiprocessors (SM), the latter are clusters of cores with multiple hardware schedulers for dispatching a fixed number of threads called warps. Each thread in the warp are committed to perform the same operation simultaneously, a feature called lock-step execution, and they are able to fetch multiple data at the same time if the addresses in the memory are aligned, a phenomena called "coalesced access". While these are good features in simple and regular memory accesses, it becomes actually a downside in case of Floyd-Steinberg because operations and accesses are neither simple nor regular. Dithering algorithm finds the closest color input and handle borders with many conditional, hence branching, instructions, which hurts the performance as the device must turn-off the portion of the threads not involved in the current branch because of the lock-step execution, moreover there's no coalesced access as the error diffusion scheme follows an irregular pattern leading to increased latency. This is the reason why the optimal theoretical algorithms presented in [5] cannot be implemented, nonetheless we try to approximately recreate the theoretical optimal situation in Parallel-FS. The function takes as input three data structures, an input buffer I_{buff} where the original image is stored, an output buffer O_{buff} where the result will be stored and error buffer E_{buff} for the temporary error diffusion values. Each row i is assigned a thread which, for each pixel (i, j), $j \in \{1, ..., I_{WIDTH}\}$ where I_{WIDTH} is the image width, reads from the input buffer and computes the output $O_{buff}(i, j)$ by finding the closest color according to the value $S = I_{buff}(i, j) + E_{buff}(i, j)$, and distribute the error $S - O_{buff}(i, j)$ to the neighbouring pixels according to the error diffusion scheme of Floyd-Steinberg, updating $E_{buff}(i, j+1)$, $E_{buff}(i+1, j)$, $E_{buff}(i+1, j-1)$ and $E_{buff}(i+1, j+1)$. The actual implementation carefully takes into consideration also borders compared to this simplified formulation. Each thread must synchronize with the previous thread in order to compute the assigned pixel result, in our implementation, in order to avoid update conflicts in the error buffer, which would require slow atomic operations, each thread waits for the previous one to be three pixels ahead with respect to his row instead of two, similarly to [21] (see Figure 4). The Figure 4 also illustrates how pixels are processed simultaneously along the diagonal (hence the name "pixel level parallelism").

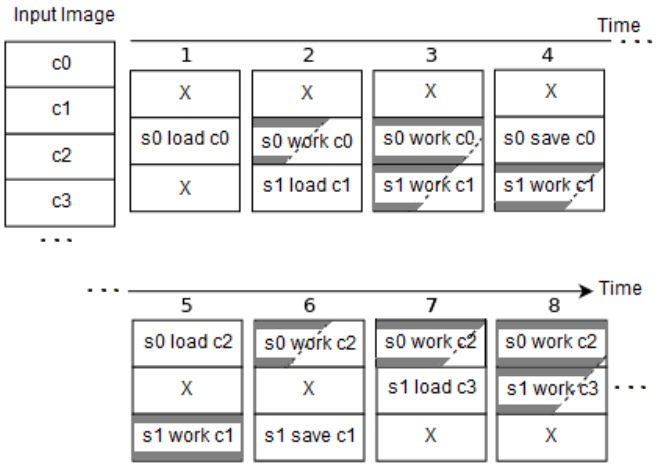


Figure 5: Visual illustration of the Double Buffering parallel execution scheme for the stream level parallelism, with two CUDA streams; sj and ci and denote CUDA stream respectively and image chunks. The "load" and "save" operations represent the data transfer from host to device input buffer I_{buff} and from device output buffer O_{buff} to host respectively and "work" consist of the invocation of Parallel-FS. It can be noted that, starting from time step 2, the memory transfer latency is completely hidden by GPU activity. "X" denotes the empty chunk, note that there is always an empty chunk because given $N_{STREAMS}$ there are $N_{STREAMS} + 1$ blocks in the data structure, so each iteration, every stream finds the results of the first row of the error buffer already calculated by the previous stream by simply sliding up.

The stream level parallelism addresses a problem more related to the physical devices limitation rather than the FS algorithm. In previous function we assumed the image to be already when the output is ready, the result is transferred back into host memory. Unfortunately, in most parallel implementations memory bandwidth represents one of the main bottleneck. The outer loop in our stored in GPU memory, however, for a task to be executed on GPU, every required data must be copied first from host memory (accessible by CPU only) to device memory (accessible by GPU only) and parallel implementation addresses this problem by hiding almost completely the memory transfer overhead between CPU and GPU wit the double buffering technique, which also removes the device memory size constraint for huge images.

Given an image of size $I_{HEIGHT} \times I_{WIDTH}$ and two input parameters N_{CHUNKS} and $N_{STREAMS}$, the first parameter define the number of chunks the image is divided into, whereas the latter represents the number of chunk to be processed simultaneously. First we define $Chunk_{HEIGHT} = \frac{I_{HEIGHT}}{N_{CHUNKS}}$ and we divide the image into N_{CHUNKS} of size $Chunk_{HEIGHT} \times N_{WIDTH}$, secondly we create three data struc-

ture of size $(Chunk_{HEIGHT} \cdot (N_{STREAMS} + 1)) \times I_{WIDTH}$ size for the input, output and error buffer respectively. Now, $N_{streams}$ CUDA stream are created, each of these concurrent execution stream are assigned an image chunk are responsible of computing the result by invoking the Parallel-FS function previously described on the assigned chunk. During the algorithm execution, every function invoked by each stream references to the same data structure but on a different memory address offset, the size of each buffer is $N_{STREAMS} + 1$ chunk blocks instead of $N_{STREAMS}$ because the the additional chunk allows to leave one chunk empty at every iterative step so Parallel-FS function avoids storing the error computed in the last row into an auxiliary buffer, minimizing the number of conditional operations. Because of the empty chunk, when a CUDA stream finishes the job, it can slide on the upper chunk in a toroidal fashion along the height, which means that the errors computed in the last row of the data structure are actually propagated in the first and the CUDA stream working in the first block moves to the last in the next outer loop iteration as it finds the error values ready to be used. The indexing and synchronization required to ensure the overall correctness of the FS algorithm details will not be discussed for the sake of simplicity. The final result is that when some CUDA streams are processing their chunk, others are transferring data from host to device memory and vice-versa, thus the memory bottleneck is hidden and the architecture is fully exploited. In Figure 5 we show the presented parallel execution scheme.

3.1 Theoretical improvement

Consider an image of size $P \times Q$, where P identifies the number of rows and Q the number of columns. We want to know the order of magnitude of the time that is theoretically needed to process the whole image. We use as a measure unit the time needed to process a single pixel.

In the serial version: $Time = P \times Q$ where $Time$ is the time for the whole process. In the parallel version, with the same image dimension, we have

$$Time = 3P + Q - 3 \text{ with } Q > P \tag{4}$$

and

$$Time = 3Q + P - 3 \text{ with } P > Q. \tag{5}$$

Considering, as said before, the time we use to process the single pixel we can assume to spend a M time to calculate the first row, but at this same time, always considering the parallel implementation, we have already calculated the triangle at the top left as shown in Figure 4. Going on to finish the second line, we only need to calculate another 3 pixels and we can proceed with a similar reasoning for all the rows in the image. So we have $Q+3\times(P-1) = Q+3P-3$. Even in an image where the number of rows and columns is swapped we have the same runtime. For these considerations we can conclude that we pass from a time that grows as $O(P \times Q)$ with the input size, in the serial case, to a time that grows linearly $O(P + Q)$ in the parallel case, all this from a theoretical point of view. This allows us to have a great saving in computational time, especially in industrial contexts where the images are of huge dimensions. This means an incredible acceleration, especially when the size of the image is large, as often happens in industrial cases.

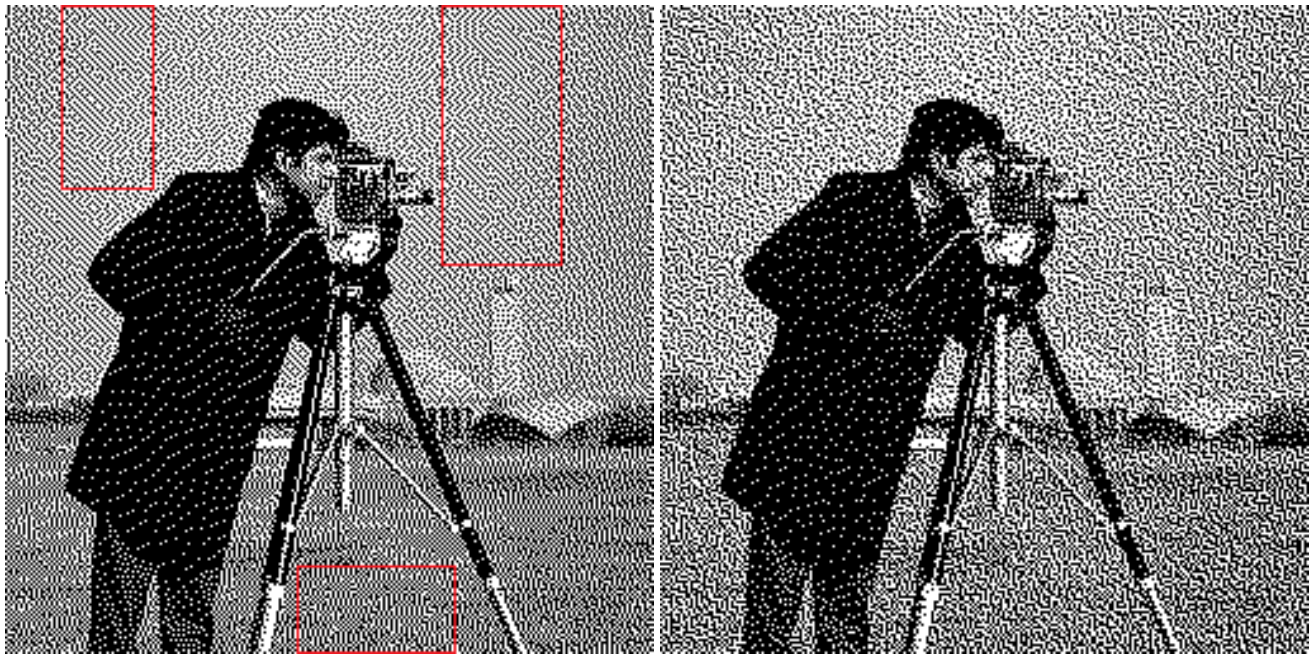


Figure 6: Cameraman image dithered with $p = 0$ on the left. The red rectangles highlight regions with visible artifacts. On the right, dithered image with optimal p value to maximize the WSNR value.

Figure 7: Lena image dithered with $p = 0$ on the left. The red rectangles highlight regions with visible artifacts. On the right, dithered image with optimal p value to maximize the WSNR value.

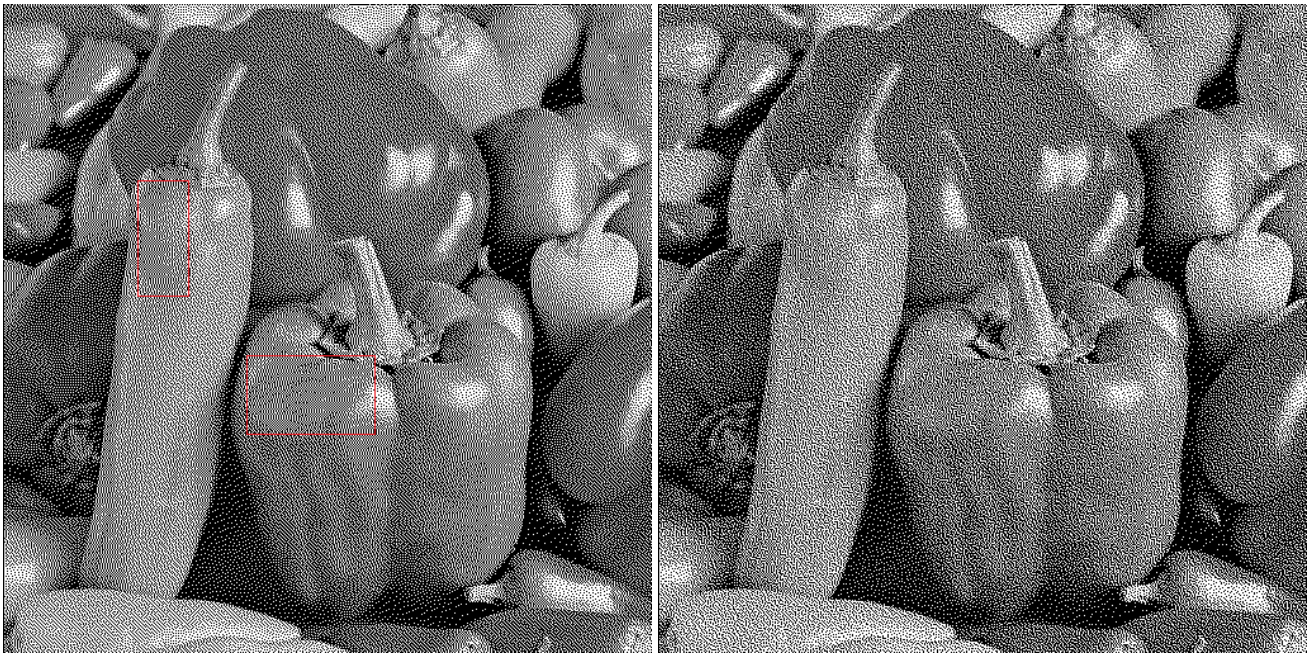


Figure 8: Peppers image dithered with $p = 0$ on the left. The red rectangles highlight regions with visible artifacts. On the right, dithered image with optimal p value to maximize the WSNR value.

4 Numerical experiments

The authors of [2] provided a sufficiently large dataset of images on which we applied our algorithm[2], the Digital Halftone Database (DHD). This dataset is composed of 196 reference images extracted from the CVG-UGR-Image database and the Genreal-100 dataset. In addition to these real images, we also assessed our approach on synthetic images. In particular, a particularly complicated image to process is the grayscale gradient, as in the Figure 10 or multi-tone chessboard. Moreover, the well known Lena, peppers and cameraman have been tested as well. Not all the images have same values of width and height since the algorithm can be applied to rectangular pictures as well. In the first experiments, we chose synthetic images. To verify the improvement with respect to the presence of artifacts,

[2] https://sites.google.com/view/sankarasrinivasan-s/research-publication/DHD-Database

we applied our algorithm to images composed of all shades of the (0 – 255) grayscale. This was done because some tones are more susceptible than others to artifact forming. In a second experiment we applied the new techniques to images often used in published works, in order to compare the results with the literature. Finally we compare against the results of [2], that are provided as results images in the same dataset. Obviously it can be observed that there is a considerable difference between the proposed stochastic approach and the simple FS. The experiments were performed on a desktop computer equipped with a CPU and GPU with the following specifications: Intel i7 7700K CPU, NVIDIA GeForce GTX 1080 GPU and 32GB RAM.

4.1 Results

With respect to image quality, we computed the measure described in Sec. 2.1 after applying FS, SFS and, for completeness, other dithering algorithms known in literature: Stucki[13], SIERRA[24] and Jarvis, Judice, and Ninke [12]. The results for the different images in Figure 10 (top row) and Figure 6, 7 and 8 are shown in Table 1 were we highlighted in bold the best WSNR values. The charts in Figure 13 show the behaviour of the WSNR value with respect to p for these images. Furthermore, we provided additional visual comparisons in Figures 11 and 12.

We compared our algorithm to the error diffusion ones presented in [2] on the same DHD dataset previously mentioned. We computed WSNR from the output images that are made available in the dataset and compared against our approach: over 196 different images we achieved the maximum WSNR value amongst all the algorithms 71% of the times, while the remaining 29% has been topped by the Filter Lite Dot Diffusion algorithm from [25].

In Table 2 we report the values of WSNR for some images of the dataset. We have to remark that, due to the different output size of the image, for some pictures we were unable to compute the WSNR value of comparison. The algorithms that are taken into account here are our FSF, Atkinson (ADD), Burkes (BDD), Fan (FDD), tradirional Floyd Steinberg (FSDD), Frankie Serra (FSIDD), Filter Lite (FLDD), Jarvis Judice and Ninke (JJNDD), Jarvis (JDD), Stucki (SDD) and Shio Fan (SFDD).

In addition to providing a theoretical discussion, we tested the different versions of the algorithm, serial and parallel, on images of different sizes. As we expected we obtained very good results: while serial time increases quadratically, parallel time increases only linearly. Let's better illustrate the trends in Figure 9.

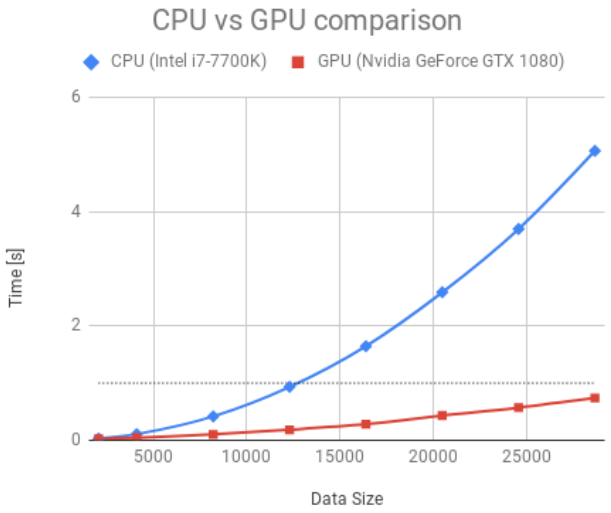


Figure 9: Comparison between the two versions: serial and parallel. Time in seconds on vertical axis, horizontally there is side dimension in pixels of a square image. In dotted line: all parallel execution time are less than one second.

Table 1: Dithering methods WSNR comparison

Image	FS	SFS			Stucki	Sierra	JJN
		$p = 0.25$	$p = 0.5$	$p = 0.75$			
Gradient	52.58	53.15	**53.34**	53.28	50.49	50.64	50.28
Lena	48.51	48.48	**48.57**	48.51	41.79	41.58	41.08
Cameramen	45.78	46.02	46.30	**46.43**	39.29	39.13	38.61
Pepper	46.13	46.39	46.93	**47.27**	39.45	39.13	38.55

Table 2: Error Diffusion WSNR comparison

Index	***FSF***	***ADD***	***BDD***	***FDD***	***FSDD***	***FSIDD***	***FLDD***	***JJNDD***	***JDD***	***SDD***	**SFDD**
1	47.50	20.00	43.61	44.33	48.23	40.90	**49.74**	26.35	40.44	41.44	44.48
3	**50.64**	24.09	44.40	44.22	46.92	41.28	47.99	28.72	40.79	41.51	44.63
5	45.49	18.80	41.72	41.08	44.91	38.80	**46.15**	23.56	38.27	39.11	41.28
7	**49.77**	20.99	43.65	45.49	47.57	41.22	49.57	27.41	40.73	41.60	45.97
17	**44.97**	15.81	39.52	40.40	42.48	36.30	43.72	20.87	35.80	36.50	40.79

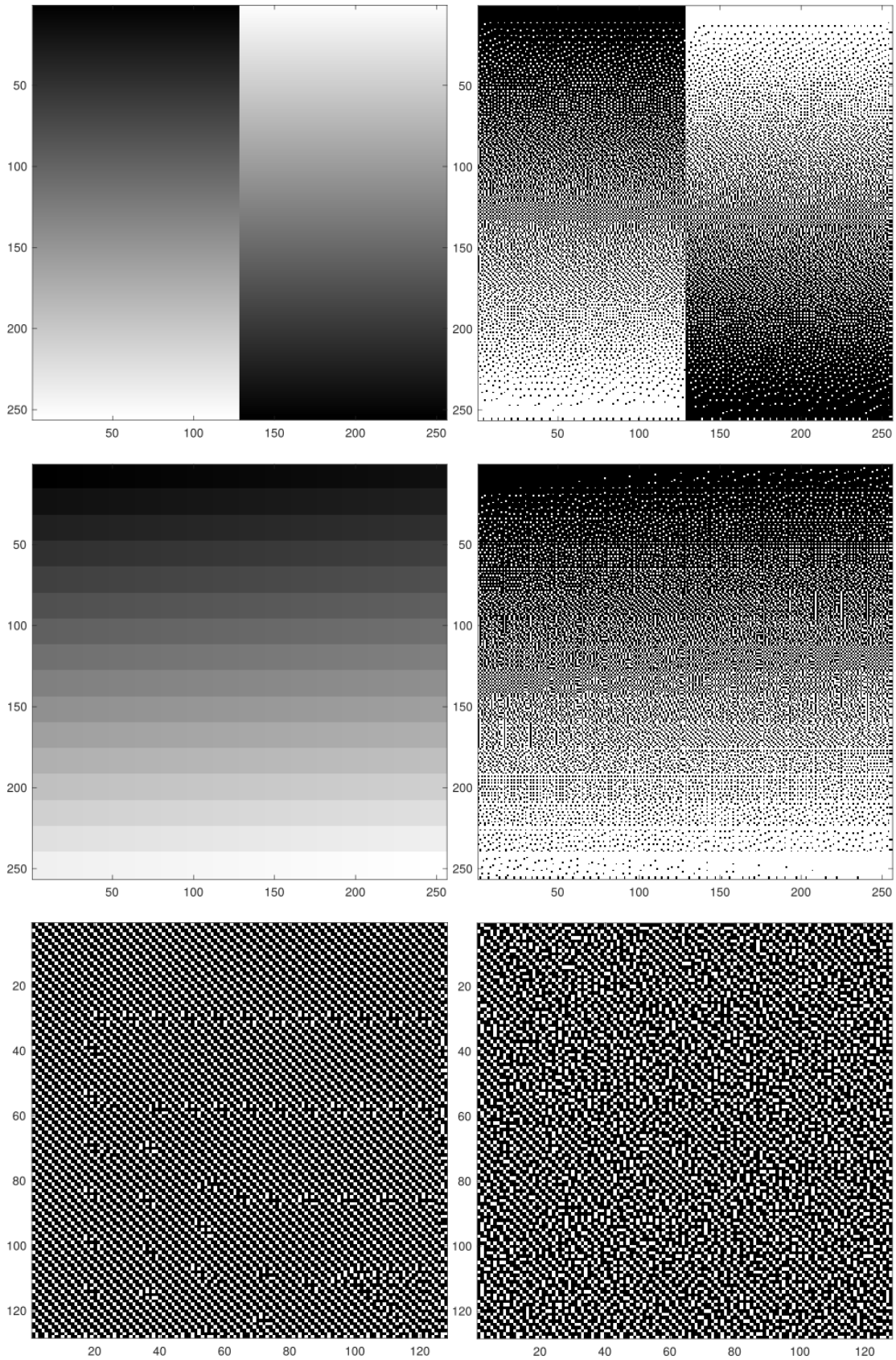


Figure 10: Inverted gradient pattern, original image (top-left) and halftoned using traditional FS (top-right): emerging artifacts are clearly discernible. In the middle row, 256 squares of all possible 1 Byte quantization values (left) ant their dithered version (right) Bottom line, particular of a single tone (value 128): on the left side artifacts arise distinctly with traditional FS, while on the right side the stochastic nature of the method allows no emerging pattern.

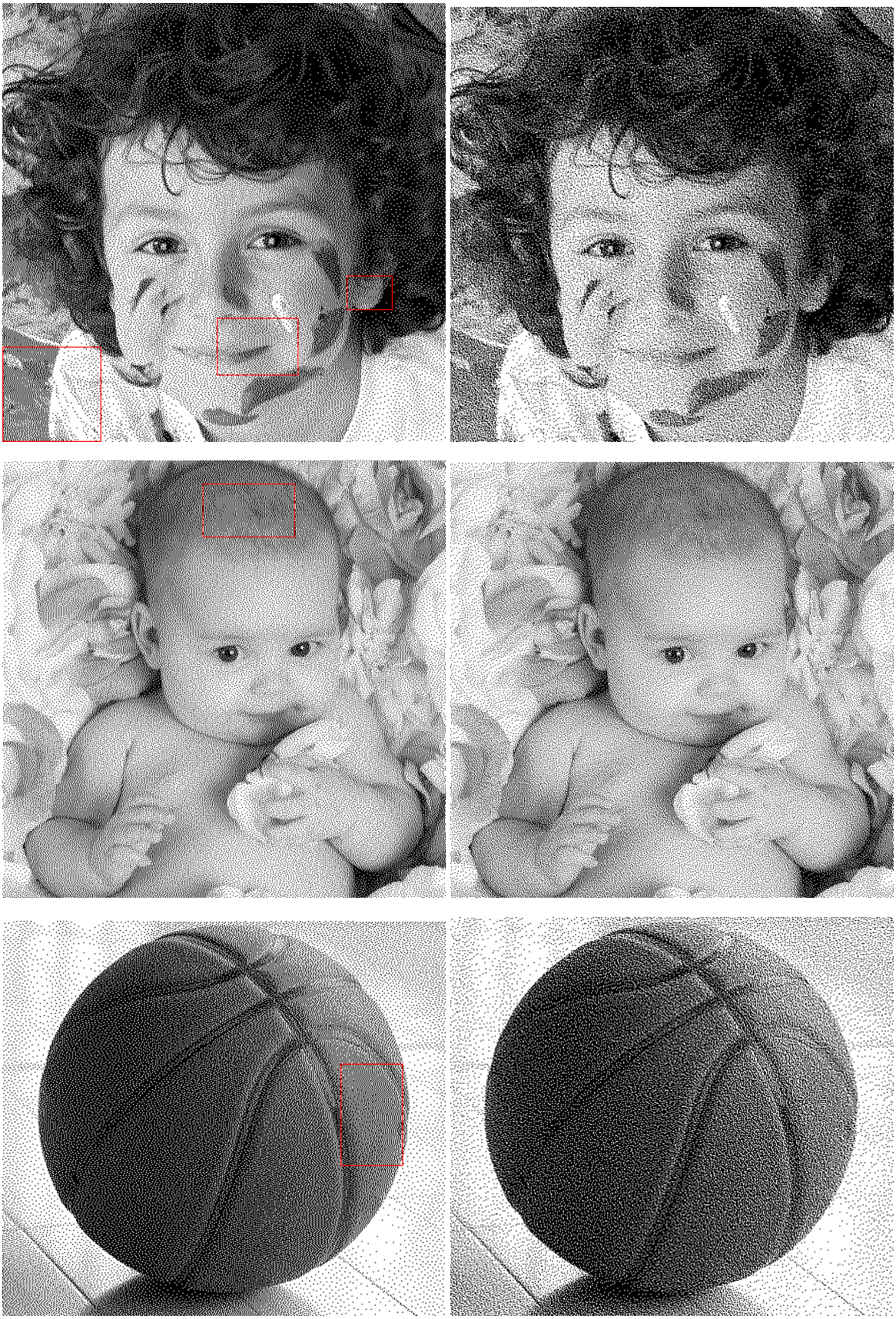


Figure 11: On the left: dithered images using FLDD for images 24, 4 and 6 of the Digital Halftone Database. The red rectangles highlight regions with visible artifacts. On the right, same images dithered using SFS with best p*.

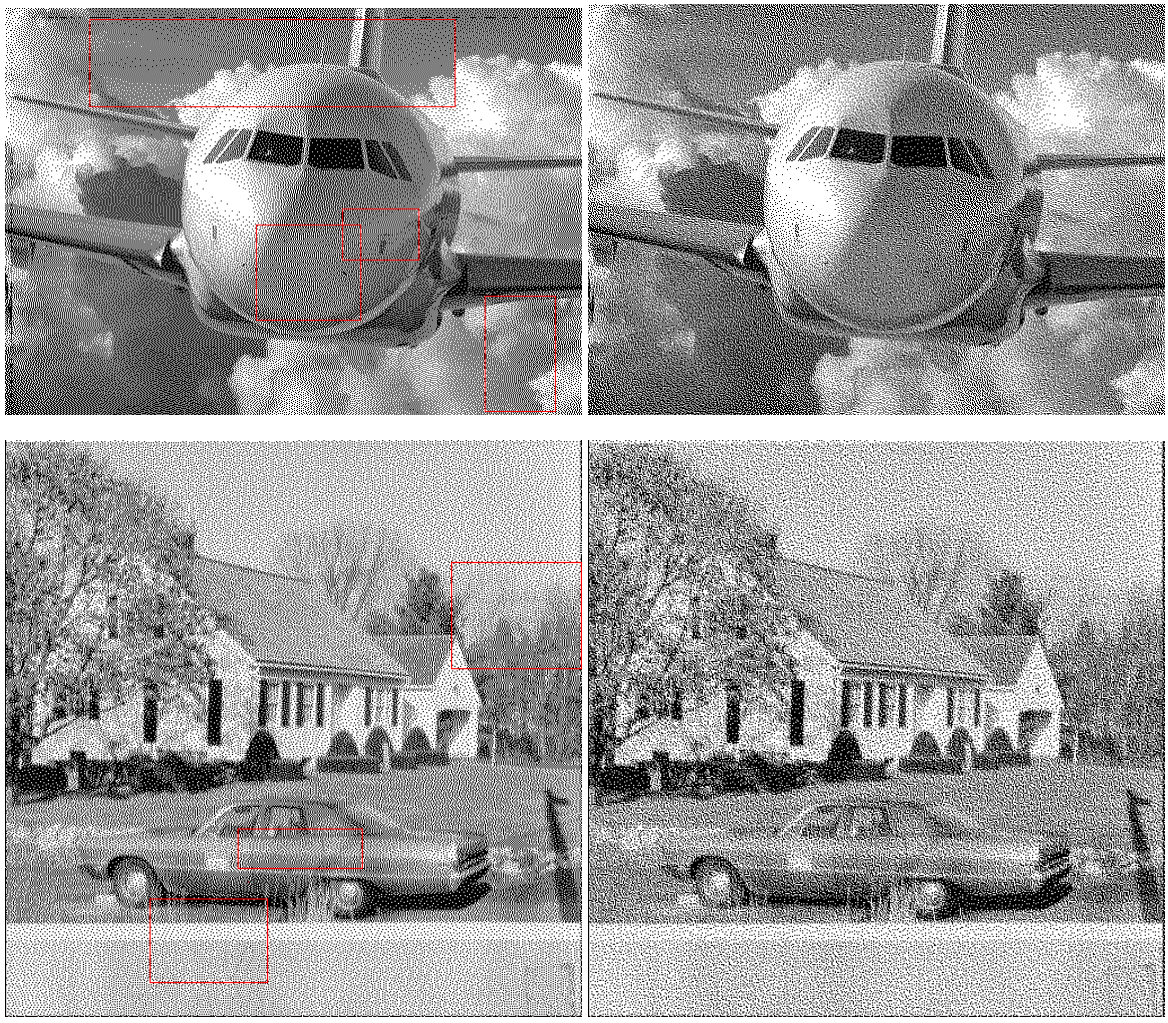


Figure 12: On the left: dithered images using FLDD for images 26 and 7 of the Digital Halftone Database. The red rectangles highlight regions with visible artifacts. On the right: same images dithered using SFS with best p*.

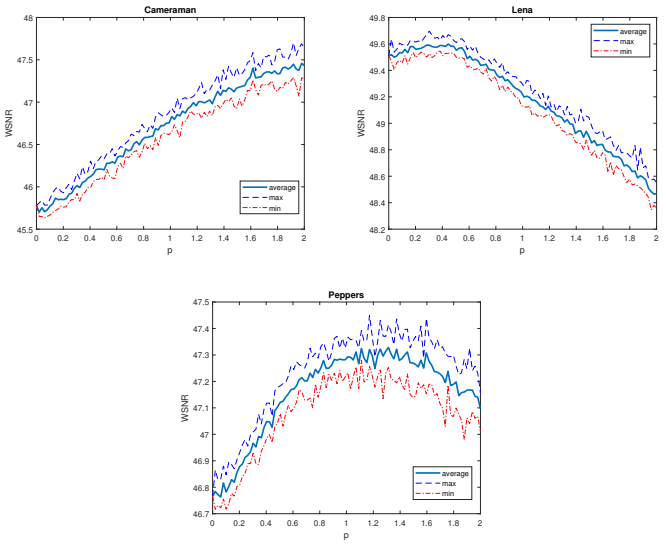


Figure 13: WSNR computed for our Stochastic Floyd-Steinberg algorithm with $p \in [0, 2]$ on the 3 well known images of Cameraman, Lena and Peppers. We proceed as for Figure 3.

5 Conclusions

In conclusion, the new algorithm we presented solves both the problems highlighted at the same time: the presence of artifacts and the slowness of execution. In particular, the transformation of the algorithm from deterministic to stochastic, with the introduction of a white noise, avoids the emergence of artifacts. The improvement can be evaluated through a specific measure that takes into account human perception called WSNR. Moreover, there is no increase in computational time; on the contrary, the parallel version we propose shows an excellent speed-up as the image size increases; which, especially in industrial applications, can be very large. Thanks to our parallel implementation, which guarantees speed to the method, we were able to fully automate the technique. So given an image we are capable of applying the best amount of stochasticity to the algorithm, thus obtaining an image as free of artifacts as possible. All the research part of the optimal parameter is done before the actual image printing process. This guarantees a method in which the operator does not have to make a decision and the printer is not occupied for a long time. In conclusion, the method is totally automatic, fast and guarantees a final image as close as possible to the initial one and suitable for printing. As future work some of the

possible extensions could be: more sophisticated methods to find optimal p, developing other error measures that take into account artifacts or geometric textures of different kinds. Another idea is to evaluate the 4, 8 and 16 gray tones versions of the dithering algorithm by proposing a technique that finds the best trade-off between image quality and number of tones.

References

[1] R. W. Floyd, L. S. Steinberg, "An adaptive algorithm for spatial gray scale," in Proceedings for the Society for Information Display, 75–77, 1976.

[2] J. M. Guo, S. Sankarasrinivasan, "Digital Halftone Database (DHD): A Comprehensive Analysis on Halftone Types," in 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 1091–1099, IEEE, 2018.

[3] Zhigang Fan, R. Eschbach, "Limit cycle behavior of error diffusion," in Proceedings of 1st International Conference on Image Processing, 2, 1041–1045, 1994, doi:10.1109/ICIP.1994.413514.

[4] M. Pedersen, F. Albregtsen, J. Hardeberg, "Detection of worms in error diffusion halftoning," Proceedings of SPIE - The International Society for Optical Engineering, **7242**, 2009, doi:10.1117/12.805555.

[5] P. T. Metaxas, "Optimal Parallel Error-Diffusion Dithering," in Proceedings of SPIE - The International Society for Optical Engineering, volume 3648, 485–494, doi:10.1117/12.334593.

[6] P. Ndajah, H. Kikuchi, M. Yukawa, H. Watanabe, S. Muramatsu, "SSIM image quality metric for denoised images," in Proc. 3rd WSEAS Int. Conf. on Visualization, Imaging and Simulation, 53–58, 2010.

[7] A. Horé, D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," in 2010 20th International Conference on Pattern Recognition, 2366–2369, 2010, doi:10.1109/ICPR.2010.579.

[8] B. Girod, "Psychovisual Aspects Of Image Processing: What's Wrong With Mean Squared Error?" in Proceedings of the Seventh Workshop on Multidimensional Signal Processing, 207–220, 1991, doi:10.1109/MDSP.1991.639240.

[9] T. Mitsa, K. L. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms," in 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, 5, 301–304, 1993, doi:10.1109/ICASSP.1993.319807.

[10] D. R. Mehra, "Estimation of the Image Quality under Different Distortions," International Journal Of Engineering And Computer Science, **8**, 17291–17296, 2016, doi:10.18535/ijecs/v5i7.20.

[11] R. Näsänen, "Visibility of halftone dot textures," IEEE Transactions on Systems, Man, and Cybernetics, **14**(6), 920–924, 1984, doi:10.1109/TSMC.1984.6313320.

[12] J. F. Jarvis, C. N. Judice, W. Ninke, "A survey of techniques for the display of continuous tone pictures on bilevel displays," Computer graphics and image processing, **5**(1), 13–40, 1976, doi:https://doi.org/10.1016/S0146-664X(76)80003-2.

[13] P. Stucki, "Image processing for document reproduction," in Advances in digital image processing, 177–218, Springer, 1979.

[14] R. Ulichney, Digital Halftoning, MIT Press, Cambridge, MA, USA, 1987.

[15] R. Ulichney, Digital halftoning, MIT press, 1987.

[16] I. Georgiev, M. Fajardo, "Blue-noise dithered sampling," in ACM SIGGRAPH 2016 Talks, 1–1, 2016.

[17] D. P. Bertsekas, J. N. Tsitsiklis, "Gradient convergence in gradient methods with errors," SIAM Journal on Optimization, **10**(3), 627–642, 2000, doi:10.1137/S1052623497331063.

[18] Y. Zhang, J. L. Recker, R. Ulichney, I. Tastl, J. D. Owens, "Plane-dependent error diffusion on a GPU," in Proceedings of SPIE - The International Society for Optical Engineering, 2012, doi:10.1117/12.906966.

[19] Y. Zhang, J. L. Recker, R. Ulichney, G. B. Beretta, I. Tastl, I.-J. Lin, J. D. Owens, "A parallel error diffusion implementation on a GPU," in Proceedings of SPIE - The International Society for Optical Engineering, 2011, doi:10.1117/12.872616.

[20] P. Li, J. P. Allebach, "Block interlaced pinwheel error diffusion," Journal of Electronic Imaging, **14**(2), 1–13, 2005, doi:10.1117/1.1900136.

[21] B. Trager, C. W. Wu, M. Stanich, K. Chandu, "GPU-enabled parallel processing for image halftoning applications," in IEEE International Symposium of Circuits and Systems (ISCAS), 1528–1531, 2011, doi:10.1109/ISCAS.2011.5937866.

[22] A. Deshpande, I. Misra, P. J. Narayanan, "Hybrid implementation of error diffusion dithering," in 18th International Conference on High Performance Computing, 1–10, 2011, doi:10.1109/HiPC.2011.6152714.

[23] Nvidia, "CUDA Programming Guide Version 10.0," .

[24] T. Helland, "Image Dithering: Eleven Algorithms and Source," http://www.tannerhelland.com/4660/dithering-eleven-algorithms-source-code/

[25] C. Labs, "Libcaca study: the science behind colour ASCII art," `http://caca.zoy.org/wiki/libcaca/study/introduction` [Accessed: Jul. 20, 2020].

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 664-675 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Improved System Based on ANFIS for Determining the Degree of Polymerization

Marcel Nicola[1,*], Marian Duță[1], Maria-Cristina Nițu[1], Ancuța-Mihaela Aciu[1], Claudiu-Ionel Nicola[1,2]

[1]*Research and Development Department, National Institute for Research, Development and Testing in Electrical Engineering – ICMET, Craiova, 200746, Romania*

[2]*Department of Automatic Control and Electronics, University of Craiova, Craiova, 200585, Romania*

ARTICLE INFO

Article history:
Received: 20 August, 2020
Accepted: 06 November, 2020
Online: 24 November, 2020

Keywords:
Degree of polymerization
2-furfuraldehyde
Fuzzy logic
ANFIS
Neural network

ABSTRACT

Transformer health assessment techniques, based on applicable standards, such as the dissolved gas analysis (DGA), through laboratory testing or online monitoring are used to analyze the symptoms of a failure which develops in transformers from an early stage. The DGA from a sample of dielectric liquid taken from the main tank generates information on the state of degradation of the active part. It was found that of all the furan derivatives which result from the degradation of insulation, 2-furfuraldehyde (2-FAL) is the only derivative which dissipates in large quantities in oil. Because of this, and due to its thermal stability as compared to other derivatives, 2-FAL is the best unit of measurement to determine and monitor the degree of polymerization (DP) of insulation. The poor accessibility of paper samples has led to difficulties in testing the ageing state of paper directly by measuring the tensile strength and the DP. It was developed methods to indirectly assess the ageing state of paper, by means of the chemical markers in oil which are associated with paper ageing. This article presents a method to determine the DP of solid insulation in transformers, which provides a faster and more accurate interpretation, as compared to the classical ones. The 2-FAL data resulting from the lab are recorded in a MySQL database, which is embedded in an intelligent system for diagnosis of DP (ISDDP) based on Adaptive Neuro-Fuzzy Inference System (ANFIS), generating at the output the report with the interpretation of the faults as word files. The automation of the DP diagnostic process will be achieved, allowing the operator to make timely decisions, to avoid any possible damages.

1. Introduction

Transformers are expensive pieces of equipment that play a significant role in electricity transmission and distribution systems. Although transformers are generally robust pieces of equipment, they also undergo failures, and there are numerous degradation mechanisms operating in components and subsystems that will limit the useful life. It is important that transformer users and distribution system administrators are properly trained to assess the condition of a group of transformers in operation, in order to take the best decisions at critical moments of possible damages, for the operations of minor faults repairing, their rehabilitation or replacement [1-4].

The assessment of the condition of the operating transformer consists in identifying the indicators that can provide accurate data on the extent of degradation of transformer components and subcomponents. The detailed health assessment of a power transformer requires extensive testing (electrical, as well as physico-chemical tests of the dielectric fluid), inspection of dataset from acquisition of other relevant data (operational history, as well as maintenance, including data of putting into service). These data are then analyzed by an expert based on the relevant standards and expert knowledge [5-8].

Most common transformer health assessment techniques apply only to the active part of the transformer. The DGA from a sample of dielectric liquid taken from the main tank generates information on the state of degradation of the active part. However, all components and subcomponents of a transformer should be considered for a full assessment. Unexpected failures of the

* Corresponding Author: Marcel Nicola, ICMET Craiova, Romania, marcel_nicola@yahoo.com

bushings and the on-load tap-changer (OLTC) can often lead to a state of transformer failure or even to its removal from the system [9-17].

The assessment of the active part of the transformer (magnetic core and windings) can be performed as follows [6], [10-16]:

- The thermal assessment concerns the faster-than-anticipated aging due to excessive temperatures resulting from the hot spot of the winding, the hot spot in the core and the defects of paper-insulated hot metals;
- The dielectric assessment tracks the process of degradation of the paper-oil insulation system due to high thermal or electrical stress, oil-paper moisture, manufacturing defects and contamination with (semi) conducting particles (solid particles);
- The mechanical assessment concerns the disintegration of the structure in the core, the windings and the solid insulation, due to one of the many reasons, such as strong electromagnetic forces resulting from the flow of the short-circuit current through the windings; the mechanical deformation of the windings due to shocks during transport, installation or during system failures, etc; and the degradation of solid insulating materials.

The oil plays a key role as an insulating medium that acts as a barrier between the windings and the tank and acts as a "cooling agent" that dissipates the heat generated in the windings and the core. It is a conservation medium which protects the windings and the solid insulation against moisture and other physical contamination.

The molecular structure of oil is subjected to electrical, mechanical and thermal stress during transformer operation, which leads to a process of decomposition, and, as a result hydrocarbon gases are generated. These gases are found in the mass of the dielectric fluid and can be quantized.

Transformer health assessment techniques, based on applicable standards, such as the DGA, through laboratory testing or online monitoring are used to analyze the symptoms of a failure which develops in transformers from an early stage [12-14].

The complexity of real world problems, the need to develop complex software systems that adapt to a dynamic environment such as the real world lead to growing interest in the use of computational intelligence techniques to develop specific applications. Computational intelligence is a subdomain of the artificial intelligence that deals with the design, application and development of computational paradigms inspired for example by biological, social and linguistic areas. The techniques covered by the computational intelligence are the fuzzy logic, the artificial neural networks and the evolutionary computation. These are mainly fuzzy systems, neural networks, cognitive systems, nature-inspired systems and hybrid intelligent systems containing these paradigms. The most common of these techniques used for the diagnostic of power transformers are the fuzzy logic and the artificial neural networks [18-27].

This article presents a method to determine the DP of solid insulation in transformers, which provides a faster and more accurate interpretation, as compared to the classical methods. The 2-FAL data resulting from the lab are recorded in a MySQL database through a client software application for writing to the database and then read through a database read client, which is embedded in the diagnostic system, generating at the output the report with the interpretation of the faults as word files. Thus, the automation of the DP diagnostic process will be achieved, allowing the operator to make timely decisions, to avoid any possible damages.

The proposed ISDDP to determine the DP of the insulation of power transformer windings based on the content of the 2-FAL derivative uses the fuzzy logic, and, for the improvement of the system, the fault learning facility is used, based on the artificial neural networks.

This article is an extension of [28] and is structured as follows: Section 2 presents the chemical markers used to determine the DP. The classical methods for the determination of the DP are presented in Section 3. Section 4 presents the implementation as fuzzy logic of the diagnostic system in the LabVIEW graphical programming environment, as well as its improvement by using artificial neural networks through an adaptive neuro-fuzzy inference system. Section 5 presents conclusions and new pursuits in future work.

2. Chemical markers used for the determination of the DP

When paper insulation ages, its electrical properties do not change significantly, but the mechanical properties are considerably diminished, due to cellulose degradation (depolymerization). During normal operation, paper cellulose will split according to the general condition of the transformer (temperature, moisture and acids). As a result, the tensile strength of the paper is reduced over time, and therefore it is important to monitor the condition of the insulation [5-8].

The poor accessibility of paper samples has led to difficulties in testing the ageing state of paper directly by measuring the tensile strength and the DP. Thus, the researchers developed methods to indirectly assess the ageing state of paper, by means of the chemical markers in oil which are associated with paper ageing.

2.1. Carbon oxides (CO and CO2)

The first studies covered the relation between the amount of carbon oxides, carbon monoxide (CO) and carbon dioxides (CO2), in oil and the DP of the insulating paper. Carbon oxides, CO2 and CO, are determined based on the DGA by applying the gas chromatography technique. As a result of thermal stresses, these gases are generated by the degradation of both paper and other parts of the transformer. In free breathing transformers, the generation of carbon oxides is influenced by free air ingress [12].

In [21], it is presented a relationship between the amount of carbon oxides, CO and CO2, dissolved in transformer oil and the DP of the insulating paper. In [17] is demonstrated that water and CO2 are the main byproducts of the thermal degradation of cellulose. Therefore, the CO2/CO ratio is normally utilized as an indicator of the thermal decomposition of cellulose [8, 9].

According to [18], the CO2/CO ratio is normally when its value is above seven, while the respective values of CO2 and CO must be greater than 5000 ppm and 500 ppm to improve the certainty factor.

In accordance with [19], the CO2/CO ratio below 3 is expected to determine a severe degradation of the paper. When the ratio

exceeds 10, it may indicate a slight overheat of 160°C. It was noted that in order to obtain safe results, correction of CO2 absorption from the atmosphere and correction of CO2 and CO background are necessary.

According to [17], high paper degradation rates are reported when the CO2/CO ratio drops below 6. However, the application of this ratio in determining the condition of the cellulose insulation is not suitable due to the long-term oxidation of the oil or as a result of free air ingress due to leakage [24]. The following additional tests are proposed, as solutions to solve this problem: the analysis of furan compounds; the determination of the DP of paper using the paper samples, to analyze the cellulose insulation.

2.2. Furan compounds

Oxidation, hydrolysis and pyrolysis of cellulose generate a group of chemical compounds which are used as a tool to analyze the condition of the cellulose insulation. Due to their solubility in oil, these chemical compounds (2-furfural; 2-acetylfuran; 5-methyl-2-furfural; 2-furfuryl alcohol; 5-hydroxymethyl-2-furfural) provide key information necessary for the diagnosis of the health of solid insulation. The limitations and applicability of the furan compounds as tools for diagnosing the state of the degradation of cellulose insulation in oil-filled transformers were constant in [6] and [7]. Studies have shown that of the five furan compounds specified in [9], 2-FAL is the most stable, by providing relevant information on the state of degradation of the solid insulation.

Their main conclusion indicates that, of the five furan compounds identified in [9], 2-FAL provides more relevant information on the paper degradation. 2-FAL has been shown to have a high degree of generation and better stability than the other furan compounds and can be extracted from oil and used to characterize the thermal decomposition of the electroinsulating paper. A high concentration of 2-FAL indicates a high level of paper degradation. The constructive differences between transformers, the thermal and electrical stresses limit the setting of typical threshold values for 2-FAL, thus hindering the diagnosis of the condition of the cellulose insulation [16], [17], [20].

Some noted disadvantages of 2-FAL, such as its low detection threshold, when the paper is thermally upgraded (TUK). TUK paper is frequently used for new transformers, has a high rate of 2-FAL generation from hemicellulose, which shows an exponential increase, so the applicability of 2-FAL as a chemical marker for paper degradation must be considered prospectively.

3. Classical methods for the determination of the DP

Cellulose insulating paper can be damaged due to the effects of hydrolysis, pyrolysis and oxidation that occur in the transformer during operation. Due to the deterioration of the cellulose insulation, the group of furan compounds appears [6].

The cellulose insulation in transformers consists of cellulose, hemicellulose, lignin and mineral substances, and their degradation results in furans (see Figure 1) [9]. The generation of furans can also be caused by pentoses. As a common point of the degradation processes, the final result is the 2-FAL [17], [20].

Following the degradation processes of solid insulation that take place in the transformer, 2-FAL has the ability to dissipate in oil and due to its stability is an analysis tool for determining and monitoring DP of cellulose insulation. The other furan compounds can be used as markers to identify the process that causes the degradation of solid insulation [17], [22].

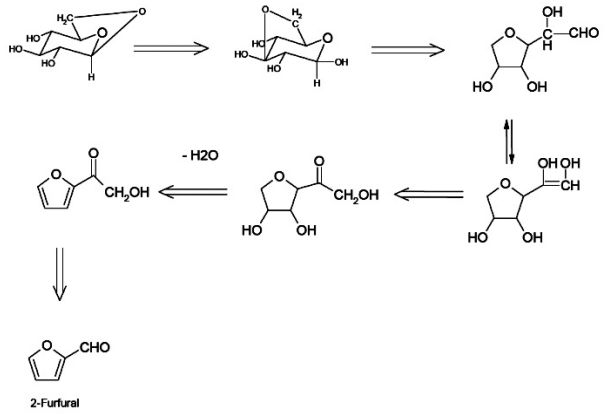


Figure 1: Determination of 2-FAL from lignin

The determination of the content of furan compounds, especially 2-FAL, was given a certain use for the diagnostic of transformers. It is believed that there is a certain correlation between the degree of polymerization of paper and the content of 2-FAL (or total furan compounds) of oil. Several researchers [25] have developed various relations for the determination of the DP, depending on 2-FAL concentration, but Pahlanvanpour's relation [15] is the most commonly used, because, when developing the formula, he relied on the states of the transformer and an assumption that 20% of the winding paper and the inner layer of the paper degrades twice as fast as the rest of the paper.

In [17] is builted a degradation model based on experimental data and field measurements in a research program conducted for CIGRE (International Council on Large Electric Systems) and found that a furfural molecule will result for every three cellulose chain splits [9], [24]. The system proposed in this paper was designed to determine the DP based on the 2-FAL ranges proposed in [6, 7] as is presented in Table 1.

Table 1: The state of the transformer according to the furan content

2-FAL content	DDP
0-0.1	1200-700
0.1-1	700-450
1-10	450-250
>10	< 250

To determine the DP, the following mathematical equations were used, which relate the DP to the content of 2-FAL:

- Chendong equation [16], [26]:

$$DP = \frac{1.51 - \log_{10} F}{0.0035} \quad (1)$$

- Scholinik et al. equation [18] , [26]:

$$DP = \frac{1.17 - \log_{10} F}{0.00288} \quad (2)$$

- Pahlavanpour equation [15] , [26]:

$$DP = \frac{800}{(0.186 \cdot F) + 1} \quad (3)$$

- De Pablo equation [17] , [26]:

$$DP = \frac{7100}{8.88 + F} \quad (4)$$

where: F represents 2-FAL concentration.

4. The architecture of ISDDP

The fuzzy logic-based approach for solving the control problems is highly recommended in very complex, nonlinear systems with uncertainties for internal or input parameters. The fuzzy logic is an approximate method by which vague knowledge, stored in a base of rules can be formally modeled. The practical implementation of the fuzzy systems is based on the advantages which allow the human experience to be transposed for the creation of the inference rules by using linguistic variables. The efficiency of the fuzzy systems or the artificial neural networks work is demonstrated by the widespread use in various fields in recent years [12], [26-30].

The fuzzy control provides a method of designing control algorithms in a user-friendly manner and provides the ability to capture human non-linear control behavior that has been shown to be suitable in many complex adjustment problems. A controller design method close to human thinking and perception reduces the time of application development and requires less trained staff in the field of regulator design. The advantage of the fuzzy tuning is the ability to program the knowledge of process operators and engineers, consisting in process operation experience, in a manner which is easy to understand and friendly. The fuzzy systems mainly have the advantage of explaining behaviors based on a set of rules and therefore their performance and development can be adjusted by modifying them [28], [30].

The neural networks are utilized to adapt the membership functions of the fuzzy systems in the applications of automatic control of the dynamic processes. Although the fuzzy logic allows the coding of expert knowledge through linguistic terms, usually it is a difficult endeavor to design and adjust the membership functions which quantitatively describe these terms. The learning techniques of the neural networks can automate the process and can therefore substantially reduce the cost and the time for the application development, leading to increased performance. To prevent the problem of knowledge acquiring, the neural networks can be designed to automatically extract the set of fuzzy rules from the disponible numerical data. Other approaches involve the use of neural networks to optimize a number of certain parameters of the fuzzy systems or to preprocess input data into the fuzzy systems [29], [30].

The general scheme for the proposed system for diagnosis of DP is shown in Figure 2. The system is achieved by using the facilities provided by LabVIEW and Matlab/Simulink graphic development environments for the implementation based on the fuzzy logic and neural networks.

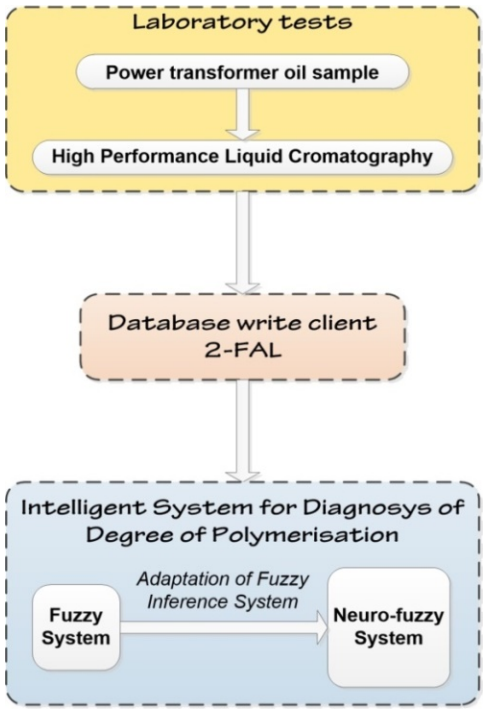


Figure 2: General scheme for the proposed system for diagnosis of DP

4.1. Fuzzy system for the determination of the DP

The system for the diagnostic of the DP of power transformers based on the fuzzy logic generates flexibility in interpreting the fault condition based on the 2-FAL data sets. The fuzzy logic diagnostic system is implemented in the Fuzzy System Designer, which is a software tool from LabVIEW graphical programming environment and basic model of the proposed fuzzy logic controller is shown in Figure 3. To overcome the difficulties in determining the boundary conditions for furan derivatives, the fuzzy logic technique is used, which allows the based rules to be implemented in a type of structure with natural language appropiate and widely used.

In the fuzzy logic, the domains correspond to the concepts of linguistic variables, so that such a variable corresponds to a domain of possible values. The input to the diagnostic system with fuzzy controller is the value of the 2-FAL derivative concentration, and the output is the result of DP value. For the 2-FAL input, the range of possible values (the set grading the membership) is between 0 and 20 and has as linguistic values the next form: {low; medium; high; very_high}. For the output, the DP is [0, 1200] and has as linguistic variables {healthy; mod_det; ext_det; end}, which represent: healthy, moderate degradation, extreme degradation, and end of life.

In order to achieve the best admissible response of the system, two types of fuzzy controllers are proposed to be used, and two types of membership functions are defined for these fuzzy controllers, namely the triangular and the trapezoidal membership functions, which correspond to the input and output of each of the two controllers. The result of the average of the two controllers of the system will be compared to the result of the fault values which is also implemented and calculated based on the classical methods presented in the equations (1) to (4).

The implementation of the two types of membership functions (triangular and trapezoidal) for the inputs and outputs of the two

logic fuzzy type controllers is shown in Figure 4 and Figure 5 respectively.

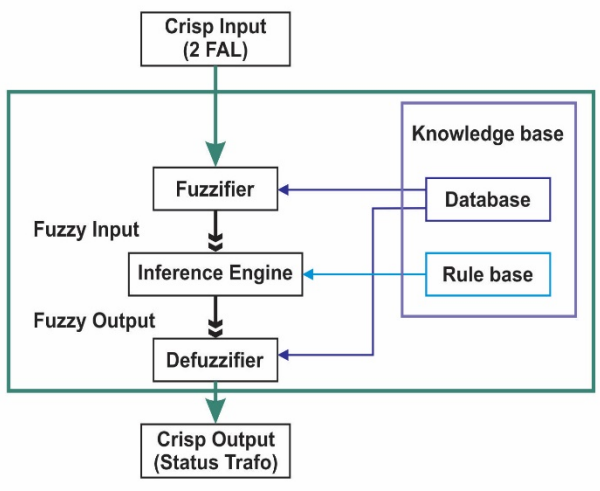


Figure 3: The proposed model of the fuzzy logic controller

Based on the 2-FAL ranges presented in Table 1, the implementation of the four triangular membership functions is carried out, and it is defined by the use of three parameters {a, b, c} covered by these ranges; the functions are described as follows:

$$triangle(x;a,b,c)=\begin{cases}0; & x\le a\\ \dfrac{x-a}{b-x}; & a\le x\le b\\ \dfrac{c-x}{c-b}; & b\le x\le c\\ 0; & c\le x\end{cases} \quad (5)$$

or, by using min and max functions:

$$triangle(x;a,b,c)=\max\left\{\min\left\{\frac{x-a}{b-a};\frac{c-x}{c-b}\right\};0\right\} \quad (6)$$

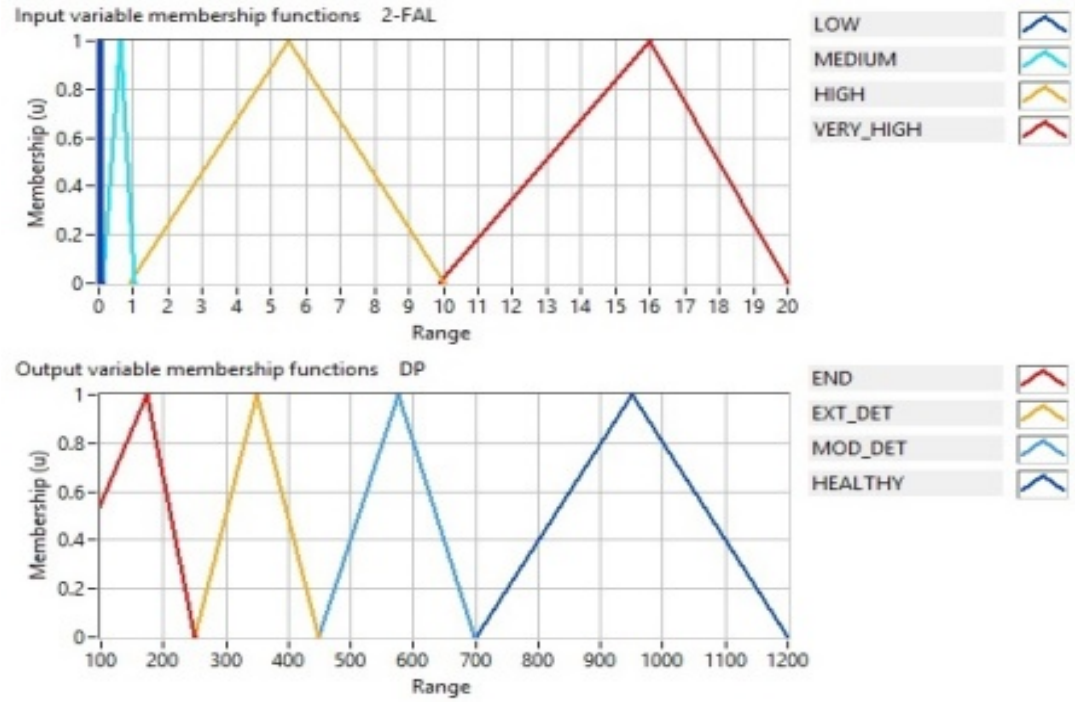
The trapezoidal membership function used for the second fuzzy controller is defined by using four parameters {a, b, c, d} for each of the ranges described in Table 1 and is expressed in the next form:

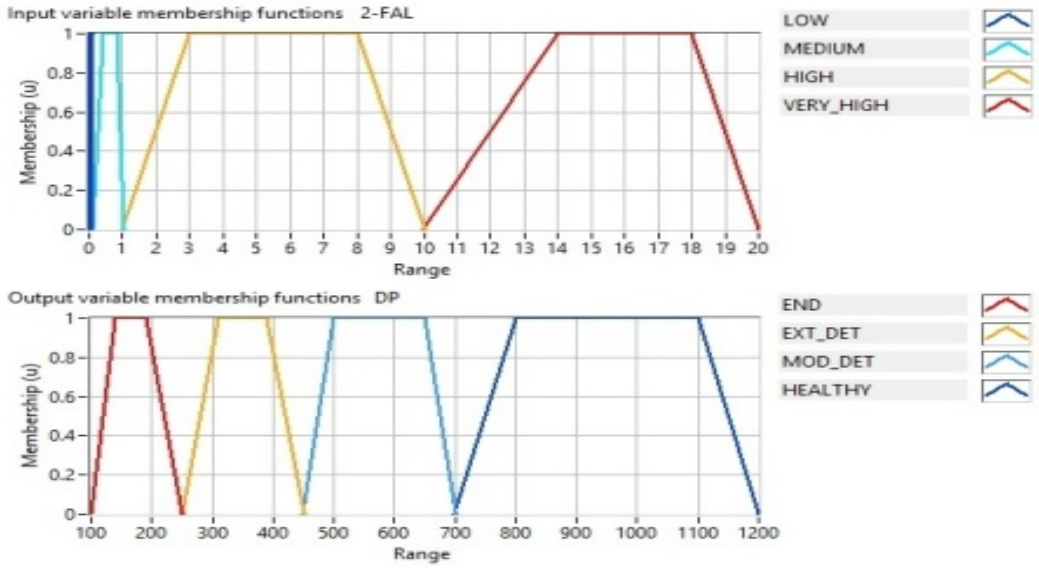
$$trapezoid(x;a,b,c)=\begin{cases}0; & x\le a\\ \dfrac{x-b}{b-a}; & a\le x\le b\\ 1; & b\le x\le c\\ \dfrac{d-x}{d-c}; & c\le x\le d\\ 0; & d\le x\end{cases} \quad (7)$$

or, by using the min and max functions:

$$trapezoid(x;a,b,c,d)=\max\left\{\min\left\{\frac{x-a}{b-a};1;\frac{d-x}{d-c}\right\};0\right\} \quad (8)$$

The defuzzification is required to transform the fuzzy output of the fuzzy controllers achieved based on the triangular and trapezoidal membership functions and averaged into a numerical representation (see Figure 5 and Figure 6). Theoretically, the fuzzy output is a fuzzy relation. The center of gravity method is the most applied practical defuzzification method and is used for the implementation of the two fuzzy controllers. The inference method used for the fuzzy logic implementation of the DP diagnostic system is the Mamadani method.





The fuzzy reasoning is an inference procedure which provides conclusions according to a set of fuzzy "if-then" rules and a set of known data and is based on the ranges specified in Table 1. The method consists in determining the crisp value for the output (DP), taking into consideration, in a balanced manner, all the influences obtained from the rules activated by the particular state of the input (2-FAL) at a given time.

Figure 6 and Figure 7 show the testing of the diagnostic system by entering a 2-FAL value for the two types of fuzzy controllers. The rule that applies for this value of the input and the obtained result of the DP output can be noted.

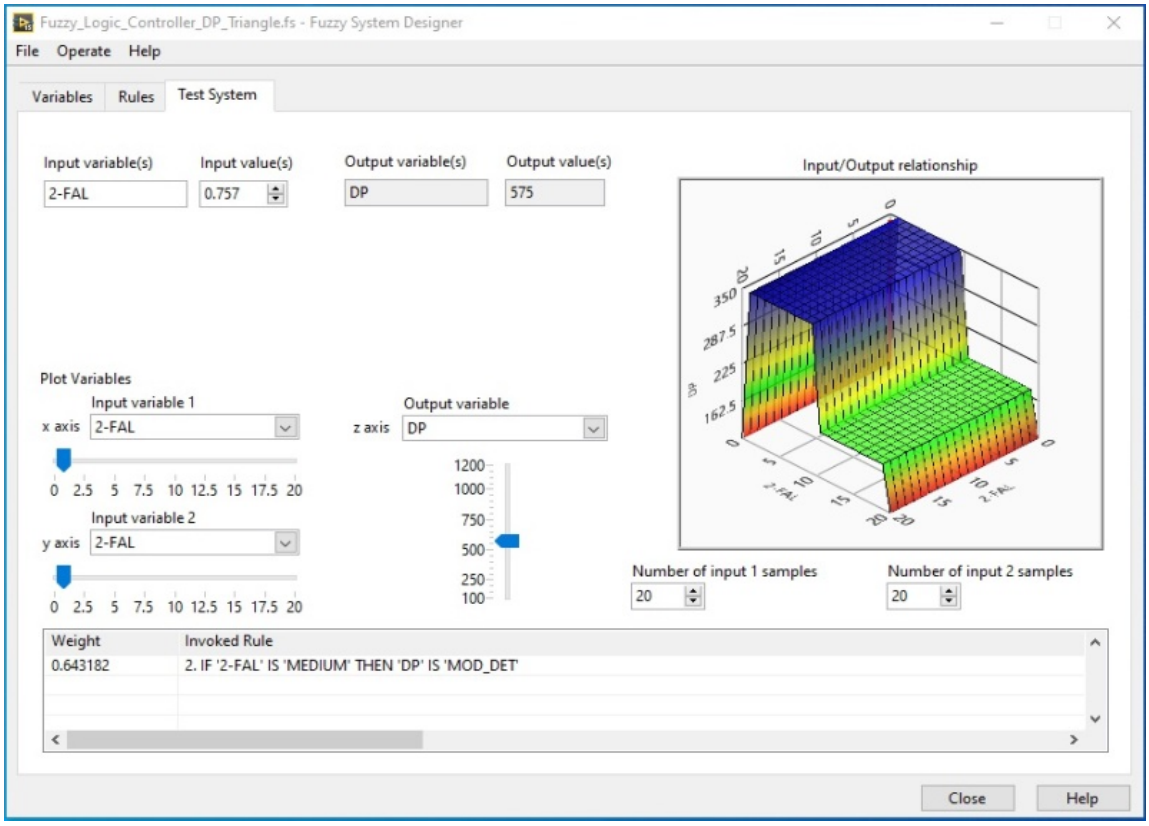


Figure 6: Surface viewer of DP result based on triangular fuzzy logic controller

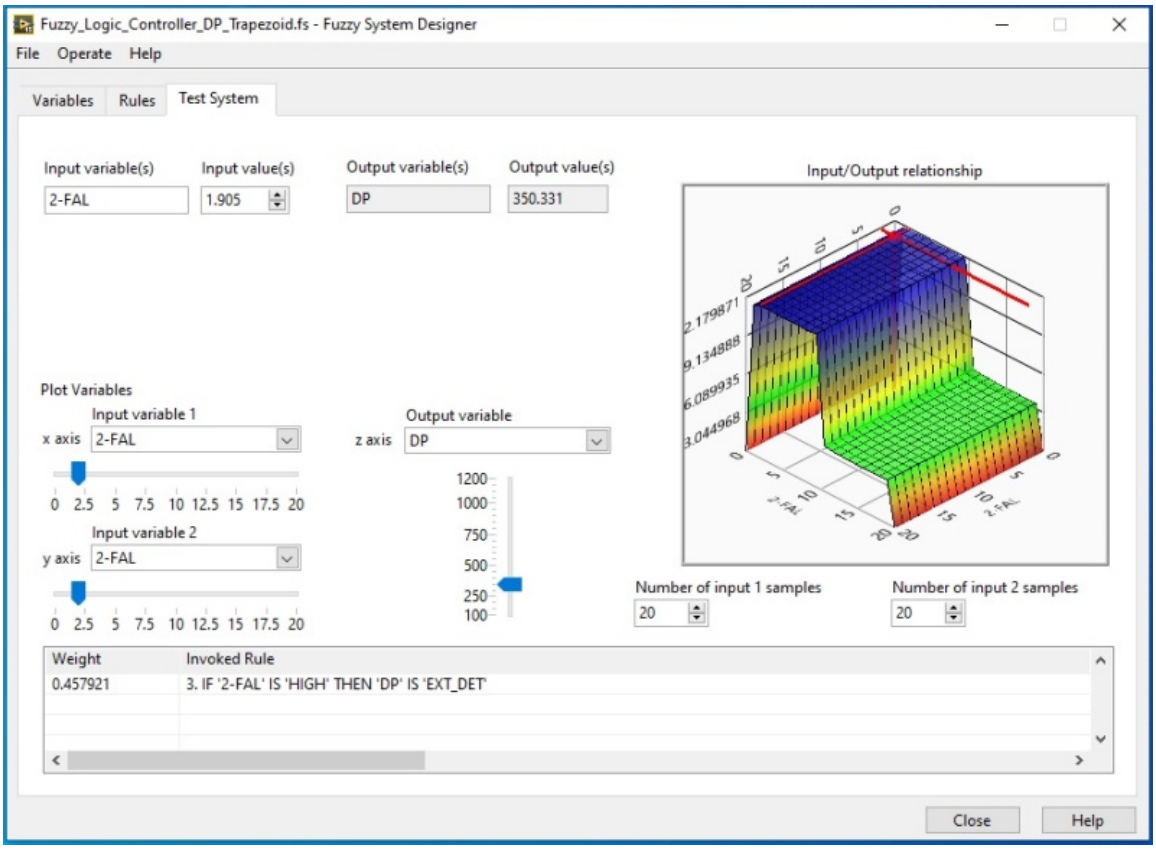
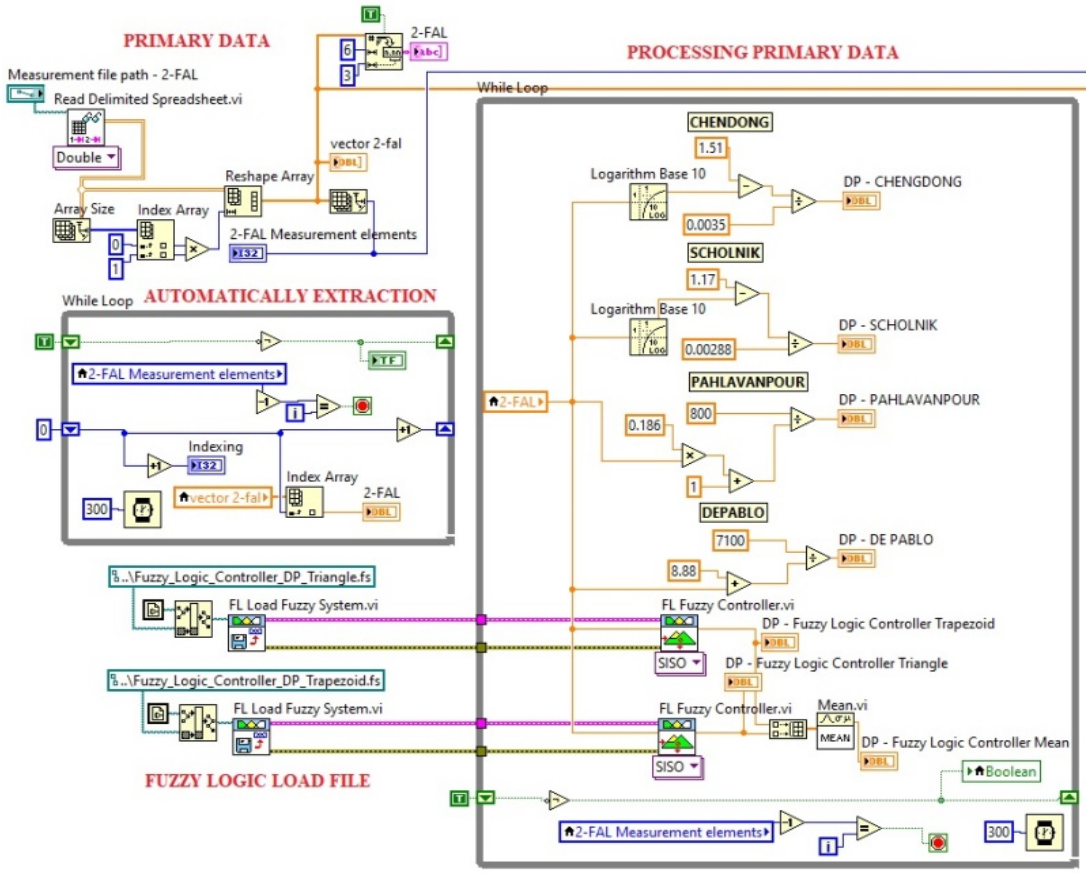


Figure 7: Surface viewer of DP result based on trapezoidal fuzzy logic controller



The software block diagram of the software application developed in LabVIEW for the main stages of the DP diagnostic software application is presented in Figure 8.

4.2. Neuro-fuzzy system for the determination of the DP

The neuro-fuzzy hybrid systems are a combination of neural networks and fuzzy logic where both techniques play a key role: the fuzzy logic provides the structure in which the learning capability of neural networks can be exploited [23], [27]. The ISDDP for power transformers is developed in the Matlab environment in the ANFIS toolbox, which is based on the adaptive neural network. The inference system used in ANFIS toolbox is of Sugeno type, also called fuzzy inference Takagi-Sugeno; it uses singleton output membership functions which for the presented application are linear functions of the input values [26, 27]. Unlike the fuzzy system described above, the proposed neuro-fuzzy diagnostic system has the capability to adapt during a learning process. Thus, by applying the hybrid optimization method, both the membership functions of the fuzzy sets which occur in the premise part of the rules and the parameters of the consequence parts of the fuzzy rules are matched, and the structure of the ANFIS model is presented in Figure 9. The dark circle (in black) on the left side represents the system input (2-FAL), and the one on the right side represents the system output (DP). The white circles on the left and right sides represents the membership functions, and the blue circles represents the rules developed by means of the logical operator "and".

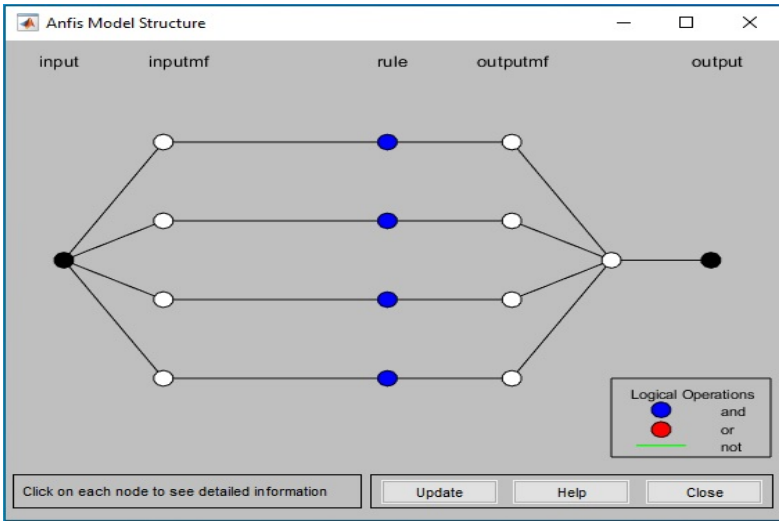


Figure 9: The model structure of the diagnostic system

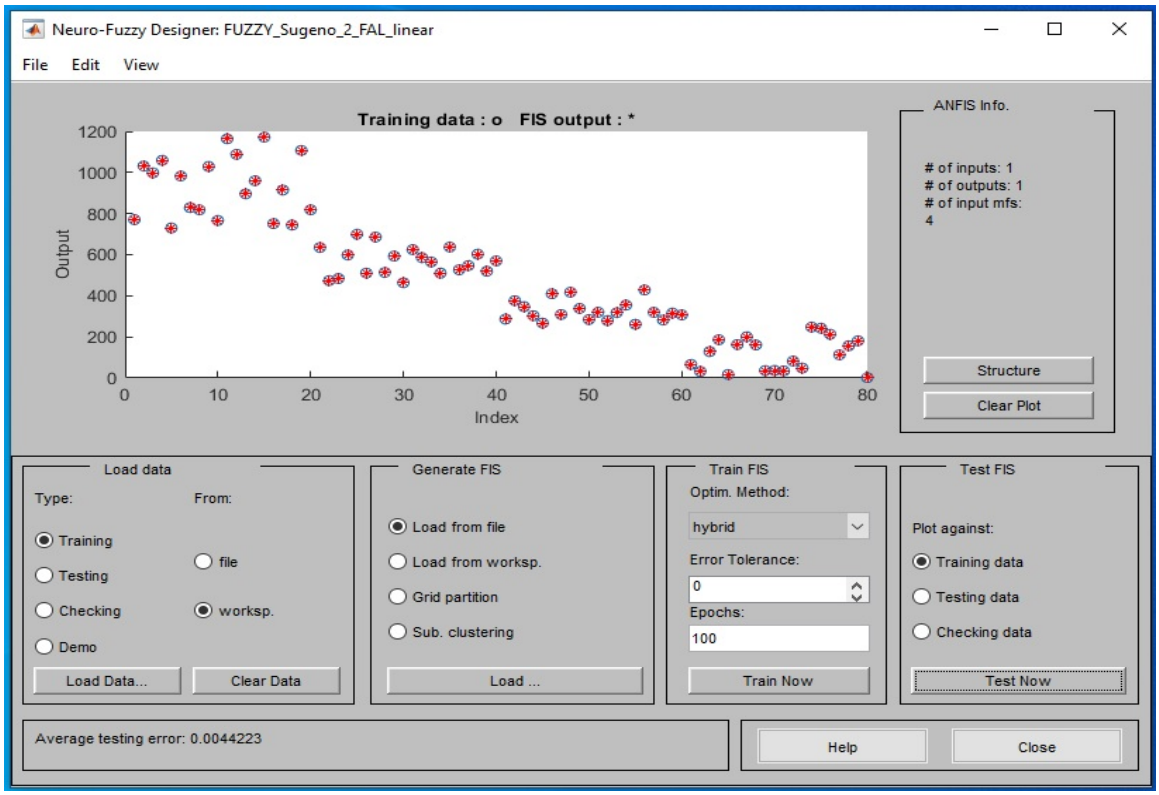


Figure 10: Training data of the ANFIS diagnostic system

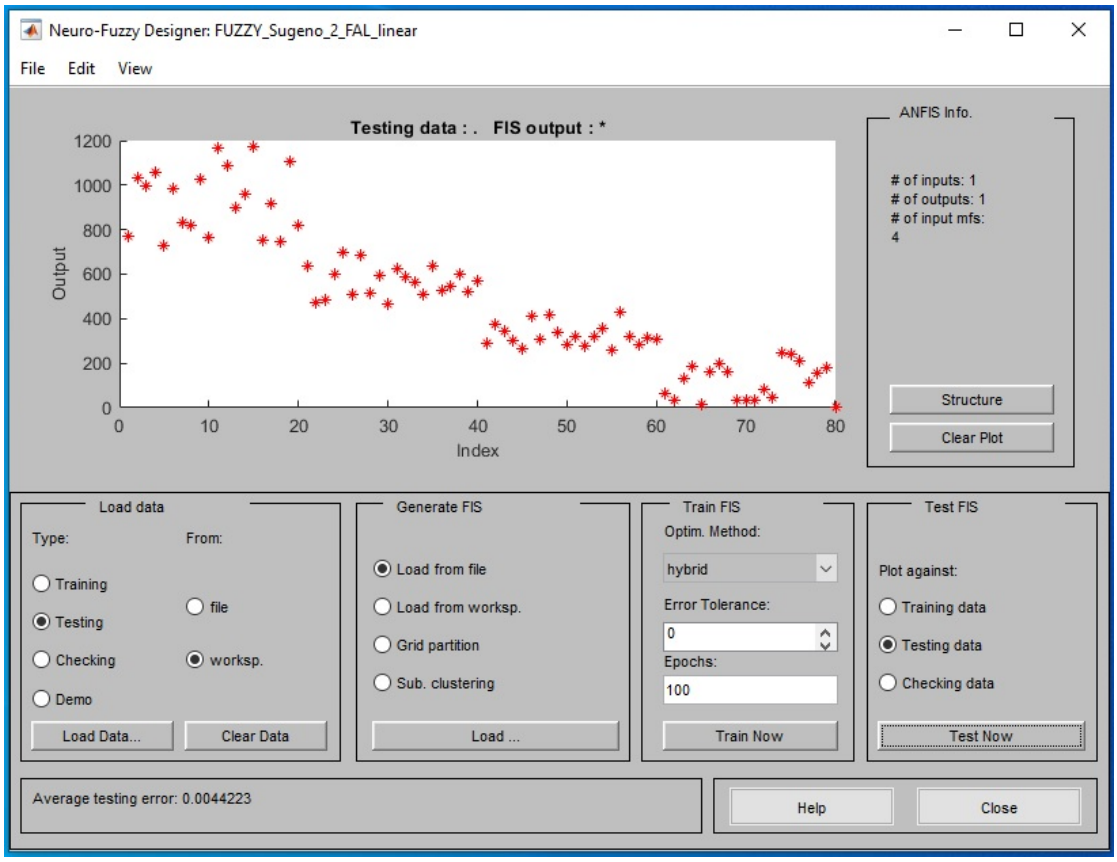


Figure 11: Testing data of the ANFIS diagnosis system

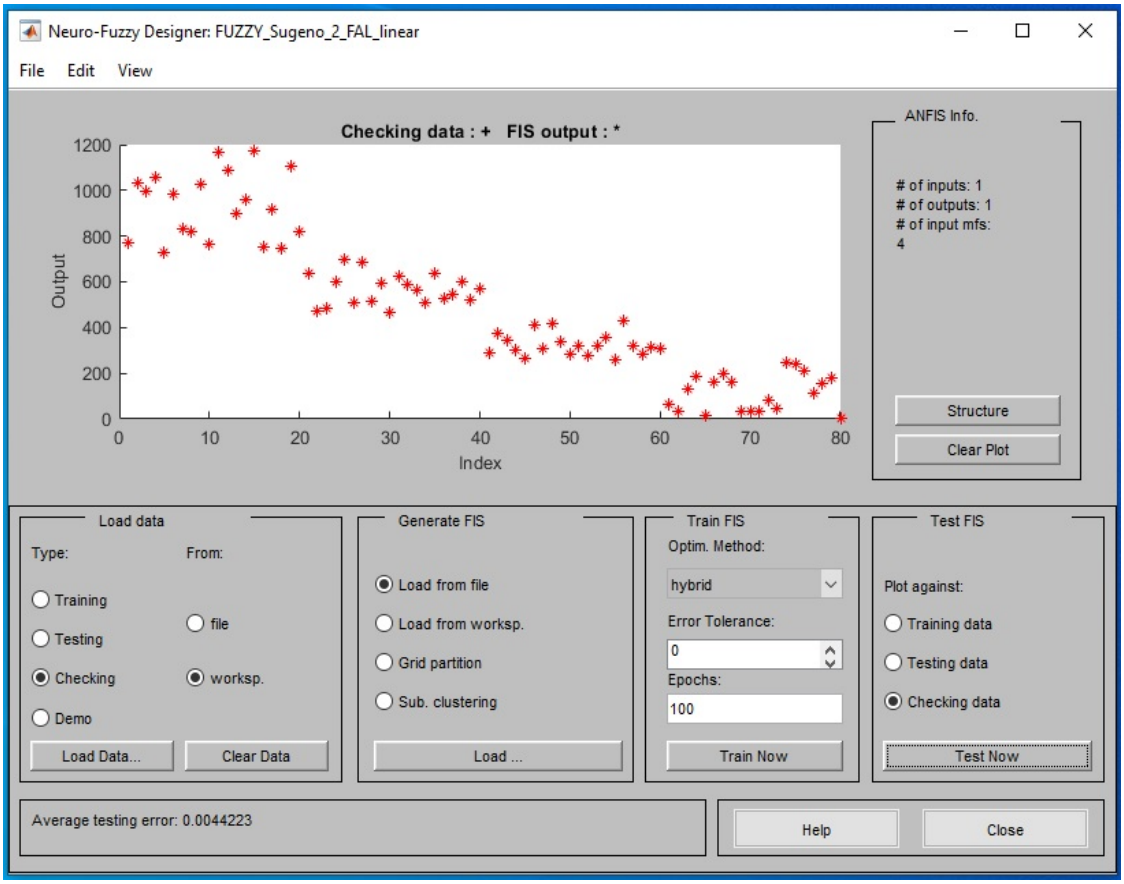


Figure 12: Checking data of the ANFIS diagnosis system

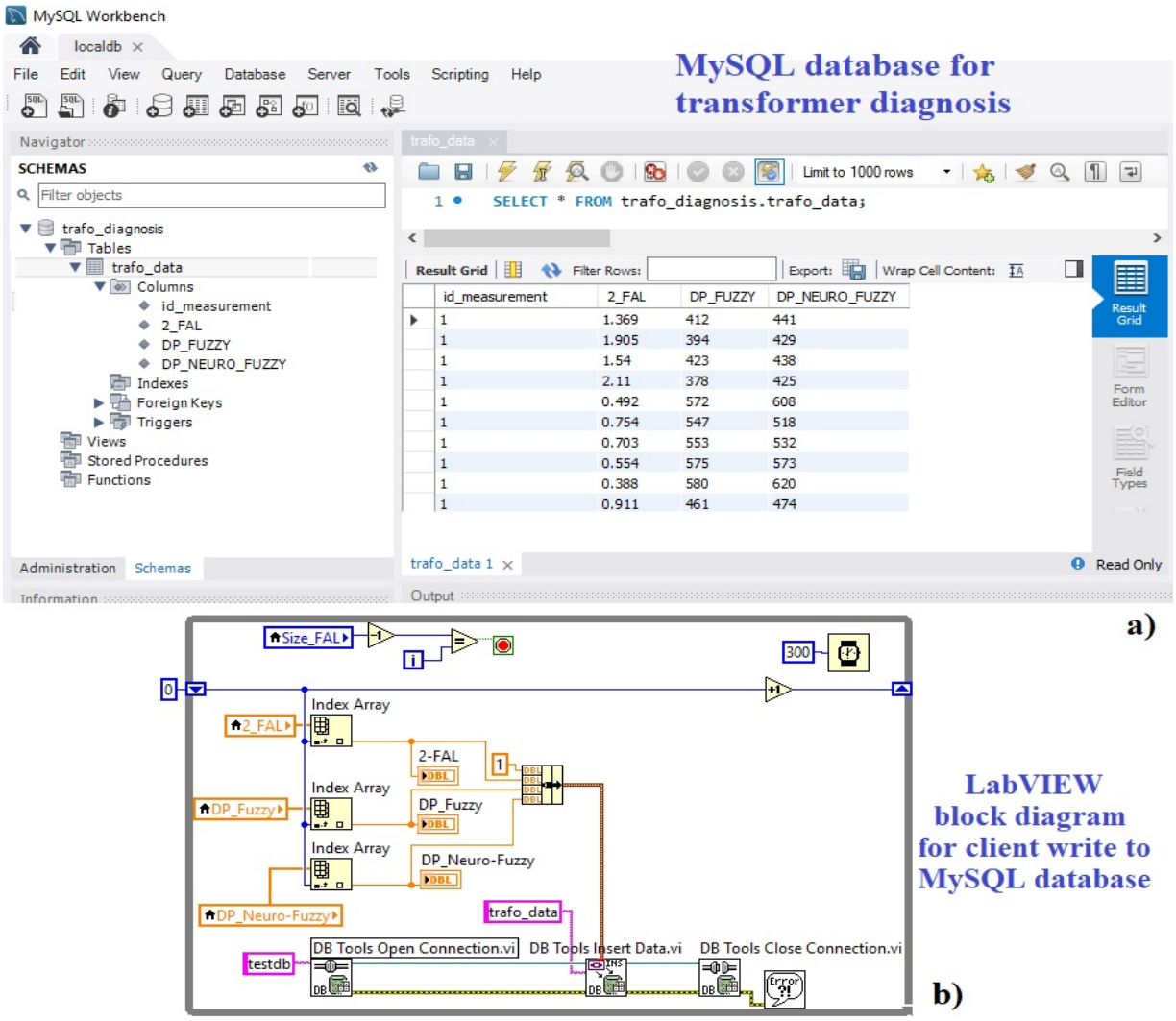


Figure 13: MySQL database connection: a) Table of results in MySQL database, b) block diagram of the client write software module into database

The data set for training, testing and checking comes from the application of classical Pahlavanpour and Chendong methods. The input/output data sets (2-FAL/DP) are entered in the ANFIS based diagnostic system, and they are trained (see Figure 10), verified (see Figure 11), and tested (see Figure 12) in 100 epochs, and the accuracy is given by the mean error obtained with a value of 0.00444223.

The data of the 2-FAL values resulting from the laboratory tests are recorded in the MySQL database by means of a client write software module into the database, and these are taken through the client write software module from the database to be provided to the ISDDP. In the Figure 13 a) is presented the MySQL database interface with the recorded 2-FAL table, and in Figure 13 b) is presented client write software module into the database.

5. Analysis and results

To demonstrate the efficiency of the proposed method and the manner in which data summarization is carried out, we present the analysis of the condition of the insulation of 10 transformers, based on 2-FAL concentration. The results obtained by using the proposed system in items 2 and 4 of Table 2 show a high DP of cellulose insulation, thus, the whole set of physical-chemical measurements of the oils of the two transformers was used. These analyses showed highly degraded wet oil, with temperature spots of >700°C in the transformer in item 2, and high-energy arc discharges, in addition to the wet oil, generated in the transformer in item 4. These discharges may be the result of oil breakdown by arcing between the coils, between terminals and earth, or arcing in the on-load tap changer along the contacts during the switching accompanied by oil leakage in the main tank.

2FAL	DP Chendong	DP Pahlavanpour	DP FUZZY	NEURO-FUZZY
1.369	442	684	412	441
1.905	392	590	394	429
1.540	377	621	423	438
2.110	338	574	378	425
0.492	519	732	572	608
0.754	466	701	547	518
0.703	475	707	553	532
0.554	504	725	575	573
0.388	548	746	580	620
0.911	442	684	461	474

Taking into account these possible problems, the two transformers were un-tanked. In the case of the first transformer, the low-voltage terminals a, and b had loose nuts and showed signs of overheating (see Figure 14), and all connections to the on-load tap-changer selector switch were also loose.

Following the untaking of the second transformer, it was found that the insulation of the flexible connections of the on-load tap-changer selector switch is completely charred, as a result of overheating caused by the arcing during switching as is shown Figure 15.

Figure 14: Low-voltage terminals a and b

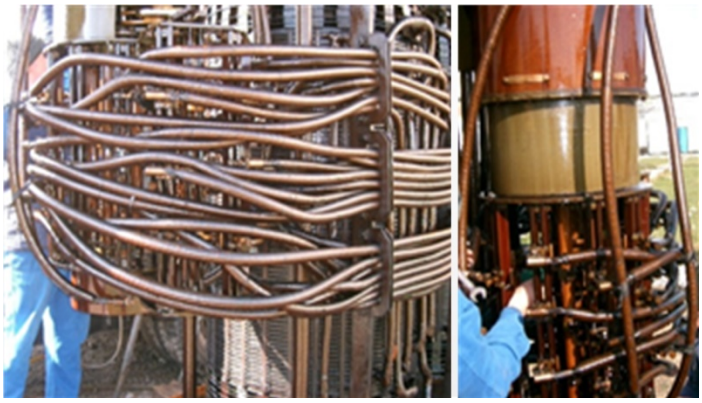


Figure 15: On-load tap-changer selector switch

The faults found and the condition of the insulation of the two transformers confirm the result of the analysis performed using the proposed system for diagnosing the degree of polymerization of solid insulation.

The faults in both transformers were remedied by replacing the damaged insulation; cleaning the terminals, nuts and flexible connection; achieving the tightening to the appropriate torque. The new cotton tape insulation was restored and all degraded insulation residue was removed from the pressboard supports. The on-load tap-changer selector switch was cleaned and the tightening of the connections on the plots of the selector switch was achieved.

The results obtained for the DP by means of the neuro-fuzzy system have a higher precision than the results obtained using mathematical equations and comply with the thresholds proposed by CIGRE according to the 2-FAL content.

The validation of the proposed system is performed by statistical means [31]. To verify the precision and the accuracy, the proposed system was applied to 80 data sets, establishing a precision of approximately 94% and an accuracy of 96% .

6. Conclusions

This article presents a method to determine the DP of solid insulation in transformers, which provides a faster and more accurate interpretation, as compared to the classical methods. The main stages in DP determination are presented, starting with the data provided by the laboratory, the use of classic methods of DP determination, ANFIS training based on a fuzzy controller, data exchange for storage and archiving in MySQL database, as well as the presentation of the validation of this system on a large number of samples, together with the comparison with the results provided by ISDDP with those obtained through un-tanked of transformer.

Thus, the automation of the DP diagnostic process will be achieved, with better results than those obtained by previous methods [26] of precision and accuracy, allowing the operator to make timely decisions, to avoid any possible damages. The main factors of comparison are the accuracy and precision. The values obtained for the improved ISDDP system are around the values presented in the current literature. In the future we propose the improvement of this system by using high-performance neural algorithms for all stages of training, testing and verification in order to improve the accuracy and precision of the system.

Acknowledgment

The paper was developed with funds from the Ministry of Education and Scientific Research - Romania as part of the NUCLEU Program: PN 19 38 01 03 and PN 19 38 02 01.

References

[1] M. Saad, E. teNyenhuis, “On-line gas monitoring for increased transformer protection,” in 2017 IEEE Electrical Power and Energy Conference (EPEC), Saskatoon, SK, Canada, 2017. doi: 10.1109/EPEC.2017.8286169

[2] G. S. Kuleshova, S. A. Yanchenko, “Online Identification of Transformer Internal Faults Based on the Parameters of the Operating Mode” in 2020 V International Conference on Information Technologies in Engineering Education (Inforino), Moscow, Russia, 2020. doi: 10.1109/Inforino48376.2020.9111735

[3] N. T. Waugh, D. D. Muir, “Improving the life cycle management of power transformers transforming data to life” in SoutheastCon 2015, Fort Lauderdale, FL, USA, 2015. doi: 10.1109/SECON.2015.7132977

[4] J. Haema, R. Phadungthin, “Development of condition evaluation for power transformer maintenance” in 4th International Conference on Power Engineering, Energy and Electrical Drives, Istanbul, Turkey, 2013. doi: 10.1109/PowerEng.2013.6635680

[5] M. Wang, A. J. Vandermaar, K. D. Srivastava, “Review of condition assessment of power transformers in service” IEEE Electrical Insulation Magazine, **18**(6), 12–25, 2002. doi: 10.1109/MEI.2002.1161455

[6] Condition assessment of power transformers, in *Cigré brochure 761*, 1–163, March 2019

[7] Mineral insulating oils. Methods for the determination of 2-furfural and related compunds, in *CEI/IEC 61198*, 1–18, June 1994

[8] Ageing of cellulose in mineral-oil insulated transformers, in *Cigré brochure 323*, 1–87, Oct. 2007

[9] J. Scheirs, G. Camino, M. Avidano, W. Tumiatti, “Origin of furanic compounds in thermal degradation of cellulosic insulating paper” Journal of Applied Polymer Science, **69**, 2541–2547, 1998

[10] C. Perrier, M. Coulibaly, M. Marugan, “Efficiency of ageing markers for different transformer insulation systems” in 2016 IEEE International Conference on Dielectrics (ICD), Montpellier, France, 2016. doi: 10.1109/ICD.2016.7547743

[11] T. V. Oommen, T. A. Prevost, "Cellulose insulation in oil-filled power transformers: part II maintaining insulation integrity and life" IEEE Electrical Insulation Magazine, **22(**2), 5–14, 2006, doi: 10.1109/MEI.2006.1618996
[12] C. I. Nicola, M. Nicola, M. C. Nițu, A. M. Aciu, "Power Transformer Fault Diagnosis Using Fuzzy Logic Technique Based on Dissolved Gas Analysis and Furan Analysis" UPB Scientific Bulletin, Series C: Electrical Engineering and Computer Science, **81**(3), 183–196, 2019
[13] S. Apte, R. Somalwar, A. Wajirabadkar, "Incipient Fault Diagnosis of Transformer by DGA Using Fuzzy Logic" in 2018 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Chennai, India, 2018. doi: 10.1109/PEDES.2018.8707928
[14] N. Mahmoudi, M. H. Samimi, H. Mohseni, "Experiences with transformer diagnosis by DGA: case studies" IET Generation, Transmission & Distribution, **13**(23), 5431–5439, 2019. doi: 10.1049/iet-gtd.2019.1056
[15] B. Pahlavanpour, G. Duffy, "Development of a rapid spectrophotometry method for analysis of furfuraldehyde in transformer oil as an indication of paper ageing" in Proceedings of IEEE Conference on Electrical Insulation and Dielectric Phenomena - (CEIDP '93), Pocono Manor, PA, USA, 1993. doi: 10.1109/CEIDP.1993.378924
[16] Chendong X., "Monitoring paper insulation ageing by measuring ageing by measuring furfural contents oil" in Seventh International Symposium on High Voltage Engineering, Dresden, Germany, 1991.
[17] M. Duval, A. de Pablo, "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases," IEEE Electrical Insulation Magazine, **17**(2), 31–41, 2001. doi: 10.1109/57.917529
[18] Guide for the interpretation of gases generated in oil-immersed transformers, in *IEEE C57.104*, 1–36, February 2009
[19] Guidance on the interpretation of dissolved and free gases analysis, in *IEC 60599*, pp. 1–15, September 2015
[20] A. B. Shkolnik, R. T. Rasor, "Statistical insights into furan interpretation using a large dielectric fluid testing database" in 2012 IEEE PES Transmission and Distribution Conference and Exposition (PES T&D), Orlando, FL, USA, 2012.
[21] M. Tamura, K. Ito, M. Honda, Y. Nakagawa, H. Sugimoto, K. Tomishig, "Direct Copolymerization of CO2 and Diols" Scientific Reports 6:24038, 1–9, 2016. doi: 10.1038/srep24038
[22] A. Abu-Siada, S. Hmood, S. Islam, "A new fuzzy logic approach for consistent interpretation of dissolved gas-in-oil analysis" IEEE Transactions on Dielectrics and Electrical Insulation, **20**(6), 2343–2349, 2013. doi: 10.1109/TDEI.2013.6678888
[23] R. A. Prasojo, K. Diwyacitta, Suwarno, H. Gumilang, "Transformer Paper Expected Life Estimation Using ANFIS Based on Oil Characteristics and DissolvedGases (Case Study: Indonesian Transformers)" Energies, **10**(8), 1–18, 2017. doi:10.3390/en10081135
[24] M. M. Nezami, S. A. Wani, S. A. Khan, N. Khera, M. A. Khan, "Fuzzy approach for residual life assessment of paper insulation" in 2017 3rd International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), Rupnagar, India, 2017. doi: 10.1109/CATCON.2017.8280178
[25] D. Urquiza, B. Garcia, J. C. Burgos, "Statistical study on the reference values of furanic compounds in power transformers" IEEE Electrical Insulation Magazine, **31**(4), 15–23, 2015. doi: 10.1109/MEI.2015.7126070
[26] M. C. Nițu, A. Aciu, C. Nicola, M. Nicola, "Fuzzy Controller vs. Classical Methods for Determining the Degree of Polymerization" in 2019 International Conference on Electromechanical and Energy Systems (SIELMEN), Craiova, Romania, 2019. doi: 10.1109/SIELMEN.2019.8905847
[27] R. A. Prasojo, E. F. Nasution, Suwarno, "Implementation of ANFIS and FIS-ID3 in Transformer Oil-Paper Assessment" in 2018 Conference on Power Engineering and Renewable Energy (ICPERE), Solo, Indonesia, 2018. doi: 10.1109/ICPERE.2018.8739422
[28] R. A. Prasojo, K. Diwyacitta, Suwarno, H. Gumilang, "Transformer paper condition assessment using Adaptive Neuro-Fuzzy Inference System model" in 2017 International Conference on Electrical Engineering and Computer Science (ICECOS), Palembang, Indonesia, 2017, doi: 10.1109/ICECOS.2017.8167141
[29] S. Forouhari, A. Abu-Siada, "Application of adaptive neuro fuzzy inference system to support power transformer life estimation and asset management decision" IEEE Transactions on Dielectrics and Electrical Insulation, **25**(3), 845–852, 2018, doi: 10.1109/TDEI.2018.006392
[30] J. Travis, J. Kring, LabVIEW for Everyone: Graphical Programming Made Easy and Fun 3rd Edition, Prentice Hall, 1996.
[31] J. P. Marques de Sá, Applied Statistics Using SPSS, STATISTICA, MATLAB and R, Springer, Berlin, Heidelberg, 2007.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 676-683 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Network Modeling with ANP to Determine the Appropriate Area for the Development of Dry Port in Thailand

Jenjira Sukmanee*, Ramil Kesvarakul, Nattawut Janthong

Department of Production Engineering, King Mongkut's University of Technology North Bangkok, Bangkok, 10800, Thailand

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 30 October, 2020
Online: 24 November, 2020

Keywords:
The Office of Transport and Traffic Policy and Planning (OTP)
Multiple Criteria Decisions Making (MCDM)
Analytic Network process (ANP)
Design Structure Matrix (DSM)

ABSTRACT

One drawback of the Multiple Criteria Decision Making (MCDM) problem with the Analytic Network Process (ANP) is that the origin of the network cannot be clearly defined. In addition, it is not possible to specify internal relationship between criteria and sub-criteria. The application of Design Structure Matrix (DSM) with the Partitioning Reachability Matrix method resulted in a clear understanding of the network's origin and was able to identify the relationship of all criteria in the system. Then, the data were analyzed by the ANP process using Super Decision Program. This will result in more reliable outcomes for selecting the best alternative with higher accuracy. This case study determined the suitable area for dry port development in 5 provinces in Thailand. Based on the analysis process mentioned above, Nakhon Ratchasima is the most suitable area to build a dry port of Thailand.

1. Introduction

Laem Chabang Port is one of the main deep-sea port for sea transportation between Thailand and other countries under the association of the Port Authority of Thailand which is located in the Laem Chabang Municipality, Si Racha and Bang Lamung, Chonburi. Presently, there are a lot of in and out containers in Laem Chabang Port. This directly affects the amount of traffic on the roads around Laem Chabang Port and the side roads over there. Therefore, it is necessary to develop a logistic transportation between Laem Chabang Port and the point of origin-endways in the country. In addition, the country's policy focuses on the development of the country's logistics as a mechanism to support the development of the country to be a center of trade, service and investment in the region. The development of rail transportation as well as various facilities and a logistics service center in the strategic path can be linked to industrial production bases and agricultural bases to Laem Chabang Port, which is the main trading gateway and important trade checkpoint of the country.

The Office of Transport and Traffic Policy and Planning (OTP) has recognized the importance of the problem mentioned above and agreed to study the dry port development master plan in accordance with the Cabinet's resolution on 4 December 2017, which has determined the levels of plans of various government agencies. In order to become the regional logistics center which is one of the important mechanisms to promote the adjustment of the mode of transportation through rail transportation. OTP can save costs and reduce traffic congestion around the Laem Chabang Port in order to become the regional logistics center in the future.

From the research data to determine the suitable area for the development of a dry port, there are 5 provinces suitable for the setting. Criteria are made up of several and inter-influence decisions for determining a site, including the interconnection of transport with rail and other transport systems, urban planning and development, and environmental suitability in terms of land and land size.

The decision to select one province as an appropriate area for the development of that port is based on many of the above criteria which is a very complicated process. It is a complex process with many criteria for decision making.

In solving multiple criteria decisions making (MCDM), the Analytic Hierarchy Process (AHP) is a technique used in human decision making which has been very popular and widely accepted at the international level. This theory was developed by Thomas L. Saaty in the 1970s [1], based on the knowledge of mathematics and psychology for helping make decisions by finding the most suitable answer and helping decision makers understand their own problems better. [2] The principle divides the problem structure

*Corresponding Author: Jenjira Sukmanee, Jenjira.s@eat.kmutnb.ac.th

into layers. The first step is to set the goal and then, criteria, sub-criteria, and alternatives respectively. Then, the analysis will be conducted for the best choice by making comparisons. Pairwise selection criteria makes it easier to decide which criteria is more important by rating according to importance or preference after giving the score prioritize the criteria. Then the analysis step to consider the alternatives one by one according to the specified criteria is conducted until all criteria are met. If the consistency rating is reasonable, the sequences can be chosen to find the best alternative. However, the AHP theory can only make decisions that are not complicated. [3] Saaty invented the Analytic Network Process theory (ANP) by making further improvements from AHP so that decisions can be made to solve complex problems. The feature of complex problems is that various decisions must come from many decisions [4-5] which have influence on each other. The system has a feedback process, time delay and Non-Linearity by ANP theory with the steps as follows.

a. Determine the structure of the problem or create a network
b. Match the comparison of vector priority matrix
c. Calculate the priority of the Super matrix model and
d. Choose the best option.

From the process above, it can be seen that creating a network is the first step and it is the most important component in solving the problem. [6-8]. ANP also has disadvantages in terms of network construction which is not clear because when starting an incorrect network, it will affect the decision in choosing the best alternative.

2. Methodology

The construction of this research is to create a model for solving complex multi criteria decision process by applying the relationship of group theory by using the Design Structure Matrix (DSM) matrix together with the Analytic Network Process (ANP) to select the best alternative. It can be divided into 2 main parts.

2.1. Creating a network

Networking from the determination of the relationship of information in the matrix format with DSM [9]. From Figure 1, the Relation Analysis by Matrix Representation is the result of creating a network by Partitioning Reachability Matrix.

It is a square matrix representing a total system correlation analysis with arrowhead directions to recheck the weight matrix of each criterion to be completely correct. If the arrow shows one incorrect correlation, it immediately affects the ranking of alternatives. The Partitioning Reachability Matrix is divided into three steps with the following details.

1. Information used in research studies which is obtained from data collection from actual operations or interviewing experts in each field of research with the information to be researched which should have enough amount covering the content that needs to be analyzed. If there is a lot of information and diverse [10]. The researcher should group similar data to be in the same group to reduce the amount of information and to be more convenient for analysis.
2. Co-relation is a method for analyzing the relationships of data with the Design Structure Matrix (DSM) theory which can be divided into steps as follows [11-13]

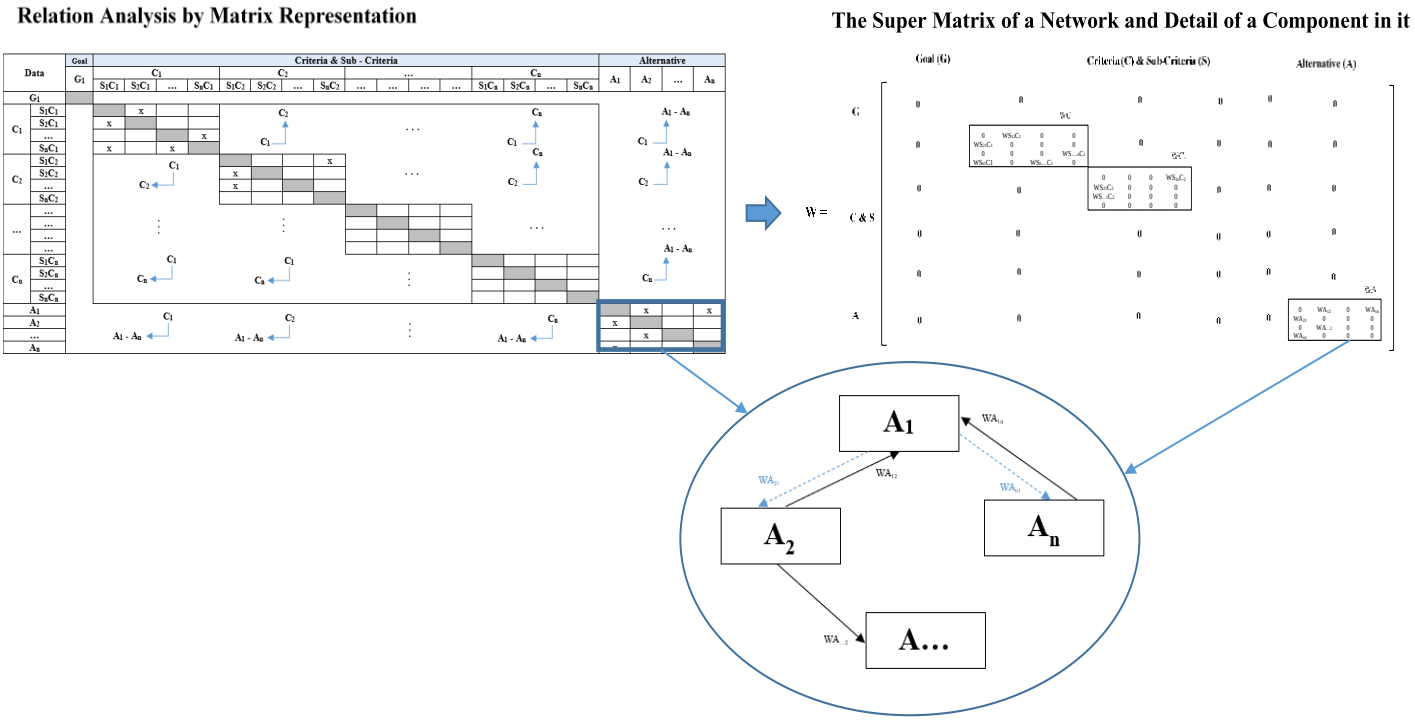


Figure 1: Key elements of the research

- Collect data from actual operations or interviewing an expert to gather information for using in defining relationships in the relationship matrix.
- Create a square matrix that is the same size as the data group to be analyzed (n x n). For example, there are 6 data items, so the generated matrix will be equal to 6 x 6 etc.
- Set the list of data to be analyzed in each row and the columns of the complete generated matrix.
- Define relationships between individual data. If there is a relationship, enter x, there is no relationship and leave it blank. However, the definition of the relationship can be divided into 2 parts which are
  - Diagonal relations mean that the data that is arranged in a row must consider vertical data such as data D and AB and C.
  - The part of the relationship, is the diagonal means that the data, is ordered in a column (Column) is a conditional condition for the data that is ordered in a horizontal row (Row), for example, the H data is the conditional data F, as shown in Table 1.

Table 1: Examples of relationship determination

		A	B	C	D	E	F	G	H
		1	2	3	4	5	6	7	8
A	1	1							
B	2	x	2						
C	3		x	3					
D	4	x	x	x	4				
E	5	x	x	x		5			
F	6	x	x	x			6		x
G	7		x	x				7	
H	8		x						8

Networking

This step is the implementation of the correlation matrix of all data to find the right working sequences using Design Matrix (DSM) Theory by Reachability Matrix Partitioning which has the following steps [14].

- Create a set of data in column j that determines the conditions for data in row i of each data by specifying that the set is equal to Set R (s) (Reachability set).
- Create a set of data in row i that needs to consider the data in column j by specifying that the set is equal to Set A (s) (Antecedent set).
- Determine the data that Set R (s) is a subset of Set A (s) and delete such items from Set R (s) and Set A (s) of all data items. Then, the said data group list was used to create the partition matrix by ordering the groups of data created in the partition matrix, the list of qualified data groups, and precedence. Set R (s) is a subset of Set A (s) which can be shown in Figure 2 and 3.

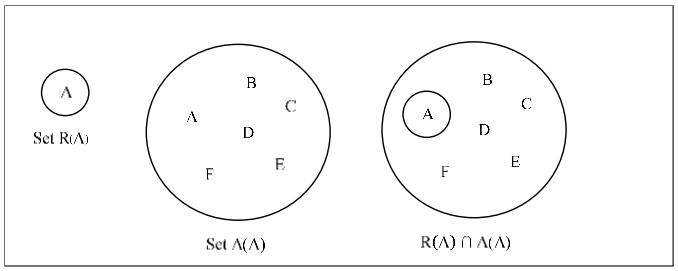


Figure 2: An example of data group based on the criteria "Set R (s) ∩ Set A (s)"

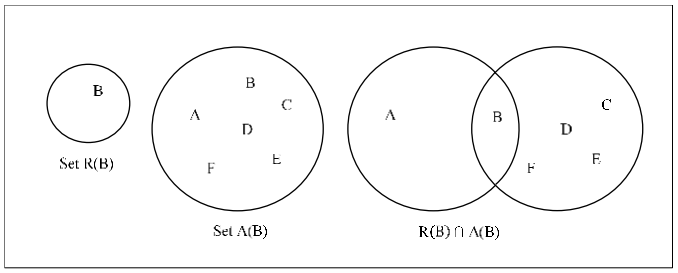


Figure 3: An example of the consideration of group B that does not qualify for "Set R (s) ∩ Set A (s)"

- Record data items one by one in the partition matrix. Adding data and relationship to the partition matrix of each item can be done by firstly, adding A to the partition matrix. Since the list contains Set R (A) as a subset of Set A (A), all Set A data can be deleted from Set R (s) and Set A (s). As a consequence, after deleting A data, Set R(s) becomes a sub-set of Set A (B). Subsequently, B data will be added into partition matrix and they will be deleted from Set R (s) and Set A (s).
- Convert Partition Matrix to Network. Taking the arranged partition data in the form of network that is ready to use in accordance with the standard of ANP.

2.2. Analysis for choosing the best Alternative

The final step is that the network that has already arranged in partition will be processed in Analytic Network Process (ANP). [15]. The researcher used Super Decision 3.2.0 to analyze data for selecting the best alternative.

3. Case Study

The case study of the Office of Transport and Traffic Policy and Planning (OTP), Ministry of Transport, Thailand shows that there is a need for developing "the dry port development plan for becoming the regional logistics center" in accordance with the cabinet resolution. Therefore, the researcher must study and review the relevant information about the dry port, [16, 17] which covers all levels including logistic behavior and the increasing quantity of products in the future. This will be used as supplementary information for setting strategic positioning and

dry port development goal to support the demand and the development of the industrial products of the country in order to meet the economic development policy, trade and transportation in the country, and within the sub-region. This can increase the competitiveness of the country. In addition, the port also helps stimulate the transportation between transportation branches. In this study, data collection and review consist of 5 parts which are:

1. Study and review of the policies of the international agreement and related laws.
2. Review of operating results infrastructure development project, transportation, and logistics related to dry port.
3. The study of data collection, concepts, experience, development, methods of model development design and management standard for dry port area.
4. Study of logistic behavior and quantity of goods moving between the point of origin / destination of transportation and the main trading gates of the country.

5. Summary of directions, policies, plans, and case studies from abroad for the purpose of determining the area suitable for the development of a dry port with the criteria for consideration and selection of areas that are suitable for the development of a dry port as follows: 1. System connection transporting with the rail system and other transportation systems, 2. urban planning and area development, 3. environmental suitability, 4. size and shape of the plot of land, and 5. cost of land procurement and development. The results of the consideration and selection of an area is suitable for the development of the dry port which is used to prioritize and analyze the feasibility of the 5 appropriate areas of the dry port as follows: 1. Khao Thong, Phayuha Khiri, Nakhon Sawan, 2. Muang Wan, Nam Phong, Khon Kaen, 3. Kut Chik, Sung Noen, Nakhon Ratchasima, 4. Don Sai and Nong Tee, Ban Pho, Chachoengsao, and 5. Wiang, Chiang Khong, Chiang Rai.

From the above information, the researcher grouped the data and analyzed to determine the suitable area for the development of the dry port of the Office of Transport and Traffic Policy and Planning (OTP) by applying the theory of grouping relations using Design Structure Matrix (DSM) matrix with the Analytic Network Process (ANP). The analysis was conducted by specifying the correlations of each criteria in the matrix and prioritized to create a network for further analysis with the theory of ANP to get accurate results in choosing the best option by starting with the steps according to the mentioned methodology.

1. Collecting data from actual operations and interviewing experts from dry ports, which can be summarized into various groups as follows in Table 2.
2. Co-relation is a method for analyzing data relationships by creating a square matrix which is the same size as the data group to be analyzed in this research, which are 11 data. Therefore, the created matrix is equal to 11 x 11 as shown in Table 3.
3. Networking: This step is to use the data relationship matrix analyzed to find the proper working sequences using Design Matrix (DSM) Theory by Reachability Matrix Partitioning Method which has the following steps.

Table 2: Grouping and details

Scenario	S1	Determine the area suitable for land port development
Criteria	C1	System connection transporting with the rail system and other transportation systems
	C2	Urban planning and area development
	C3	Environmental suitability
	C4	Size and shape of the plot of land
	C5	Cost of land procurement and development
Alternative	A1	Nakhon Ratchasima
	A2	Nakhon Sawan
	A3	Chachoengsao
	A4	Chiang Rai
	A5	Khon Kaen

a. Define Set R (s) (Reachability set) and collect Set R (s) members of each data item as shown in Table 4.
b. Define Set A (s) (Antecedent set) and collect Set A (s) members of each data item as shown in Table 5.
c. Determine the data that Set R (s) is a subset of Set A (s) and delete those items from Set R (s) and Set A (s). At the same time, the said data group list is used to create the partition matrix by ordering the groups of data created in the partition matrix, the list of qualified data groups, and precedence. Considering that Set R (s) is a subset of Set A (s) can be shown in Table 6.

Table 3: The analyzed relationship matrix

		Group	**Scenario**	**Criteria**					**Alternative**				
		No.	1	2	3	4	5	6	7	8	9	10	11
Group	No.	Data	S1	C1	C2	C3	C4	C5	A1	A2	A3	A4	A5
Scenario	1	S1	1	x	x	x	x	x	x	x	x	x	x
Criteria	2	C1	x	2	x			x					

	3	C2	x	x	3	x	x	x					
	4	C3	x	x	x	4	x						
	5	C4	x			x	5	x					
	6	C5	x		x		x	6					
Alternative	7	A1	x						7				x
	8	A2	x							8			
	9	A3	x								9		
	10	A4	x									10	
	11	A5	x						x				11

Table 4: Total relationship of Set R(s) members

Group	Data	No.	Set name	Reachability Set Set R(s)
Scenario	S1	1	set R(1)	{1,2,3,4,5,6,7,8,9,10,11}
Criteria	C1	2	set R(2)	{1,2,3,6}
	C2	3	set R(3)	{1,2,3,4,5,6}
	C3	4	set R(4)	{1,2,3,4,5}
	C4	5	set R(5)	{1,4,5,6}
	C5	6	set R(6)	{1,3,5,6}
Alternative	A1	7	set R(7)	{1,7,11}
	A2	8	set R(8)	{1,8}
	A3	9	set R(9)	{1,9}
	A4	10	set R(10)	{1,10}
	A5	11	set R(11)	{1,11}

Table 5: Set A(s) members

Group	Data	No.	Set name	Antecedent Set Set A(s)
Scenario	S1	1	set A(1)	{1,2,3,4,5,6,7,8,9,10,11}
	C1	2	set A(2)	{1,2,3,4}
	C2	3	set A(3)	{1,2,3,4,6}
Criteria	C3	4	set A(4)	{1,3,4,5}
	C4	5	set A(5)	{1,3,4,5,6}
	C5	6	set A(6)	{1,2,3,5,6}
Alternative	A1	7	set A(7)	{1,7,11}
	A2	8	set A(8)	{1,8}
	A3	9	set A(9)	{1,9}
	A4	10	set A(10)	{1,10}
	A5	11	set A(11)	{1,7,11}

Table 6: Set R(s) is a subset of Set A(s)

Group	Data	No.	Reachability Set R(s)	Antecedent Set A(s)	Product Set R(s) ∩ Set A(s)
Scenario	**S1**	**1**	{1,2,3,4,5,6,7, 8,9,10,11}	{1,2,3,4,5,6,7, 8,9,10,11}	{1,2,3,4,5,6,7, 8,9,10,11}
	C1	**2**	{1,2,3,6}	{1,2,3,4}	{1,2,3}
	C2	**3**	{1,2,3,4,5,6}	{1,2,3,4,6}	{1,2,3,4,6}
Criteria	**C3**	**4**	{1,2,3,4,5}	{1,3,4,5}	{1,3,4,5}
	C4	**5**	{1,4,5,6}	{1,3,4,5,6}	{1,4,5,6}
	C5	**6**	{1,3,5,6}	{1,2,3,5,6}	{1,3,5,6}
	A1	**7**	{1,7,11}	{1,7,11}	{1,7,11}
	A2	**8**	{1,8}	{1,8}	{1,8}
Alternative	**A3**	**9**	{1,9}	{1,9}	{1,9}
	A4	**10**	{1,10}	{1,10}	{1,10}
	A5	**11**	{1,11}	{1,7,11}	{1,7,11}

3.1. Recording data items in the partition matrix

Delete the items of Set R(s) and Set A(s) and store them in partition matrix until finished. Then, put the data in partition matrix orderly as shown in Table 7 to Table 11.

Table 7: Partition arrangement No. 1 after deleting number 1

Group	Data	No.	Reachability Set R(s)	Antecedent Set A(s)	Product Set R(s) ∩ Set A(s)
	C1	2	{2,3,6}	{2,3,4}	{2,3}
	C2	3	{2,3,4,5,6}	{2,3,4,6}	{2,3,4,6}
Criteria	C3	4	{2,3,4,5}	{3,4,}	{3,4,5}
	C4	5	{4,5,6}	{3,4,5,6}	{4,5,6}
	C5	6	{3,5,6}	{2,3,5,6}	{3,5,6}
	A1	7	{7,11}	{7,11}	{7,11}
	A2	8	{8}	{8}	{8}
Alternative	A3	9	{9}	{9}	{9}
	A4	10	{10}	{10}	{10}
	A5	11	{11}	{7,11}	{7,11}

Table 8: Partition Arrangement No. 2 after deleting numbers 8, 9, 10 and 11

Group	Data	No.	Reachability Set R(s)	Antecedent Set A(s)	Product Set R(s) ∩ Set A(s)
	C1	2	{2,3,6}	{2,3,4}	{2,3}
	C2	3	{2,3,4,5,6}	{2,3,4,6}	{2,3,4,6}
Criteria	C3	4	{2,3,4,5}	{3,4,}	{3,4,5}
	C4	5	{4,5,6}	{3,4,5,6}	{4,5,6}
	C5	6	{3,5,6}	{2,3,5,6}	{3,5,6}
Alternative	A1	7	{7,11}	{7,11}	{7,11}

Table 9: Partition arrangement No. 3 after deleting number 7

Group	Data	No.	Reachability Set R(s)	Antecedent Set A(s)	Product Set R(s) ∩ Set A(s)
	C1	2	{2,3,6}	{2,3,4}	{2,3}
	C2	3	{2,3,4,5,6}	{2,3,4,6}	{2,3,4,6}
Criteria	C3	4	{2,3,4,5}	{3,4,}	{3,4,5}
	C4	5	{4,5,6}	{3,4,5,6}	{4,5,6}
	C5	6	{3,5,6}	{2,3,5,6}	{3,5,6}

Table 10: Partition arrangement No. 4 after deleting numbers 2 and 3

Group	Data	No.	Reachability Set R(s)	Antecedent Set A(s)	Product Set R(s) ∩ Set A(s)
Criteria	C3	4	{4,5}	{4}	{4}
	C4	5	{4,5,6}	{4,5,6}	{4,5,6}
	C5	6	{5,6}	{5,6}	{5,6}

Table 11: Partition Arrangement no. 5 after deleting numbers 4 and 5

Group	Data	No.	Reachability Set R(s)	Antecedent Set A(s)	Product Set R(s) ∩ Set A(s)
Criteria	C5	6	{6}	{6}	{6}

3.2. Convert partition matrix into network

Convert the data after partition arrangement into the form of network that is ready to use in accordance with the standards of ANP.

Table 12: The matrix with new relationship order arrangement

	No.	1	8	9	10	11	7	2	3	4	5	6
No.	**Data**	**S1**	**A2**	**A3**	**A4**	**A5**	**A1**	**C1**	**C2**	**C3**	**C4**	**C5**
1	**S1**		x	x	x	x	x	x	x	x	x	x
8	**A2**	x										
9	**A3**	x										
10	**A4**	x										
11	**A5**	x										
7	**A1**	x										
2	**C1**	x							x	x		
3	**C2**	x						x		x		x
4	**C3**	x							x		x	
5	**C4**	x							x	x		x
6	**C5**	x						x	x		x	

4. Result

Table 12 shows the matrix after converting into a network by analyzing the relationship of the criteria. The value x in the table shows the relationship with each other. All collected relationships are shown in Figure 4 and 5 for further use in the ANP.

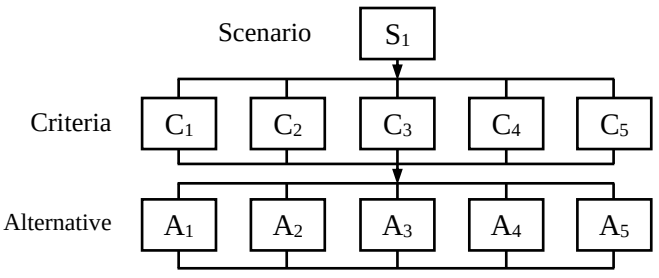


Figure 4: The hierarchical structure of ANP feedback system

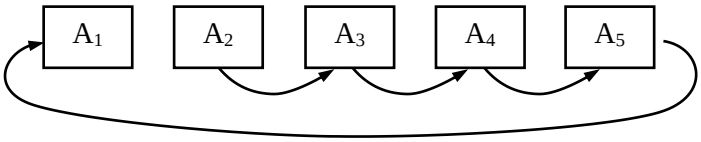


Figure 5: Internal relations of alternative

From the created network above, the researcher continued to analyze with ANP theory by using Super Decision version 3.2.0 as a tool to assist in the analysis to find the best option in determining the suitable area for the development of dry ports, i.e. C1. the connection of the transportation system with the rail system and other transportation systems, C2. the city planning and area development, C3. the suitability of the environment, C4. the size and the shape of land plots, and C5. the costs of land and development. There are 5 options to consider, which are A1. Nakhon Ratchasima, A2. Nakhon Sawan, A3. Chachoengsao, A4. Chiang Rai, and A5. Khon Kaen. Through the analyzing, the results can tell which province is the most appropriate area for dry port development.

From the Figure 6, it can be seen that the network consists of 5 criteria which are C1. the connection of the transportation system with the rail system and other transportation systems, C2. the city planning and area development, C3. the suitability of the environment, C4. the size and the shape of land plots, and C5. the costs of land and development.

A score is computed by multiplying the weight with the level of criteria relationship and the alternative intensity as shown in Table 13. The alternative rankings are shown in Table 14.

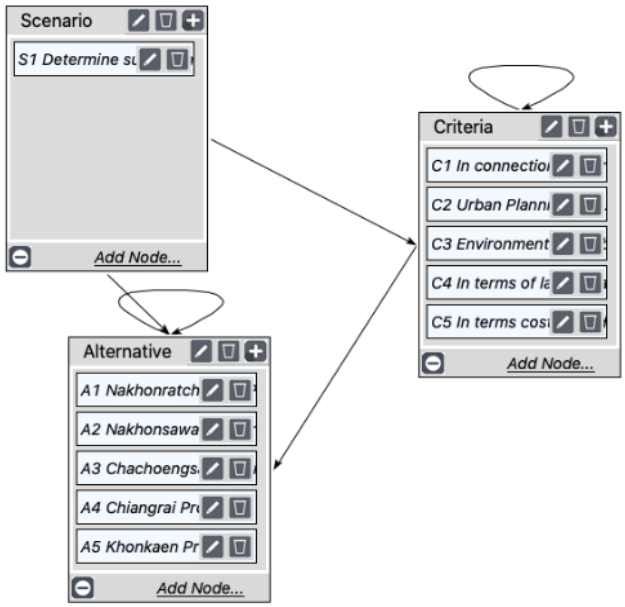


Figure 6: Network model from the ANP Super Decision Software

From the results of the analysis, it can be concluded that the most suitable area for dry port development is A1 Nakhon Ratchasima, followed by A5 Khon Kaen, A4 Chiang Rai, A3 Chachoengsao and A2 Nakhon Sawan respectively.

Table 13: Ranking of areas under each particular criterion

	Alternative					Criteria					Scenario
	A1	**A2**	**A3**	**A4**	**A5**	**C1**	**C2**	**C3**	**C4**	**C5**	**S1**
A1	0.00000	0.00000	0.00000	0.00000	1.00000	0.56524	0.32050	0.40007	0.28124	0.27972	0.49977
A2	0.00000	0.00000	0.00000	0.00000	0.00000	0.07340	0.08180	0.05420	0.56250	0.05578	0.06564
A3	0.00000	1.00000	0.00000	0.00000	0.00000	0.16736	0.23141	0.32736	0.38827	0.44020	0.24350
A4	0.00000	0.00000	1.00000	0.00000	0.00000	0.02854	0.04579	0.04822	0.05407	0.05418	0.03586
A5	0.00000	0.00000	0.00000	1.00000	0.00000	0.16546	0.32050	0.17014	0.22018	0.17012	0.15523
C1	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.11920
C2	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000	0.00000	0.00000	0.27706
C3	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000	0.00000	0.40040
C4	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.06849
C5	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.13484
S1	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Table 14: Ranking of alternatives

Alternatives	Total	Normal	Ideal	Ranking
A1 Nakhonratchasrima Province	0.3308	0.4229	1.0000	1
A2 Nakhonsawan Province	0.0205	0.0263	0.0621	5
A3 Chachoengsao Province	0.1152	0.1473	0.3484	4
A4 Chiangrai Province	0.1292	0.1651	0.3905	3
A5 Khonkaen Province	0.1865	0.2384	0.5637	2

5. Conclusion

This case study to find the most suitable area for dry port development in Thailand employed the simulation of Analytic Network Process (ANP) together with Design Structure Matrix (DSM) of Reachability Matrix using Partition Matrix. It can be concluded that the most appropriate area for the development of dry port in Thailand is Nakhon Ratchasima Province. The network model has been established from rearranging the relationship and considering the relationship related to the network and then used to create a hierarchical network. The applying of Design Structure Matrix with Analytic Network Process results in clearer network sources which shows interaction relationship of the whole system. The result is very effective in finding the best alternative better than using only the Analytic Network Process. However, if there are more criteria used for decision-making, we can use a clustering method for networking with DSM apart from the Partitioning Reachability method, which can be a further study. It is hoped that the applied method above will be useful for those interested in studying the source of the network before analyzing with ANP. For those who are interested in analyzing the correlation of individual criteria in the system will find it useful to increase the efficiency of ANP analysis as well.

References

[1] T. L. Saaty, Theory and Applications of the Analytic Network Process: Decision Making with Benefits, Opportunities, Costs, and Risks, 3rd RWS Publications, 2005.

[2] H. Qudrat-Ullah, J.M. Spector, P.I. Davidson, Complex Decision Making Theory and Practice, Springer New England Complex Systems Institute Book Series, 2008.

[3] K. Z. Edmundas, et al., "Hybrid multiple criteria decision-making methods: a review of applications for sustainability issues" Economic Research-Ekonomska Istraživanja.; **29**(1), 857-887, 2016. https://hrcak.srce.hr/file/253438

[4] T. L. Saaty, "Decision making with the analytic hierarchy process" Int. J. Services Sciences,; **1**(1), 83-98, 2008.

[5] S. Apiradee, "Multi-criteria decision making: Comparison between SAW, AHP and TOPSIS concept and methods" Princess of Naradhiwas University Journal.; **8**(2), 180-192, 2016.

[6] K. Kaku, T. Honma. & M. Fukuda., "An Application of AHP/ANP to A Wildfire Management Project to Help Mitigate Global Warming" in 2009 Proceeding of the International Symposium on the Analytic Hierarchy Process, 2009.

[7] G. Kou, D. Ergu, Y. Peng, Y. Shi, Data Processing for the AHP/ANP, Springer-Verlag Berlin Heidelberg, 2013.

[8] B. D. Rouyendegh, G. F. Canb, "Selection of working area for industrial engineering students" in 2012 Procedia - Social and Behavioral Sciences **31**, 15 – 19, 2012.

[9] T. R. Browning, "Use of Dependency Structure Matrices for Product Development Cycle Time Reduction" in 1998 Proceedings of the Fifth ISPE International Conference on Concurrent Engineering:Research and Applications, Tokyo, Japan, July 15-17, 1998.

[10] T. R. Browning, "Applying the design structure matrix to system decomposition and integration problems: a review and new directions" IEEE Transactions on Engineering Management, **48**(3), 292 – 306, 2001. Doi: https://doi.org/10.1109/17.946528

[11] J. Nattawut, "The Development of Matrix based Scheduling for Customized Product Development Project" Princess of Naradhiwas University Journal.; **5**(3), 47-65, 2013. Doi: https://doi.org/10.1108/14637151111105562

[12] J. Nattawut, "A Methodology to Reducing Process Complexity Using Integration Definition for Function Modeling and Dependency Structure Matrix" Princess of Naradhiwas University Journal.; **7**(2), 48 - 61, 2015.

[13] D. V. Steward, "The design structure system: A method for Managing the design of complex systems" IEEE Transactions on Engineering Management, **28**(3), 71 – 74, 1981. https://doi.org/10.1109/TEM.1981.6448589

[14] A. A. Yassine, "An Introduction to Modelling and Analyzing Complex Product Development Processes the Design Structure Matrix (DSM) Method" Quad. di Manag. (Italian Manag. Rev.) 9, 1 – 17, 2004.

[15] C. Thanaphon, I. Mongkon, "Selection of airport logistics park location in Thailand" Kasem Bundit Engineering Journal **9**(3), 45-60, 2019.

[16] F. Saruchera, "Determinants of effective high-risk cargo logistics at sea ports: A case study" in Open Access Journal of Transport and Supply Chain Management, 1-13, 2020. https://doi.org/10.4102/jtscm.v14i0.488

[17] S. Jenjira, K.Ramil, J. Nattawut, "A Methodology of Network Modeling of Risk Prioritization in Hazardous Product Transportation" IEEE International Conference on Industrial Engineering and Engineering Management, art. no. 8978702, 955-959, 2019. https://doi.org/10.1109/IEEM44572.2019.8978702

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 684-688 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Social Influence Factor of e-Tourism Application Case Study University Student

Kristianus Oktriono[1], Surjandy Surjandy[2,*], Meyliana Meyliana[2], Michele Carolina[2], Stephanie Stephanie[2]

[1]*Language Center, Tourism Destination, Faculty of Humanities, Bina Nusantara University, Jakarta, 11480, Indonesia*

[2]*Information Systems Department, School of Information Systems, Bina Nusantara University, Jakarta,11480, Indonesia*

ARTICLE INFO

Article history:
Received: 06 September, 2020
Accepted: 18 October, 2020
Online: 24 November, 2020

Keywords:
e-Tourism
e-Tourism apps
Social Influence
Causal Research
Explanatory Research

ABSTRACT

The income generated from the tourism sector exhibits the main driving force for national revenues in Indonesian. Based on data from the Ministry of Tourism in Indonesia, it turns out that the potential tourist category originated from the cluster of college students or school-aged cohort. In this line, social media signifies a similar age of internet user based on the Indonesian Internet Service Provider Association (IISPA). In this sense, previous research underlined that all students are active users of the smartphone. Therefore, this study aims to explore how smartphones influence the student's daily utilization in supporting applications of e-Tourism based in Indonesia. The study was conducted using descriptive analysis, causal or explanatory research and SPSS as a tool of data analysis. Operationally, the Bivariate technique serves as the means to answer research questions. As a result, this study revealed 18 relationships that affect students in utilizing smartphones for e-Tourism activities in Indonesia. This research contributes dominantly for the industry in determining the use of a tourism-based application eventually.

1. Introduction

Tourism is one of the mainstay economic sectors of the Indonesian government [1]. Based on the results of research conducted by the Central Bureau of Statistics Indonesia, it was found that most users in tourism are clustered from aged 15-25 years. The respondents can be categorized as a college-aged cohort. Another thing related to tourism is the number of applications available to support tourism activities such as to purchase airline tickets, hotels, and trains such as Traveloka, PegiPegi, and Tiket.com. In this part, e-tourism applications seize advantages and increase engagement of the user, provider, and the system. Reflecting this condition, IISPA or AJPII also published previous research reports stating that students are active internet users in Indonesia [2]-[4]. However, the scarcity of the research is found regarding the student role in tourist [3]. In this part, these phenomena attract researchers to delve in the influence of friends (close-person), family-as undertaken in previous studies [5], and advertisements for the use of applications that support tourism activities [6]. The hypotheses in this research are formulated as follows:

- The influence for students in using smartphones for e-Tourism application
  H0: No relationship was found
  H1: Relationship found
- There is a social influence for students in the use of smartphones on e-Tourism applications
  H0: No relationship was found
  H1: Relationship found

Figure 1 exhibits the research design that includes three major elements, i.e. the students, social influence, and e-tourism applications. The students consist of university students. In this context, the aspects of social influence factor will be analyzed based on e-tourism application. The growth of technology and its advancement in tourism will influence the stakeholders in many ways. In this frame, the activity related to technology application is categorized to describe the merits and challenges towards the engagement of the users. The study highlights the significance of advertising to bolster travelling, the experience of tourist, both individual and family that becomes a determining factor in the use of applications eventually.

2. Literature Review

In this section, the authors explain the relevant and prominent literature, such as understanding of e-Tourism, causal research/explanatory research, e-tourism apps, and respondent

*Corresponding Author: Surjandy Surjandy, surjandy@binus.ac.id

data in this study.

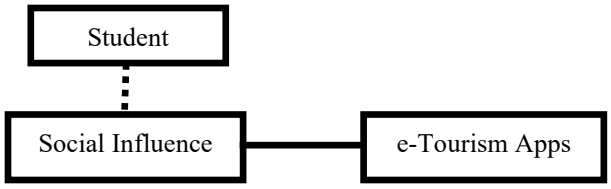


Figure 1: e-Tourism Research Design

2.1. E-Tourism

Technology contributes significantly to human existence in terms of efficiency enhancement, improving service and customers' experience in travelling. In this line, technology has revolutionized tourism in a paradigm shift. Technically, e-tourism represents the use of electronic media or internet-based applications to bolster the tourism process such as ticket reservations, hotel reservations, and so that customers can experience comfortable. [7].

2.2. Casual Research/Explanatory Research

Causal research or explanatory research is research that is generally implemented by marketing or sales to examine the relationship and the level of influence between two factors [7]. This research perused at several prominent factors such as student background (gender, age, screen width) and the influence of peers, family, and advertisements on the use of e-Tourism support applications.

2.3. E-Tourism Application

The e-Tourism application is an application engaged to assist tourists in carrying out tourism activities such as purchasing airline tickets, train tickets, hotel reservations, booking vehicle rentals, and other related amenities.

Table 1: Data of Respondent

No.	Description	N	%
1	Gender (A1)		
	Female	107	33.4
	Male	213	66.6
2	Age (A2)		
	1. <15 Years Old (YO)	0	0
	2. 16 – 20 YO	300	93.8
	3. 21 – 25 YO	15	4.7
	>26	5	1.6
3	Screen Size of Smartphone (A3)		
	1. <4"	27	8.4
	2. 4"-5."	99	30.9
	3. 5"-6."	151	47.2
	>6"	43	13.4

2.4. Respondent

This research uses a technique commonly used that is snowball sampling by utilizing services from Google based on Google Form. The questionnaire is disseminated through the internet and social media. In terms of population, this study involves the number of 320 respondents with backgrounds, as presented in table 1. The respondents can be categorized as a college-aged cohort.

3. Methodology

This descriptive study uses the SPSS application as a tool in analyzing data, and the Bivariate Technique is used to answer research questions. There are several steps needed in this research before analyzing the data, such as validation of the questions made. A test of the data reliability and after all, this is fulfilled, the analysis can be done.

3.1. Validation of Questions

In this part, researchers validate the questions using SPSS tools. The way to validate is to calculate the Corrected item-total Correlation (CI-TC) score. If the value of the CI-TC is higher than the Pearson r Table, the questions made are declared valid. In other words, in best practice, if the CI-TC value> 0.5, we ensure that the question is valid. For more details, table 2 described the questions validation.

Table 2: Questions Validation

No.	Description	CI-TC	Status
1	People influence my behaviour to use the travelling app (SI1)	0.673	Valid
2	People who are important to me think that I should use the travelling app (SI2)	0.732	Valid
3	I do travel through the app because many people do so (SI3)	0.665	Valid
4	My friends are supportive of the use of travelling mobile app (SI4)	0.628	Valid
5	My Family are supporting the use of travelling mobile app (SI5)	0.585	Valid
6	Advertisement strengthens me in using a travelling mobile app (SI6)	0.630	Valid

3.2. Data Reliability

In this section, researchers conducted a test of the reliability of the data obtained, namely by looking at the scores from Cronbach's Alpha [8]. The guidelines underscore that if the Cronbach's alpha value is greater than 0.6, the data is reliable[8]. Table 3 explains this information in detail.

3.3. Bivariate Test Result

This section presents the results of calculations based on the bivariate method to check the relationship between 2 factors [5]. Table 4 exhibits the relationship for more details. In this stage, the hypothetical results proposed that:

H0 is rejected; it signifies the relationship between factors or H1 is accepted

H1 rejected; It indicates that there is no relationship between factors.

The p-value is examined to find out whether there is a relationship [9]. In this frame, a relationship is obtained if the p-value <0.05. Doubtless, table 4 demonstrates the relationship between factors. Related to these factors, the value '*' denotes alpha 5% or 95% confidence level and for '**' refers to alpha 1% or 99% confidence level. The term of NC illustrated that there is no relationship (no correlation).

Table 3: Reliability of Questions

No.	Description	Cronbach's Alpha	Status
1	People influence my behaviour to use a travelling app (SI1)	0.831	Reliable
2	People who are important to me think that I should use a travelling app (SI2)	0.820	Reliable
3	I do travel through the app because many people do so (SI3)	0.833	Reliable
4	My friends are supportive of the use of travelling mobile app (SI4)	0.840	Reliable
5	My Family are supporting the use of travelling mobile app (SI5)	0.847	Reliable
6	Advertisement strengthens me in using a travelling mobile app (SI6)	0.840	Reliable

Table 4: Correlation Factors

No.	Correlation Factors	Pearson Corr. & p-value	H0	H1
1	A1 & SI1	-0.118* 0.035	Reject	Accept
2	A1 & SI2	NC	Accept	Reject
3	A1 & SI3	NC	Accept	Reject
4	A1 & SI4	NC	Accept	Reject
5	A1 & SI5	NC	Accept	Reject
6	A1 & SI6	NC	Accept	Reject
7	A2 & SI1	NC	Accept	Reject
8	A2 & SI2	NC	Accept	Reject
9	A2 & SI3	NC	Accept	Reject
10	A2 & SI4	0.134* 0.017	Reject	Accept
11	A2 & SI5	NC	Accept	Reject
12	A2 & SI6	NC	Accept	Reject
13	A3 & SI1	NC	Accept	Reject
14	A3 & SI2	-0.123* 0.027	Reject	Accept
15	A3 & SI3	NC	Accept	Reject
16	A3 & SI4	NC	Accept	Reject
17	A3 & SI5	NC	Accept	Reject
18	A3 & SI6	NC	Accept	Reject
19	SI1 & SI2	0.673** 0.000	Reject	Accept
20	SI1 & SI3	0.621** 0.000	Reject	Accept
21	SI1 & SI4	0.400** 0.000	Reject	Accept
22	SI1 & SI5	0.379**	Reject	Accept
23	SI1 & SI6	0.500** 0.000	Reject	Accept
24	SI2 & SI3	0.626** 0.000	Reject	Accept
25	SI2 & SI4	0.515** 0.000	Reject	Accept
26	SI2 & SI5	0.453** 0.000	Reject	Accept
27	SI2 & SI6	0.516** 0.000	Reject	Accept
28	SI3 & SI4	0.430** 0.000	Reject	Accept
29	SI3 & SI5	0.401** 0.000	Reject	Accept
30	SI3 & SI6	0.486** 0.000	Reject	Accept
31	SI4 & SI5	0.640** 0.000	Reject	Accept
32	SI4 & SI6	0.497**	Reject	Accept
33	SI5 & SI6	0.470** 0.000	Reject	Accept

3.4. Relationship Strength

This section explains how strong the influence of the relationship between the two factors that occur, to examine how much influence (relation) exists [9] and it is represented from the magnitude of the absolute value of the Pearson correlation. The information of the values between 0.1 to 0.3 indicates the level of strength of the relationship is small, while the value of 0.3 to 0.5 is categorized as medium. The values above 0.5 states a strong relationship. To check the degree to which strength, the researchers square the Pearson correlation value and multiply it by 100 to obtain the percentage of the relationship strength as exhibited in Table 5.

Table 5: The Relationship Strength

No.	Factor	|r|	Tension	$\|r\|^2$	%
1	A1 & SI1	0.118	Low	0.014	14%
2	A2 & SI4	0.134	Low	0.018	18%
3	A3 & SI2	0.123	Low	0.015	15%

No.	Factor	|r|	Tension	$\|r\|^2$	%
4	SI1 & SI2	0.673	High	0.453	45.3
5	SI1 & SI3	0.621	High	0.386	38.6
6	SI1 & SI4	0.400	Medium	0.16	16%
7	SI1 & SI5	0.379	Medium	0.144	14.4%
8	SI1 & SI6	0.500	High	0.25	25%
9	SI2 & SI3	0.626	High	0.392	39.2%
10	SI2 & SI4	0.515	High	0.265	26.5%
11	SI2 & SI5	0.453	Medium	0.205	20.5%
12	SI2 & SI6	0.516	High	0.266	26.6%
13	SI3 & SI4	0.430	Medium	0.185	18.5%
14	SI3 & SI5	0.401	Medium	0.161	16.1%
15	SI3 & SI6	0.486	Medium	0.236	23.6%
16	SI4 & SI5	0.640	High	0.409	40.9%
17	SI4 & SI6	0.497	Medium	0.247	24.7%
18	SI5 & SI6	0.470	Medium	0.206	20.6%

4. Result and Discussion

In this section, researchers explain the findings of the study. The calculation result values of this study are presented in Tables 4 and Table 5. This study highlights 18 relationships between factors and the magnitude of influence are as follows:

- A1 (Gender) & SI1 (People influence my behaviour to use a travelling app). It obtained the opposite relationship (minus). In this case, this study says that men (A1) tend not to be affected by the behaviour of others in using applications for travelling app.
  The magnitude of the strength of influence expressed the position of 14% or low influence. The study found a similar result with the previous research [3]
- A2 (Age) & SI4 (My friends are supportive of the use of travelling mobile app). It attained that a direct or positive relationship which states that the higher the age (old), the greater the influence of friends to use the travelling app. The magnitude of the influence strength between factors is 18% or low influence.
- A3 (Screen Size) & SI2 (People who are important to me think that I should use a travelling app). It confirmed the relationship is not direct (negative) which means that the greater the width of the smartphone layer that is used the smaller or fewer people feel it is important for me always to use the travelling apps. The magnitude of the strength of influence between factors of 15% or low influence.
- SI1 (People influence my behaviour to use the travelling app) & SI2 (People who are important to me think that I should use a travelling app). It revealed a direct relationship which means the more people influence me to use the travelling app, the greater the important person also states that I have to use the travelling app. The magnitude of influence strength between factors is 45.8% or strong influence.
- SI1 (People influence my behaviour to use a travelling app) & SI3 (I do travel through the app because many people do so). It underscored a direct relationship, which means that more people influence one to use the travelling app, the more often I travel using the application — the magnitude of the strength of influence between factors of 38.6% or strong influence.
- SI1 (People influence my behaviour to use a travelling app) & SI4 (My friends are supportive of the use of travelling mobile app). It detected a direct relationship, which means that more people influenced me to use the travelling app, so more friends gave support to use the travelling app. The magnitude of the influence strength between factors is 16% or medium influence.
- SI1 (People influence my behaviour to use a travelling app) & SI5 (My Family are supporting the use of travelling mobile app). It highlights directional relationships where the higher the influence of other people influencing me to use the travelling app, the higher the support from family in using the travelling mobile app. The amount of influence strength is 14.4% or medium influence.
- SI1 (People influence my behaviour to use a travelling app) & SI6 (Advertisement strengthens me in using the travelling mobile app). It emphasized a directional or positive relationship that states that the stronger the influence of people on my behaviour using travelling apps, the more advertisements affect me to use travelling mobile apps. The magnitude of the influence strength of 25% or a strong influence.
- SI2 (People who are important to me think that I should use a travelling app) & SI3 (I do travel through the app because many people do so). This highlights a direct or positive relationship which means that the more people who have an essential relationship with me think I should use the travelling application, the higher I use the travelling application to travel. The magnitude of the effect strength is 39.2% or strong influence.
- SI2 (People who are important to me think that I should use a travelling app) & SI4 (My friends are supportive of the use of travelling mobile app). It exhibits a direct or positive relationship which means that the greater the people who have significant relationships, the higher the friends provide support to use travelling apps. The amount of influence strength is 26.5% or strong influence.
- SI2 (People who are important to me think that I should use a travelling app) & SI5 (My Family are supporting the use of travelling mobile app). It accentuates a direct or positive relationship which means that the higher the person who has an essential relationship with me states I have to use the application, the higher the support from family to use the travelling mobile app. The amount of influence strength is 20.5% or medium influence.
- SI2 (People who are important to me think that I should use a travelling app) & SI6 (Advertisement strengthens me in using the travelling mobile app). It explained a direct or positive relationship which means that the higher the person who has an essential relationship with me states that I must use the application, the higher the influence of advertising to use the travelling app. The magnitude of influence strength is 26.6% or strong influence.

- SI3 (I do travel through the app because many people do so) & SI4 (My friends are supportive of the use of travelling mobile app). It stated a direct relationship or positive which means that the higher the person who has an essential relationship with me stated that I have to use the application, the higher the influence of friends who support the use of travelling applications. The magnitude of the effect strength is 18.5% or medium influence.
- SI3 (I do travel through the app because many people do so) & SI5 (My Family are supporting the use of travelling mobile app). It described a direct or positive relationship which means the higher I use the application because many people do the same thing, the higher the family also provides support. The magnitude of the effect strength is 16.1% or medium influence.
- SI3 (I do travel through the app because many people do so) & SI6 (Advertisement strengthens me in using the travelling mobile app). It signifies a direct or positive relationship which means the higher the use of the application because other people also use the higher the effect of advertising. The amount of influence is 23.6% or medium influence.
- SI4 (My friends are supportive of the use of travelling mobile app) & SI5 (My Family are supporting the use of travelling mobile app). It explained a direct or positive relationship which means the higher the influence of friends supporting using the application, the higher the family support for using the application. The magnitude of the effect is 40.9% or strong influence.
- SI4 (My friends are supportive of the use of travelling mobile app) & SI6 (Advertisement strengthens me in using the travelling mobile app). It affirms a direct or positive relationship which means that the stronger the influence of friends to support using the travel application, the higher the effect of advertising to use the travelling app. The amount of influence is 24.7% or medium influence.
- SI5 (My Family are supporting the use of travelling mobile app) & SI6 (Advertisement strengthens me in using the travelling mobile app). It asserts a direct or positive relationship, which means that the higher the influence of the family in supporting using a travel application, the higher the effect of advertising to use the travelling app. The amount of influence is 20.6% or the medium of influence.

5. Research Limitation and Implication

This study provides an overview of implications regarding the influence of friends or colleagues, family, advertisements in utilizing e-tourism applications to support travel activities. In this part, the need to increase of adopting the website in the tourism industry serves as the means to develop the industry performance at the same time [10]. It contributes to stakeholders in developing attractive e-tourism. It adds more value to users and developers in increasing application engagement. Koo et al. stated that travellers gathered the most popular information from different available social media through the internet [11]. In this line, the marketing aspect requires strategies to promote the tourism destination and it recommends the e-tourism platform [12].

The limitations of this study are related to have not implemented the full model such as TAM or UTAUT to review the comprehensive picture. The populations are originated from domestic tourists in Indonesia.

6. Conclusion and Future Research

The conclusion of this study emphasizes the importance of advertising to support travelling, the experience of users both individuals and from families becomes a determining factor in the use of applications eventually. In this case, youth tourism plays important role in the contribution to the tourism industry because of its positive social and economic impacts [13]. The multifaceted influence of tourism increases the promotion and efficiency of the tourism product [14]. The research dominantly indicated that the more people (university student) influence someone to use the travelling app, the greater the important person also states that someone must use the travelling app. Furthermore, it underscored that more people influence one to use the travelling app, the more often one travels using the application. The strength relationship appears in SI1 & SI6, SI2 & SI3, SI2 & SI4, SI2 & SI6, SI4 & SI5. The future research will investigate the other factors that influence and prioritize the user in using the e-tourism application.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Thank you for Bina Nusantara University for the support.

References

[1] Asosiasi Penyelenggara Jasa Internet Indonesia, Potret Zaman Now, Pengguna & Perilaku Internet Indonesia, Apjii, 1–7, 2018.
[2] B. Dictionary, What is causal research? Definition and meaning, 2019.
[3] I. Bizirgianni, P. Dionysopoulou, "The Influence of Tourist Trends of Youth Tourism through Social Media (SM) & Information and Communication Technologies (ICTs)," Procedia - Social and Behavioral Sciences, **73**(February 2013), 652–660, 2013, doi:10.1016/j.sbspro.2013.02.102.
[4] S. Herayanti, P.A. Wikanatha Sagita, N.P. Eka Mahadewi, "Faktor-Faktor Yang Mempengaruhi Keputusan Wisatawan Menggunakan Situs Tripadvisor Di Kawasan Pariwisata Kuta Selatan-Bali," Jurnal IPTA, **6**(2), 149, 2018, doi:10.24843/ipta.2018.v06.i02.p09.
[5] IBM, Bivariate Correlations, 2019.
[6] P. Irfan, A. Apriani, "Analisa Strategi Pengembangan E-Tourism Sebagai Promosi Pariwisata Di Pulau Lombok," ILKOM Jurnal Ilmiah, **9**(3), 325–330, 2017, doi:10.33096/ilkom.v9i3.164.325-330.
[7] H. Oppewal, "Causal Research," Wiley International Encyclopedia of Marketing, 2010, doi:10.1002/9781444316568.wiem02001.
[8] UCLA, What does Cronbach's alpha mean?, 2019.
[9] Kristin Yeager, LibGuides: SPSS Tutorials: Pearson Correlation, 2019.
[10] V. Pujani, Meuthia, A.V. Iboo, "Adopsi E- tourism pada Industri Perhotelan di Kota Padang : Identifikasi Klaster Berdasarkan Karakteristik Fitur Website," TEKNOSI, **02**(03), 47–56, 2016.
[11] K. Chulmo, J. Youhee, H. Heejeong, C. Namho, "A structural model for destination travel intention as a media exposure: belief-desire- intention model perspective," International Journal of Contemporary Hospitality Management, **28**(7), 2016.
[12] A. Murtadho, M. Rifki Shihab, "ANALISIS SITUS E-TOURISM INDONESIA: STUDI TERHADAP PERSEBARAN GEOGRAFIS, PENGKLASIFIKASIAN SITUS SERTA PEMANFAATAN FUNGSI DAN FITUR," Journal of Information Systems, **7**(1), 13–25, 2011.
[13] World Tourism Organization, "Global Report on The Power of Youth Travel," Affiliate Members Global Report, **13**, 60, 2016.
[14] E. Stiakakis, M. Vlachopoulou, "The impact of social media on travelers 2.0," Tourismos, **12**(3), 48–74, 2017.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 689-698 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Hand-Based Biometric Recognition Technique - Survey

Katerina Prihodova*, Miloslav Hub

Faculty of Economics and Administration, University of Pardubice, 532 10, Czech Republic

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 01 November, 2020
Online: 24 November, 2020

Keywords:
Hand geometry
Biometric recognition
Hand-base authentication

ABSTRACT

In today's modern world, reliable automatic personal recognition is a crucial area of discussion, primarily because of the increased security risks. A large number of systems first require the recognition of a person before they can access their services. Biometric recognition can be used, which can be understood as automatic identification or automatic verification of persons based on physiological or behavioural characteristics. Examples of biometric characteristics may be the fingerprint, face, signature, hand geometry. There are many approaches, and each has its pros and cons. However, there is currently no current extensive research and evaluation of hand-based biometric systems. Given the user-friendliness of these systems and the needs of society, this article aims to compare different methods of biometric recognition based on hand, so that the article reduces the entry barrier into this area of research. Furthermore, the article aims to determine the research gap and suggest possible directions for research in the future.

1. Introduction

This paper is an extension of work originally presented in International Conference on Information and Digital Technologies [1]. In today's modern society, there is an increasing emphasis on working with information. Information is one of the most valuable assets we can own, so it needs to be protected. Among the basic requirements for their security are their confidentiality, integrity and availability. The requirement of confidentiality information requires that only authorized entities have access to it. One way to ensure the confidentiality of information is through authentication, during which the identity of the entity to which the relevant rights are subsequently assigned is verified [2].

Authentication can be defined as: “A method of verifying a user's identity on a system to control access to the system or resources.” [3] Authentication can be performed using knowledge (password, pin), an authentication object (token, mobile phone), or using what we are physically and mentally (biometric characteristics). The last option is used by biometric recognition systems, which have come to the fore, especially in recent years, thanks to scientific and technical progress.

Biometric systems are applications of biometric technologies that allow automatic identification or verification of a specific natural person, using physicavl (e.g. face, fingerprint) or behavioural (e.g. voice, gait) characteristics. Biometric characteristics must be universal (the characteristics must exist for all persons), they must have high external variability (high diversity for different persons) and low internal variability (low diversity for one person).

Identification ("One-To-Many Matching") is understood in this article as the process of comparing a set of biometric features obtained from a scanned biometric sample with all reference templates stored in a database. If there is a reference template in the database that sufficiently corresponds to the set of biometric features, the biometric system identifies the person by the identity associated with this reference template, otherwise the person is not identified [4].

Verification ("One-To-One Matching") is understood in this article as the process of verifying that the person being examined is who he or she claims to be. The set of biometric features obtained from the scanned sample is compared with one own biometric template, which is stored in a database. If the set of biometric features sufficiently matches this reference template, the identity is verified, otherwise it is not verified [4].

Biometric recognition of people is based on their biometric characteristics, and it is their unique physiological and behavioural characteristics. Physiological characteristics are those characteristics with which we were born and include, e.g. iris [5, 6], retina [7], face [8], shape of the outer ear [9], fingerprints [10], palmprints [11] and footprints [12], fingers geometry [13], hand geometry, [14], topography (layout and shape) of wrist veins

*Corresponding Author: Katerina Prihodova, st30351@student.upce.cz

[15], DNA [16]. Behavioural characteristics include, e.g. voice [17], writing [18], signature [19], and gait [20].

From the biometric characteristics, distinguishing repeatable biometric features can be obtained for biometric recognition. Biometric features are numbers or designations obtained from biometric samples that are used for subsequent comparison. Biometric samples are an analogue or digital representation of biometric characteristics before the actual extraction of biometric features [21].

The article aims to provide a comprehensive survey with the inclusion of most recent research papers up to 2020, covering a total of 117 journal publications, conference proceeding, patents, and papers. The aim is also lower the entry barrier to this research area by providing a comprehensive reference for novices. Offer a wide range of comparisons in diverse research angles and perspectives in hand geometry recognition. Identify current trends in this area and the need for future research.

2. Development of Hand-Based Biometric Systems

The first modern biometric systems used the hand geometry as a biometric characteristic, and then it was an authentication system in the verification mode (the user submitted an identity and it was subsequently confirmed or rejected). Robert Miller dealt with the issue of hand-based biometric systems in the 1960s and based on research, and he created a device at the Stanford Research Institute, which he patented in 1971. This device performed the pattern recognition by purely mechanical means, where four spring bars slid towards the user's fingers and subsequently the length of the user's fingers was measured. If the finger lengths matched the pattern embossed on the user's identity card, a simple switch was activated, which was used to operate the electric lock [4, 22].

The first commercial hand scanner was the Identimat, based on the patent Robert Miller's as mentioned above [22]. The measurement of the hand was performed by scanning the photocells located under grooves on which the user placed the hand (fingers). The light source was light with a power of 1000 watts. A reading head with a magnetic stripe was attached to this scanning mechanism, to which a user's magnetic card was attached during verification. If the pulses on the card matched the signal times from the photocell sensors within the threshold value, the user's identity was verified. This system was successfully used in various applications. The first use was of a military nature, and it was the security of nuclear weapons. Later, Identimat was used to access US nuclear power plants and also to control employee attendance at an investment company on Wall Street. The second development line of Identimat was already two-dimensionally oriented, measuring not only the length but also the width of all fingers on one hand [4, 23].

The history of the practical implementation of hand geometry recognition is further associated with the name David Sidlauskas and his company Recognition System Inc., which he founded in 1986. This company was the first in the world to start developing and selling modern hand scanners. In the mid-1980s, David Sidlauskas patented the first biometric system to use a three-dimensional hand image. The patented system consists of a camera and an optical measuring board. The camera has a plan view and a side view of the hand. The identification code is entered using the keypad [24]. Based on this patent, the first commercially successful biometric system using the hand geometry of the Handkey ID3D scanner was created.

BioMet Partners, Inc. developed a two-finger version of the Digi-2 hand scanner and was launched in 1995. The scanner uses a CCD camera to scan only the index finger and middle finger and obtain their 3D image [4].

In the 1990s, biometric hand scanners were used in the INSPASS program, a program at selected US airports that accelerated the check-in of frequent travellers. These persons could apply for inclusion in a program in which their hand geometry was scanned and then obtained an identification card with a reference template. These persons were able to complete check-in quickly at selected airports with the help of installed hand scanners [25].

Another hand-scanner recognition project was an Israeli project that automated motion control on the Israeli-Palestinian border. During the creation of the biometric system, suitable biometric characteristics were selected. Biometric characteristics face, hand geometry and fingerprints were considered. A multi-biometric system was chosen that combined face recognition and hand geometry, which was chosen mainly because most people crossing the border are workers whose fingerprints are not reliable for biometric verification [26].

In terms of security, hand scanners were also used at the 1996 Atlanta Olympics, which controlled access to the Olympic Village. Hand scanners were also used by the Colombian legislature, which voted safely through them [27]. Until 1998, only two scientific articles in the field of hand-based biometric systems were published, namely: "Vital Signs of Identity" a "A Performance Evaluation of Biometric Identification Devices." [2, 23, 28] From 1998 to 2006, there was an almost exponential increase in the number of publications in this area [29].

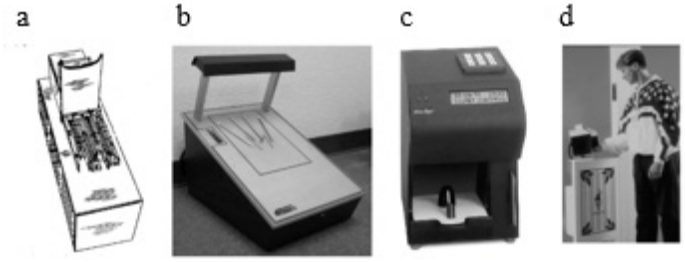


Figure 1: a) Drawing of mechanical hand reader [22] b) Identimat [30] c) two-finger scanner from Biomet [31] d) utilization hand geometry scanner at the Atlanta Olympics [32]

3. Current methods in hand-based biometric systems

Current hand-based biometric systems can be divided into the following groups according to biometric characteristics [33]:

- hand geometry and handshape,
- palmprint,
- fingerprint,
- topography (finger, hand).

At the beginning of the research of hand-based biometric systems, hand geometry was used, e.g. widths, lengths, angles [35-37], later the silhouette of the hand was added as another biometric characteristic [38-40].

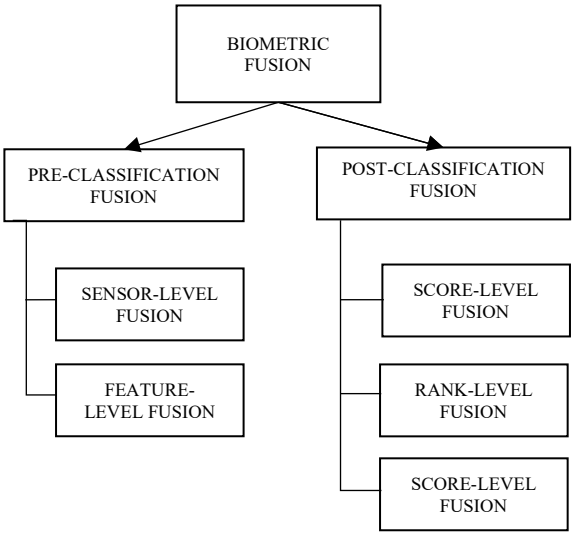


Figure 2: Fusion at various levels of multi-biometric systems (source: [34])

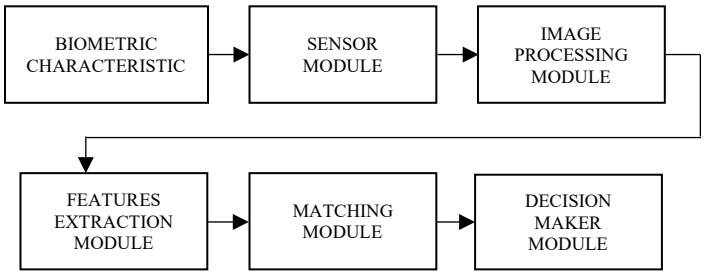


Figure 3: General model of biometric system

Currently, there are also multi-biometric systems that use a combination of multiple biometric characteristics. A palmprint [41-45] or bloodstream topography [46-48] is used in combination with hand geometry or silhouette. The topography of bloodstream is measured either on the palm or back of the hand when the whole hand is scanned. Systems have also been developed that combine more than two characteristics, such as hand geometry, palmprint, and fingerprint [49, 50]. Multi-biometric systems achieve higher accuracy [51] than biometric systems using only one biometric characteristic. Newly emerging biometric systems also use as a characteristic the heartbeat captured on the finger [52], pores on the hands [53], or hand gestures [44].

Combining multiple biometric characteristics in one biometric system requires fusion, which is the interconnection of information. Fusion can be achieved at different levels of the biometric system. Fusion levels can be divided into two large groups: pre-classification fusion and post-classification fusion. Pre-classification fusion include sensor-level fusion and feature-level fusion while post-classification fusion includes score-level fusion, rank-level fusion rank and decision-level fusion [34]. The individual fusion levels are shown in Figure 2.

The biometric recognition system has several modules: the sensor captures the biometric characteristic, followed by image processing, extraction of identification features, comparison with the biometric template and decision. The individual modules of the biometric system are shown in its general model in Figure 3.

3.1. Sensor

CCD chip cameras, a light-sensitive image detector used by digital cameras, camcorders, and webcams, e.g. are most commonly used to capture biometric characteristics [24, 53]. The scanner can also be a biometric sensor [42, 49, 54]. The sensor captures either a 2D [55-57] or a 3D [48, 58, 59] image of the hand. 3D images are obtained using two cameras, with the help of mirrors, or they can be obtained using a 3D digitizer [60].

Lighting also plays an essential role in obtaining a biometric sample. Some systems have their own light source, mainly commercial biometric scanners. Most sensors acquire images of the hand in visible light, but some also use infrared radiation [47, 61], which achieves a higher accuracy of biometric recognition [62].

Today, commercially available hand readers use spacer pins that firmly determine the position of the hand when scanning [48, 58, 63]. As spacer pins can deform the silhouette of the hand [64], systems without spacer pins have come to the fore in scientific circles [53, 65, 66]. The latest possibilities are contactless systems [38, 61, 67, 68], where, in addition to the spacer pins, direct contact with the reader is not required. These systems are considered to be the most user-friendly, mainly due to the elimination of the need for physical contact.

3.2. Image processing

After obtaining the image of the hand, it is necessary to process the acquired image. The first step in image processing is image pre-processing. It is a set of methods that are applied to an image because only a minimal percentage of hand images are taken under optimal conditions. Almost no captured image is completely noise-free, not entirely focused, obtained in optimal lighting conditions. For this reason, filters are used to sharpen the image or to remove noise. It is also possible to change the brightness using geometric transformation methods. [69]

The second step of image processing is image segmentation, i.e. dividing the image into segments with common properties. The aim of segmentation in hand-based biometric systems is to separate the hand from the rest of the image automatically. The more hand the image is obtained in more limited conditions, the easier the segmentation and can be done by faster methods. Segmentation methods can be divided into groups [69]:

- statistic methods,
- methods based on edge detection,
- hybrid methods (morphological),
- knowledge methods.

If the hand image is acquired on a scanner, the background is constant, and segmentation is relatively simple, then the segmentation method of image thresholding can be used. Image thresholding is the simplest statistical segmentation method based on pixel brightness evaluation. During thresholding, problems arise with rings, bracelets, or dirty skin [70-72].

Segmentation of the hand image can also be performed using edge detection methods. Edges are detected by edge detectors based on significant differences in the values of neighbouring pixels [37, 73].

There are also multi-stage segmentation methods where multiple methods are combined. These multi-stage methods can detect a hand from a crowded background. An example of such a method is a method that first detects the skin by colour (knowledge) and then detects the shape of the hand [74].

Another method of segmentation and at the same time localization of biometric features is AAM (Active Appearance Model) [55]. The AAM method presents the hand using an appearance model, which is created from a shape model and a hand texture model based on a statistical method - principal component analysis (PCA). The advantage of the method is the complexity of the information it contains about the hand [75].

The disadvantage of some segmentation methods is that their use breaks the silhouette of the hand. Mathematical morphology is a mathematical tool that can solve this problem and connect disconnected fingers. Mathematical morphology is also used to remove rings, bracelets or watches [56, 76, 77].

If the position of the hand is not fixed when acquiring the image, it is necessary to align the image of the hand after segmentation within the image processing. This can be done using landmarks, which can be extremes (fingertips, finger valleys) [36, 57, 78], of the centre of gravity, or a fixed reference point on the wrist [79].

Image processing is performed to facilitate the extraction of the region of interest. Biometric features are then extracted from the area of interest.

3.3. Extraction of biometric features

The image processing is followed by the extraction of biometric features. Extraction of biometric features is defined in [21] as*: "A process applied to a biometric sample to isolate and obtain a repeatable output of characteristic numbers or designations that can be compared with features extracted from other biometric samples."* For the biometric characteristic of hand geometry, biometric features are geometric measurements, such as finger lengths and widths, finger area, radii of circles on fingers and palms, and angles between fingers. Figure 4 shows 17 geometric measurements, but some systems use a different number of measurements, most often 13-50 measurements [35, 49, 80].

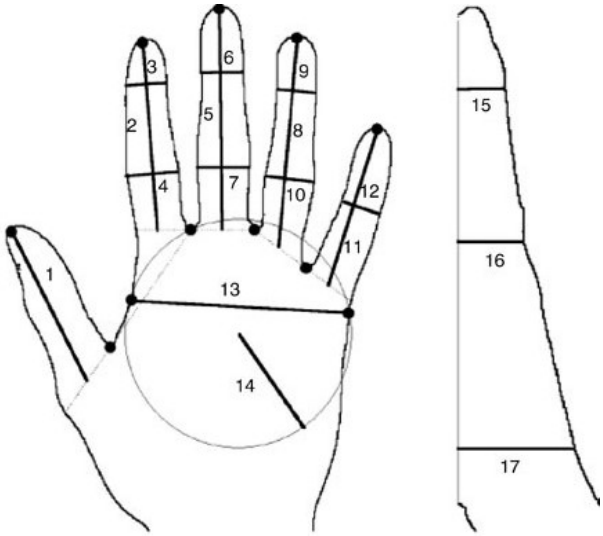


Figure 4: Geometric measurements (source: [81])

More information than hand measurements is provided by the silhouette of the hand, which is used by many systems [40,70]. These systems use biometric features to extract generic or customized feature extractors. For example, customize shape extractors [40, 82, 83] or figure features extractors algorithms Scale Invariant Feature Transform (SIFT) [82], LBP [84]. They are also used palm-based textures [85]. The silhouette is also distinguishable by its contour pixels; it represents the real shape of the hand [67].

Contours are less explored than geometric features, especially since geometric features are easier to determine from a normalized hand image. The issue of normalization of various hand positions and accurate localization of fingertips and finger valleys in unrestricted and contactless systems requires further research. [67]

Some systems extract biometric features of hand geometry and hand shape at the same time, extractors are specially adapted to the shape and geometry of the hand [86, 79].

Recent research uses deep neural networks, such as convolutional neural networks [87, 88] or Siamese networks, to extract biometric features of the hand [87].

3.4. Comparison with the reference template

In this phase, a set of biometric features obtained from a recognized person's biometric sample is compared with a reference template or templates (stored in a database). The score is then generated using the following methods:

- statistic methods,
  - minimum distance (Euclidean metric, Hamming metric, Chebyshev metric),
  - Kullback–Leibler divergence,
  - correlation coefficient,
  - Gaussian mixture models (GMM),
- machine learning,
  - support vectors machine (SVM),
  - k-nearest neighbours,
  - Bayesian networks,
  - neural networks.

Statistic methods

Statistical methods of comparing a set of biometric features with a reference template when recognizing a person based on the geometry and shape of the hand include a comparison by a minimum distance. If M reference templates are stored in the database x_{1R}, x_{2R}, …, x_{MR}, the minimum distance of the hand biometric sample x from the reference template is calculated as follows:

$$\min_{\forall s} \|x_{sR} - x\| \quad (1)$$

x_{sR}…… reference templates

s …… number of reference templates

x …… biometric features set (vector)

When recognizing a person based on hand geometry and hand shape, various metrics are used to determine the distance, such as the Euclidean metric [37, 56, 70], the Hamming metric [39], or the Chebyshev metric [80].

- Euclidean metric is defined by the relation:

$$\rho_E(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2} \quad (2)$$

- Hamming's metric, sometimes called Manhattan, is defined by relations:

$$\rho_H(\mathbf{x}_1, \mathbf{x}_2) = \sum_{i=1}^{n}|x_{1i} - x_{2i}| \quad (3)$$

- Chebyshev's metric is defined by the relation:

$$\rho_C(\mathbf{x}_1, \mathbf{x}_2) = \max_{\forall i}|x_{1i} - x_{2i}| \quad (4)$$

The Kullback – Leibler divergence can also be used to express "distance", which expresses the degree of dissimilarity between two probability distributions [36]. The correlation coefficient can also be used as a measure of similarity [81, 89].

Gaussian mixture models are located at the interface of statistical and machine methods (GMM). It is a pattern recognition method that is based on the GMM's ability to approximate any probability density using a mixture of Gaussian densities. The probability of the vector of characteristics x_n is estimated as a weighted sum of Gaussian densities. The result is, therefore, the probability with which the pattern (hand image) belongs to a given class. This method gives better results in the field of recognizing a person based on hand geometry than methods working with distance [37, 59].

The Gaussian mixture model is defined by the relation:

$$p(\vec{x}|u) = \sum_{i=1}^{M} \frac{c_i}{(2\pi)^{L/2}|\Sigma_i|^{1/2}} exp\left\{-\frac{1}{2}(\vec{x} - \overrightarrow{u_i})^T {\Sigma_i}^{-1}(\vec{x} - \overrightarrow{u_i})\right\} \quad (5)$$

u ……identity of the person to whom the hand image is assigned with probability p

x ……. vector of biometric features

c_i …… weight of individual Gaussian models

μ_i …… average vector model

Σ_i …… covariance matrix

M …… number of models

L ……. dimension vector of biometric features

Machine learning

Machine learning is a group of algorithms that allow a computer system to learn. Mittchell [90] d defined machine learning as follows: *"A computer program is said to learn from experience E concerning some classes of T and the measured power P, if its performance P improves with experience E."*

A large and used group of machine learning are algorithms called support vectors or support vector machines (SVM). These are classifiers that use boundaries in image space. SVM aims to find a superstructure that divides the symptom space. SVM used Vinodkumar and Srikantaswamy as a classifier in a multi-biometric system, where the biometric characteristics were hand geometry, fingerprints and palmprint [91]; however, they have also been used in many other hand-based multi-biometric systems [41, 42, 49, 92].

A frequently used machine learning method in hand-based biometric systems is the nearest neighbour (k-NN) method [54, 93]. It is one of the simplest classification methods. Bayesian networks [93, 94] or random forest [84] can also be used for comparison with the pattern.

Some types of neural networks, multilayer perceptron (MLP) can also be used as classifiers [95]. Generalized regression neural network (GRNN) [96] and convolutional neural network (CNN) were used as classifiers that were not preceded by classical features extraction, classification was performed using SOFTMAX [86].

3.5. Decision

In most systems, the final decision on a user's identity is based on a threshold that is chosen so that the biometric system meets security and throughput requirements. The score calculated in the previous system module is compared with the threshold value. If the score is higher than the threshold, the user is identified or verified, but if the score is lower than the threshold, the user is not identified or verified. The outcome of decisions for hand-based biometric systems is significantly influenced by user training, with the right habits of users, the score increases (agreement of the compared samples) [97, 98].

4. Quantitative evaluation of biometric system

Possible identification results are shown in Table 1.

Table 1: Possibilities of recognition decision

		Decision	
		Accept	Reject
Actual output	Authorized user	Correct acceptance	Incorrect rejection
	Unauthorized intruder	Incorrect acceptance	Correct rejection

Source: [2]

Based on the decision, the performance of biometric systems can be quantified. Selected quantified identifiers are [99].

False Rejection Rate – FRR

The probability that the biometric system does not recognize the authorized user will lead to an erroneous rejection, i.e. a type 1 error. If an error of this type occurs, the user can try to prove his identity again. The FRR is calculated according to the formula:

$$FRR = P(FR/OU) = \lim_{N_{OU}\to\infty} \frac{N_{FR}}{N_{OU}} \quad (6)$$

FR… occurrence of a rejection error

OU … recognition (verification or identification) by an authorized user

N_{FR}… number of false rejections

N_{OU}… number of attempts by authorized users to recognize

False Acceptance Rate – FAR

The probability that the biometric system will accept an unauthorized intruder will lead to incorrect acceptance, i.e. a type 2 error.

The FAR calculation is considered according to the formula:

$$FAR = P(FA/NN) = \lim_{N_{NN}\to\infty} \frac{N_{FA}}{N_{NN}} \quad (7)$$

FA… occurrence of an acceptance error

NN… recognition by an unauthorized intruder

N_{FA}… number of incorrect acceptances

N_{NN}… the number of attempts by an unauthorized intruder to recognize

The FAR and FRR rates depend on the threshold and are also inversely proportional. The higher the threshold, the more secure the biometric system, but the more common the type 1 error occurs and the user-friendliness decreases. On the contrary, the lower the threshold value, the more user-friendly the biometric system, but more often an unauthorized intruder, i.e. a type 2 error, is admitted, and the security of the system is reduced. The threshold value must always be set according to the requirements that are placed on the biometric system.

Equal Error Rate – EER

Value where the error rates of both FAR and FRR are equal, (i.e. FAR = FRR). With such a setting, the system incorrectly rejects and incorrectly identifies the same number of people. The lower the EER value, the higher the accuracy of the system. It is the most widely used indicator for evaluating and comparing biometric systems.

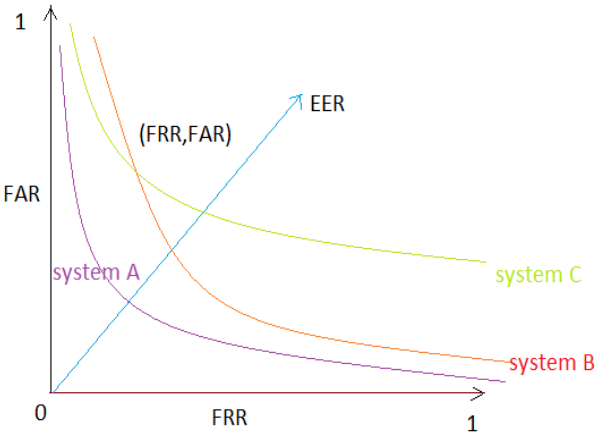


Figure 5: ROC Curves [100]

Receiver Operating Characteristic – ROC

The probability's curve that is given by the relationship between the probability of false acceptance (FAR) on the x-axis and the corresponding probability of correct acceptance on the y-axis for all allowable threshold values. Using ROC curves, multiple models can be compared with each other. Comparison and evaluation of multiple ROC curves are performed using the Area Under Curve (AUC) areas, which are located below the individual ROC curves [101]. Figure 5 shows the ROCs of the three systems, system A being the best system under the circumstances.

EER, FRR, FAR, ROC was selected for this work. They are a good indicator and the most frequently used tool for evaluating and managing biometric systems.

In articles dealing with biometric systems, we also encounter the concept of accuracy. The accuracy is calculated according to the formula:

$$\text{accuracy} = \frac{CD}{NR} \quad (8)$$

CD … correct decision (correct rejection and correct acceptance)

NR… number of recognition attempts

5. Comparison

The most commonly used physiological characteristics in biometric are fingerprint, face, hand geometry, and iris [102]. Table 2 compares biometric systems that use commonly biometric characteristics based on EER. Of these biometric characteristics, most often the lowest EER is achieved, biometric systems working with the iris, but as can be seen in Table 2, other systems, such as fingerprints, also achieve very low EER. However, in addition to error rates compilation speed, acqurability, privacy, cost and ease of use, should also play a role in choosing a biometric system [102].

Table 2: Comparison of selected hand other biometric systems

Sources	Biometric Characteristic of Recognition	Size of the Database	Performance
Chen et al. [103]	Iris	3 500	EER = 0.43
Yingzi et al. [104]	Iris	610	EER = 0.0295
Wijaya [105]	Face	2 268	EER = 0.0457
Marvadi [106]	Face	427	EER = 0.1525
Liu et al. [107]	Fingerprint	350	EER = 0.0042
Jain & Feng [108]	Fingerprint	449	EER = 0.081
Varchol & Levický [109]	Hand geometry	408	EER = 0.0462
Jain et al. [110]	Hand geometry	500	EER= 0.06

Table 3 summarizes the essential characteristics of some systems and methods that appeared in the journal and conferences. Some systems have been tested in identification mode, some in verification mode. Only some papers report EER, FAR / FRR or Accuracy values, some authors evaluate systems using ROC curves. The selection in Table 3 seeks to compare a representative sample of hand-base systems by characteristics (Hand-base Biometric Characteristic, Mode, Techniques Applied for Features Extraction, Techniques Applied for Recognition, Size of the Database, Performance).

Older systems use geometric measurements as features and are mostly in verification mode [110, 115]. Newer systems [111, 117] use a combination of multiple biometric characteristics such as hand shape and hand texture and use deep learning methods. Today's systems are mainly in identification mode, and FRR or EER has less than 1%.

Table 3: Comparison of selected hand based biometric system

Sources	Biometric characteristic / Mode	Techniques Applied for Features Extraction	Techniques Applied for Recognition	Size of the Database	Performance
Afifi [111]	Hand texture and shape / identification	CNNs + LBP	Support vector machine (SVM)	11 076	EER = 0.009
Kumar & Zang [112]	Hand geometry / identification	discretization of hand-geometry features	SVM	1 000	Accuracy = 0.94
Shanmukhappa & Sanjeevakumar [113]	Hand geometry /identification	Feature vector construct by hand image graph	SVM (Radial basic function)	1 440	FRR = 0.0205
Jain et al. [110]	Hand geometry/ verification	Geometric measurements	Euclidean distance	500	EER= 0.06
Varchol & Levický [109]	Hand geometry/ verification	Geometric measurements	Hamming distance	408	EER = 0.0973
Villegas et al. [114]	Hand geometry /identification	Wavelet Features	Nearest Neighbour	120	FAR = 0.11 FRR = 0.1
Varchol & Levický [109]	Hand geometry/ verification	Geometric measurements	Gaussian Mixture Model	408	EER = 0.0462
Burques et al. [115]	Hand geometry/ verification	Geometric measurements	Distance measure	12 800	EER = 0.0016
Charfi et al. [116]	Hand shape/identification	Scale Invariant Feature Transform (SIFT)	SIFT Matching	1 170	EER = 0.0586
Yoruk et al. [29]	Hand shape /identification	Independent Component Analysis (ICA)	L1/L2 norm of the difference of the feature vectors	458	Accuracy = 0.9731
Yoruk et al. [29]	Hand texture /identification	Principal Component Analysis (PCA)	L1/L2 norm of the difference of the feature vectors	458	Accuracy = 0.9791
Yoruk et al. [29]	Hand texture /identification	Angular radial transform (ART)	L1/L2 norm of the difference of the feature vectors	458	Accuracy = 0.976
Prihodova & Hub [117]	Hand texture and shape / identification	CCN (GoogLeNet)	CCN (Softmax)	456	FRR= 0.001

6. Research gap

It was not possible to cover all the literature dealing with hand biometric systems, but we included a representative sample of modern methods. However, we were able to identify some research gaps. To make hand-based biometric systems more efficient in real-world applications, the following challenges need to be overcome:

A large number of users – In the real world, it is often necessary to recognize millions of identities. However, the variability between some individuals may be minimal.

Quality of input data – Thanks to the progress of sensing sensors, the input image obtained in a real application under standard lighting conditions is correctly exposed. However, the image quality deteriorates in poor lighting conditions. Other problems we face despite the significant progress of the sensors are accurate colour capture and noise.

The permanence of biometric characteristics – Hand-based biometric characteristics have high stability over time but are not 100 %. For the hand geometry and hand shape, the topography of the bloodstream is primarily a problem of biological ageing and injury.

Attacks on systems – Hand-based biometric systems must face attacks at various levels in real-world applications. One of them is an attack on an input device, where the attacker forged a biometric sample, as a defence against this type of attack is the detection of liveliness.

Datasets – Small datasets are also a problem in this area of research, most current systems are tested on small datasets.

Costs – Costs should always be considered, and efforts should be made to make systems more accessible.

7. Discussion

Given the current state of hand-based biometric recognition, which was described in previous chapters of the article and due to the unresolved challenges in the field, one of the possibilities for further research is the use of a multi-biometric system, which uses hand geometry and thermal properties of hand. The thermal characteristic can be used in the form of blending two images (thermogram and image from the visible part of the electromagnetic spectrum), this blending creates a multispectral image. The multispectral image provides more significant variability between certain individuals than the mere geometry of the hand. A multispectral image is also advantageous in low lighting conditions, the quality of input data decreases less because the thermogram is not dependent on lighting conditions. When using a multispectral image, some types of attacks on biometric systems are also eliminated, because the thermal map of the hand, which is captured by the thermogram, is complicated to imitate and at the same time liveness is detected. At the same time, the multi-biometric system should increase recognition accuracy.

Another research option is also a multi-biometric system using hand geometry or palmprint in combination with skin colour (soft biometrics). This multi-biometric system could help to solve the problem of low variability between specific individuals.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

This article was supported by grant No. 2020_018 supported by the Student Grant Competition.

References

[1] K. Prihodova, M. Hub. "Biometric Privacy through Hand Geometry - A Survey" International Conference on Information and Digital Technologies (IDT). IEEE; 395-401, 2019. https://doi.org/10.1109/DT.2019.8813660
[2] M. Hub. "Bezpečnost a ochrana informací v prostředí internetu" (book in Czech) Pardubice: University of Pardubice, 2013.
[3] T. Doseděl. "Počítačová bezpečnost a ochrana dat." (book in Czech) Brno: Computer Press, 2004. ISBN 80-251-0106-1.
[4] A. Jain, P. Flynn, A. Ross. "Handbook of biometrics" New York: Springer, c2008.
[5] A. W. K. Kong, D. Zhang, M.S. Kamel. "An Analysis of IrisCode" IEEE Transactions on Image Processing. **19**(2), 522-532, 2010. DOI: 10.1109/TIP.2009.2033427.
[6] D. Nguyen, T. Pham, Y. Lee, K. Park. "Deep Learning-Based Enhanced Presentation Attack Detection for Iris Recognition by Combining Features from Local and Global Regions Based on NIR Camera Sensor" Sensors. 2018, **18**(8). DOI: 10.3390/s18082601.
[7] S. M. Lajevardi, A. Arakala, S. A. Davis, K. J. Horadam. "Retina Verification System Based on Biometric Graph Matching" IEEE Transactions on Image Processing. **22**(9), 3625-3635, 2013. DOI: 10.1109/TIP.2013.2266257.
[8] D. Wang, H. Lu, M-H Yang. "Kernel collaborative face recognition" Pattern Recognition. **48**(10), 3025-3037, 2015. DOI: 10.1016/j.patcog.2015.01.012.
[9] Ž. Emeršič, V. Štruc, P. Peer. "Ear recognition: More than a survey" Neurocomputing., **255**, 26-39, 2017. DOI: 10.1016/j.neucom.2016.08.139.
[10] A. K. Jain, J. Feng, K. Nandakumar. "Fingerprint Matching" International Journal of Computer, **43**(2), 36-44, 2010.
[11] N. Nedjah, R. S. Wyant, L. De Macedo Mourelle. "Efficient biometric palmprint matching on smart-cards for high security and privacy" Multimedia Tools and Applications. **76**(21), 22671-22701, 2017. DOI: 10.1007/s11042-016-4271-8.
[12] V. D. Ambeth Kumar, M. Ramakrishnan. "Multiple Feature Extraction for FootPrint Image" National Journal of System and Information Technology. 2012, **5**(2), 142, 2012.
[13] L. Zhu, S. Zhang. "Multimodal biometric identification system based on finger geometry, knuckle print and palm print" Pattern Recognition Letters. **31**(12), 1641-1649, 2010. DOI: 10.1016/j.patrec.2010.05.010.
[14] B. Aghili, H. Sadjedl. "Personal Authentication Using Hand Geometry" In: 2009 International Conference on Computational Intelligence and Software Engineering. IEEE, 2009, 2009, s. 1-4. DOI: 10.1109/CISE.2009.5363947.
[15] Y. Wang, D. Zhang, Q. Ql. "Liveness detection for dorsal hand vein recognition" Personal and Ubiquitous Computing. **20**(3), 447-455, 2016. DOI: 10.1007/s00779-016-0922-z.
[16] D. Frumkin, A. Wasserstrom, A. Davidson, A. Grafit. "Authentication of forensic DNA samples" Forensic Science International: Genetics. 2010, **4**(2), 95-103. DOI: 10.1016/j.fsigen.2009.06.009.
[17] "Voice biometrics growth likely" Biometric Technology Today. 2009, **17**(6), 3-4. DOI: 10.1016/S0969-4765(09)70085-2.
[18] T. Scheidat, M. Kalbitz, C. Vielhauer. "Biometric authentication based on 2D/3D sensing of forensic handwriting traces" IET Biometrics. 2017, **6**(4), 316-324. DOI: 10.1049/iet-bmt.2016.0127.
[19] T. Hafs, L. Bennacer, M. Boughazi, A. Nait-Ali. "Empirical mode decomposition for online handwritten signature verification" IET Biometrics. 2016, **5**(3), 190-199. DOI: 10.1049/iet-bmt.2014.0041.
[20] T. K. M. Lee, M. Belkhatir, S. Sanei. "A comprehensive review of past and present vision-based techniques for gait recognition" Multimedia Tools and Applications. 2014, **72**(3), 2833-2869. DOI: 10.1007/s11042-013-1574-x.
[21] Information technology - Biometric performance testing and reporting - ISO/IEC 19795-1. City, 2006.
[22] R. Miller. "Finger dimension comparison identification system" USA. US3576538 (A). Written down 27.4.1971.
[23] B. Miller. "Vital signs of identity [biometrics]" IEEE Spectrum. 1994, **31**(2), 22 30. DOI: 10.1109/6.259484.
[24] D. Sisdlouskas. "3D hand profile identification apparatus" 1988. USA. US4736203A. Written down 5.4.1988.
[25] Immigration and naturalization service passenger accelerated service system pilot program. Office of the Inspector General. Washington, DC.
[26] Technology assessment. Edtion ed. Washington, DC (PO Box 37050, Washington 20013): US General Accounting Office, 2002.
[27] D. Maltoni. "Handbook of fingerprint recognition" 2nd ed. London: Springer, c2009.
[28] X. Fan, H. Zhang, C. Leung, C. Miao. "Comparative study of machine learning algorithms for activity recognition with data sequence in home-like environment" In: 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI). IEEE, 2016, 2016, s. 168-173. DOI: 10.1109/MFI.2016.7849484.
[29] E. Yoruk, H. Dutagaci, B. Sankur. "Hand biometrics. Image and Vision Computing" 2006, **24**(5), 483-497. DOI: 10.1016/j.imavis.2006.01.020.
[30] D. P. Sidlauskas, S. Tamer. "Hand Geometry Recognition" 2008, In: Jain A.K., Flynn P., Ross A.A. (eds) Handbook of Biometrics. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-71041-9_5
[31] R. Rak, V. Matyáš, Z. Říha. "Biometrie a Identita Člověka: Ve Forenzních a Komerčních Aplikacích" (book in Czech) 1electronic edition ed. Prague: Grada; 2008.
[32] B. Lindstrom "Hand Geometry A Biometric Technology with a Pedigree BiometriTech June 2003 Bruce Lindstrom, Regional Sales Manager" 2003. https://slideplayer.com/slide/5032579/
[33] J. A. Unar, W. C. Seng, A. Abbasi. "A review of biometric technology along with trends and prospects" Pattern Recognition. 2014, **47**(8), 2673-2688. DOI: 10.1016/j.patcog.2014.01.016.
[34] A. Ross "An introduction to multibiometrics." IEEE; 2007.
[35] B. Aghili, H. Sadjedi. "Personal Authentication Using Hand Geometry" In: 2009 International Conference on Computational Intelligence and Software Engineering. IEEE, 2009, 2009, s. 1-4. DOI: 10.1109/CISE.2009.5363947.
[36] G. Fouquier, L. Likforman, J. Darbon, B. Sankur. "The Biosecure Geometry-Based System for Hand Modality" In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07. IEEE, 2007, 2007, I-801-I-804. DOI: 10.1109/ICASSP.2007.366029. ISBN 1-4244-0727-3.
[37] R. Sanchez-Reillo, C. Sanchez-Avila, A. Gonzalez-Marcos. "Biometric identification through hand geometry measurements" IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000, **22**(10), 1168-1171. DOI: 10.1109/34.879796.
[38] De Santos Sierra, A. J. G. Casanova, C. S. Avila, V. J. Vera. "Silhouette-based hand recognition on mobile devices" In: 43rd Annual 2009 International Carnahan Conference on Security Technology. IEEE, 160-166, 2009. DOI: 10.1109/CCST.2009.5335548.
[39] C. Oden, A. Ercil, B. Buke. "Combining implicit polynomials and geometric features for hand recognition" Pattern Recognition Letters. 2003, **24**(13), 2145-2152. DOI: 10.1016/S0167-8655(03)00087-4.
[40] E. Yoruk, E. Konukoglu, B. Sankur, J. Darbon. "Shape-based hand recognition" In: IEEE Transactions on Image Processing. IEEE, 2006, **15**(7), s. 1803-1815. DOI: 10.1109/TIP.2006.873439.

[41] S. A. Angadi, S. M. Hatture. "Biometric Person Identification System: A Multimodal Approach Employing Spectral Graph Characteristics of Hand Geometry and Palmprint" International Journal of Intelligent Systems and Applications. 2016, 8(3), 48-58. DOI: 10.5815/ijisa.2016.03.06.

[42] M. Ferrer, C. Travieso, J. Alonso. "Multimodal Biometric System based on Hand Geometry and Palm Print Texture" In: Proceedings 40th Annual 2006 International Carnahan Conference on Security Technology. IEEE, 2006, 2006, s. 92-97. DOI: 10.1109/CCST.2006.313435.

[43] A. Kirthika, S. Arumugam. "Texture and colour intensive biometric multimodal security using hand geometry and palm print. "International Journal of Advances in Engineering & Technology, **4**(1), 226-235, 2012.

[44] T. O. Oladele, K. Adeniyi, T. O. Aro. "Framework for User Authentication at a Distance for Mobile Phones Using Contactless Hand-based Multimodal Biometric System" Journal of Computer Science and Control Systems. ProQuest Central, 2019, **12**(1), 24-27, 2019.

[45] W.-C. Wang, W.-S. Chen, S.-W. Shih. "Biometric recognition by fusing palmprint and hand-geometry based on morphology" In: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2009, 2009, s. 893-896. DOI: 10.1109/ICASSP.2009.4959728.

[46] P. Gupta, S. Srivastava, P. Gupta. "An accurate infrared hand geometry and vein pattern-based authentication system" Knowledge-Based Systems.**103**, 143-155, 2016. DOI: 10.1016/j.knosys.2016.04.008. ISSN 09507051.

[47] E. C. Lee, H. Jung, D. Kim. "New Finger Biometric Method Using Near Infrared Imaging" Sensors. **11**(3), 2319-2333, 2011. DOI: 10.3390/s110302319.

[48] C. Oden, A. Ercil, B. Buke. "Combining implicit polynomials and geometric features for hand recognition "Pattern Recognition Letters.**24**(13), 2145-2152, 2003. DOI: 10.1016/S0167-8655(03)00087-4.

[49] M. A. Ferrer, A. Morales, C. M. Travieso, J. B. Alonso. "Low Cost Multimodal Biometric Identification System Based on Hand Geometry, Palm and FingerPrint Texture" In: 2007 41st Annual IEEE International Carnahan Conference on Security Technology. IEEE, 2007, 2007, s. 52-58. DOI: 10.1109/CCST.2007.4373467.

[50] F. Yang, B. Ma. "A New Mixed-Mode Biometrics Information Fusion Based-on Fingerprint, Hand-geometry and Palm-print" In: Fourth International Conference on Image and Graphics (ICIG 2007). IEEE, 2007, 2007, s. 689-693. DOI: 10.1109/ICIG.2007.39.

[51] A. Ross, A. Jain. "Information fusion in biometrics. Pattern" Recognition Letters. 2003, **24**(13), 2115-2125. DOI: 10.1016/S0167-8655(03)00079-5.

[52] M. Kim, W.-Y. Kim, J. Paik. "Optimum Geometric Transformation and Bipartite Graph-Based Approach to Sweat Pore Matching for Biometric Identification" Symmetry. 2018, **10**(5), 175. DOI: 10.3390/sym10050175.

[53] M. Adán, A. Adán, A. S. Vázquez, R. Torres. "Biometric verification/identification based on hands natural layout" Image and Vision Computing. 2008, **26**(4), 451-465. DOI: 10.1016/j.imavis.2007.08.010.

[54] S. Gonzeilez, C. M. Travieso, J. B. Alonso, M. A. Ferrer. "Automatic biometric identification system by hand geometry" In: IEEE 37th Annual 2003 International Carnahan Conference onSecurity Technology, 2003. Proceedings. IEEE, 2003, s. 281-284. DOI: 10.1109/CCST.2003.1297573.

[55] R. Gross, Y. Li, L. Sweeney, X. Jiang, W. Xu, D. Yurovsky "Robust Hand Geometry Measurements for Person Identification using Active Appearance Models" In: 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems. IEEE, 2007, 2007, s. 1-6. DOI: 10.1109/BTAS.2007.4401936.

[56] N. Charfi, H. Trichili, A. M. Alimi, B. Solaiman. "Personal recognition system using hand modality based on local features" In: 2015 11th International Conference on Information Assurance and Security (IAS). IEEE, 2015, 2015, s. 13-18. DOI: 10.1109/ISIAS.2015.7492764.

[57] S. Xu, M. Li, J. F. Ding, Y. Q. Cui. "Personal Identification by Fusing Hand Shape Geometry and Palmprint Features" Applied Mechanics and Materials, 278-280, 1228-1231, 2013. DOI: 10.4028/www.scientific.net/AMM.278-280.1228.

[58] A. Iula, G. Hine, A. Ramalli, F. Guidi. "An Improved Ultrasound System for Biometric Recognition Based on Hand Geometry and Palmprint" Procedia Engineering.**87**, 1338-1341, 2014. DOI: 10.1016/j.proeng.2014.11.709.

[59] R. Sanchez-Reillo. "Hand geometry pattern recognition through Gaussian mixture modelling" In: Proceedings 15th International Conference on Pattern Recognition. ICPR-2000. IEEE Comput. Soc, 2000, s. 937-940. DOI: 10.1109/ICPR.2000.906228.

[60] V. Kanhangad, A. Kumar, D. Zhang. "Human hand identification with 3D hand pose variations" In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. IEEE, 2010, 2010, s. 17-21. DOI: 10.1109/CVPRW.2010.5543236.

[61] J.-M. Guo, C.-H. Hsia, Y.-F. Liu, J.-C. Yu, M.-H. Chu, T.-N. Le. "Contact-free hand geometry-based identification system. Expert Systems with Applications" 2012, **39**(14), 11728-11736. DOI: 10.1016/j.eswa.2012.04.081.

[62] A. Morales, M. A. Ferrer, J. B. Alonso, C. M. Travieso. "Comparing infrared and visible illumination for contactless hand based biometric scheme." In: 2008 42nd Annual IEEE International Carnahan Conference on Security Technology. IEEE, 2008, 2008, s. 191-197. DOI: 10.1109/CCST.2008.4751301.

[63] A. Jian, N. Duta. "Deformable matching of hand shapes for user verification" In: Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348). IEEE, 1999, s. 857-861. DOI: 10.1109/ICIP.1999.823019.

[64] M. Ferrer, A. Morales, C. M. Travieso, J. B. Alonso. "Influence of the pegs number and distribution on a biometric device based on hand geometry" In: 2008 42nd Annual IEEE International Carnahan Conference on Security Technology. IEEE 2008, 221-225, 2008. DOI: 10.1109/CCST.2008.4751305.

[65] G. Boreki, A. Zimmer. "Hand Geometry: A New Approach for Feature Extraction" In: Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05). IEEE, 2005, s. 149-154, 2005. DOI: 10.1109/AUTOID.2005.33.

[66] M. Faundez-Zanuy. "Biometric verification of humans by means of hand geometry" In: Proceedings 39th Annual 2005 International Carnahan Conference on Security Technology. IEEE, 2005, 2005, s. 61-67. DOI: 10.1109/CCST.2005.1594816.

[67] A. Bera, D. Bhattacharjee, M. Nasipuri. "Finger contour profile-based hand biometric recognition" Multimedia Tools and Applications. 2017, **76**(20), 21451-21479. DOI: 10.1007/s11042-016-4075-x.

[68] J. Doublet, O. Lepetit, M. Revenu. "Contactless Hand Recognition Based on Distribution Estimation" In: 2007 Biometrics Symposium. IEEE, 2007, 2007, s. 1-6. DOI: 10.1109/BCC.2007.4430547.

[69] M. Španěl, V. Beran. "Obrazové segmentační techniky: Přehled existujících metod" (book in Czech) Brno, Brno University of Technology, 2006.

[70] G. Amayeh, G. Bebis, A. Erol, M. Nicolescu. "Peg-Free Hand Shape Verification Using High Order Zernike Moments." In: 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06). New York: IEEE, 2006, s. 40-40. DOI: 10.1109/CVPRW.2006.155.

[71] A. M. Badawi, M.S. Kamel. "Free style hand geometry verification system" In: 2003 46th Midwest Symposium on Circuits and Systems. IEEE, 2003, s. 324-327, 2003. DOI: 10.1109/MWSCAS.2003.1562284.

[72] M. Golfarelli, D. Maio, D. Malton. "On the error-reject trade-off in biometric verification systems" IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997, **19**(7), 786-796. DOI: 10.1109/34.598237.

[73] J. Mintalvao, L. Molina, J. Canuto. "Robust hand image processing for biometric application" Pattern Analysis and Applications. 2010, **13**(4), 397 407, 2010. DOI: 10.1007/s10044-010-0185-7.

[74] A. Bapat, V. Kanhangad. "Segmentation of hand from cluttered backgrounds for hand geometry biometrics" In: 2017 IEEE Region 10 Symposium (TENSYMP). IEEE, 2017, 2017, s. 1-4. DOI: 10.1109/TENCONSpring.2017.8070016.

[75] J. Král, "Rozpoznávání výrazu práce" Brno, 2010. Thesis. Technical University Brno, 2010.

[76] N. Covavisaruch, P. Prateepamornkul, P. Ruchikachorn, P. Taksaphan. "Personal verification and identification using hand geometry" ECTI Transactions on Computer and Information Technology (ECTI-CIT), **1**(2), 134-140, 2005.

[77] V. Kanhangad, A. Kumar, D. Zhang. "Combining 2D and 3D hand geometry features for biometric verification" In: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2009, 39-44, 2009. DOI: 10.1109/CVPRW.2009.5204306.

[78] X. Jiang, W. Xu, L. Sweeney, Y. Li, R. Gross, D. Yurovsky. "New Directions in Contact Free Hand Recognition" In: 2007 IEEE International Conference on Image Processing. IEEE, 389-392, 2007. DOI: 10.1109/ICIP.2007.4379174.

[79] S. Sharma, S. R. Dubey, S. K. Singh, R. Saxena, R. K. Singh. "Identity verification using shape and geometry of human hands" Expert Systems with Applications. Elsevier, 2015, **42**(2), 821-832, 2015. DOI: 10.1016/j.eswa.2014.08.052.

[80] T. Fotak, P. Koruga, M Baca. "Trends in Hand Geometry Biometrics" In: Central European Conference on Information and Intelligent Systems. Varazdin: Faculty of Organization and Informatics Varazdin, 2012, 319-324, 2012.

[81] D. L. Woodard, P. J. Flynn. "Finger surface as a biometric identifier. Computer Vision and Image Understanding" Elsevier, 2005, **100**(3), 357-384, 2005. DOI: 10.1016/j.cviu.2005.06.003.

[82] N. Charfi, H. Trichili, A. M. Alimi, B. Solaiman. "Novel hand biometric system using invariant descriptors" In: 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR). IEEE, 2014, 2014, 261-266. DOI: 10.1109/SOCPAR.2014.7008016.

[83] W. Kang, Q. Wu. "Pose-invariant hand shape recognition based on finger geometry" IEEE Transactions on Systems, Man, and Cybernetics: Systems **44**(11), 1510–1521. 2014.

[84] A. Ungureanu, S. Salahuddin, P. Corcoran. "Towards unconstrained palmprint recognition on consumer devices: A literature review", 2020. DOI:10.1109/ACCESS.2020.2992219

[85] R. S. Choraś, M. Choraś. "Hand shape geometry and palmprint features for the personal identification" 2006. IEEE Proc. of 6th Intl Conf. on Intelligent Systems Design and Applications, 1085–1090, 2006.
[86] A. Bera, D. Bhattacharjee. "Human Identification Using Selected Features from Finger Geometric Profiles" IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2020, **50**(3), 747-761. DOI: 10.1109/TSMC.2017.2744669. ISSN 2168-2216.
[87] A. Ungureanu, S. Salahuddin, P. Corcoran. "Towards unconstrained palmprint recognition on consumer devices: A literature review" 2020. DOI:10.1109/ACCESS.2020.2992219.
[88] K. Příhodová. "Convolutional neural networks in hand-based recognition system" In: 34th International Business Information Management Association Conference. 2019, 4744-4750.
[89] A. Kumar, D. C. M. Wong, H. C. Shen, A. K. Jain. "Personal Verification Using Palmprint and Hand Geometry Biometric" In: Audio- and Video-Based Biometric Person Authentication. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, 2003-6-24, s. 668-678. Lecture Notes in Computer Science. DOI: 10.1007/3-540-44887-X_78.
[90] T. Mitchell, "Machine learning" Boston: McGraw-Hill, 1997.
[91] Vinodkumar, R. Srikantaswamy. "Fusion of fingerprint, palmprint and hand geometry for an efficient multimodal person authentication system" In: 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT). IEEE, 2016.
[92] J. J. Fuertes, C. M. Travieso, M. A. Ferrer, J. B. Alonso. "Intra-modal biometric system using hand-geometry and palmprint texture" In: 44th Annual 2010 IEEE International Carnahan Conference on Security Technology. IEEE, 2010, 2010, 318-322. DOI: 10.1109/CCST.2010.5678697.
[93] A. Kumar, D. Zhang. "Hand-Geometry Recognition Using Entropy-Based Discretization. IEEE Transactions on Information Forensics and Security" 2007, **2**(2), 181-187. DOI: 10.1109/TIFS.2007.896915.
[94] X.Fan, X. H. Zhang, C. Leung, C. Miao. "Comparative study of machine learning algorithms for activity recognition with data sequence in home-like environment" In: 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI). IEEE, 2016, 2016, s. 168-173. DOI: 10.1109/MFI.2016.7849484.
[95] M. Ferrer, J. Fabregas, M. Faundez, J. B. Alonso, C. Travieso. "Hand geometry identification system performance" In: 43rd Annual 2009 International Carnahan Conference on Security Technology. IEEE, 2009, 2009, s. 167-171. DOI: 10.1109/CCST.2009.5335545.
[96] O. Polat, T. Yildirim. "Hand geometry identification without feature extraction by general regression neural network" Expert Systems with Applications. 2008, **34**(2), 845-849. DOI: 10.1016/j.eswa.2006.10.032.
[97] S. J. Elliott, B. Senjaya, E. P. Kukula J. M. Werner, M. Wade. "An evaluation of the Human Biometric Sensor Interaction using hand geometry" In: 44th Annual 2010 IEEE International Carnahan Conference on Security Technology. IEEE, 2010, 2010, 259-265. DOI: 10.1109/CCST.2010.5678710.
[98] E. P. Kukula, B. P. Gresock, S. J. Elliott, N. W. Dunning. "Defining Habituation using Hand Geometry" In: 2007 IEEE Workshop on Automatic Identification Advanced Technologies. IEEE, 2007, 2007, 242-246. DOI: 10.1109/AUTOID.2007.380627.
[99] ISO/IEC 19795-1. Information technology: Biometric performance testing and reporting. 2016.
[100] M. R. Rajput and G. S. Sable, "IRIS biometrics survey 2010–2015," 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, 2016, 2028-2033, DOI: 10.1109/RTEICT.2016.7808195.
[101] J. Továrek, "Využití umělé inteligence pro vícenásobnou bezkontaktní biometrickou autentizaci" (in Czech) Ostrava, 2017. Ph. D Thesis. VSB - Technical University of Ostrava, 2017.
[102] R.M. Bolle, J.H. Connell, Pankanti S., Ratha N.K. and Senior A.W. The Common Biometrics. In: Guide to Biometrics. Springer Professional Computing. Springer, New York, NY. 2004.
[103] R. Chen, X.R. Lin, T.H. Ding "Iris segmentation for non-cooperative recognition systems" IET Image Process., **5** (2011), 448-456, 2011.
[104] Y. Du, E. Arslanturk, Z. Zhou, C. Belcher. "Video-Based Noncooperative Iris Image Segmentation" IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), **41**(1):64-74, 2011.
[105] I G. P. Suta Wijaya. "Face Recognition Using Map Discriminant on YCbCr Color Space." 2018.
[106] D. Marvadi, C. Paunwala, M. Joshi, A. Vora. "Comparative analysis of 3D face recognition using 2D-PCA and 2D-LDA approaches" IEEE; 2015;2016.
[107] F. Liu, G. Liu, Q. Zhao, L. Shen. "Robust and high-security fingerprint recognition system using optical coherence tomography" Neurocomputing (Amsterdam). 2020. 402:14-28, 2020.
[108] F. Liu, J. Liang, L. Shen, M. Yang, D. Zhang, Z. Lai. "Case study of 3D fingerprints applications" PLoS ONE. 2017; **12**(4) 0175261.
[109] P. Varchol and D. Levický, "Using of hand geometry in biometric security systems", Radio Eng., **16**(4), 82-87, 2007.
[110] A.K. Jain, A. Ross and S. Pankanti, "A prototype hand geometry-based verification system", Proc. Second Int. Conf. Audio and Video based Personal Authentication, 166-171, 1999.
[111] M. Afifi. "11K Hands: Gender recognition and biometric identification using a large dataset of hand images" Multimedia tools and applications, **78**(15), 20835-54, 2019.
[112] A. Kumar, D. Zhang "Hand-Geometry Recognition Using Entropy-Based Discretization" IEEE Transactions on Information Forensics and Security. 2007, **2**(2),181-7, 2007.
[113] S. Angadi, S. Hatture "Hand geometry-based user identification using minimal edge connected hand image graph" IET computer vision, **12**(5),744-52, 2018.
[114] O. Vergara Villegas, H. J. Ochoa Dominguez, V. G. Cruz Sanchez, L. Ortega and H. Madero. "Biometric Human Identification of Hand Geometry Features Using Discrete Wavelet Transform" Discrete Wavelet Transforms - Biomedical Applications. InTech, 09-12, 251-266. 2011.
[115] J. Burgues, J. Fierrez, D. Ramos, M. Puertas and J. Ortega-Garcia, „Detection Invalid Samples in Hand Geometry Verification via Geometric Measurements", 2010 International Workshop on Emerging Techniques and Challenges for Hand-Based Biometrics, Istanbul, 2010, 1-6, DOI: 10,1109 / ETCHB.2010.5559296.
[116] N. Charfi, H. Trichili, A. M. Alimi and B. Solaiman, "Novel hand biometric system using invariant descriptors," 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR), Tunis, 2014, 261-266, doi: 10.1109/SOCPAR.2014.7008016.
[117] K. Prihodova and M. Hub. "Hand-Based Biometric System Using Convolutional Neural Networks". Acta Informatica Pragensia, **9**(1), 48-57. 2020. doi: 10.18267/j.aip.131.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 699-709 (2020)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Automatic License Plate Detection and Recognition for Jordanian Vehicles

Khalil Mustafa Ahmad Yousef *, Bassam Jamil Mohd, Yusra Abd-Al-Haleem Al-Khalaileh, Ahlam Hani Al-Hmeadat, Bushra Ibrahim El-Ziq

Department of Computer Engineering, Faculty of Engineering, The Hashemite University, P.O.Box 330127, Zarqa 13133, Jordan

ARTICLE INFO

Article history:
Received: 19 August, 2020
Accepted: 16 November, 2020
Online: 24 November, 2020

Keywords:
ANPR (automatic number plate recognition)
Jordanian license plates
OCR (optical character recognition)
rectangularity
make of the vehicle

ABSTRACT

Nowadays, automatic number plate recognition (ANPR) is very important especially in the era of smart cities and intelligent transport systems. Fully automated number plate detection and recognition system helps in reducing time, error, and cost for tracking of vehicles and for recording traffic violations. The main goal of this paper is to design a low cost fully automated number plate detection and recognition system targeting the Jordanian license plates. Several problems (e.g., cost, wasted efforts, manual intervention, and possible errors) were identified in the currently used Jordanian number plate recognition for recording traffic violations. We hope that the proposed system would mitigate such problems. The proposed system performs two main tasks. First, it automatically detects and recognizes the license plate number of a given Jordanian vehicle using a robust metric; the rectangularity measurement, and identifies the vehicle's type (e.g., governmental, private, visitor, public, etc.). Second, it recognizes a selected number of trained classes for the make of the vehicle whenever applicable. The experimentation results and the performance evaluations compared to other ANPR approaches show that proposed system achieves the best performance among the tested systems with a plate detection accuracy of 95%, OCR recognition accuracy of 94.68%, make recognition accuracy of 86.84%, and an overall ANPR accuracy of 89% excluding the make results.

1 Introduction

The system of automatic license plate number detection and recognition, known as ANPR (automatic number-plate recognition), is considered a great application and an example of the computer vision technology. ANPR is widely being used in, but not limited to, radar cameras, car park management and traffic management [1]. In Jordan, ANPR is being used in recording traffic violations (e.g., over-speeding), however it has several problems. In particular, some of the ANPR steps: cropping of the license plate and optical character recognition (OCR) are still being done manually [2]. This might be due to the high cost of a fully customized commercial ANPR system for detecting and recognizing the Jordanian license plates. Therefore, the main goal of this paper is to propose a low cost ANPR system that is customized for automatically detecting and recognizing the numbers of the Jordanian license plates. We aim to reduce and avoid human labor, and the possible human errors. Having human involved ANPR steps might result in misidentifying a vehicle plate numbers and data, which can cause erroneous tickets.

The main contributions of this work are as follows:

- Propose a low cost fully automated and customized Jordanian ANPR system.
- Propose and utilize an optimized calculation methodology of the rectangularity metric [3] to efficiently detect the Jordanian license plates.
- Build a database of Jordanian vehicles to be used as a benchmark for evaluating the performance of the proposed system.
- Compare the proposed system to other ANPR systems with the help of the built database for the Jordanian vehicles.
- Incorporate the recognition of the vehicle type and make into the system.

The rest of the paper is organized as follows. Section 2 presents the related work. Section 3 discusses in details the proposed system. Section 4 presents the experimentation results. Section 5 provides a discussion about the proposed system and the obtained results. Section 6 concludes the paper and provides some future work directions.

*Corresponding Author: Khalil Mustafa Ahmad Yousef, Department of Computer Engineering, Faculty of Engineering, The Hashemite University, P.O.Box 330127, Zarqa 13133, Jordan, Contact No: +962796669967 & Email: khalil@hu.edu.jo

2 Related Work

Generally speaking, any ANPR system consists of four main steps as listed below, which are also shown in Figure 1.

1. Capturing an image of a vehicle.
2. Detecting vehicle's license plate.
3. Segmenting out the license plate objects or characters.
4. Recognizing the segmented license plate characters using OCR (Optical Character Recognition).

Figure 1: ANPR steps

Based on the ANPR steps above, in the literature, several works have been proposed to address and optimize each step. Additionally, there is a significant work in combining the ANPR system with detecting and recognizing the make of the vehicle. In what follows, some of the related work regarding the ANPR systems and the vehicle make recognition systems are briefly presented.

In [4] and [5], the authors proposed an ANPR system, where mean shift segmentation is used to identify candidate plate regions. Then, the authors trained and used a Mahalanobis linear classifier with respect to three features to accurately detect number plate region. Namely, they used rectangularity [3], which is defined by the authors as "a measurement that represents how well an object fits its minimum enclosing rectangle", aspect ratio and edge density features.

In [6], the authors proposed an Iranian ANPR system, where edge detection and morphological operators are used to identify license plate candidate regions. The percentage of rectangularity and color filter of each candidate plate region are used to accurately detect the plate region. Multilayer Perception (MLP) neural network is finally used for recognizing the numbers of the detected plate.

In [7], the authors proposed an Indian ANPR system. License plate detection is based on using Sobel filter edge detector, morphological operations and connected component analysis. The segmentation and recognition of plate characters, which are in different Indian languages, are performed based on training a Support Vector Machine (SVM) classifier.

In [2], the authors proposed a Jordanian ANPR system that is based on SIFT features to detect Jordanian license plates and recognize them. Template matching technique is used to locate special marks in the license plate to segment out the license plate. Finally, OCR is used to recognize the characters or numbers from the segmented plate.

In [8], the authors surveyed several keypoint features under the feature-based automatic license plate detection techniques. They categorized the vehicle detection techniques into motion and appearance based, which included sophisticated probabilistic modeling and feature extraction. Also, the authors surveyed and discussed vehicle recognition utilizing vehicle attributes like license plate, color, type, and logo.

In [9], the authors applied a bag of features method to the car make and model recognition (VMMR). The sparse coding technique based on square mapped gradients (SMG) feature descriptors is used to learn a dictionary of codewords. The authors then apply supervised K-means or SVM with SMG to recognize the car make and model. Finally, the authors compared the performance of the two approaches.

In [10], the authors proposed an automatic license plate recognition system to trap traffic violations. For example, to trap offenders disobeying the traffic rules in an enclosed space (e.g., campus).

In [11], the authors proposed an ANPR-based vehicle tracking system, which monitors vehicle's Euclidean distance from a visible calibration line. License plate detection was based on the fact that license plate has an inherent property of having greater transitions in pixel values compared to that of its surroundings.

In [12], the authors proposed a number plate detection and localization system based on segmentation and morphological operators to segment out license plate from the captured image. Specifically, the authors used histogram checking, and vertical and horizontal edge detection methods for the segmentation of the license plate. The experimentation dataset consisted of 350 license plates from different countries including Jordan. The reported efficiency rate from the proposed system is 93.43%.

In [13], the authors proposed a Jordanian ANPR system, where the license plate was detected and localized based on a multistage algorithm using the aspect ratio and the edges ratio. Back propagation neural network was used for recognizing the license plate characters (OCR). The average success rate from the proposed system is 90% of 240 high resolution images of size 1188×960 pixel.

Very recently, in [14], the authors proposed an ANPR system to recognize the Bangladeshi characters in vehicle license plates. Morphological operation and histogram analysis with a dynamic threshold of vertical and horizontal projection profiles were applied in order to detect plate area. Character recognition is carried out using Support Vector Machine (SVM) classifier, where extracted Histogram of Oriented Gradient (HOG) features were used as input. The system was applied to 630 images of Bangla license plates of different categories of vehicles and achieved 91% accuracy in the extraction of plate ROI.

After surveying the literature and to the best of our knowledge, the work of [2, 12, 13] is the only published work that examines the detection and recognition of the license plates of the Jordanian vehicles. Therefore, the work presented in this paper extends the literature regarding the Jordanian license plates detection and recognition. In fact, the proposed work improves the automatic Jordanian license plate detection and recognition proposed in [2] by using another feature, which considers both the appearance shape of the license (rectangularity and aspect ratio) and the license plate content for detecting the license plate. Furthermore, the recognition of vehicle logo and make was also incorporated. It is worth mentioning the notes concerning the work by authors in [12] and [13]:

- The processing time for detecting the license plate and recognizing the plate numbers in both [12] and [13] was not reported at all, whereas in the proposed work it is being reported.
- The dataset of [13] consisted of high resolution images with-

out presenting any sample images. In the proposed work many sample images are provided from the created and employed dataset, which mostly consisted of mid to low resolution images, but better than 320 × 240.

- Neither the work in [12] nor in [13] presented a fully comprehensive ANPR system, where both the license plate type (private, public, government, etc.) and the vehicle's make are being recognized similar to what are being implemented in the proposed work.

Additionally, it is noteworthy to mention that the proposed work shares some similarity with the work presented in [4] and [5]. The proposed system utilizes the rectangularity and aspect ratio metrics in detecting the license plate (Jordanian plate in our case), where it achieved the best plate detection accuracy compared to using the rectangularity methodology proposed in [4] and [5]. This will be discussed in more details in the next section.

3 Proposed System

This section presents the details of the proposed Jordanian ANPR system. Taking into consideration the main ANPR steps presented in Figure 1, in the designed and implemented ANPR system, the following steps were considered as shown in Figure 2. The system was fully implemented in MATLAB R2018a [15], where the graphical user interface (GUI) is shown in Figure 3.

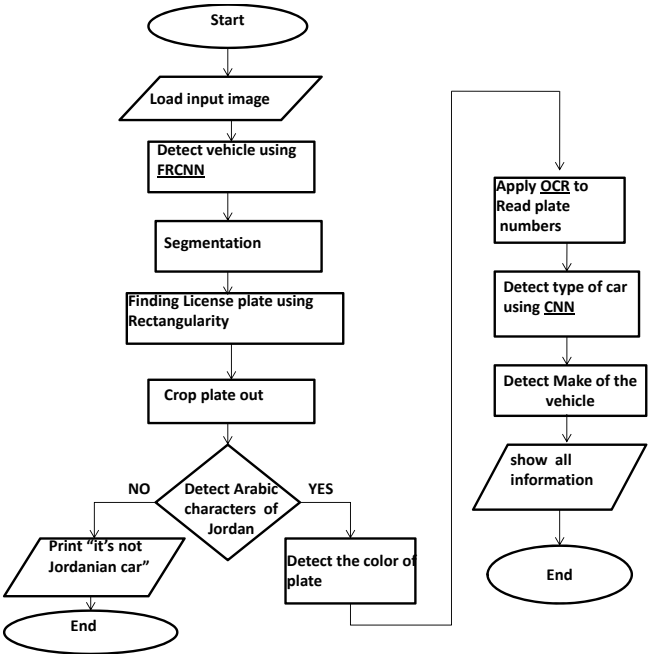


Figure 2: The proposed ANPR steps

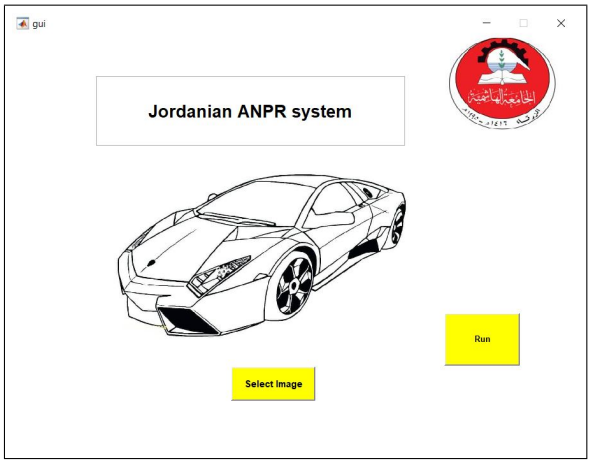


Figure 3: The graphical user interface of the proposed Jordanian ANPR system

3.1 Input image

The very first step in the system is to capture a vehicle image. An off-the-shelf webcam and a mobile phone camera were used to capture a set of database images to evaluate the system's performance. Additionally some of the images were obtained from the Jordanian Opensooq website [16]. In general, the captured images should be of good illumination conditions whenever possible, capture either the back or the front of the vehicle, and have a good resolution (e.g., better than 320 x 240). The database consisted of 100 images, which corresponds to different illumination conditions, different backgrounds, and different perspectives of the captured Jordanian vehicles (i.e., angle and distance between the camera and the vehicle location). Figure 4 shows 16 randomly selected input images chosen from the created dataset. Several preprocessing steps are then being performed on each input image that included histogram equalization to help in reducing varying day light effect and for contrast enhancement.

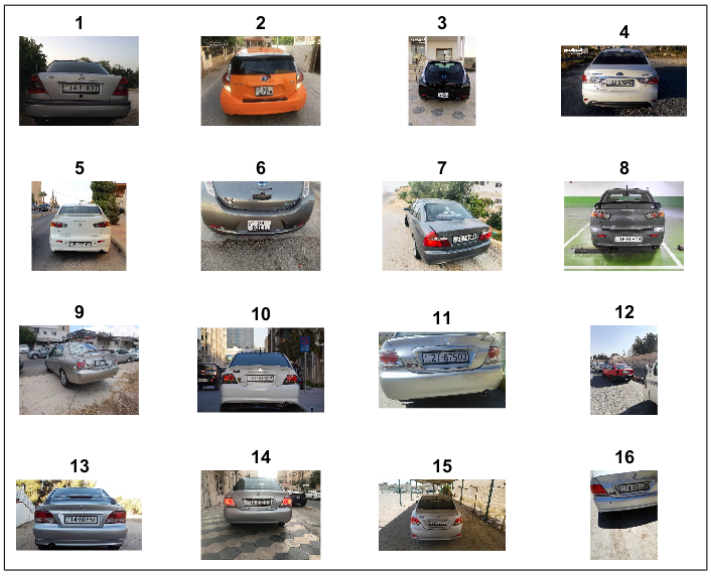


Figure 4: 16 randomly selected input images chosen from the created dataset

3.2 Vehicle detection

The second step is to detect the vehicle from the input image. Faster Regional Convolution Neural Network (FRCNN) [17], which is an objecte detection built-in MATLAB function, was trained to detect and determine the vehicle and its location in the image. FRCNN was chosen since it is significantly faster in training and testing sessions over other detection methods (e.g., R-CNN). Additionally, it is considered as a highly successful framework for generic object detection [18].

The process of using FRCNN is done iteratively among different zooming versions or scales of the input image. This iterative procedure stops, when a vehicle is found or the input image is zoomed to the maximum. In this case (the latter), it may be that no vehicle is found and thus the ANPR system will give an error message indicating such a state and halts. Otherwise (i.e., once the vehicle is found), a blue bounding box around the detected vehicle is drawn. The detected vehicle is then cropped and converted to gray scale. Figure 5 shows an example of this step.

Figure 5: An example of a given input image and the detected cropped vehicle

3.3 License plate detection

The third step is to detect the license plate from the cropped gray scale vehicle image, which is one of the most important ANPR steps. Initially, the fast marching segmentation method [19] is executed to find the best license plate candidate regions. Precisely, this segmentation method will segment the vehicle image into several connected components or parts. The segmented growing parts are initially initialized based on a uniformly chosen seeds within the cropped vehicle image as shown in Figure 6. One of those parts must be the license plate. Figure 7 shows some of the segmented parts as it was very difficult to visualize all candidate segmented parts for all seeds in the same image; as the segmented parts may overlap and therefore the segments will not clearly be viewed. To accurately identify the correct license plate region or part, the following analysis is performed:

- First, the shape of the Jordanian license plate is rectangle with two styles: European and American as shown in Figures 8 and 9, respectively. The European style places the numbers in one row. The American style places the numbers in two rows.
- Next, rectangularity measurement is used [3]–[5] to first identify the segmented parts whose shape is best described by a rectangle, but of certain aspect ratio. This is the aspect ratio, which is computed based on the ration between the height and width of either the American or European styles of the Jordanian license plate. Figure 10 correctly shows identified and detected license plate part based on the rectangularity measurement metric.
- Consequently, a red bounding box is drawn around the part, which satisfies the above constrain or criterion, and crop it as shown in Figure 11.

It is important to emphasize that one of the main contributions in this paper is the methodology used in computing the rectangularity measurement, and thus that for detecting the license plate more efficiently and precisely. Specifically, the presented methodology is quite different and efficient compared to that described in [4], [5] as explained next.

Figure 6: Uniformly chosen seeds in the vehicle cropped image to initialize the segmentation process for the plate detection step.

Figure 7: Some samples of the segmented vehicle parts after the fast marching segmentation during the license plate detection step.

Figure 8: European-style Jordanian plate.

Figure 9: American-style Jordanian plate.

Figure 10: Correctly identified and detected license plate part based on the rectangularity measurement metric.

Figure 11: Cropped license plate after being rotated as to maximize the rectangularity calculation

3.3.1 *Contribution in Calculating the Rectangularity Measurement*

The methodology for computing the rectangularity in [4], [5] requires that:

- each candidate license plate region is to be treated as a rigid object,
- computing its area A_0 by counting the total number of pixels in the region,
- finding its minimum enclosed rectangle (MER) fitted to its boundary by iteratively rotating this object in one degree steps in the range $[-15°, 15°]$ to account for possible tilted angles of the captured vehicles,
- computing the area of MER of this object at each incremental rotation,
- identifying the angle at which the MER goes through the minimum value (A_{MER}), and
- finally computing the rectangularity R at this angle as given in 1:

$$R = \frac{A_0}{A_{MER}} \tag{1}$$

However, the methodology does not require rotating each candidate plate region in a range of steps, which we believe is an important step to perform in order to efficiently detect the correct license plate with high confidence. However, performing this step iteratively and incrementally might be very time consuming and might not produce precise rectangularity calculation. Alternatively,the following steps are performed, which constitute one of the main contributions of this work:

- Each candidate license plate region or object is checked to make sure it has a convex shape as was proposed in [20]. In case it is not convex, it is converted to a convex shape so that spaces in the outwards lines of the object (if any) are connected and filled. This process uses a set of morphological operation (dilation, followed by filling, and lastly followed by erosion).
- Next, ellipse fitting is applied to each convex shaped object. This helps us determine the major and minor axis's of the fitted ellipse of the object.
- The object is then rotated as to remove the rotation angle, which is the angle between the elongation axis (major axis for the fitted ellipse) and the X-axis. Indeed, this is the angle at which the MER goes through the minimum value (A_{MER}).
- Then the area A_0 of the rotated object (the total number of pixels in the convex shape) is computed.
- Finally, the rectangularity R using Equation 1 is calculated.

3.3.2 *Rectangularity Analysis*

The methodology presented in [4], [5] was compared with the proposed methodology. It is important to mention to the best of our knowledge that we could not find any work other than [4], [5] which used the rectangularity metric as a major step in detecting and localizing the license plate similar to what is proposed in this paper. And, since this step is really the most critical step and the heart of any ANPR system, we believe that using the methodology of [4] and [5] as one of the baseline comparisons is still a fair choice especially with the proposed enhancement given in the previous subsection, which we believe shows an advancement to the literature in this direction. In the comparison, the dataset of 100 images is used to compute the detection accuracy of correctly finding the license plate region and the average rectangularity calculation time. The results are presented in Table 1. The results indicate that the proposed rectangularity calculation is more accurate, precise and faster. In fact, the analysis showed that proposed rectnagularity measurement values were very close to one for the candidate license plate objects compared to around 0.7 based on [4], [5]. Additionally, it is worth noting that the rectangularity measurement calculation under different $\Delta\theta$'s $\in [-15, \ldots, 15]$ as in [4], [5] might not be unique. The analysis shows that the same rectangularity value can be obtained under multiple different $\Delta\theta$'s, which can be problematic and result in misidentifying the correct license plate region.

Table 1: Rectangularity analysis

Rectangularity methodology	Average plate detection accuracy	Average rectangularity calculation time (sec)
[4], [5]	65%	0.2785
The proposed methodology	95%	0.0315

3.4 Jordanian license plate Verification

Upon the detection of a possible license plate region, it remains to verify that it is indeed a Jordanian plate. To do so, the label part is examined. It is located to the left side of the Jordanian license plate as shown in Figure 12. It contains the English and Arabic characters of the word Jordan. Therefore, it was decided to use template matching technique based on computing the score of the cross correlation coefficients between two binary windows; (the binary templates of the Arabic characters of the word Jordan and the binary connected components objects extracted from the anticipated label area of the identified license plate candidate). Following are the performed template matching steps:

- converting the label area into the binary scale using Otsu's method.
- segmenting out all possible characters or objects within the label area by performing connected component analysis on them.
- resizing these segmented objects to match the size of the stored binary templates of the Arabic characters of the word Jordan
- exhaustively computing the correlation coefficient between each object and the templates.
  - if the computed correlation coefficients are above heuristically computed threshold, then the detected license plate is declared to be a Jordanian one.
  - otherwise, it gives an error message that it could not find a Jordanian plate and halts.

3.5 Determine license plate type

The fifth step is to determine the type of the Jordanian vehicle. This information is color-coded (see Table 2) in the label area of the license plate. Thus, in the proposed ANPR system, we analyze the color of the detected license plate label area in the original input image. Consequently, based on the analyzed color (most frequently occurred color value) and the recognized code part from the next ANPR step (numbers in first line in American license plate and the digits before the dot in European license plate), we determine the type of the vehicle. Figure 13 shows some examples of the detected Jordanian license plate types from the ANPR system. Table 3 summarizes the types found in the created dataset.

Figure 12: Label image part.

(a) White Label Area: Private Plate

(b) Red Lable Area: Government Plate

(c) Yellow Label Area: Temporary plate

(d) Green Label Area: Public plate

Figure 13: Types of Jordanian license plate [21].

Figure 14: License plate area that excludes the label region

Table 2: Color coded Jordanian license plate types [22]

The two reserved digits indication: selected list		License plate color indication
1 – 5	Government vehicles	Red
6	Aqaba free zone vehicles	Red
7	Diplomatic vehicles	Yellow
8-9	Temporary entrance vehicles	Yellow
10 – 21	Passenger vehicles	White
41 – 42	Light goods vehicles	White
44	Tractors	Red
46	Motorcycles and Scooters	White
50	Taxi	Green
56	Small buses	White
58	Coaches	White
71	Trailer	Yellow

Table 3: Summary of the license plate types found in the dataset

Type of License Plate	Number of License Plates
Private, White	96
Diplomatic, Yellow	2
Taxi, Green	1
Temporary Entrance, Yellow	1
Total	100

3.6 Recognition of license plate numbers (OCR)

The sixth step is to recognize the license plate numbers. First, all possible characters within the detected license plate binary image

are extracted by performing connected component analysis within the plate region. Each extracted connected component or object is resized to match the size of the stored binary templates of the numeric plate numbers (from 0 to 9) that are graphically shown in Figure 15. The typeface used to generate such templates is known as Mandatory [23], which is the same typeface that is used for the United Kingdom vehicle registration plates.

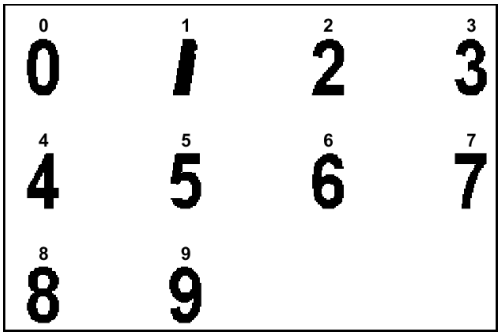


Figure 15: The stored binary templates (inverted in order to best be viewed) of the numeric plate numbers that are desired to be recognized (from 0 to 9)

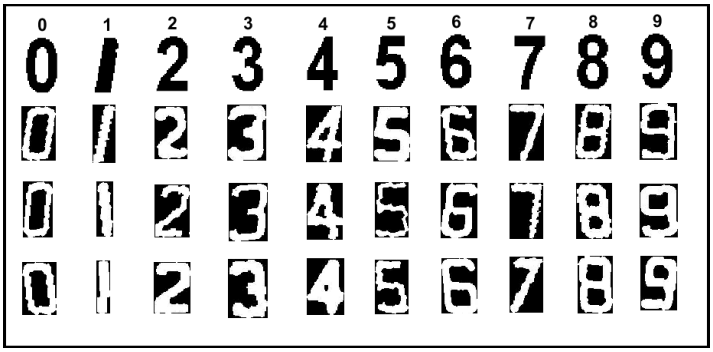


Figure 16: More binary templates of the numeric plate numbers that are desired to be recognized (from 0 to 9)

Finally, the correlation coefficient between each object and the templates is exhaustively computed to recognize the plate numbers. It is noteworthy to mention that, in order to increase the OCR recognition accuracy, the number of stored binary templates for the character set '0123456789' was increased as shown in Figure 16. This resulted in having better matching of the segmented objects and thus high confidence in the associated recognition results. Additionally, the license plate area that excludes the label region (Figure 14) is the OCR region of interest because the segmentation of possible numbers or objects is performed in it. As stated above, the Jordanian license plate splits into two different styles: European and American. Thus, the performed connected component analysis considers the properties and characteristics (distribution of the numbers) of the two types. It also considers the following facts:

- The width and length of all numbers are the same.
- The distance between numbers is nearly the same, except the code part (numbers in first line in American license plate and the digits before the dot in European license plate), which determines the type of vehicles such as: 50: taxi, 36: private bus, etc.).
- All numbers in the license plate have the same color, which is black. Additionally, the background color for the license plate (except label part) is white.

A green bounding box is drawn around each recognized number in the license plate region as shown in Figure 17. After applying the OCR, the code of the license plate is computed by calculating the distance between all numbers. The code part determines the type of the vehicle as we mentioned before and listed in Table 2.

Figure 17: OCR results: remove background from the license plate, segment-out the possible numbers, recognize them, and draw green bounding box around each recognized number.

Table 4: Summary of the vehicle makes found in the dataset

Vehicle Make Index	Vehicle Make	Number of Vehicles
1	Mercedes	11
2	Toyota	10
3	Nisan	7
4	Hyundai	10
5	Mitsubishi	39
6	Opel	3
7	Audi	1
8	Ford	8
9	Mazda	1
10	Lexus	2
11	Daewoo	1
12	Jaguar	1
13	Jeep	1
14	Kia	2
15	Honda	1
16	Peugeot	1
17	Chrysler	1
Total number of Images		100

3.7 Recognizing the make of the vehicle

The seventh and final step of the proposed ANPR system is to determine the make of the vehicle. In this step, the make of the vehicle is recognized by localizing and recognizing the vehicle's logo. Generally speaking, the fact that the vehicle's logo is approximately located in the region above the detected license plate can be relied on, which can be a weak assumption. However, in this presented step, it is desirable to prove the concept of being able to add such feature (i.e., recognizing the make of the vehicle) easily. Thus, the vehicle's logo in the segmented parts from the third step is searched for. The segmented parts are just above the detected license plate region. Once a logo was detected, a yellow bounding box is drawn around it. To help recognize those detected logos, several logo images from the Internet were collected. The images were specifically related to five auto makers MERCEDES, MITSUBISHI, KIA, NISSAN and TOYOTA. The list could be expanded easily. It should be mentioned that the dataset consisted of 17 vehicle makes or brands as summarized in Table 4, and the recognition focus was just put on five of them as indicated above. After that, using the collected logo dataset, a convolutional neural network (CNN) is trained with the help of MatConvNet MATLAB toolbox to recognize the five makes of the vehicles as listed above. While testing this ANPR step on the input images dataset, it is assumed the case where none of the logo images of the vehicles would match the makes from which the CNN classifier is trained on as an "UNKNOWN MAKE". Figure 18 shows the recognized make for the vehicle that was shown in Figure 10.

Figure 18: The recognized make for the vehicle that was shown in Figure 10

To conclude this section, Figure 19 shows and summarizes the output of all stages of the proposed ANPR system. As it is clear from Figure 19:

- The detected vehicle is identified by a blue bounding box.
- The detected plate is identified by red bounding box.
- The detected logo of the vehicle is identified by yellow bounding box.
- The detected license plate is a Jordanian one and its type is private; since it has a white color and its code is recognized as "24".
- The recognized license plate numbers are "12692", where each number is bounded by a green bounding box.
- The recognized make of the vehicle is "Mitsubishi".

4 Experimental results

This section presents the experimental results of the proposed ANPR system. A 100 images were tested as mentioned in the previous section. The average execution time for all tested images was found to be 8.0047 sec on average (ANPR steps 2-6) on a computer machine that has 16 GB of RAM and two quad core 2.5 GHz i7 processor. The system implementation was purely sequential and thus a single core was utilized. Table 5 presents the results (success rates or accuracies). The reported accuracy for each of the rows in Table 5 was computed for each step separately assuming correct results from previous step. As it can be seen from Table 5, the final license plate recognition accuracy from the OCR step was 94.68%. Consequently, the overall ANPR accuracy is 89% given the fact that the license plate detection step failed in 5 images, the license plate verification step failed in 1 image, and the OCR step failed in 6 images for a total of 11 images out of 100 (this is shown in the last row of Table 5). Finally, the accuracy for recognizing the make of the vehicles was 70.21% as given in the second row of Table 6 with an average execution time of 0.0096 sec. However the actual make accuracy was observed to be 86.84% as shown in the last row of Table 6. We should note that images without valid logos and unclear logos were manually excluded. Therefore, by excluding those images from the make evaluation dataset, the above accuracy was recomputed.

Table 5: The experimental results of the proposed ANPR system

ANPR Steps	Num. of input images	Num. of success images	Num. of fail images	Accuracy	Average execution time
Step 2: Vehicle detection	100	100	0	100%	0.3394 sec
Step 3: License plate detection	100	95	5	95%	3.6327 sec
Step 4: License plate verification	95	94	1	98.95%	3.5924 sec
Step 5: License plate type	94	94	0	100%	0.00089 sec
Step 6: OCR	94	89	5	94.68%	0.4393 sec
ANPR system with all Steps 2-6	100	89	11	89%	8.0047 sec

We have compared the proposed system with the following ANPR systems: the ANPR system, which is available online at [24] and the ANPR system proposed in [2] (our previous work). For this comparison, we only used the 94 images out of the created dataset for which the license plates were successfully detected and verified as listed in Table 5. The result of this comparison is shown in Table 7, which shows that the proposed system achieved the best performance in terms of the recognition accuracy.

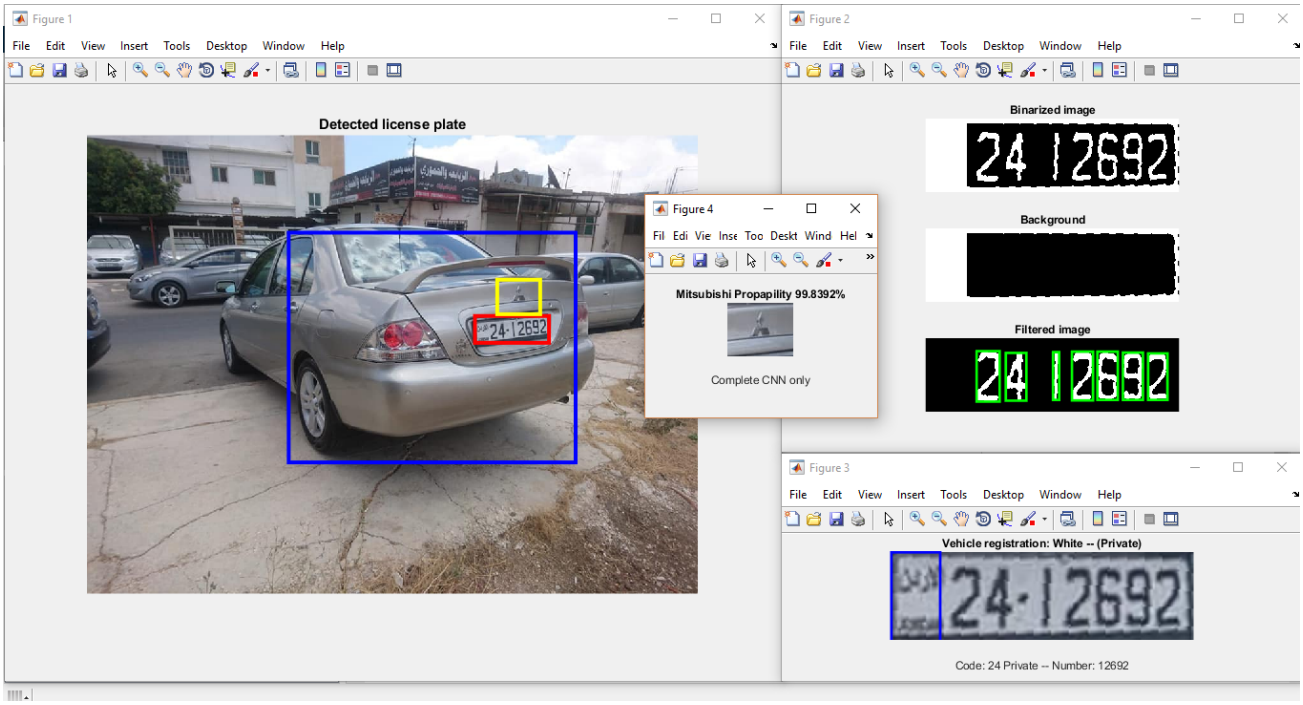


Figure 19: Summary of the output of all the steps of the proposed ANPR system.

Table 6: The make recognition results of the proposed ANPR system

ANPR Steps	**Num. of input images**	**Num. of success images**	**Num. of fail images**	**Accuracy**	**Average execution time**
Step 7: Make of the vehicle	94 (Successfully detected and verified license plate images)	66	28	70.21%	0.0096 sec
Step 7: Make of the vehicle	76 (Same as previous, but excluding the images that have unclear/missing logos)	66	10	86.84%	0.0096 sec

Table 7: Comparison results with other systems

ANPR Systems	License Plate Recognition Accuracy
`https://www.anpronline.net/demo.html`	56%
SIFT-based ANPR system [2]	75%
Our proposed system	94.68%

5 Discussion

As it can be inferred from Table 5, some of the steps need to be optimized either in terms of the average execution time or the accuracy of detection and recognition. For example, the average execution time of the license plate detection step was around 3.6327 sec and it was the largest among the rest of the steps. This is caused by the largest number of candidate segmented regions that were needed to check the license plate for; Rectangularity measurement and aspect ratio. Specifically, this is controlled by the number of the initialized seeds which is equal to $256 = 16 \times 16$ uniformly spaced grid as was shown in Figure 6. Despite that being said, it is important to highlight that our average calculation time of the rectangularity measurement per each candidate license plate part was better compared to the work of [4], [5] as was shown in Table 1 with a detection accuracy of 95%. Studying the factor of reducing the number of initialized seeds to reduce the license plate detection time is planned as a future work and how would that affect the detection accuracy.

Another observation from Table 5 is that the next highest average execution time was the license plate verification step. We believe that this is related to the way in which the verification process was performed as large number of small sized binary segmented windows were exhaustively compared against the binary templates of the English and Arabic characters of the word Jordan. Moreover, we carefully investigated the single failure case from this step, and it turned out to be caused by the bad resolution cropped license plate. The produced binary image of the cropped license plate is as shown in Figure 20.

Additionally, based on the results obtained from the OCR step, it was found that the main causes for the relatively low OCR recognition rate are related to one or more of the following factors:

- The numbers on some of license plates were not so much clear in the captured input images or may have bad illumination conditions; such as the example shown in Figure 21.
- Some of the numbers have certain painting defects that cause the OCR algorithm to not recognize them correctly such as the example shown in Figure 22 for the number ‘6’.

These kind of problems result in a bad segmentation of the plate characters and thus bad input to the OCR algorithm. Therefore, bad OCR recognition results.

Figure 20: The single case where the license plate verification step failed due to the bad resolution of the cropped license plate; the English and Arabic words of Jordan can not be found in the binary image.

Figure 21: A blurring problem in one of the numbers in the detected license plate (number 6) that caused the OCR step to fail

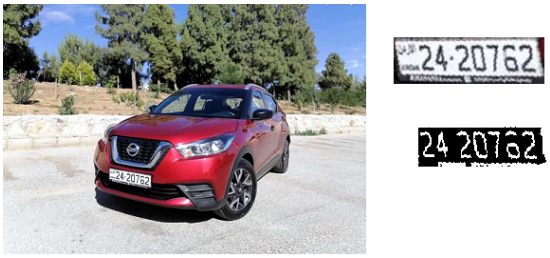


Figure 22: A painting defect problem in one of the numbers in the detected license plate (number 6) that caused the OCR step to fail

Moreover, the lowest recognition accuracy was for the step of recognizing the vehicle's make. This was due to the assumption in localizing the logo of vehicle, which was assumed to always be found in the upper regions above the license plate. We believe that the main challenge was that the location of the logo above the license plate is not indeed consistent among all the vehicles, especially the tested ones (there were too much variations). Therefore, as was mentioned in Section 4 and after analyzing the captured input images, it was found that there were several images without any logo as can be seen in Figure 23, and in some of the images the logo was not clear due to the bad illumination conditions or due to the shadow effect as seen in Figure 24. It turned out that the number of images in the created dataset that possess these kind of problems was found to be 18. As a result of this observation and in order to test the performance of the trained make recognition subsystem, we excluded those 18 images from the make evaluation dataset (94 images). Consequently, the make recognition acucray was observed to be 86.84% as was shown in last row of Table 6.

For the final cost of the proposed ANPR system in terms of the hardware needed to run, it is clear that it only needs a computer machine and a camera to capture input images of the vehicles. Thus, considering the fact that cameras are now almost available everywhere (e.g., mobile phone cameras and surveillance cameras) and that much of the processing needed can be done on the cloud, the estimated cost of the proposed system could be as low as zero.

Figure 23: Four sample vehicle images, where the logo of vehicle that indicates its make is missing

Figure 24: Two sample vehicle images, where there is bad illumination conditions around the logo of each vehicle

Last but not least, although the proposed Jordanian ANPR system in this paper was shown to provide very good recognition results, however it may not work or produce correct results for all cases of the input images. Indeed it needs to be further analyzed under bad illumination conditions, blurriness in image, dirtiness or painting or physical defects problems in the numbers and characters on the license plate, and under the bad weather conditions (e.g., foggy and

rainy conditions). About the speed performance of the proposed system, it is important to emphasis that the estimated time usually depends on the scene complexity of each input image, input image resolution, and the underlying speed of the running machine.

6 Conclusions and Future Work

The paper proposes a fully automated license plate recognition system targeting the Jordanian license plates as a case of study. The proposed system uses and optimizes the calculation of the rectangularity measurement to better identify the best license plate candidate region, which is considered one of the main contributions of the paper. The proposed Jordanian ANPR system was tested on a database that consisted of 100 Jordanian vehicles and compared it to two ANPR systems from the literature. Consequently, proposed system achieves and reports the best performance among the compared systems in terms of the OCR recognition rate, which was 94.68%. We hope that the presented system will be adopted in the very near future in Jordan.

The presented system still needs to be optimized in different aspects as was discussed in Section 5, which is slated for future work. For example, it is desirable to study the effect of the selected number of seeds during the license plate detection step on license plate detection accuracy and the average detection time. Additionally, it is useful to analyze and further experiment the proposed system under bad illumination and weather conditions, blurriness in image, and under physical defects problems in the numbers and characters of the license plate. Moreover, examining better methods for recognizing the make of the vehicle and increasing the number of trained classes for the recognition of other makes are another area of future work.

Conflict of Interest The authors declare no conflict of interest.

References

[1] L. Keilthy, "ANPR system performance," Parking trend international, 2008.

[2] K. M. Ahmad Yousef, M. Al-Tabanjah, E. Hudaib, M. Ikrai, "SIFT based automatic number plate recognition," in 2015 6th International Conference on Information and Communication Systems (ICICS), 124–129, IEEE, 2015, doi:10.1109/IACS.2015.7103214.

[3] K. R. Castleman, Digital Image Processing, Prentice Hall, 1996.

[4] W. Jia, H. Zhang, X. He, "Mean shift for accurate number plate detection," in Third International Conference on Information Technology and Applications, 2005. ICITA 2005., volume 1, 732–737, IEEE, 2005, doi: 10.1109/ICITA.2005.176.

[5] W. Jia, H. Zhang, X. He, "Region-based license plate detection," Journal of Network and computer Applications, **30**(4), 1324–1333, 2007, doi: 10.1016/j.jnca.2006.09.010.

[6] S. Rastegar, R. Ghaderi, G. Ardeshipr, N. Asadi, "An intelligent control system using an efficient License Plate Location and Recognition Approach," International Journal of Image Processing (IJIP) Volume (3), **3**(5), 252–264, 2009.

[7] M. Shidore, S. Narote, "Number plate recognition for indian vehicles," IJCSNS International Journal of Computer Science and Network Security, **11**(2), 143–146, 2011.

[8] R. A. S. Ma'moun Al-Smadi, Khairi Abdulrahimand, "Traffic Surveillance: A Review of Vision Based Vehicle Detection, Recognition and Tracking," International Journal of Applied Engineering Research, **11**(1), 713–726, 2016.

[9] I. Kamal, K. Housni, M. Y. Hadi, "Car Make and Model Recognition using Sparse Coding and Bag of Features," in Proceedings of the 2nd international Conference on Big Data, Cloud and Applications, 79, 1–5, ACM, 2017, doi: 10.1145/3090354.3090435.

[10] P. Agarwal, K. Chopra, M. Kashif, V. Kumari, "Implementing ALPR for detection of traffic violations: a step towards sustainability," Procedia Computer Science, **132**, 738–743, 2018, doi:10.1016/j.procs.2018.05.085.

[11] K. Indira, K. Mohan, T. Nikhilashwary, "Automatic License Plate Recognition," in Recent Trends in Signal and Image Processing, 67–77, Springer, Singapore, 2019, doi:10.1007/978-981-10-8863-6_8.

[12] A. A. Moustafa, M.-I. R. M. Jaradat, "A new approach for license plate detection and localization: Between reality and applicability," International Business Research, **8**(11), 13, 2015, doi:10.5539/ibr.v8n11p13.

[13] H. Alhaj Mustafa, S. Hassanin, M. Al-Yaman, "Automatic Jordanian License Plate Recognition System Using Multistage Detection," in 2018 15th International Multi-Conference on Systems, Signals Devices (SSD), 1228–1233, IEEE, 2018, doi:10.1109/ssd.2018.8570408.

[14] R. Islam, M. R. Islam, K. H. Talukder, "An efficient method for extraction and recognition of bangla characters from vehicle license plates," Multimedia Tools and Applications, 20107-20132, 2020, doi:10.1007/s11042-020-08629-8.

[15] The Mathworks, Inc., Natick, Massachusetts, MATLAB version 9.4.0.813654 (R2018a), 2018.

[16] Available online: `https://jo.opensooq.com/en/cars/cars-for-sale` (accessed on 3 Jan 2020).

[17] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, **39**(6), 1137 – 1149, 2017, doi: 10.1109/tpami.2016.2577031.

[18] Q.-x. Zhang, G.-h. Lin, Y.-m. Zhang, G. Xu, J.-j. Wang, "Wildland Forest Fire Smoke Detection Based on Faster R-CNN using Synthetic Smoke Images," Procedia engineering, **211**, 441–446, 2018, doi:10.1016/j.proeng.2017.12.034.

[19] J. A. Sethian, "A fast marching level set method for monotonically advancing fronts," Proceedings of the National Academy of Sciences, **93**(4), 1591–1595, 1996, doi:10.1073/pnas.93.4.1591.

[20] P. L. Rosin, "Measuring shape: ellipticity, rectangularity, and triangularity," Machine Vision and Applications, **14**(3), 172–184, 2003, doi:10.1007/s00138-002-0118-6.

[21] Available online: `http://www.plateshack.com/y2k/Jordan/jordany2k.htm` (accessed on 3 Jan 2020).

[22] Available online: `https://en.wikipedia.org/wiki/Vehicle_registration_plates_of_Jordan` (accessed on 3 Jan 2020).

[23] Available online: `https://en.wikipedia.org/wiki/Mandatory_(typeface)` (accessed on 3 Jan 2020).

[24] "ANPRonline," Available online: `https://anpronline.net/index.html` (accessed on 3 Jan 2020).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 710-717 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Adaptive Identification Method of Vehicle Model for Autonomous Driving Robust to Environmental Disturbances

Yohei Yamauchi*, Mitsuyuki Saito

Systems Engineering, Graduate School of Information Sciences, Hiroshima city University, Hiroshima 731-3194, Japan

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 11 November, 2020
Online: 24 November, 2020

Keywords:
Autonomous Driving
Vehicle Model
Adaptive Identification
Modeling Error
Neural Network

ABSTRACT

Many recent studies on autonomous driving have focused on model-based control. A number of studies has addressed that simple models such as the Kinematic Bicycle Model are easier to design controls for autonomous driving systems. However, such a simple vehicle model has a weakness in that it is subject to modeling errors. This is because it does not take into account the nonlinear characteristics due to road conditions and driving conditions (environmental disturbances: road friction coefficient, large steering, acceleration, sideslip, etc.) Therefore, the purpose of this study is to identify vehicles with high accuracy and in real time, adapting to environmental disturbances.
This study propose a vehicle model based on the Kinematic Bicycle Model. The nonlinear characteristics of the vehicle are represented by the deviation of the front wheel steering angle of the Kinematic Bicycle Model. This deviation is trained and estimated online using a three-layer Neural Network. In other words, the AI is adaptive learning of modeling errors caused by nonlinear characteristics of the vehicle.
This paper presents an example of model-based control using model predictive control.

1. Introduction

In recent years, study and development of autonomous driving has been conducted in the automobile industry, IT companies, and universities in each country. In study on autonomous driving, some autonomous driving systems that combine Artificial Intelligence (AI) and model-free control methods is proposed [1,2]. However, it is considered that such the autonomous driving system is difficult to obtain system stability and reliability in unknown environments. Therefore, fusion technology of AI technology and model-based control has gained much importance in study on autonomous driving [3,4]. Model-based control is a control method in which a control target is represented by a mathematical model and optimal control input is determined based on the model. Model-based control is widely used in various industries [5]. It has problem that the control performance cannot be exhibited when the model is different from the actual dynamics. Additionally, the more complex the controlled object, the more complex the model and the more the amount of calculation. There is a limit to the number of computing units that can be equipped in an autonomous vehicle. Therefore, the model used for autonomous driving is required to be a simple model with less calculation amount. This paper proposes a simple and highly accurate method for vehicle identification (partially published in [6]).

Several studies agree that simple models, such as Kinematic Bicycle Model [7] and linear single-track model [8], are easier to design controllers for autonomous driving systems [9,10]. These vehicle models do not include nonlinear characteristics due to road conditions and driving conditions (environmental disturbances: road friction coefficient, large steering, acceleration, sideslip, etc.). Hence, the accuracy may be deteriorated due to a modeling error between the actual vehicle and the vehicle model. In order to consider the nonlinear characteristics of the vehicle, vehicle models that includes model equations such as tires and suspensions in the vehicle model has also been proposed [11,12]. However, since these vehicle models include multiple models expressions in the vehicle model, the structure of the vehicle model is complicated. It is inferred that if these are used in an autonomous driving system, it may impose calculated load on the computing unit and impair the real-time performance of the system. In other words, it is important for the vehicle model used for autonomous driving controllers to accurately model the vehicle in real time, even if there are environmental disturbances. This paper proposes a vehicle model based on the Kinematic Bicycle Model [7] in order to represent vehicle behavior simply and with high accuracy. The

*Corresponding Author: Yohei Yamauchi, Hiroshima City University, +81-82-830-1749, yamauchi-ics.info@tomatolynx4.sakura.ne.jp

Kinematic Bicycle Model does not include nonlinear characteristics due to acceleration and deceleration or large steering, etc. Therefore, an error may occur between the actual gravity center position of the vehicle and the gravity center position calculated by the vehicle model. The behavior of the actual vehicle and the behavior calculated by the model are different due to the position error of the center of gravity, and the modeling error becomes large. Therefore, this method considers the vehicle model in which the center of gravity is fixed at the center of the wheelbase of the Kinematic Bicycle Model and the modeling error is expressed by the deviation of the front wheel steering angle. In addition, this study uses Neural Network to adaptively identify vehicle model by training and estimating the deviation. This paper verifies the usefulness of the proposed method through simulation experiments using vehicle motion analysis software (CarSim: Virtual Mechanics). In this study, simulations were performed in situations closer to actual driving conditions than in [6] (Section 4). Since this method models the vehicle while determining the control input in real time, it does not exist as a modeling technology alone and must be combined with model-based control. This paper shows an example using Model Predictive Control (MPC) as an example of model-based control to show the usefulness of the method. The method requires accurate location information acquisition. Since it is expected that the measurement accuracy will improve with the development of GNSS (Global Navigation Satellite System) in the future, the simulation is performed assuming that accurate position information can be obtained.

In summary, there are two aspects of the proposed approach that are particularly unique. The first is that the structure of the model is simple and easy to identify. In conventional models, several parameters must be identified in advance, but only one parameter is required in this study in advance. This means that the controller design of autonomous driving could be simplified by relieving the task of examining cornering stiffness and tire parameters in advance. Second, by focusing on the coordinates of the center of gravity, the approach can analyze the entire vehicle as nonlinear motion. Online learning may be able to respond to changes in vehicle mass (due to the number of passengers and loads) and road surface. It is notable that the method is robust to environmental disturbances and easy to identify.

This paper sets up the issue in Section 2. Section 2.1 introduces the conventional method and Section 2.2 describes our proposed identification method in detail. This paper also presents and discuss the simulation results in Sections 3 and 4. Section 3 mainly considers the effects of acceleration, deceleration and steering on the vehicle's nonlinear characteristics, while Section 4 considers the situation with road surface changes. And Section 5 concludes this paper.

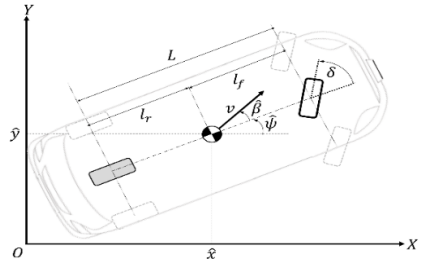


Figure 1: Kinematic Bicycle Model on the XY coordinate

2. Statements of The Issue

This section will set the issue for the proposed method. Section 2.1 introduces simple two-wheel models and accurate nonlinear models to clarify the problem. Section 2.2 details the proposed method for solving the problem.

2.1. Conventional study of vehicle models

2.1.1. Two-wheel model with simple structure

Typical vehicle models used for model-based control include simple two-wheeled models such as the Kinematic Bicycle Model [7] and linear single-track model [8]. These vehicle models are based on the assumption that the state quantities are observed instantaneously, and some conditions (e.g. constant speed, left and right tire characteristics are equal, roll and pitching motions are ignored) are set to represent vehicle dynamics in a simplified way. These vehicle models have simple structure, and thus the turning radius can be easily calculated. Therefore, they can be easily introduced to the controller design of autonomous driving systems. However, these do not take into account various nonlinear characteristics due to environmental disturbances (road friction coefficient, large steering, acceleration, etc.), which can cause modeling errors between the actual vehicle and vehicle models. This paper uses the Kinematic Bicycle Model as an example of the simple two-wheel model to test its accuracy. The model diagram of the Kinematic Bicycle Model is shown in Figure 1. Equations (1-4) show the model equations of the vehicle model. The velocity $v[k-1]$(m/s) and the front wheel steering angle $\delta[k-1]$(rad) are inputs, and the center of gravity coordinates of the vehicle model $(\hat{x}[k],\ \hat{y}[k])$, the direction of the vehicle model $\hat{\psi}[k-1]$(rad), and the sideslip angle around the center of gravity $\hat{\beta}[k-1]$(rad) are outputs. $(x[k-1], y[k-1])$ is the coordinates of the center of gravity observed one step ago.

$$\hat{x}[k] = x[k-1] + v[k-1]\Delta t \cdot \cos\left(\hat{\psi}[k-1] + \hat{\beta}[k-1]\right) \tag{1}$$

$$\hat{y}[k] = y[k-1] + v[k-1]\Delta t \cdot \sin\left(\hat{\psi}[k-1] + \hat{\beta}[k-1]\right) \tag{2}$$

$$\hat{\psi}[k-1] = \hat{\psi}[k-2] + \frac{v[k-2]\Delta t}{l_r}\sin\hat{\beta}[k-2] \tag{3}$$

$$\hat{\beta}[k-1] = \tan^{-1}\left(\frac{l_r}{l_f + l_r}\tan\delta[k-1]\right) \tag{4}$$

here k is the current time, Δt is the sampling time, $l_{f(r)}$(m) are the distance from the front (rear) wheel axle to the center of gravity, and L(m) is the wheel base.

The trajectory of the vehicle model without these nonlinear characteristics (Figure 1) is confirmed. In this case, experiments and verifications should be performed using actual vehicles, but verifications are performed by simulation experiments that are easy to analyze and verify and that can accurately acquire the

vehicle state. Specifically, this study used the Driving Simulator in Figure 2. The Driving Simulator is a Windows PC with a vehicle motion numerical analysis software (CarSim) and a game handle device (Logitech) connected. The PC used in the simulation are as follows: Windows10 64bit, CPU: Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz, and installed memory (RAM):8GB.

Figure 2: Driving Simulator

A simulation course [Slalom] was created on CarSim as shown in Figure 3. The driver drove this course between two pylons lined up in the course, gradually increasing the velocity as shown in Figure 5. The running trajectory at that time is the solid line in Figure 4 (Vehicle's running trajectory). The velocity v obtained as the vehicle data at that time is shown in Figure 5, and the front wheel steering angle δ is shown in Figure 6. The velocity v and the front wheel steering angle δ are input to the vehicle model (1-4) and the running trajectory is calculated as shown by the broken line in Figure 4 (Equation of Vehicle Model). Figure 7 shows the position error between the observed trajectory (x, y) and the trajectory calculated by the vehicle model $(\hat{x}, \hat{y})$. From Figure 4 and Figure 7, there is a maximum position error of about 0.15m in the running trajectory of the actual vehicle and the running trajectory calculated using the vehicle model of (1-4). This is thought to be due to the nonlinear characteristics (tire deformation, expansion and contraction of suspension, etc.) caused by acceleration, deceleration and steering during running. It is consider that the front wheel steering angle δ and the vehicle traveling direction do not match due to the influence of the nonlinear characteristic. This deviation affects the modeling error. In addition, this simulation is based on the assumption of asphalt surface. If it is snow or ice road, the deviation increases further. This is because the effect of the road surface is not taken into account in this model.

Figure 3: Simulation course[Slalom]

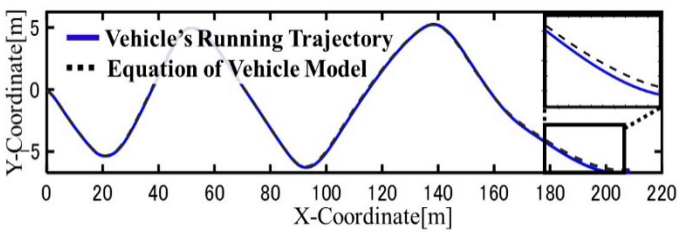


Figure 4: Driving trajectory

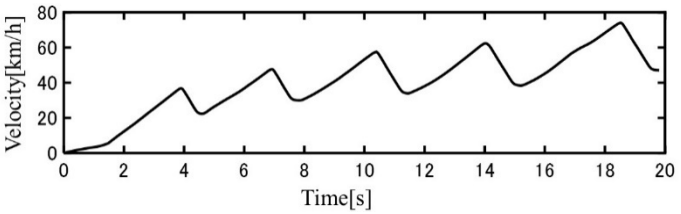


Figure 5: Velocity v

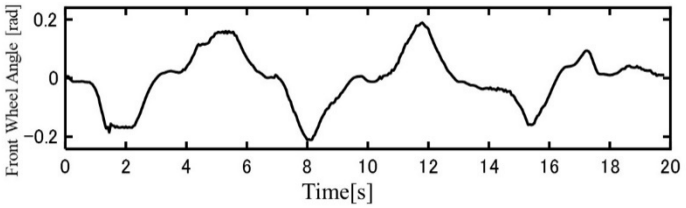


Figure 6: Front wheel steering angle δ

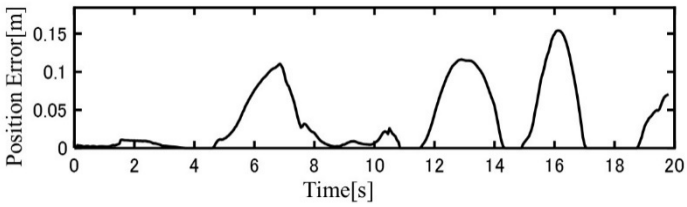


Figure 7: Position error

2.1.2. *Example of non-linear vehicle models that accurately represents vehicle behavior*

As shown in 2.1.1, simplifying the vehicle behavior may increase the modeling error. Hence, there are several conventional studies that use nonlinear vehicle models to represent the nonlinear motion of vehicles in detail. In literature [13], an autonomous driving system combined with a nonlinear vehicle model and MPC is proposed. The nonlinear model is a combination of two-wheel model and nonlinear tire model. This literature shows good results even on compacted snow surface with a low coefficient of friction. In this literature, two tire models were prepared beforehand, one for asphalt and the other for compacted snow, and were tested on each surface. In other words, the experiment is based on the assumption that the road friction coefficient is known. This means that the road friction coefficient, which changes from time to time, must be known.

To solve these problems, a combination of adaptive Model Predictive Control and tire-stiffness estimator [14] has been proposed [15]. This method estimates the tire stiffness from the tire-stiffness estimator. It is able to estimate tire stiffness in situations where the road surface changes and select the optimal road friction coefficient and tire parameters. However, the relationship between the chosen parameters and tire stiffness must be known. In order to find out the relationship between the two, it is necessary to conduct field tests or using a testbench beforehand, which may change depending on the degree of tire wear and other factors such as ageing. In addition, these literatures focused only on tire nonlinearity and did not mention nonlinear vehicle motion due to changes in vehicle mass (due to the number of passengers and loads.), suspension, body stiffness and other effects. These nonlinear motions can also lead to modeling errors. This paper proposes a vehicle model for online learning of nonlinear characteristics of the vehicle by focusing on the change of the vehicle's center of gravity position. By focusing on the change of

the center of gravity, it is possible to model not only the tires but also the nonlinear characteristics of the entire vehicle. Furthermore, online learning eliminates the hassle of pre-testing and allows you to deal with disturbances such as vehicle mass that change with each drive

2.2. Vehicle Model to Estimate Modeling Error

As described in Section 2.1.1, due to the nonlinear characteristics of automobiles, deviation occurs between the front wheel steering angle δ and the actual running direction of the vehicle. The actual direction of travel of the vehicle is defined as the front tire steering angle $\hat{\delta}$. Here, the front wheel tire steering angle $\hat{\delta}$ is an angle that includes nonlinear characteristics due to environmental disturbances and vehicle dynamics. The front wheel steering angle δ is the angle that the front wheels are facing, which can be calculated by the steering wheel angle δ_{SW}. In order to accurately represent the behavior of the vehicle, it is necessary to include in the vehicle model the deviation between the direction the front wheels are facing and the direction the vehicle is actually going, in other words, the deviation between the front wheel steering angle δ and the front wheel tire steering angle $\hat{\delta}$. However, it is difficult to directly observe and theoretically obtain the deviation. Therefore, this deviation is named the modeling error $\hat{\alpha}$ and is defined as the front tire steering angle $\hat{\delta}$ as (5).

$$\hat{\delta}[k-1] = \delta[k-1] + \hat{\alpha}[k-1] \tag{5}$$

$$\hat{\beta}[k-1] = \tan^{-1}\left(\frac{l_r}{l_f + l_r}\tan\hat{\delta}[k-1]\right) \tag{6}$$

Since the front wheel tire steering angle $\hat{\delta}$ is defined as the actual direction in which the vehicle is traveling, (4) is modified as in (6). In other words, our proposed vehicle model is (1-3,5,6). The vehicle model is as shown in the Figure 8. The vehicle model needs to identify the distance from the front (rear) wheel axle to the center of gravity $l_{f(r)}$. The position of the vehicle's center of gravity changes from moment to moment during driving. This is because acceleration, deceleration and large steering causes nonlinear motion in the vehicle, including the tires and suspensions. It is difficult to determine the exact position of the vehicle's center of gravity. Therefore, in this study, the position of the center of gravity of the vehicle is fixed at the center of the wheelbase ($l_{f(r)} = L/2$), and the identification error of the center of gravity position is corrected by $\hat{\alpha}$.

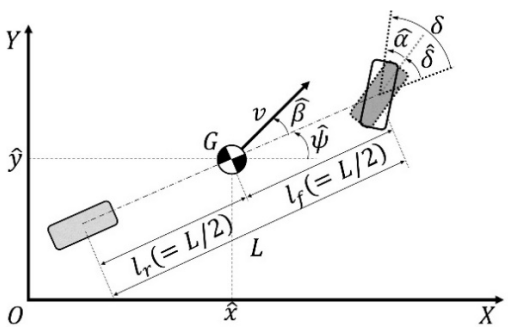


Figure 8: Vehicle model including modeling error $\hat{\alpha}$

Here, a method for estimating the modeling error $\hat{\alpha}$ is described. Since the modeling error $\hat{\alpha}$ is the parameter representing nonlinear motion due to acceleration and deceleration, steering, and road surface changes, it has nonlinearity and is expected to change from moment to moment. Therefore, this study proposes the method for estimating the model error $\hat{\alpha}$ in real time while deriving the control input by model-based control.

This paper considers the system that uses MPC to derive the front wheel steering angle δ and velocity v that are control inputs. MPC is a control law that derives the optimal control input while predicting its future behavior using a predictive model representing the dynamics of a control object. MPC solves an open-loop optimal control problem from the current time to finite horizon for each control period. MPC is an attractive method for controlling autonomous vehicles because it can consider the dynamics and constraints of the controlled object and the ability to adapt to driving scenarios [16-18]. Figure 9 shows the system configuration.

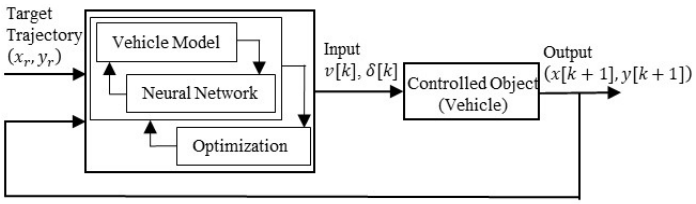


Figure 9: Block diagram of the proposed system

This system uses a Neural Network to train online the nonlinear characteristics of a vehicle that cannot be considered in the vehicle model. The trained Neural Network is used to control the vehicle while estimating the unknown parameters of the vehicle model. Figure 10 shows the flowchart of this system.

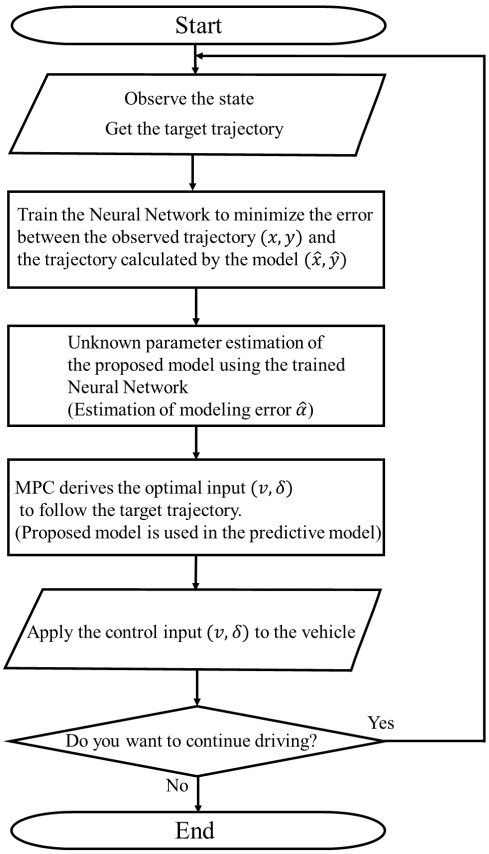


Figure 10: Flowchart of the adaptive identification method for vehicle

The control goal of MPC is to derive the control input that matches the output with the target trajectory. The control objective is to derive the control input ($\delta[k]$ and $v[k]$) that matches the output ($x[k+1]$, $y[k+1]$) with the target trajectory ($x_r[k+1], y_r[k+1]$). This system derives control inputs that minimize the cost function (7).

$$J(\boldsymbol{\delta}(k), \boldsymbol{v}(k)) = \sum_{i=1}^{H} [(\hat{x}[k+i] - x_r[k+i])^2 + (\hat{y}[k+i] - y_r[k+i])^2] \quad (7)$$

H is the prediction horizon. The prediction model (Vehicle Model and Neural Network in the Figure 9) predicts the vehicle trajectory ($\hat{x}[k+i]$, $\hat{y}[k+i]$), $1 \le i \le H$. Since the trajectory depends on the future input, the input sequence $v[k+i], \delta[k+i], 0 \le i \le H-1$ are derived so that the predicted trajectory ($\hat{x}[k+i]$, $\hat{y}[k+i]$), $1 \le i \le H$ approaches the target trajectory ($x_r[k+i], y_r[k+i]$), $1 \le i \le H$ of the obtained input sequence, only the first $v[k], \delta[k]$ are used as the actual inputs.

From here, the Neural Network that trains and estimates the modeling error $\hat{\alpha}$ is described. From Figure 4 to Figure 6, it can be confirmed that the position error (modeling error) increases as the velocity v increases and the front wheel steering angle δ increases. In other words, the position error is considered to depend on the velocity v and the front wheel steering angle δ. The parameter required to correct this position error is the modeling error $\hat{\alpha}$. This modeling error $\hat{\alpha}$ is considered to include nonlinearity. This study uses a 3-layer Neural Network with 2 inputs and 1 output for estimation. This is because the nonlinear system is modeled with high accuracy and the load on the computer memory is reduced as much as possible. The relationship between the input and output of the Neural Network is shown in (8-10). In this paper, I_1 and I_2 are inputs, c_1-c_4 are thresholds, and $w1$ and $w2$ are weighting factors. The input value of the hidden layers are s_1-s_4, and the sigmoid function is used for the output value h_1-h_4 of the hidden layers. Akaike's Information Criterion (AIC) is used to determine the number of hidden layers. The input I is the front wheel steering angle δ and the velocity v, and the output O represents the modeling error $\hat{\alpha}$.

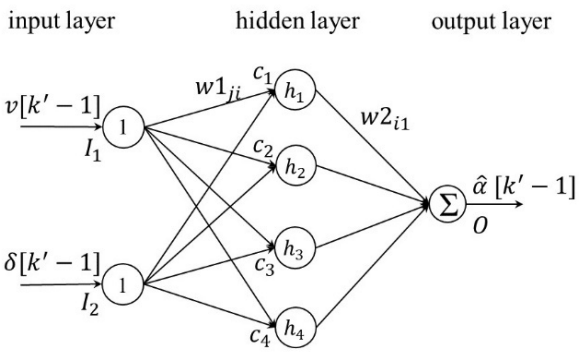
$$s_i[k'-1] = \sum_{j=1}^{2} w1_{ji}[k'-1] \cdot I_j[k'-1] + c_i[k'-1] \quad (8)$$

$$h_i[k'-1] = \frac{1}{1+\exp(-s_i[k'-1])} \quad (9)$$

$$\hat{\alpha}[k'-1] = \sum_{i=1}^{4} w2_{i1}[k'-1] \cdot h_i[k'-1] \quad (10)$$

The observed values of the position coordinates are given to the Neural Network as instruction signal, and online training is performed so as to minimize the cost function (11).

$$I = \sum_{k'=k-W}^{k} \{(\hat{x}[k'] - x[k'])^2 + (\hat{y}[k'] - y[k'])^2\} \quad (11)$$



$(\hat{x}, \hat{y})$ represents the position coordinates of the vehicle model, and (x, y) represents the position coordinates of the actual vehicle. W represents the window width. This neural network is trained so that the position coordinates of the vehicle model $(\hat{x}[k'], \hat{y}[k'])$ from k to W match the position coordinates of the actual vehicle $(x[k'], y[k'])$. Altogether, the parameter $\hat{\alpha}$, which represents the nonlinear properties, is estimated from the position coordinates.

By training and estimating the modeling error due to the nonlinear characteristics online, the behavior of the vehicle can be accurately represented in situations such as acceleration and deceleration, large steering and road surface changes. Because the vehicle is identified in real time, it may be able to respond to changes in vehicle weight, such as changes in the number of passengers. Furthermore, our identification method only uses the wheelbase L as the setting parameter of vehicle model. This means that different types of vehicles can be identified by only changing the wheelbase L. Conventional vehicle models have set parameters (e.g., cornering stiffness, vehicle mass, etc.), which vary for each vehicle. The key feature of this method is that there is only one configuration parameter.

3. Simulation of Fixed Road Surface

In this section, a simulation comparing the Kinematic Bicycle Model (1-4) with the proposed model (1-3,5,6) is described. As in Section 2, the simulation was performed using CarSim installed in the Driving Simulator. It verified whether the center of gravity coordinates of the proposed vehicle model $(\hat{x}, \hat{y})$ matches the center of gravity coordinates of the actual vehicle (x, y). The Kinematics Bicycle Model (1-4) and the proposed vehicle model (1-3,5,6) were given the velocity v and front wheel steering angle δ as inputs, and the trajectory was calculated. The accuracy is checked by comparing the calculated trajectory with the actual vehicle trajectory. The input data and the actual vehicle trajectory are obtained by driving the simulation course shown in Figure 3, which was created in the Driving Simulator (Figure 2) as a driving course with acceleration, deceleration, and steering. This study assumes that the 27 degree of freedom vehicle model in CarSim is the actual vehicle. This simulation assumes a dry asphalt surface (surface friction coefficient $\mu = 0.85$) and drive a B-Class hatchback vehicle (Figure 12).

The driver repeatedly accelerated and decelerated between the two pylons in the course [Slalom] shown in Figure 3. Figure 13 and Figure 14 show the v and δ.The solid line in Figure 15

shows the running trajectory. In this paper, the trajectory is used as the actual vehicle trajectory. The dashed lines in Figure 15 show the trajectory when the velocity v (Figure 13) and the front wheel steering angle δ (Figure14) were given as inputs to the proposed model. The front wheel tire steering angle $\hat{\delta}$ is shown in Figure 16. The calculated position error between the calculated trajectory and the actual vehicle trajectory is shown in Figure 17 and the estimated modeling error $\hat{\alpha}$ is shown in Figure 18. As shown in Figure 17, the maximum positional error is 0.05m, which is considered to be within the practical range.

Figure 12: B-Class hatchback vehicle in CarSim

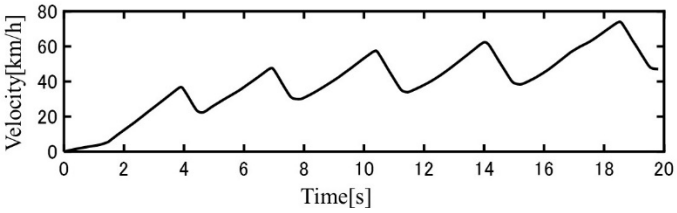


Figure 13: Velocity v □ input□

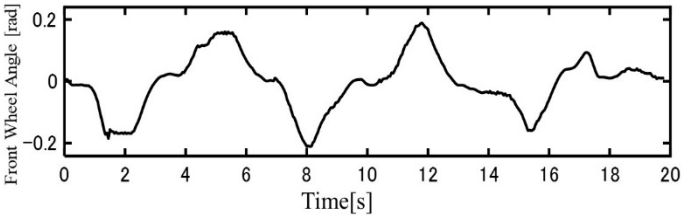


Figure 14: Front wheel steering angle δ □ input□

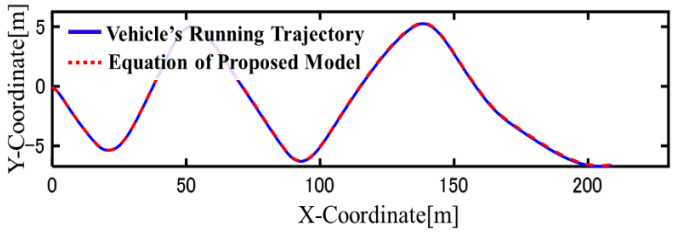


Figure 15: Driving trajectory［Slalom］

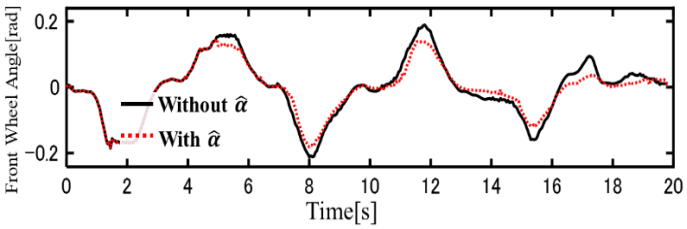


Figure 16: Front wheel steering angle δ and front wheel tire steering angle $\hat{\delta}$

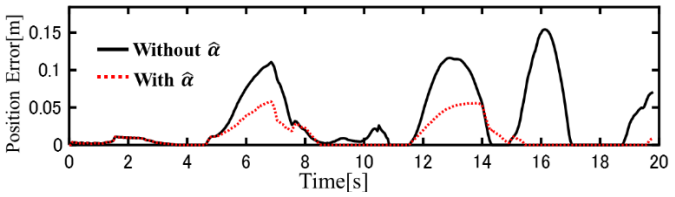


Figure 17: Position error

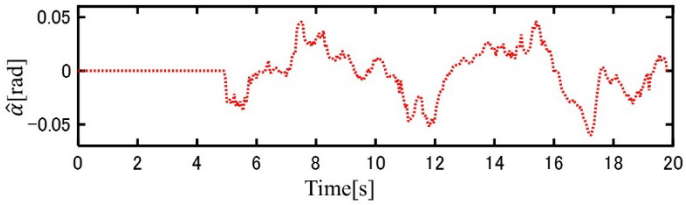


Figure 18: Modeling error $\hat{\alpha}$

The results show that the behavior of the vehicle can be identified with high accuracy. This means that the proposed model (1-3,5,6) can contribute to the controller design of autonomous driving systems using model-based control. However, at this stage, this study has only validated a single driver driving a B-Class hatchback in CarSim several times around the track in a simulation experiment and have obtained good results. In order to prove the effectiveness of the proposed method, it is necessary to conduct similar tests on various courses and vehicle models, and this is a subject for future study.

4. Simulation of Road Surface Change

This section presents additional examples of situations that more closely resemble actual driving situations in order to verify the usefulness of the proposed model. As in Section 3, the experiments were conducted using the Driving Simulator shown in Figure 2. The course used is shown in Figure 19. This course was designed to simulate a mirror burn. Mirror burn is a phenomenon in which the surface of the road is polished by the traffic and becomes very slippery at a part of the intersection. This course was driven by the vehicle (B-class hatchback) in CarSim. This course is designed to have surface friction coefficient μ=0.2 at the center of the intersection and μ=0.5 outside the center of the intersection. This simulation assumes driving on the left side of the road because it is based on Japanese roads. The trajectory of the vehicle on this course is treated as the center of gravity coordinates of the actual vehicle (x, y). In addition to vehicle dynamics, this simulation allows us to verify whether the vehicle can adapt to changing road conditions. The trajectory of the vehicle $(x[k],\ y[k])$ while driving on the course is shown by the solid line in Figure 22. The dotted lines in Figure 20 to Figure 25 indicate the boundary of the surface friction coefficient. Figure 20 and Figure 21 show the v and δ.The solid line in Figure 22 shows the running trajectory. In this paper, the trajectory is used as the actual vehicle trajectory. The dashed lines in Figure 22 show the trajectory when the v (Figure 20) and δ (Figure 21) were given as inputs to the proposed model. The front wheel tire steering angle $\hat{\delta}$ is shown in Figure 23. The calculated position error between the calculated trajectory and the actual vehicle trajectory is shown in Figure 24 and the estimated modeling error $\hat{\alpha}$ is shown in Figure 25. As shown in Figure 24, the maximum positional error is 0.01m, which indicates that the proposed model is able to adapt to the changes in the road surface.

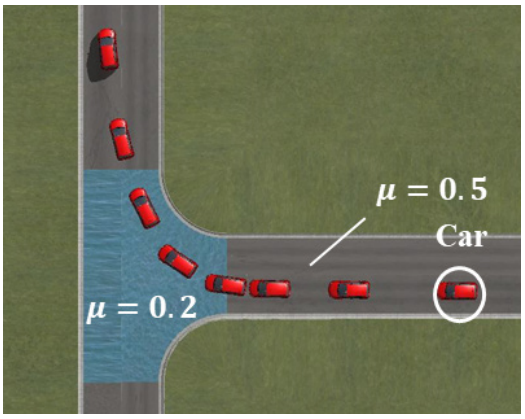


Figure 19: Simulation course［Mirror Burn］

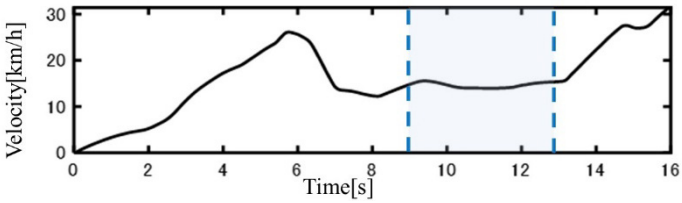


Figure 20: Velocity v（Input）

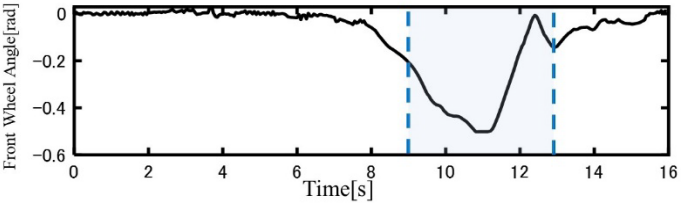


Figure 21: Front wheel steering angle δ □ Input□

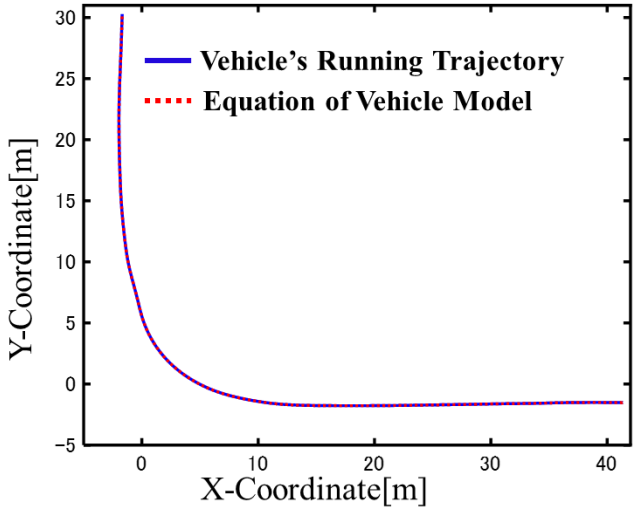


Figure 22: Vehicle's driving trajectory［Mirror Burn］

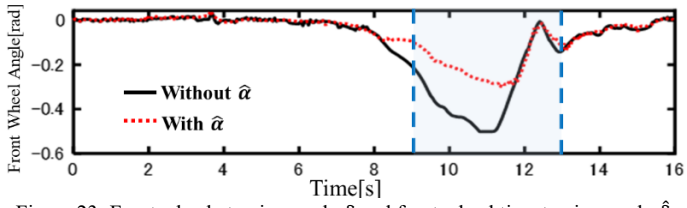


Figure 23: Front wheel steering angle δ and front wheel tire steering angle $\hat{\delta}$

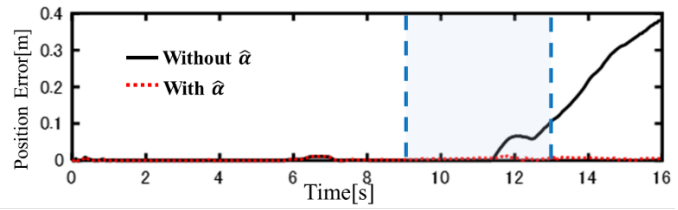


Figure 24: Position Error

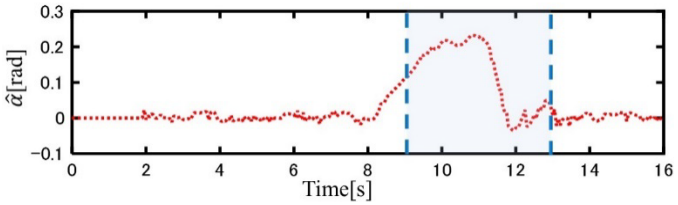


Figure 25: Modeling Error $\hat{\alpha}$

5. Conclusion

The purpose of this paper is to identify an autonomous vehicle with high accuracy in real time. This paper proposed a simple vehicle model that represents the error in the center of gravity between the actual vehicle and the vehicle model as the deviation $\hat{\alpha}$ of the front wheel steering angle. This paper also proposed the method to estimate the $\hat{\alpha}$ in real time using neural network, and simulation experiments using CarSim showed the usefulness of the method in situations that require acceleration and deceleration, large steering, and road surface changes. The authors emphasize that the method can represent the nonlinear characteristics of the vehicle as it is learning online and that the only parameter to be identified in advance is the wheelbase. In other words, this study can eliminate the process of identifying multiple parameters beforehand and contribute to the design of control controllers that is robust to ever-changing environmental disturbances.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. Bojarski, D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. Jackel, M. Monfort, U. Muller and J. Zhang, et al. "End to end learning for self-driving cars." arXiv preprint arXiv:1604.07316, 2016. https://arxiv.org/abs/1604.07316

[2] F. Codevilla, M. Müller, A. López, V. Koltun and A. Dosovitskiy, "End-to-end driving via conditional imitation learning." in 2018 IEEE International Conference on Robotics and Automation (ICRA), BNE, AU, 2018. https://doi.org/10.1109/ICRA.2018.8460487

[3] B. Paden, M. Čáp, S. Yong, D Yershov and E. Frazzoli. "A survey of motion planning and control techniques for self-driving urban vehicles." IEEE Transactions on intelligent vehicles, **1**(1), 33-55, 2016. https://doi.org/10.1109/TIV.2016.2578706.

[4] T Ariyoshi. "パーソナルカー 自動運転システムとそのめざす価値." SICE 57(7), 488-492, 2018, in Japanese. https://doi.org/10.11499/sicejl.57.488

[5] S.Adachi. "物理と情報と制御." Calsonic Kansei technical review, 11, 6-14. 2014, in Japanese.

[6] Y. Yamauchi, M. Saito and T. Ono "Adaptive identification method of vehicle modeling according to the fluctuation of road and running situation in autonomous driving", 2019 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), HIJ, JP, 2019. https://doi.org/10.23919/SICE.2019.8859845

[7] R. Rajamani, Vehicle Dynamics and Control, ser. Mechanical Engineering Series. Springer, 2011.

[8] R Bosch "Automotive Handbook" 8 ed. John Wiley & Sons Ltd, 2011.

[9] J. Kong, M. Pfeiffer, G. Schildbach and F. Borrelli. "Kinematic and dynamic vehicle models for autonomous driving control design." in IEEE Intelligent Vehicles Symposium (IV), SEL, KR, 2015. https://doi.org/10.1109/IVS.2015.7225830

[10] P. Polack, F. Altché, B. d'Andréa-Novel and A. Fortelle. "The kinematic bicycle model: A consistent model for planning feasible trajectories for autonomous vehicles?." in IEEE Intelligent Vehicles Symposium (IV), SEA , USA, 2017. https://doi.org/10.1109/IVS.2017.7995816

[11] K.Kabe and N. Miyashita" A study of the cornering power by use of the analytical tyre model." Vehicle System Dynamics, **43**(sup1), 113-122, 2005. https://doi.org/10.1080/00423110500139999

[12] Y. Mitsuhashi and S. Fujii. "Validation of Fiala model for motorcycle with FEM tire model.(article in Japanese with an abstract in English)" Transactions of the Society of Automotive Engineers of Japan, **44**(1), 75-80,2013. https://doi.org/10.11351/jsaeronbun.44.75

[13] Y. Uno, R. Quirynen, S. Di Cairano, B. Karl. "Vehicle Integrated Control by Nonlinear Model Predictive Control in GNSS Based Autonomous Driving System." IJAE, **50**(4), 1176-1181, 2019, in Japanese. https://doi.org/10.11351/jsaeronbun.50.1176

[14]] K. Berntorp and S. Di Cairano, "Tire-stiffness and vehicle-state estimation based on noise-adaptive particle filtering," IEEE Transactions on Control Systems Technology, **27**(3), 1100-1114, 2018. https://doi.org/10.1109/TCST.2018.2790397

[15] B. Karl, R. Quirynen, and S. Di Cairano. "Steering of autonomous vehicles based on friction-adaptive nonlinear model-predictive control." 2019 American Control Conference (ACC), PA, USA, 2019. https://doi.org/10.23919/ACC.2019.8814448

[16] F. Borrelli, P. Falcone, T. Keviczky, J. Asgari and D. Hrovat. "MPC-based approach to active steering for autonomous vehicle systems." International journal of vehicle autonomous systems **3**(2-4), 265-291 2005. https://doi.org/10.1504/IJVAS.2005.008237

[17] M. Bujarbaruah, X. Zhang, H. Tseng and F. Borrelli. " Adaptive mpc for autonomous lane keeping." 14th International Symposium on Advanced Vehicle Control (AVEC), BJS, CN, 2018. https://arxiv.org/abs/1806.04335

[18] X. Qian, I. Navarro, A Fortelle and F. Moutarde. "Motion planning for urban autonomous driving using Bézier curves and MPC." in 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), RIO, BR, 2016. https://doi.org/10.1109/ITSC.2016.7795651

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 718-723 (2020)

www.astesj.com

Second International Virtual Conference on Multidisciplinary Research 2020

ASTESJ

ISSN: 2415-6698

Cognitive Cybernetics in the Foresight of Globalitarianism

Zdenko Balaž[1], Krystian Wawrzynek[2,*]

[1]*VERN University, Iblerov trg, Zagreb, 10000, Croatia*

[2]*Zagreb University of Applied Sciences, Zagreb, 10000, Croatia*

ARTICLE INFO

Article history:
Received: 23 April, 2020
Accepted: 07 October, 2020
Online: 24 November, 2020

Keywords:
Cognitive
Cybernetics
Totalitarianism
Globalitarianism
Education
Captology

ABSTRACT

This paper presents the results of the research conducted with the help of cognitive cybernetics about the "mass" factor from the theory of totalitarianism. According to the expert system model, "big data" analysis sought to discover knowledge for assessing the future status of digital social connectivity. Originally developed models and methods of cognitive and computer research and processing of "eminent text" with the help of "convolution" from the theoretical background of many works about the past totalitarianism, recognize the same characteristics of emerging globalitarianism. Using search correlation, algorithms have confirmed a suspicion that intelligent interactive technologies impact the changes of the human psychophysical structure, through the digital social network of globalitarianism. Applying new intelligent interactive technologies without being familiar with their deeper impact, could plausibly make people accept them as part of themselves. People's digital obsession with the internet in a "global village", is increasingly similar to blind obedience, of the sympathizers and followers, gathered around past totalitarianism's grand leaders. There is a visible correlation between such technological-integration engagement, and the loss of social intelligence, leading to a conclusion that the future interactions between people and intelligent technologies will turn the history into the implosion of current events. History today is already pretty accelerated, due to its own mass digital integration and interaction, turning it into an illusion. There is a danger that globalitarianism, as an elusive end of the history, will escape cyclical times, due to unavoidable repetitiveness.

1. Introduction

According to the above-mentioned sample literature, Žarko Paić's book: "The talk about totalitarianism can never be neutral nor objective, like talking about something that forever belongs to the distant past. Man in his historical being is a creature of the existential sense of responsibility and also has his endless mission. What's left, in the end, is a radical settling of accounts with the possibilities that the past as a traumatic event constantly defines our future with a sense of guilt and responsibility. By taking over the legacy of the past something much bigger than just an unconditional ethical obligation is also taken along. This is the last lesson in which this world is becoming a place for searching the truth without fear of the consequences. And that's the only thing that makes sense. And therefore is the hardest."

*Corresponding Author: Krystian Wawrzynek, krystian.wawrzynek@gmail.com

Based on the principles of cognitive cybernetics, searching for truth could start by creating a system of history placed in time-space, in which from today's perception humans' integral actions are used for the purpose of technology. The first step for that system is to define time and its thinking models, and later to begin with cybernetic observation. Hannah Arendt, in her book (selected sample literature…), points out how being a human and being free, (the event of liberty- lat. *constitutio libertatis*), is the one and the same thing. A human being is the only creature in the world capable of initializing an action which could trigger a change characterized by unpredictability, uniqueness, and singularity of its activity.

In globalitarianism, these characteristics are lost, and as such, in this world of actions, are inevitably and unnecessarily subsided into the transformation of the events into the life itself. Through globalization, a man is brought into a state presented to him as a limit of his existence, and while venturing into modern technology

projects, and without the ability to think for himself, he doesn't even realize he is not free. In this digital sociability, people could become followers of an unknown leader, who gains their obedience due to being unnoticeably lured. Revealing a lack of freedom is managed by the help of cognitivistic and cybernetic interoperability, in processing thematic contents.

2. Historical activities

From the Cybernetics point of view, history as a system of actions aspires towards a contingency of events, which in historical aspect means that it strives towards stability. Civilization as a system, when it is coming closer to a perfect operationalization, is on the brink of its breakdown. Observing history in space as a real factor in which it is happening, and in time as an abstract factor, an event happening in running time finds itself in a void between history and future, presenting a mindboggling reality that new information is already outdated. Sensory experience in space/time within which it "instantly happens", as a "situational psychology" in "position grouping", "is the essence of sustainability in the encounter as a constellation" [1-4].

The effect of today's intelligent technology is so strong that the continuity of today's historical era is becoming inhumane machinery. Authors of the selected representative writings [5-6], on historical example of totalitarianism, point out that a man through right activities can take control of what is not allowed and what is needed to be done. The criteria for doing good are selflessness and thoughtfulness which could produce positive indications in behavior towards others. Common sense is not common to everybody, but it helps people to fit into a community and allows them to communicate in situations expressed by senses. Common sense judgments are community judgments, where the sound mind is the only sense capable to act. Capability and possibility of human (no)actions are disclosed in this paper by cognitive cybernetics, as something past and terrible (totalitarianism), but also as potentially dangerous still to come (globalitarianism). Figure 1.

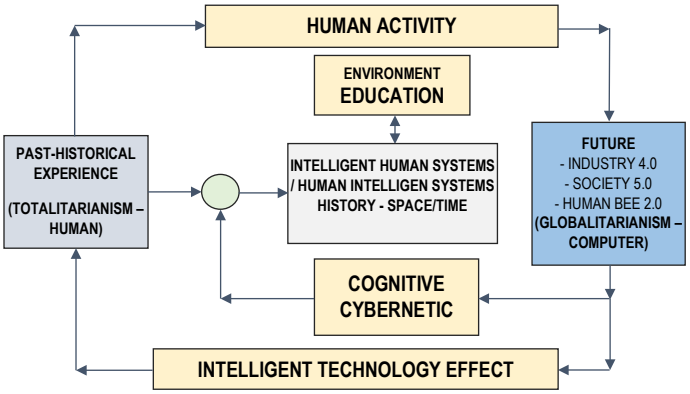


Figure 1: Space history of the system with cognitive cybernetics feedback and efficiency of human and intelligent technologies

For a human activity, which could be mental or physical, one needs a will which must be free and reasonable, otherwise, it wouldn't exist. Free will behaviour is able to identify itself with a command, but also with predictable unpleasant feelings of force and self-deception. With a lack of will, self-deception stops surpassing the feelings, in order to achieve a command. Therefore while searching for answers about history's time, it would be useful to search for them in its historical experiences but also in predictions of its future.

It is noticed that history's time is slowed down by indifference and amazement and those are the exact parameters of the historical activity that the cognitive cybernetics which also confirms its importance in globalitarism, and which has been precisely discussed in the literature [7-8].

2.1. Time

Based on a philosophical perception of historical experience, (Fink, Gadamer, Paić, Wilber, Husserle,…), time (lat. *tempus*) stands for a way and a flow. According to Plato, it is a motion picture of eternity, and a factor of all changes in the world, and every change happens suddenly (gr.εξαφνικότητα) when one form transforms into another. The link between time and unexpectedness is amazement, which for an instant turns into an ecstasy, represented in the word "now". If time is authentic, then it gets a sense of being, from its truth in the original time and belongs to the topology of the being. Historical time makes sense in narrating factual events, while primal or original time, i.e. prehistoric (gr.- *αρχικά*), is lost as soon as it is brought down to a level of "just now". A paradox of time as repetition is a pseudo synthesis of the past and the future, like "eternal occurrence" of the exact same thing that is presumably objective necessity for freedom of actions. The chronological sense of time calculated by a clock is a measurement for living a human life. The time paradox is the objective time in a technical sense, which is eternal present and eternal future, and infinity and eternity. Generally, time's metaphysical structure is an everyday life present in relation to the time essence, and zone of infinity and temporality, as well as a linear cyclical flow of events as existence. Time's psychological structure raises the "self" from the abyss to the highest spheres of being and existence, in the form of life's creative potential [9-11].

Time as a meeting event and a mediation medium occur in possessive attitude towards things, and as temporality, it is an original, existential, hermeneutic phenomenon of engagement in the reconstruction of fundamental issues of time and being. Time as an abyss of freedom is not astronomical nor calendar, because it is unlimited in a sense that even now is a limitless line of dots. Time understood as an ability to act in a certain date in time, marked using time measurements based on important world's events, is stretched between carefully securing a passing of the present while awaiting a transformation of the past, which is far from a simple flow of equal time segments. That is the future that under no circumstances should bet the destruction of history. Time must not be a moment of anything but freedom, to act based on non-exclusivity, discovery, and validation, which are foundations for cybernetic observation and realization, bringing us to a conclusion that time must not ever become virtual [12].

2.2. Cognitive cybernetics

Cognitive cybernetics is a new scientific discipline based on foundation of cognitive sciences and cognition as a human mental potential, that connects cybernetic understanding of everything as a system (technical, social, natural…), with the new intelligent

technologies that by cognition of interactions could point out to captologic anthropomorphosis of present-day human [13-15].

Based on available sources of information, today's science is becoming more and more empirical science about a man and everything that could become his personal technical experience achieved through new technologies. With its help, a human is placed in the world that adequately processes multiple productions. All that is accomplished everywhere based on and in the interest of scientific development of certain human capacity determined and lead by new cybernetics [16-18]. It determines a human as a social-working being and it is a theory that covers the possibility of planning and organizing human work. The language of cybernetics is transformed into a news exchange medium by which the whole world and a human's place in it, is determined and controlled. It is important to respond cognitively because only human engagement which must be thoughtful (thought, will, reasoning - cognition as human energy) and productive (working, producing and acting), can ensure the effect of unbiassed power of freedom. Based on that, two terms introduced from the totalitarianism, represent a foundation on which lies a hypothesis that identifies the meaning of the two globalitarianism's factors which are observed through a prism of today's digital social connectivity: (1) Latent structure of mass conscience; and (2) Permanent terror of irrational rationality. The results of previous research processed in published papers, [12-18] confirm the possibility of implementation in other areas. Considering that a way/goal of an action is much more important than any conscience manipulation, the modern ideology of mass digital society directs attention into the virtual reality which appears as a cultural totalitarian reality [19-23].

3. Globalitarianism as a system

In all totalitarian regimes, morality collapsed and became a mere habit and a custom with catastrophic consequences. Those who were the culprits of totalitarianism and those who remained quiet and accepted the situation, although they had an opportunity to speak up, but chose not to, are equally responsible. Moral behavior is not understood by itself, because a man is influenced by an example. And he reacts based on his own judgments. Using the new intelligent and interactive technologies in digital social networking, a human becomes less and less responsible towards himself. If he lost self-respect and started to despise himself, he wouldn't be able to act in his own interest and towards his own self-improvement anymore, and therefore he would not be able to take care of others.

If global is observed as a system (lat. *globalus),* which relates to the whole planet, with stress on absolutely strict centralistic and birocratic unity as unitarism (lat. *unitus*), then acronym is globalitarianism.

The ideology of intelligent technologies in today's digital society is becoming the means/purpose of a regime and it will cause much worse things then brutality of terror because the historical memories are fading away. Digital social connectivity created by intelligent technologies is close to a violence policracy. That Lars Rensmann's term is connected to a totalitarian state as a pluralistic non-state and paradoxically to self-understanding of everything that doesn't have an analogy in history. A problem occurs when every system (economic, political, cultural) is observed as totally rational and technically perfect, and states as communities of the global world organize the entire life of their citizens and depoliticize members of the birocratic management machine. The modern societies have become mass machinery of consumption fabrication. Technology and science together with a mass consumption culture make those societies dehumanized production plant which is going to use the system of globalitarianism as a process.

3.1. Modeling

When using a search algorithm to select dominant axioms from the Glasser's Choice Theory (CT) which is described by 10 axioms [24], then they could be presented as "Model 4P" where 4Ps are: need, action, behavior, and acceptance (cro. *potreba, pokretanje, ponašanje, prihvaćanje*). The model could also be applied to an evolved captological "3P" model, where 3Ps are: persuasive, permissive, and pervasive.

In "3P" model persuasive is captological, (*Captology = Computer As Persuasive Technology*), and that is presented through the Hermeneutic circle where the Choice Theory will be complemented by four Wilber's modernity quadrants. The circle would consist of a sequence of information, control, time, and freedom. A coincidence between the Hermeneutical circle from the Choice Theory [24], and the level of Wilber's halons [25], where I, "ME" stands for intentional, "WE" for cultural, "THIS" for social and "THAT" for behavioral, is integrated in a role of persuasiveness in totalitarianism, which model will be used to provide adequate parameters in Figure 2.

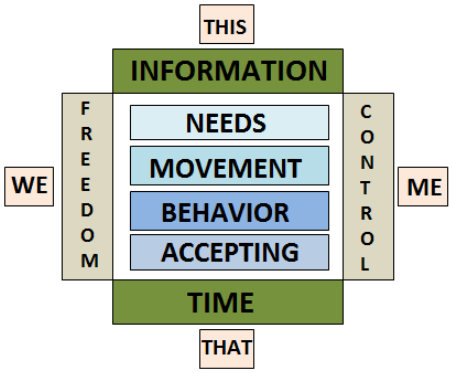


Figure 2: The dominant axioms of Glasser's Theory of Choice and Wilber's modernity for the treatment of totalitarianism with cognitive cybernetics

Processing with the help of the cognitive and cybernetic inter-operability is based on the technical possibilities of the practical search system with adapted communication which are basically:

Syntactic: because they allow systems to communicate and exchange data in specific formats and communication protocols (XML; SQL standards, ASCII formats, including lower format levels of alphabetic signs). Letters treated as a multi-referential sign point out to the order, which is related to the alphabet syntax, and is different from one manifested, and follows the combinations and permutations logic, which in a word and a text display crypto semantic traces, and acquire a visual profile in their interpretation and shape. That way their construction, while inscribed in the order of meaning, determines the literal manifestation of the analyzed text that carries a specific meaning. Storing it into a database provides a memory which preserves the

media, communication, transparency, and uniqueness of the characters. The letters are given the power to repeat in memory what has previously been established, guaranteed and unquestioned.

Semantic: as it allows for automated and meaningful interpretation of exchanged data for accurate formatting of results. The final use of the exchanged information of all involved parties is enabled by a reference model. The exchanged content is unambiguously defined by the original template PETAP, (hrv. *Pomoćni Eminentni Tekst Aplikativne Paradigme*) [23-30]. That template as its name suggests is "eminent text" that: a) serves for a presentation of the fixed sense; b) opens / reveals possibilities for understanding; c) cannot be predicted in advance; and d) has a character of actuality. The actuality emerges as an example of a probable occurrence of reality, as created in cognitive processing by reading and thinking by which a fixed experience of understanding is gained. That way presenting of the fixed thought content means "the presence of an existing" which could objectively meet the normative requirement. It is the same product that is repeatedly appearing in the search stages focusing on it in order to achieve verified meaning and understanding. The fixed content is conditioned and limited by the same thing, and when it was saved (without an ability to escape), that condition modeled on the hermeneutical approach becomes cognitive learning of its own character manifestation saved through technical and general functional forms. Superior cognitive interoperability refers to the cognitive collaboration of different systems, individuals and organizations which besides technical systems cognitive engineering is also taking into consideration social, political, organizational, chronological, biological, psychological and plausible factors. The processing modeled on computer metaphors and terminology is at the end completed with the original algorithmic model MOKSA (hrv. *Model Opservacije Konvolucijsko Separacijskom Analogijom*). According to this model, a mathematical function of convolution with all its characteristics of linearity, commutativity, and associativity is calculated similarly to correlation by mirroring one of the functions [31].

The application of the search and processing tools based on the expert system principle results in the available databases and knowledge bases. Searches that begin by analyzing and specifying requirements for knowledge bases and databases with well-defined need are crossing over into creating applications attempting a search using the learning agents' experience. An action initially begins in an unknown environment, so learning elements are used in the search process and the performance elements are formed by the evaluator and task generator. As the search gains more knowledge, the action becomes better and safer. The Data Collection Model has enabled upgrading in accordance with the acquired knowledge and needs, which were also empirically enabled by using relations and user forms in inputting and reviewing databases and formatting reports. Presenting and fulfilling the sense of the search is close to the cybernetic forensics and a base for presented understanding, which is fulfilled over the time, therefore searching and processing is to be completed by the cognitive synthesis and by reinforcing the concluding expertise.

3.2. Role of Persuasion

Persuasion (lat. *persuasio*) is connected to a form of communication seeking to influence judgments, attitudes, beliefs, or actions of an individual. As part of the persuasion acronym captology (lat. *Persuasibilibus = lured),* refers to being lured by the intelligent interactive computer technologies which also represents the form of communication. A form of communication that conveys information through the media represents propaganda. Extensive researches reveal that some forms of propaganda serve to inform, to persuade or in some more complicated synesthetic way to occupy human life in order to draw attention. The word propaganda dates from 1626 when a Christian community *Sacra congregatio Christiano nomini propaganda* has appeared, founded with the purpose of expanding Catholicism and regulating church relations in non-Catholic countries. More recently, since 1920, it has been used as methods of broad political industrial consciousness and a way of articulating ideological field of action although it is not an ideology itself. Propaganda linked to the development of technical (new) media, for which it has been proven to serve as a massive manipulation of feelings and attitudes.

This fact was disclosed in 2000 based on the research carried out by Manstead, Fischer, Kardum, and Larsen [32]. By setting up a mood control system, described by the cognitive cybernetic system, Larsen found that personality is a dynamic system within an individual which by its psychophysical parameters can adapt to a specific environment, and can also be affected by the environment. Every current state of the individual is his current mood, which Larsen found is modeled after potential mechanisms of individual differences that are in fact his mood regulators. An individual pursuit to achieve and maintain a good and comfortable mood and a positive attitude include the existence of negative feedback in the regulation model. Any action to reduce a difference appearing in the disadvantage of the good mood could be influenced by the strong cultural matrix in which the medium acts, and which could contain the immediate environment disturbances. This is particularly apparent in the totalitarianism where the masses believe in a new culture of organized lies. The problem is a double morality due to the personality philistine profile dominating in such an environment. Additionally the destruction of people's privacy and personal moral continually redefines individual identities imposed by the society as an escape from the social identification. Recruiting from such a mass of seemingly uninterested people, who many consider too apathetic and have already given up on them, allows for an introduction of completely new methods in propaganda and for ignoring of adversary arguments.

The future model as a system functions based on an assumption that the desired mood is to be achieved by comparing the current mood with the one to be achieved, and could be misleading because of not recognizing the desired state due to strong influence of Parameters, (Pi), Mechanisms, (Mi) and Strategies (Si).

The strength of the media is exactly in its appearing as affectively relevant environment, which is confirmed by McLuhan's 1960's anthropological technological determinism [33] in Figure 3.

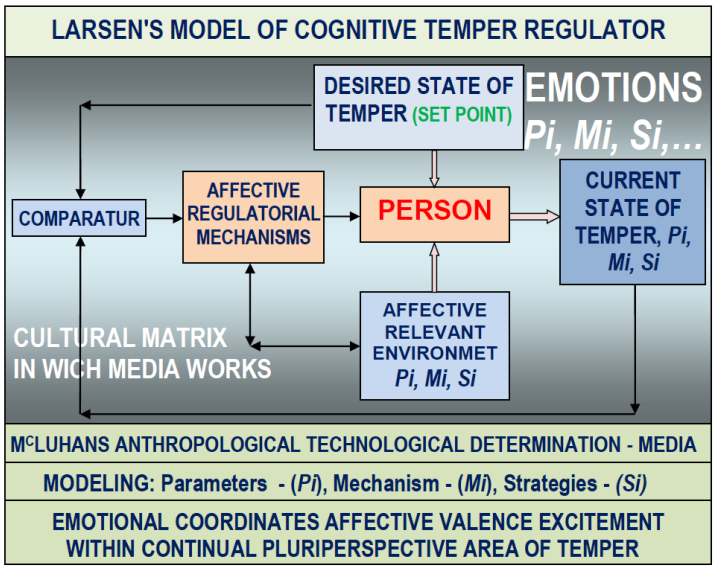


Figure 3: Model of cognitive temper regulation by Larsen and McLuhan's anthropological technological determinism

3.3. Mass Man's Parameters

The early 19th century's predictions pointing to the emerging of a mass man and the arrival of a mass era came true in the twentieth century. Besides the psychology of the European mass man, the profile of the mass man, gained increased importance, by unexpected and unpredictable phenomena such as:

- a radical loss of personal interest in everything that surrounds him,
- cynical and decadent indifference when encountering disasters or other personal catastrophes and even the death itself,
- the passionate tendency towards the most abstract concepts as life ideals,
- general disgust towards any events,
- loss of criteria even towards the most obvious rules of common sense.

Such an apathy, as a feature of totalitarianism, is today's plague called anhedonia. Search implementation based on the results of the modern analytical ontology is successful primarily because of the theory of general categories, which uses a data format to also present a concept. The formats/concepts that are used are for example person, personality, event, and event as a happening. The transition from a possibility to reality is an ontology of real events for which ontological research expands even to the relationships that people have with each other.

Therefore the conceptual modeling has encompassed characteristics determination of the presence of space and time using a classic relational model, whose basic concepts are structure, constraints and set of operations.

All existing real foundations from the early thirties of the past century are attributes of today's virtual realities that are rapidly streaming towards globalitarianism. The research-based premises have proven to be identical principles of establishing a framework for sustainability and goals of totalitarianism and globalitarianism in Figure 4.

TOTALITARISM	GLOBALITARISM
LEADER / initiator	INTERNET / initiator
PARTY / nebulizer	SIMULAKRUM / nebulizer
TOTALNARIAN STATE – MASS followers, supporters	"GLOBAL VILLAGE" – MASS followers, supporters
PROPAGANDA	PERSUASION

IDENTICAL PRINCIPALS OF ESTABLISHED FRAMEWORK

FASCINATION – as a social phenomenon
FANATISM – overly fascinated socjety
IDEALISM – experiences of created community
SOCIAL MOVEMENT FRAMEWORK – followers, identification
CONFORMITY – modern opinion chaos

IDENTICAL RESILIENCE FACTORS

- lack of critical thought
- unyielding persistence
- videospread convictions
- true followers unselfishness
- exceeding every real experience
- cancelling every immediate personal interest
- increasing apologetic – nostalgic approach
- increasing psychological – nostalgic attitude
- modern opinion chaos
- general condemning of moral norms
- crown mentality

IDENTICAL GOALS

- **SUPERFICIAL DEPOLITIZATION AND NEUTRALIZATION OF POLITICS**
- **LATENT STRUCTURE OF MASS CONSCIOUSNESS MANIPULATION**
- **MODERN IDEOLOGY OF MASS SOCIETY WITH TARGETED ATTENTION**
- **SOCIETY AS A TOTAL MONITORED GROUP FOUNDED ON WORK AND TECHNICS**

Figure 4: Comparison of totalitarianism and globalitarianism considering establishment premise, sustainability and goals

Instead of asking a question whether the ideology of terror, control, totalitarianism, etc., has a chance in the future, the successful search has proved that the "man of the mass" has, in fact, a problem of losing the social intelligence and functional connectivity due to technological integration. In the example of the ontological creation of the concept of persons, personalities, and events, the comparative studies of Buyse and Larsen, 1979 and Dunbar and Spores, 1995, confirm the structure of a man as a social being created through contact at the expense of social intelligence. Described examples of totalitarianism confirm that policies were consciously developed by a steadily increasing number of sympathizers with the controlled number of movement

members (parties). Totalitarianism's theoretical background modeling *today's digital social connectivity* (3SC), (hrv. *današnja društveno digitalna sprega* -3DS) has confirmed these premises.

4. Conclusion

The real function of the totalitarian matrix is the ingeniously finding members' illusion towards sympathizers. Actually, the movement has carefully recruited its members by establishing an internal gradual hierarchy in order to avoid indoctrination. The outside world viewed as "normal", considers sympathizers as innocent fellow citizens in the non- totalitarian society. In that formation an anonymity greatly contributes to the unusual phenomenon of the mass easily manipulated, because of being too lazy or not brave enough, and always under the pressure of constant "moving", because it is known that the movement has no plan, but as the word itself says, "must move".

The problem is multiplied by the fact that today's digital social connectivity became imperative in creating curricula, plans, and strategies of modern education. There are very few studies which could differentiate the effectiveness of this novelty, and even less professional and scientific reviews of verified problems [34, 35].

On the contrary, today's digital social connectivity has negative effects on the younger population whom it deprives of very much needed mental creativity [36, 37]. It has been emphasized in the introduction that the temporality of history has been slowed down by indifference and astonishment, which are some of the selected parameters studied by cognitive cybernetics, in revealing the man of the mass globalitarianism [38, 39]. This problem repeatedly occurs in philosophical reflections and futuristic predictions [40]. Statistical data collected for the Big Data model through historical and current reality as results of the research conducted with the help of cognitive cybernetics about the "mass" factor from the theory of totalitarianism are the subject of processing for the next scientific paper. Using search correlation, models and algorithms have confirmed a suspicion that intelligent interactive technologies impact the changes of the human psychophysical structure, through the digital social network of globalitarianism.

Conflict of Interest and Funding

No conflict of interest was reported by the authors. This work was supported by the AEROING d.o.o. Zagreb, Croatia, Europe and ELOS d.o.o. Split, Croatia, Europe as the partners and co-founders of the Institute for Cognitive Cybernetics Eastern Europe, Zagreb, Croatia.

Thanks

Special thanks to Mrs. Marijana Agičić (Saamis Immigration Services Association, 659-3rd Street, Medicine Hat, Alberta-Canada) who carefully read this paper and contributed to her advice and editing in English.

References

[1] U.V Detlev, Psychologie und Welt, W. Kohlhammer, 1972.
[2] M. Cipra, Temelji ontologije, Tehnički centar NZMH, 2003.
[3] C. Duhigg, Moć navike Zašto činimo što činimo i kako se promijeniti, Mozaik knjiga, 2013.
[4] G. Figal, Der Sinn des Verstehens, Beiträge zur hermeneutischen Philosophie, Philipp Reclam, 1996.
[5] Ž. Paić, Totalitarizam?, Meandarmedia, 2015.
[6] H. Arendt, Izvori totalitarizma, Kikagraf, 2015.
[7] Z. Balaž, K. Meštrović, Politehnička kognitivna kibernetika, Manualia Polytechnici Studiorum Zagrabiensis, 2018.
[8] Z. Balaž, M. M. Ninčević, Kognitivna kibernetika, Računska racionalnost misaonih sustava, GK, 2019.
[9] M. Heidegger, Ontology (Hermeneutics and Factuality), Bröcker-Oltmanns, 1988.
[10] H. G. Gadamer, The Legacy of Europe, Parnas, Press, Targa, 1997.
[11] Ž. Paić, Sfere egzistencije: Tri studje o Kierkegaardu, Matica Hrvatska, 2017.
[12] Z. Balaž, "Integration of Cognitive Cybernetics into Intelligent Human Systems?", 1st International Conference on Intelligent Human System Integration, 2018.
[13] Z. Balaž, "Brain - A Flexible Self-Adaptive Entity", Public lecture and workshop conducted on 13th Brain week in Croatia, at Electrical department of Zagreb University of Applied Sciences, 2014.
[14] Z. Balaž, "Brain - Communication Through Rhetoric and Artificial Intelligence", Public lecture and workshop conducted on 14th Brain week in Croatia, at Electrical department of Zagreb University of Applied Sciences, 2015.
[15] Z. Balaž, Brain – "Cooperation or Bidding with Intelligent Systems", Public lecture and workshop conducted on 15th Brain week in Croatia, at Electrical department of Zagreb University of Applied Sciences, 2016.
[16] Z. Balaž, "The Brain lost in the space of persuasive no-placelessness", Public lecture and workshop conducted on 16th Brain week in Croatia, at Electrical department of Zagreb University of Applied Sciences, 2017.
[17] Z. Balaž, "Brain – Part of cognitive cibernetics against artificial intelligence", Public lecture and workshop conducted on 17th Brain week in Croatia, at Electrical department of Zagreb University of Applied Sciences, 2018.
[18] Z. Balaž, "Brain – Cognitive cibernetics against digital demention", Public lecture and workshop conducted on 18th Brain week in Croatia, at Electrical department of Zagreb University of Applied Sciences, 2019.
[19] M. Heidegger, Kraj filozofije i zadaća mišljenja, Rasprave i članci, Filozofska biblioteka, 1996.
[20] Z. Balaž, "Cognitive Cybernetics - a Future Society 5.0", International Symposium P&P, 2017.
[21] Z. Balaž, M. Haun, "Cognitive Cybernetics – A Future that is Started", 34th International Symposium on New Technologies, 2017.
[22] Z. Balaž, B.-I. Balaž, Human Computer Interaction, Cognitive Cybernetic & Captological Education, Whioce Publishing Pte. Ltd., 2018.
[23] Z. Balaž, S. Desin, "Kognitivna kibernetika i globaritarizam", 38 Munarodni elektroinženjerski simpozij, EIS 2019- Dani Josipa Lončara, 2019.
[24] W. Glasser, Choice theory: a new psychology of personal freedom, Alinea, 2000.
[25] K. Wilber, A Theory of Everything, Gorin, 2004.
[26] E. Fink, Einleitung in die Philosophie, Franz-A. Schwarz, 1985.
[27] Ž. Paić, Sloboda bez moći, Politika u mreži entropije, Udruga Bijelo val, 2013.
[28] Ž. Paić, Traume razlika, Biblioteka Intermedia, Meandarmedia, 2007.
[29] Ž. Paić, Politika identiteta, Kultura kao nova ideologija, Antibarbarus, 2005.
[30] E. K. Wilber, Brief History of Everything, Gorin, 2005.
[31] Z. Balaž, K. Meštrović, Inteligentni i ekspertni sustavi u elektroenergetici, Tehničko veleučilište u Zagrebu, Elektrotehnički odjel, Manualia Polytechnici Studiorum Zagrabiensis, 2015.
[32] R. Larsen, D. Buss, Personality psychology, Denis Bratko Naklada Slap, 2008.
[33] H. M. McLuhan, Understending Media - Media as Man Extensions, Golden Marketing, 2008.
[34] M. Spitzer, Digitalna demencija - Kako mi i naša djeca silazimo s uma, Ljevak d.o.o., 2018.
[35] T. Karsenti, S. Collin, Why are New Teachers Leaving the Profession? Results of a Canada-Wide Survey, Education Faculty, University of Quebec and Montreal, 2013.
[36] N. Carr, Plitko - Što Internet čini našem mozgu, Jesenski i Turk, 2011.
[37] N. Carr, Stakleni kavez, Jesenski i Turk, 2016.
[38] J. Baudrillard, Le Pacte de Lucidité ou L'Intelligence du Mal, Ljevak d.o.o. 2006,
[39] J. Baudrillard, Simulakrumi i simulacije - Biblioteka Psefizma, 2001.
[40] Ž. Paić, Posthumano stanje: Kraj čovjeka i mogućnost druge povijesti, Litteris, 2011.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 724-730 (2020)

www.astesj.com

Second International Virtual Conference on Multidisciplinary Research 2020

ASTESJ

ISSN: 2415-6698

Ultra Wide Band-based Control of Emulated Autonomous Vehicles for Collision Avoidance in a Four-Way Intersection

Jashandeep Bhuller[1], Paolo Dela Peña[1], Vladimir Christian Ocampo II[1], Julio Simeon[1], Lawrence Materum[1,2,*]

[1]*De La Salle University, 2401 Taft Ave., Malate, Manila, 0922, Philippines*

[2]*Tokyo City University, 1-28-1 Tamazutsumi, Setagaya, Tokyo, 158-8557, Japan*

ARTICLE INFO

Article history:
Received: 11 September, 2020
Accepted: 07 October, 2020
Online: 24 November, 2020

Keywords:
Autonomous Systems
Ultra Wide Band
Self-driving Vehicles

ABSTRACT

Traditional crossroads use traffic lights, but may cause delays due to its stop and wait process. Moreover, autonomous vehicles are being made, although it is not yet mass-produced. This work presents a design and implementation of an agile system that communicates and coordinates autonomous vehicles entering the intersection by adjusting their speeds and maneuvers to avoid collisions, allowing them to weave and pass through intersection traffic and avoiding them to completely stop, leading to an increase in the throughput of the intersection, and therefore reducing congestion. A centralized computer implements agile communication and controls the system based on a first-come-first-served policy. The hardware implementation emulates four prototype vehicles modeled as radio-controlled cars integrated with an ultra wide band (UWB) localization technology. The cars were used to simulate the passage without collision in a typical four-way intersection. Pozyx UWB modules were chosen for tracking the vehicles due to its compatibility with Arduino as well as its numerous other functionalities and accuracy relative to other indoor-positioning systems. Results show that Pozyx could track autonomous vehicles. The study resulted in a system that can control self-driving cars through the intersection with key measures. Moreover, the characterization of Pozyx in this hardware implementation was observed.

1. Introduction

Traffic problems have risen with the increase in vehicles, users, roads, and crossroads. Numerous traffic delays and accidents occur in various locations, and one of the most concerning of these is crossroads. Smith [1] said that almost forty percent (40%) of all crashes in the United States involves crossroads, and about 165,000 accidents occur annually in crossroads caused by red-light runners. Crossroads are designed to lessen accidents, traffic congestion, and increase efficiency, but these become a source of problems. Erring drivers, vague navigation signs, and blind curves are some of the reasons for these. Although the installation of stoplights and various signs at crossroads are seen as solutions to these problems, these devices have not been as effective. These devices are not so good for traffic calming, and this meant that these may make the flow of vehicles better but not necessarily safer for drivers [2]. Accidents increased when traffic signals were introduced, and these devices are not the solutions to less dangerous roads [3]. Proposals to solve these problems include the introduction of autonomous vehicles (AV) and intelligent intersection management, but there have been very few implementations in actual situations. Communication between AVs and road infrastructure is also important to ensure road safety [4]. Moreover, AVs can be enhanced to intelligent vehicles by adding high technology features like efficient 3D road map exchange in real-time for navigation and localization [5]. The study plans to address the problems of accidents and traffic congestion at crossroads using prototype AVs with attached Pozyx whose speeds and turns are controlled by a centralized control communications system. Pozyx [6]–[10] modules are ultra-wide band-based devices for localization. In this paper, vehicles are referred to as autonomous, which is also known as self-driving cars. The use of Pozyx was

*Corresponding Author: Lawrence Materum, 2401 Taft Ave., Manila, Philippines, +6325244611, lawrence.materum@dlsu.edu.ph

also characterized in terms of accuracy and delay for tracking and localization of AVs.

In crossroads, vehicles must coordinate their movements to pass through without collisions. In an ideal situation, vehicles would only have to accurately adjust their speed and direction to create a sufficient duration for the intersecting vehicle to weave through without stopping. Furthermore, the vehicles should be able to process and execute these maneuvers with high precision and minimal delay to avoid collisions. Unfortunately, the current system in crossroads is traffic light-based where vehicles have to stop and wait for intersecting vehicles to pass. Vehicles are needed to stay idle when their road is asked to stop, interrupting their journey. At present, almost all vehicles are operated by humans, but in the future, the proliferation of AVs can be seen, and this is the context of this study. The focus of this paper is on the use of Pozyx for communication and localization of autonomous vehicles through a central computer system so that they can pass in crossroads without collision.

This research on autonomous intersection management helps address traffic-related problems and congestion at crossroads. The modified and programmed autonomous vehicles would be able to adjust their speeds at crossroads without stoplights resulting in a more efficient flow of traffic and a decrease in traffic congestion and accidents. The software algorithm in the centralized system determines the speed of each car passing the crossroad and calculates the appropriate angle of turn for each of these vehicles so that collisions would be avoided. The centralized system communicates with each of the vehicles to allow the flow of traffic at crossroads to be orderly and continuous, and thus, would save time. With the increase in research about autonomous and intelligent vehicles, this study is an add-on with its contributions such as agile control software for collision avoidance of AVs using a first-come-first-served (FCFS) tile-based approach with graphical user interface (GUI), and its hardware adaptation of collision avoidance in small-scaled miniature AVs with attached Pozyx for tracking and localization.

It is assumed that connection quality between the vehicles and the central computer is ideal in this study. The crossroad is also assumed to be initially free of obstacles such as foreign objects and road hazards. Four scaled-down controllable vehicles running simultaneously in a four-way two-lane crossroad utilizing a clover-shaped track free of obstacles were included in the scope of the hardware implementation. Moreover, the scope of the study includes a control system that monitors the velocity of the vehicles and performance measures of the system through its vehicle throughput. The study is delimited by the following: human-driven vehicles are not factored in, the vehicles do not abide by the SAE Level 5 definition of driving automation [11], and the cars cannot recognize foreign obstacles which are any object that is either not an element of the track or another object not running as part of the system.

The system has two main parts: the software algorithm and hardware control system. Its overview is illustrated through a block diagram shown in Figure 1. The system is described as agile since it is meant to quickly react and adapt to vehicles entering the crossroad and managing it according to the other vehicles. Moreover, it is then described to be collision restrained as it would avoid letting cars collide in the crossroad. This means that if there are multiple vehicles with current trajectories expected to collide, the algorithm adjusts their velocities to avoid this potential outcome. The software runs the crossroad algorithm and the communications interface. The control communications system deals with the controllers of the scaled-down motorized cars which are referred to as 'vehicles.' The microcontrollers would have their own wireless communications module that uses either Bluetooth or Wi-Fi. Moreover, the tracking and location features were added through Pozyx. This control communications system is responsible for taking in data from the different sensors of the vehicle and relaying the information to the agile control communications. They also control the motors for the instructions sent back. The actual hardware of the project is focused on the vehicles and their sensors.

The system is tested to measure its vehicle throughput. In each trial, the authors adjusted the number of vehicles active and running on the track to change the load on the crossroad. Varying the load on the crossroad would account for real-life situations such as rush hours. The recorded delay is the time it takes from sending a reservation request to exiting the crossroad. They then adjusted the average speed of the cars and then further recording the delay of the cars, as well as the chances of collisions. Using the data gathered from the trials, linear scaling was performed to see how the data would project to real-world cars and crossroads.

This paper is organized then as follows. Section 2 gives further details of the design of the overall system, and its corresponding components. Section 3 provides the results of the design, and lastly, Section 4 concludes the paper with recommendations for future work.

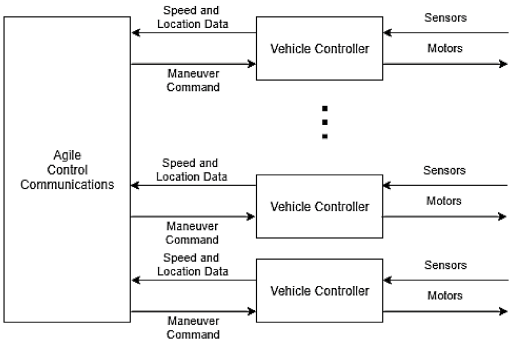


Figure 1: Block diagram of the control system.

2. System Design and Methodology

2.1. Pozyx

Pozyx is a UWB-based real-time location system utilizing both hardware and software to be able to provide accurate positioning information with an advertised accuracy of up to ten (10) centimeters compared to other traditional indoor location tracking systems, such as Wi-Fi or Bluetooth which are accurate within several meters [6]–[9], [12]. This localization is achieved through the use of UWB technology combined with smart filters as well as machine learning, thus, allowing positioning to be done even in indoor environments without a clear line of sight to the reference points. To obtain precise and accurate positioning, two-way ranging is done between the Pozyx tag and at least three

reference points, or Pozyx anchors, followed by trilateration. These anchors are ideally spread out as much as possible and placed at the corners of the space in which location tracking will be performed to get the most reliable data. The anchor may be reprogrammed to work as a tag as well and does not need an Arduino to function as it is designed to operate on its own.

The Pozyx tag is an Arduino compatible shield which makes use of Pozyx anchors to provide accurate positioning as well as other types of data such as its orientation using its 9-axis inertial measurement unit. It is also equipped with different sensors along with a UWB transceiver based on the DW1000 UWB-chip from Decawave which, in addition to performing ranging, can also be used for wireless messaging. Figure 2 displays the pinout of a Pozyx tag, whereas Figure 3 for the anchor. Pozyz interfaces with Arduino through I2C communication. It can also interface with Raspberry Pi through a micro-Universal Serial Bus (micro-USB).

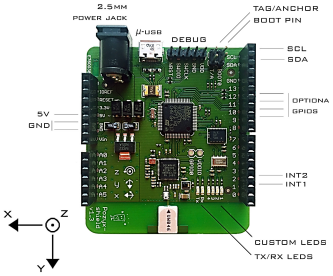


Figure 2: Pozyx Tag Pinout [6]

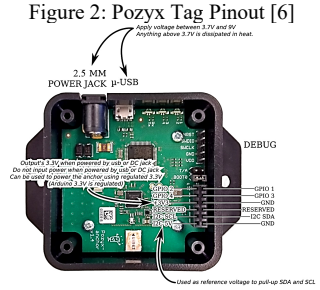


Figure 3: Pozyx Anchor Pinout [6]

For this study, Pozyx was chosen to be the basis for tracking the RC Cars due to its compatibility with Arduino as well as its numerous other functionalities and accuracy relative to other indoor-positioning systems. In addition to being easy to set up for most indoor environments, it also provides different possible configurations to optimize its performance to meet the desired need whether it be accuracy or speed. Figure 4 provides visualization on how the location of a Pozyx tag is obtained about the anchors as well as to the Master Pozyx tag. It may be seen that each Pozyx tag does not automatically provide its location as it only does so when it is asked first. Once asked, it will then calculate its location using the four Pozyx anchors as a reference at which point, the tag then finally sends its position data back to the Pozyx master.

2.2. Arduino Uno

The Arduino Uno acts as the brain of the RC car as it processes commands received through Bluetooth as well as utilizes the connected Pozyx tag to perform the necessary maneuvers to ensure that it follows its intended path.

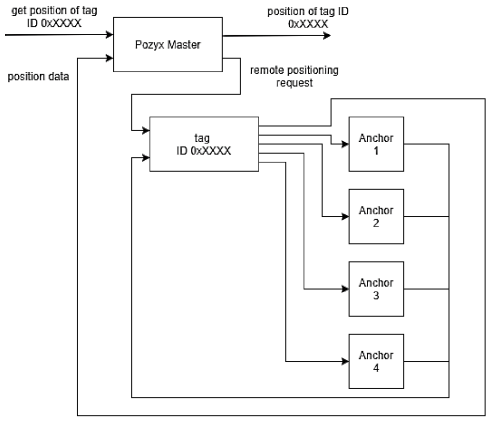


Figure 4: Pozyx Block Diagram

2.3. Radio-Controlled Car

A Radio-Controlled (RC) car is a battery-powered model car which can be controlled by remote control, mostly done through radio, hence, the term ”radio-controlled”. These RC cars represent a scaled-down model of real cars. A typical RC car has a receiver that captures radio signals from its remote control. To simulate the algorithm and adapt it into hardware, RC cars are used. In this research, the cars were modified by adding different modules to their parts. A location module is added to track the car’s position, and the RC communication is replaced with a sophisticated communication module. A microcontroller is added to manage the modules and control the car. The movement of the RC car is driven by motors. A motor driver is used to interface the microcontroller and the motors providing motion to the RC car. The motor driver in the system is powered using a power bank separate from the microcontroller circuit. The motor driver chosen for RC cars was the 2-Channel Tiny DC Motor Driver from e-Gizmo based on Allegro’s A3906 low voltage dual DC motor driver IC.

2.4. FCFS Software

The tile-based first-come-first-served (FCFS) policy, developed by Dresner and Stone [13], is the basic reservation system. ACC applies the FCFS approach for the reservations of vehicles, meaning that vehicles that requested a reservation will be entertained first. When another vehicle requests for a reservation, it will be checked first from vehicles that had made successful reservations earlier. A centralized computer provides coordination by communicating with AVs and managing their reservations. When a vehicle sends a request, the ACC simulates as if the vehicle will pass through the intersection. The reservation is tile-based, in which collisions are detected if a vehicle occupies a tile reserved by another vehicle. If at least two vehicles occupy the same tile, which would result in a collision, the request is rejected. Otherwise, it is accepted. The significance of adopting the FCFS is its versatility because it can be extended to be used in many different purposes in autonomous intersection management.

2.5. Car Simulation

The car runs on a program independent of the ACC. For reservations to commence, the car software and the ACC, which employs the FCFS algorithm, must exchange information. The

ACC follows the principle of Autonomous Intersection Management [14]. The simulation of the cars would maintain minimal communication complexity. After the reservation, it is the responsibility of the car to meet the reservation at the correct time and place.

2.6. Pozyx in RC Car

In modifying the RC car, an Arduino was added. The Arduino was interfaced to a Bluetooth Module through Universal Synchronous-Asynchronous Receiver-Transmitter (USART), and two motor drivers that power the motors. Two 5 V, 3600 mAh power banks, energize the Arduino and the motors. Moreover, a Pozyx tag is attached to the Arduino, which communicates with it through I2C. It does the job of accurately locating and tracking the cars through UWB. The Pozyx localization system employs four reference anchor modules, with one of the anchors interfaced with an Arduino that is connected via a serial interface to a computer that runs the simulation software. The Pozyx tag relays the location of the car where it was attached. It also orients the car at an angle from 0◦ to 360◦. The Bluetooth BLE HM-10 module receives commands which the Arduino module then decodes. The decoded instructions determine which of the two motor drivers activates the corresponding motor driver depending on the received instruction. The two motor drivers present in each RC car are responsible for controlling both its steering and acceleration depending on the instruction decoded by the Arduino with one motor driver handling the steering motor and the other handling the throttle motor. The components are interconnected as shown in the block and circuit diagrams of the RC car illustrated in Figure 5 and 6.

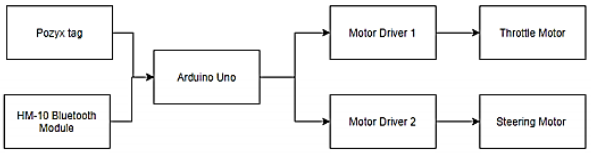


Figure 5: RC Car Block Diagram

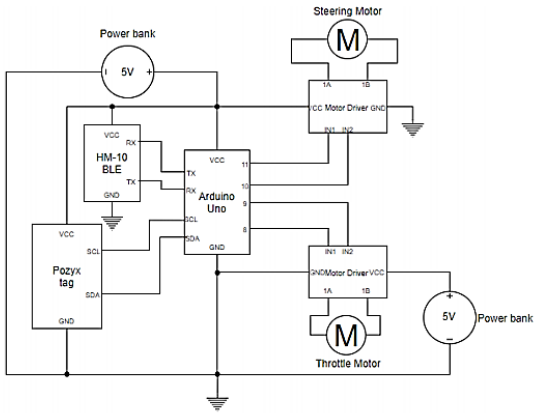


Figure 6: RC Car Circuit Diagram

2.7. Communication Protocol

The communication protocol is the medium at which the software running on the computer can interface with the Pozyx modules as well as issuing commands to the AVs. The study has opted to make a standard protocol that handles the execution of functions as well as sending data to and from the central computer software and the hardware controllers. Although there are several uses for the protocol their format is similar regardless of whether they are between the data collection of the Pozyx or the command issuing for the cars. Between the computer program and the connected microcontrollers that handle Pozyx and Car communication, the format of their messages are as follows: preamble, length, message type, and data. The preamble is a predetermined set of bytes used to indicate to the receiver that a message is incoming. Next is the length byte. It is a single byte used to indicate the length of the entire message. This can vary depending on the number of bytes used for sending data if applicable. Additional details can be found at [6].

2.8. Scaling Method

The method used to establish the scaling between the RC car and an actual vehicle is based on the turning radius of the RC car instead of its dimensions. This is because it was observed that its steering angle is much larger than that of life-sized cars which resulted in the RC car taking wider turns than what was expected based on its dimensions. As a result, the dimensions of the test track are based on the turning radius of the RC cars, the resulting dimensions of the car, as well as its speed in simulation, would be proportional to the ratio between its turning radius and the average turning radius of most common midsized passenger cars today of 17 feet or 5.18 meters [15]. After the computation based on Equation 1, a scale factor of 15.94 was obtained. The scaling process is illustrated in the block diagram in Figure 7.

$$\text{scale factor} = \frac{\text{actual turning radius}}{\text{RC car turning radius}} \quad (1)$$

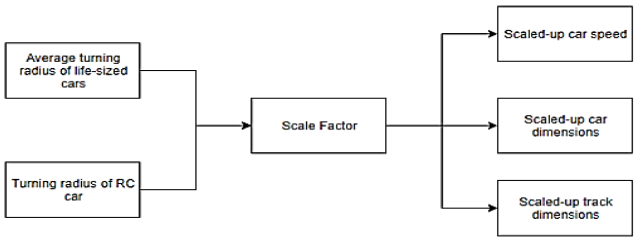


Figure 7: Scaling Process Block Diagram

2.9. Materials

The materials used are summarized in Table 1.

Table 1: Materials

Component	Quantity
Arduino UNO	4
Pozyx developer tag	5
Pozyx creator anchor	4
3600 mAh power bank	8
2-Channel Tiny DC Motor Driver	8
RC Car	4
HM-10 BLE Module	8

2.10. System Block Diagram

The entire system block diagram is shown in Figure 8. The inputs of the system are user interface (UI) inputs, track dimensions, and car initial states. The center shows the various processes that run in the software which process all the data from the inputs and produce the corresponding outputs. The main outputs which correspond to the key measures of the project are UI output, car motion, number of collisions, and duration in the crossroad.

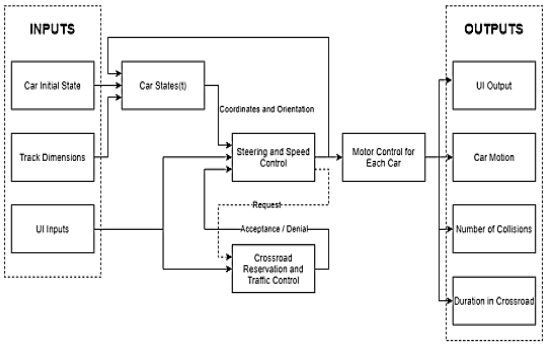


Figure 8: System Block Diagram

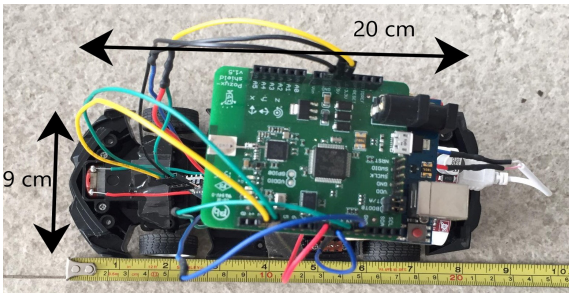


Figure 9: Dimensions of Scaled-Down Model Car

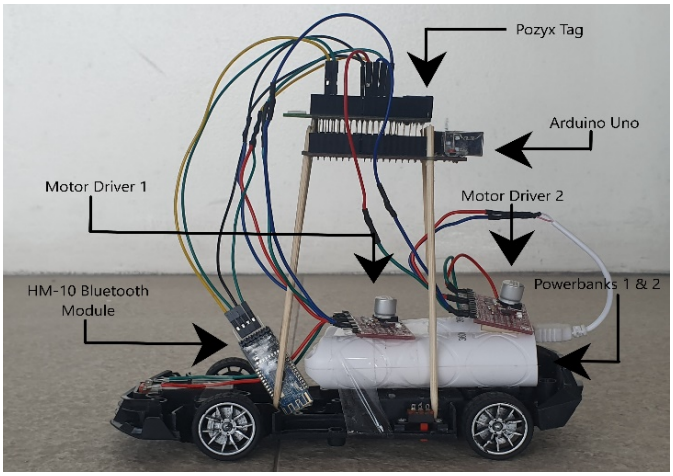


Figure 10: Scaled-Down Model Car Side View

3. Results, Analysis, and Discussions

3.1. Prototype Car

The resulting scaled-down model as shown in Figure 9 is twenty (20) cm in length and five cm wide, as it small enough to not require a large test track while big enough to mount all the components needed. The car was able to fit all the components in its shell but would likely negatively impact the Pozyx tag's connection to the nearby anchors due to the shell preventing a good line of sight. Discovering that Pozyx had more accurate results if it was elevated higher from the ground, long thin sticks were used to prop it up. The car goes through different types of movements and vibrations during testing, and thus, the components and their connections to each other must be secured. Therefore, connecting wires to motor drivers must be directly soldered, as seen in Figure 10. The motor drivers and power banks were also attached. The resulting characteristics of the four prototype cars are shown in Table 2. With the same throttle level, the four cars have different speeds. This variation in average speeds is less than ideal when compared to the perfect scenario of having all four cars perform the same at each throttle level. It was also observed that the measured average car speeds on later trials were generally slower compared to the earlier trials. This indicated that their speeds decrease as they are used due to the car's power banks being discharged.

Table 2: Car Speed Characteristics

Throttle Level	Average Speed (m/s)			
	Car 1	Car 2	Car 3	Car 4
100	1.456	1.318	1.281	1.486
90	1.394	1.235	1.218	1.398
75	1.09	0.942	0.919	1.222
60	0.878	0.57	0.722	0.985
50	0.618	0	0.294	0.666

The four prototype cars running simultaneously in their physical track with specified dimensions are shown in Figure 11.

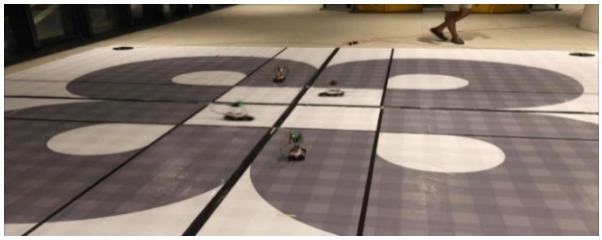


Figure 11: Test track

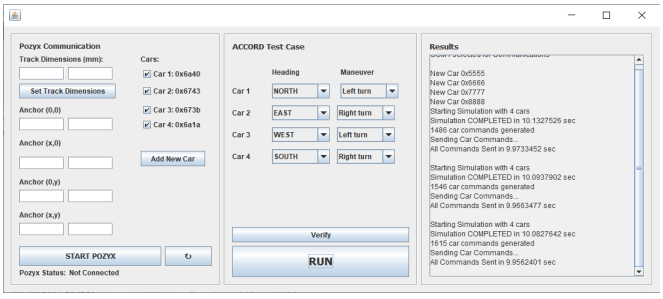


Figure 12: Graphical User Interface

3.2. User Interface for the Control

The graphical user interface (GUI) of the system provides the users with an easy way to set up and initialize the hardware and

setting up the parameters for each test case. The latter function specifically assigns the initial locations of the car as well as injects them with the turn maneuvers that they will execute later during run time. As shown in Figure 12, the GUI window is divided into three sections, Pozyx Communication, Test Case, and results.

3.3. Control Capability Results with Pozyx

Pozyx is used to locate and track cars in real-time. It gives position and orientation to the system which helps it determine the speed and steering required by each car in the track. One of the most important functionalities of the Pozyx is providing the software program with real-time accurate position and orientation for the program that determines the speed and steering for each car in the track. However, the Pozyx updates are too slow in comparison to the speed of the cars reducing the accuracy of tracking the cars as they cross the intersection. The issues encountered with the Pozyx are about location accuracy and delay in updates. The company that sold the Pozyx modules listed a location accuracy of 5 to 10 centimeters while the update rate for each tag was 16 milliseconds. However, the data obtained in the study did not match these characteristics, leading to numerous problems in the performance of the system and the run of each car in the track. The authors performed tests using one Pozyx tag at a time followed by multiple tags to determine the accuracy of Pozyx and the consistency of its performance. The important factors that were observed were the accuracy and delay in the results provided by Pozyx. For each test, the Pozyx tags were placed in a location within the Pozyx anchors. The location provided by Pozyx was compared to the actual location to evaluate the accuracy of Pozyx. Fifteen tests were performed for one Pozyx tag at a time, while one test was performed that used four Pozyx tags simultaneously. The authors were able to note some issues regarding the Pozyx accuracy by testing the Pozyx tags. In the tests done for one tag at a time, the error for accuracy showed a range of 10 to 40 cm, and there were some cases where the delay reached more than a second between updates. There were also cases when no data location was shown for an update, and some of the location data freeze between updates. It was observed that there was a different location for the tags in each test. These location points were chosen based on critical points that tested the functionality of Pozyx in terms of accuracy in data location. Moreover, the test was done on four tags and the delay was more than 0.2 seconds and the error is from 20 to 50 cm for some tags. The graphs of the differences and deviations are shown in Figure 13 and Figure 14, respectively. In the system, accuracy and delay are very important factors in determining the speed and steering for each car. Hence, when the system receives location data that is not accurate, there will be many problems in the performance of the system and movement of the cars. In addition to receiving larger location inaccuracies than was expected, there were also numerous instances of outputting negative location values implying that the tag being located was out of bounds when it was not in reality. This can be seen in Figure 15 wherein Pozyx outputs Y coordinates that are nearly half a meter off in some instances from its true value of 300. Frequent positioning errors as shown in Figure 16 can also have a severe impact on the system in terms of adding to delay as it takes more than a second for Pozyx to recover from the aforementioned error and report a more accurate value. This is further exacerbated when four Pozyx tags are being located simultaneously as the already large average delay between updates of 200 milliseconds is made worse due to even larger delays between updates.

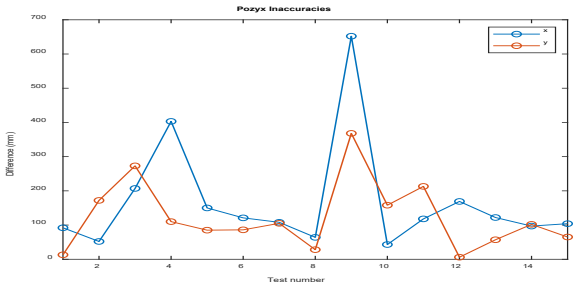


Figure 13: Pozyx Inaccuracies by Differences

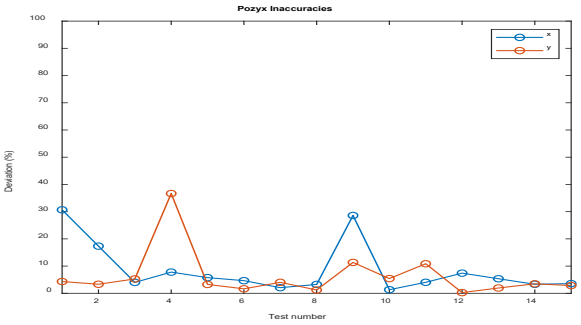


Figure 14: Pozyx Inaccuracies by Deviation

```
Time: 5474 POS ID 0x6743, x(mm): 5234, y(mm): -246, z(mm): 0
Time: 5523 POS ID 0x6743, x(mm): 5207, y(mm): -227, z(mm): 0
Time: 5573 POS ID 0x6743, x(mm): 5172, y(mm): -213, z(mm): 0
Time: 5623 POS ID 0x6743, x(mm): 5151, y(mm): -204, z(mm): 0
Time: 5671 POS ID 0x6743, x(mm): 5111, y(mm): -157, z(mm): 0
Time: 5722 POS ID 0x6743, x(mm): 5011, y(mm): -35, z(mm): 0
Time: 5770 POS ID 0x6743, x(mm): 4916, y(mm): 73, z(mm): 0
Time: 5820 POS ID 0x6743, x(mm): 4964, y(mm): 13, z(mm): 0
```

Figure 15: Pozyx Negative Values

```
Time: 5193 POS ID 0x6743, x(mm): 2525, y(mm): 2492, z(mm): 0
ERROR positioning on ID 0x6743, error code: 0x0
Time: 6273 POS ID 0x6743, x(mm): 2535, y(mm): 2498, z(mm): 0
ERROR positioning on ID 0x6743, error code: 0x0
Time: 7351 POS ID 0x6743, x(mm): 2556, y(mm): 2505, z(mm): 0
```

Figure 16: Pozyx Delay

4. Conclusions and Recommendations

This work involves software design and hardware adaptation of collision avoidance of autonomous vehicles simultaneously crossing an intersection. Hardware implementation was done by using small-scaled RC cars with attached Pozyx where test cases were done. The accuracy of the Pozyx was tested and verified. Upon running the various test cases to verify the system with different car placements and maneuvers, it was observed that the Pozyx module had issues in location accuracy and delay between updates that impede in determining the accurate location of each car in the track.

Recommendations include incorporation of machine learning in AV and autonomous intersection management, implementation of software simulation in different programming languages, an

extension of hardware implementation on large-scale vehicles, and to some extent, real vehicles and crossroads. Moreover, it is suggested to use uninterrupted, fast, and accessible communication technologies. Communication among AVs with the central computer system can be extended by using accurate and efficient road map data exchange and cloud databases. Expanding the model of the car's movement from linear acceleration to a more dynamic model would greatly improve the accuracy of the resultant predetermined coordinates, and consequently, more appropriate car commands. Besides, more sensors that can keep track of attributes not necessarily related to motion would be a useful feature.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

De La Salle University is acknowledged for supporting this work.

References

[1] E. A. Smith, Statistics on Intersection Accidents. 2018. https://www.autoaccident.com/statistics-on-intersection-accidents.html

[2] R. Ewing, Traffic Calming State of the Practice. 1999. https://nacto.org/wp-content/uploads/2012/06/Ewing-Reid-1999.pdf

[3] N. Kazis, To Get Safer Streets, Traffic Lights and Stop Signs Aren't the Answer. 2011. https://nyc.streetsblog.org/2011/04/26/to-get-safer-streets-traffic-lights-and-stop-signs-arent-the-answer

[4] K. F. Chu, E. R. Magsino, I. W.-H. Ho, and C.-K. Chau, "Index Coding of Point Cloud-Based Road Map Data for Autonomous Driving," in 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), 2017, 1–7, 2017. doi: 10.1109/VTCSpring.2017.8108280.

[5] I. W.-H. Ho, S. C.-K. Chau, E. R. Magsino, and K. Jia, "Efficient 3D Road Map Data Exchange for Intelligent Vehicles in Vehicular Fog Networks," IEEE Transactions on Vehicular Technology, **69**(3), 3151–3165, 2020, doi: 10.1109/TVT.2019.2963346.

[6] "Home | Pozyx." https://www.pozyx.io/. 2020.

[7] J. Pereira, H. Gomes, S. Mendes, R. Santos, S. Faria, and C. Neves, "Long range RFID indoor positioning system with passive tags," in Industry 4.0–Shaping the Future of The Digital World: Proceedings of the 2nd International Conference on Sustainable Smart Manufacturing (S2M 2019), 9–11 April 2019, Manchester, 112, 2020. doi: 10.1201/9780367823085-21.

[8] Y.X. Wang and C.-L. Chang, "ROS-base Multi-Sensor Fusion for Accuracy Positioning and SLAM System," in 2020 International Symposium on Community-centric Systems (CcS), 1–6, 2020. doi: 10.1109/CcS49175.2020.9231442.

[9] V. Di Pietra, N. Grasso, M. Piras, and P. Dabove, "Characterization of a Mobile Mapping System for Seamless Navigation," The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, **43**, 227–234, 2020. doi: 10.5194/isprs-archives-XLIII-B1-2020-227-2020.

[10] C. Siebert, P. R. DeStefano, and R. Widenhorn, "Teaching physics with a local positioning system," The Physics Teacher, **57**(6), 428–429, 2019. doi: 10.1119/1.5124294.

[11] SAE International and S. O.-R. A. V. S. COMMITTEE, "Surface Vehicle Information Report: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems," Jan. 2014.

[12] K.-M. Mimoune, I. Ahriz, and J. Guillory, "Evaluation and improvement of localization algorithms based on uwb pozyx system," in 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 1–5, 2019. doi: 10.23919/SOFTCOM.2019.8903742.

[13] K. Dresner and P. Stone, "A Multiagent Approach to Autonomous Intersection Management," Journal of Artificial Intelligence Research, **31**, 591–656, 2008.

[14] T.-C. Au, S. Zhang, and P. Stone, "Autonomous Intersection Management for Semi-Autonomous Vehicles," Handbook of Transportation, 88–104, 2015.

[15] Vehicle Turning Paths Dimensions and Drawings. Dimensions.Guide https://www.dimensions.guide/collection/vehicle-turning-paths-radius

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 731-737 (2020)

www.astesj.com

Second International Virtual Conference on Multidisciplinary Research 2020

ASTESJ

ISSN: 2415-6698

Minimizing Collisions of Self-Driving Cars by a Control System Using Predetermined Two-Dimensional Grid Localization

Jashandeep Bhuller[1], Paolo Dela Peña[1], Vladimir Christian Ocampo II[1], Julio Simeon[1], Lawrence Materum[1,2,*]

[1]*De La Salle University, 2401 Taft Ave., Malate, Manila, 0922, Philippines*

[2]*Tokyo City University, 1-28-1 Tamazutsumi, Setagaya, Tokyo, 158-8557, Japan*

ARTICLE INFO

Article history:
Received: 10 September, 2020
Accepted: 07 October, 2020
Online: 24 November, 2020

Keywords :
Self-driving Vehicles
Autonomous Systems
Command and Control

ABSTRACT

Most crossroads use a traffic light to communicate between human-driven vehicles by conveying to the drivers when to stop and when to proceed in an organized fashion. However, the stop and wait nature of traffic lights cause slower mobility, leading to congestion. This paper presents a localization scheme based on two-dimensional (2D) flat geometry adaptation of collision avoidance of autonomous vehicles simultaneously crossing an intersection. Using a first-come-first-served policy with tile retrieval process, the collision reduction performance was evaluated through a unit test. Four self-driving prototype cars were made with the desired functionality of controlling the speed and maneuver. Through the predetermined 2D grid locales, simulations indicate the system that can control the self-driving cars through a crossroad with zero or minimal collisions using key measures such as time duration in the intersection, average speeds of the cars, and collision count.

1. Introduction

An increase in the number of cars, roads, and crossroads over the years brought problems related to traffic flow. Numerous accidents and traffic delays occur in various locations, and one of the most concerning of these is crossroads. Smith [1] claimed that almost 40 percent of all crashes in the United States involve crossroads, and about 165,000 accidents occur annually in crossroads caused by red-light runners. Though crossroads are designed to lessen accidents and traffic congestion, these become sources of problems. Some of the reasons for these are confusing navigation signs, erring drivers, and blind curves. Though the installation of stoplights and various signs at crossroads are perceived as solutions to these problems, these devices have not been effective. These devices are not so suitable for traffic calming, which may make vehicles' flow better but not necessarily safer for drivers [2]. Accidents increased when traffic signals were introduced, and these devices are not the solutions to less dangerous roads [3]. Proposals to solve these problems include introducing autonomous vehicles (AVs) and intelligent intersection management, but there have been very few implementations in actual situations. The authors plan to address the problems of accidents and traffic congestion at crossroads using prototype vehicles whose speeds and turns will be controlled by a centralized control communications system. In this paper, prototype vehicles are referred to as autonomous, that is, self-driving cars. Ho, et al. [4] studied AVs with added high technology features like efficient 3D road map data exchange in real-time for navigation and localization. The study [5] worked on an efficient information dissemination system of 3D point cloud road map data (3DMADS) for intelligent vehicles and roadside infrastructure integrated into a vehicular fog computing architecture, which was validated with empirical mobility traces, 3D Light Detection and Ranging (LIDAR) data, and an experimental multi-robotic testbed. Moreover, communication between AVs and road infrastructure is vital to ensure road safety. A study [5] proposed the Road Map Data Encoding and Dissemination System (REDS), which considers roadside vehicle information through index coding of point cloud-based road map data. REDS was evaluated in a four-way junction scenario and was shown to improve overall data efficiency.

In an ideal situation, an autonomous vehicle's path from its position to its destination would be straightforward, with no interruptions or obstacles. However, current roadways require paths to intersect with other roads, and therefore, vehicles coming

*Corresponding Author: Lawrence Materum, 2401 Taft Ave., Manila, Philippines, +6325244611, lawrence.materum@dlsu.edu.ph

from different parts of the crossroad must coordinate their movements to pass through without collisions. In an ideal situation, vehicles would only have to adjust their speed and direction enough to create sufficient duration for the crossing vehicles to weave through without stopping. The vehicles should be able to process and execute these maneuvers with high precision and minimal delay to avoid collisions. Unfortunately, the present system in place in crossroads is traffic light-based, where vehicles have to stop and wait for crossing vehicles to pass. Vehicles are needed to stay idle when their road is asked to stop, interrupting their journey. Currently, almost all vehicles are operated by humans, but AVs' proliferation can be seen in the future, and this is the context of this paper. One primary objective of this work is to design and implement computer control to avoid a collision in crossroads without traffic lights.

The assumptions in implementing this paper are as follows: the connection quality between the vehicles and the central computer is ideal, and the crossroad is initially free of obstacles such as foreign objects and road hazards. The scope includes an implementation covering a four-way two-lane crossroad, scaled-down controllable vehicles, utilization of a clover-shaped track, at least four cars running simultaneously, and a control system that monitors the velocity of the vehicles, and performance measures of the system through its vehicle throughput. The delimitations are: the paper does not factor in human-driven vehicles, the vehicles do not abide by the SAE Level 5 definition of driving automation [6], and the cars cannot recognize foreign obstacles, which are any object that is either not an element of the track or another object not running as part of the system.

The collision control is primarily based on a control system comprised of the crossroad algorithm and the communications interface. The algorithm is responsible for taking in the vehicles' telemetry data and deciding the correct course of actions in terms of steering and speed control. The paper adopted a first-come-first-served (FCFS) policy with an added tile retrieval process to the system. The communication interface handles the sending and receiving of data and commands between the vehicles. On top of the algorithm and the communications modules, the authors developed a data recovery tool that can be used to calculate useful statistics such as delay, number of cars per second, the car's real-time velocity during testing, and the like. This tool is used to monitor the car and check how the algorithm adjusts its speed. The data gathering tools contribute to measuring the system's performance and were interfaced with the communications module.

This paper is organized then as follows. Section 2 specifies the system and its simulation. Section 3 gives an analysis of the results. Finally, Section 4 concludes the paper.

2. System Design and Simulation

2.1. FCFS

The control system monitors the intersection and the vehicles crossing it. Vehicles communicate with the control system to obtain a reservation and cross the intersection without colliding with other vehicles. It is implemented using a tile-based approach and the FCFS policy. The control system applies the FCFS for vehicles' reservations, meaning that vehicles that requested a reservation will be entertained first. When another vehicle requests a reservation, it will be checked first from vehicles that had made successful reservations earlier. FCFS determines if vehicles will be reserved or not.

Conversely, FCFS results in determining which vehicles will cross the intersection first. Vehicles are entertained based on the order of requests sent; that is, the first vehicle that sends a request is served first, regardless of speed. The tile-based FCFS, developed by Dresner and Stone [7], is the basic reservation system. The significance of adopting the FCFS is its versatility because it can be extended to many different purposes in autonomous intersection management (AIM). The intersection model used is an n-by-n grid of tiles. Thus, the approach being "tile-based." In this model, n is called the granularity of the policy or the intersection's resolution. In implementing FCFS, the tile retrieval process is added. Using data structures and modular arithmetic, the tiles are retrieved starting from the vehicle's bounds, then to the inside.

2.2. Reservations

When the vehicle sends a request for a reservation, the control system receives the request and simulates the vehicle's path when it crosses the intersection. As the vehicle crosses the intersection in the control system, it occupies a set of tiles, which are then checked against tiles that had been reserved by other vehicles. If no more than two vehicles occupy the same tile, no collision will occur, and hence, the control system issues the reservation, and the vehicle, accepting parameters from the control system such as speed and acceleration, may now cross the intersection. Otherwise, the request of the vehicle is rejected. Reservations are managed by a map where the reserved path through the intersection corresponds with the ID of the reserved vehicle.

2.3. Car Simulation

The car runs on a program independent of the control system. For reservations to commence, the car and the control system must exchange information. The control system follows the principle of AIM [8]. The simulation of the cars would maintain minimal communication complexity. After the reservation, it is the car's responsibility to meet the reservation at the correct time and place.

2.4. Predetermined 2D Grid

A localization scheme, which the authors refer to as a predetermined 2D grid, is an alternative solution for determining the cars' location. The cars' path is determined by their states, such as their current steering and throttle power. In other words, the locations are preprogrammed into the vehicles instead of fetching locations in real-time. The car commands that correspond to an appropriate maneuver are defined by specific coordinates in the track and the car's motion state. Each car is given an initial state consisting of x and y coordinates, speed, orientation, and route direction (left turn, right turn, or straight). With each pair of x and y coordinates, there is a corresponding car command. In this way, accurate navigation results for cars crossing the intersection. These car objects are then passed through the system that

simulates the cars for a given time. The results are a predetermined set of coordinates and a set of car commands over time.

2.5. Track Generation

The physical track (Sec. 2.9) on which the cars would be navigating was recreated in the simulation with the same dimensions and geometry. The crucial role of the simulated track is to provide cars with a driving line. To make the simulation versatile enough to accommodate multiple shapes and sizes of the track needed for testing, the track is subdivided into multiple track segment classes. The Track Segment class is the basic building block when constructing a full track for the simulation. Given the x and y locations, the track segment can return the distance to the nearest point of the driving line and the direction at that point. Each segment shape has a different method for calculating the distance from the driving line and direction.

2.6. Steering Adjustments

For steering, the car is concerned with its lateral distance from the driving line and its current orientation with respect to the driving line. Using these two data points, the car can either steer right, left, or straight. The goal for the car is to be centered on the driving line as much as possible. If the car deviates from the center in a specific direction, it is programmed to correct itself by steering in the opposite direction. It accomplishes this by simply checking if its distance from the centerline is higher than an absolute threshold value at which it will then start to steer in the direction towards the center of the lane.

2.7. Throttle Adjustments

Throttle adjustments are based on what is in front and behind that car, as well as the reservations. When a car is approaching a crossroad, it cannot merely adjust its speed based on the cars in front of it. The other cars inside or approaching the intersection must be taken into consideration. With that said, the car would leave the decision to maintain or decrease its speed to the reservation given by the control system. The control system evaluates two conditions that determine the throttle adjustment. If there is already an existing reservation, it should maintain its current speed since its present course is already approved. On the other hand, a rejection of a reservation request would indicate that there would be a collision if the car were to maintain its speed, and thus, would decrement its speed and send another request to the control system.

2.8. Communications Protocol

Between the computer program and the connected controllers that handle localization and car communication, the format of their messages are as follows: preamble, length, message type, and data. The preamble is a predetermined set of bytes used to indicate to the receiver that a message is incoming. The length is a single byte used to indicate the length of the entire message. This message can vary depending on the number of bytes used for sending data if applicable.

2.9. Road Track Design

The track design is based on a two-lane four-way crossroad that accommodates at least four cars. As shown in Figure 1, the important parameters are road shape, lines, and dimensions in the track design. The size of the track, the distance of the lane, and the intersection dimension are limited to the turning radius of the self-driving car.

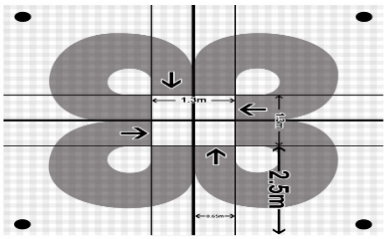


Figure 1: Test Track Map. The scaling was based on the turning radius of the car wheels.

2.10. Scaling Method

The scaling method followed the turning radius of the self-driving car instead of its dimensions. This scaling was done because the simulated car's steering angle was much larger than that of life-sized cars, resulting in the self-driving car taking wider turns than expected based on its dimensions. The dimensions of the test track, the self-driving cars' resulting dimensions, and their speeds in the simulation would be proportional to the ratio between its turning radius and the average turning radius of most common midsized passenger cars of 17 feet or 5.18 meters [9]. A scale factor of 15.94 was obtained from (1). Figure 2 shows the scaling process.

$$\text{scale factor} = \frac{\text{actual turning radius}}{\text{self-driving car turning radius}} \quad (1)$$

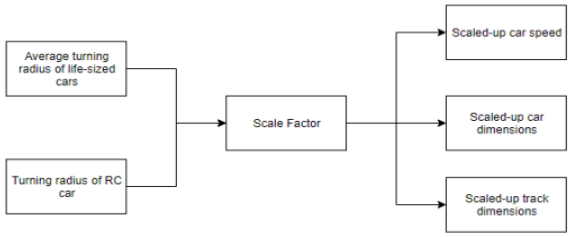


Figure 2: Scaling Process Block Diagram

2.11. System Program Flowchart

The system program flowchart is presented in Figure 3.

2.12. Evaluation

The intersection contains four entry lanes. It has no start and endpoints, and thus, it makes sense to treat the intersection as cyclic when it comes to determining the test cases. Four cars are to be arranged at the entry lanes of the intersection. The variables that determine the number of test cases are numerous, and some have infinite possibilities, such as speed, acceleration, arrival time to the intersection, and many more. Other variables are countable, such as the headings of the entry (north, south, east, west) and actions of maneuver (straight, left turn, and right turn), and these were the characteristics used.

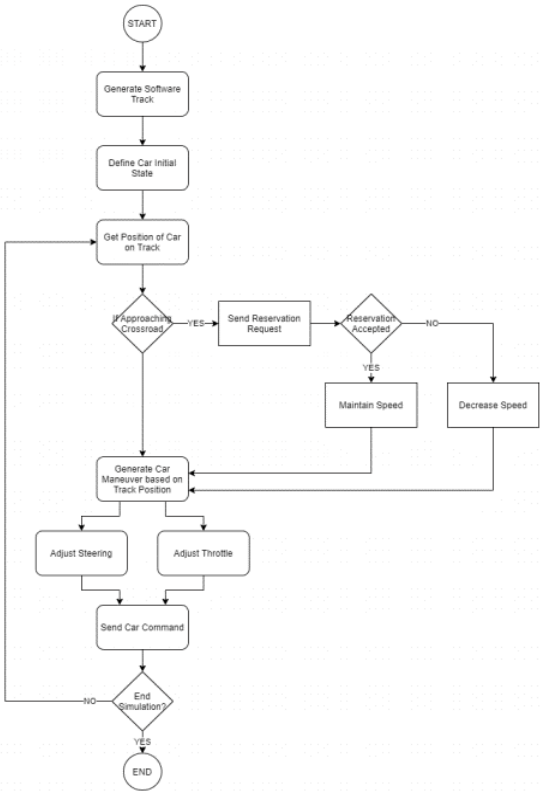


Figure 3: System Program Flowchart

3. Results and Discussions

3.1. Crossroad Reservation Queueing

When the car sent its request for a reservation, the control system either gave an acceptance or denial. Depending on the result, the car either maintained its speed or reduced it. Each time the reservation method is called, the control system saved an entry of the reservation details and results. The most notable data in the reservation logs are the arrival time, where the Car reaches the crossroad and the car's current throttle. An example of a summary of the test cases' logs can be found in Table 1.

Table 1: Crossroad Reservation Logs

Time Stamp (sec)	Car	Reserved	Arrival Time (sec)	Throttle
24.361	Car 4 - 0x6a1a	TRUE	25.397	70
24.833	Car 1 - 0x6a40	FALSE	25.524	70
24.868	Car 2 - 0x6743	TRUE	25.992	70
24.937	Car 3 - 0x673b	TRUE	25.88	70
25.153	Car 1 - 0x6a40	TRUE	26.024	66

3.2. Car Paths

After giving an initial state for each car, the system generates a predetermined 2D grid that contains all the cars' coordinates throughout the simulation. A path of each car can then be drawn from the set of coordinates to visualize the cars' movements.

The markers along the line represent a unique entry of coordinates from the predetermined 2D grid. In every iteration of the simulation, the car's location and other attributes are incremented by the amount of time elapsed between iteration. Following the trend of these markers shows the path and, most importantly, the cars' action. Analyzing the path of Cars 1 and 2 in Figure 4 shows left turn maneuvers and cars 3 and 4 showing the right turns. It can be concluded that the predetermined car paths match up with the desired test case.

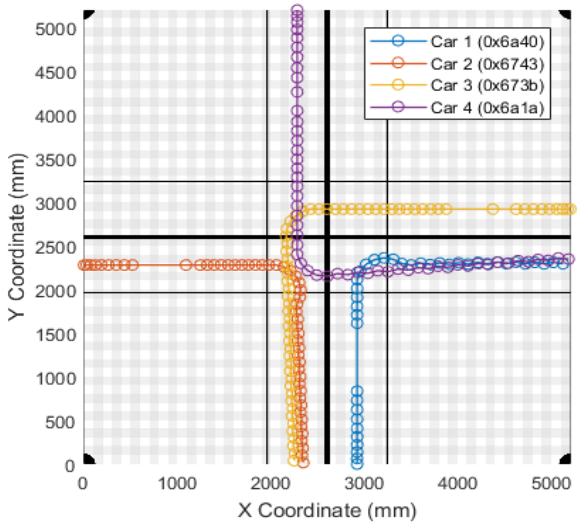


Figure 4: Car Paths for Test Case 4

3.3. Comparison of 2D Grid Simulation Results with Actual Results

The locations of the Predetermined 2D grid must be compared to the physical motions of the cars to see if they can follow the simulated paths. The range of variation in the speeds affected the time in crossroad for the physical cars. Since the simulated cars could achieve speeds far greater than the physical cars, they had several instances of spending less time in the crossroad. More often than not, the physical cars would spend more than one second in the crossroad while most predetermined cars have sub-one-second times. The predetermined and physical average speeds before and after the crossroads are presented in Table 2 and Table 3. The accuracy of the car's movements depends on their locations predicted by the predetermined 2D grid coinciding with the actual locations on the track.

Table 2: Predetermined Average Speeds Before and After the Crossroad

	Average Speed Before (m/s)				Average Speed After (m/s)			
Test Case	Car 1	Car 2	Car 3	Car 4	Car 1	Car 2	Car 3	Car 4
1	0.766	0.725	0.805	1.008	0.823	0.818	0.854	1.083
2	0.767	0.727	0.806	1.012	0.823	0.818	0.854	1.083
3	0.969	0.538	0.807	1.019	1.016	0.818	0.854	1.083
4	0.969	0.728	0.807	1.019	1.016	0.818	0.854	1.083
5	0.967	0.727	0.806	1.012	1.016	0.818	0.854	1.050

Table 3: Average Physical Speeds Before and After the Crossroad

	Average Speed Before (m/s)				Average Speed After (m/s)			
Test Case	Car 1	Car 2	Car 3	Car 4	Car 1	Car 2	Car 3	Car 4
1	1.206	0.938	1.152	1.124	1.311	1.127	0.995	1.404
2	1.085	0.934	1.027	1.034	1.152	0.958	0.891	1.331
3	1.036	0.934	0.971	1.015	1.325	0.975	0.856	0.993
4	1.030	0.971	0.860	0.988	1.252	0.921	0.929	0.953
5	1.125	0.963	0.777	1.054	1.048	0.853	0.942	1.109

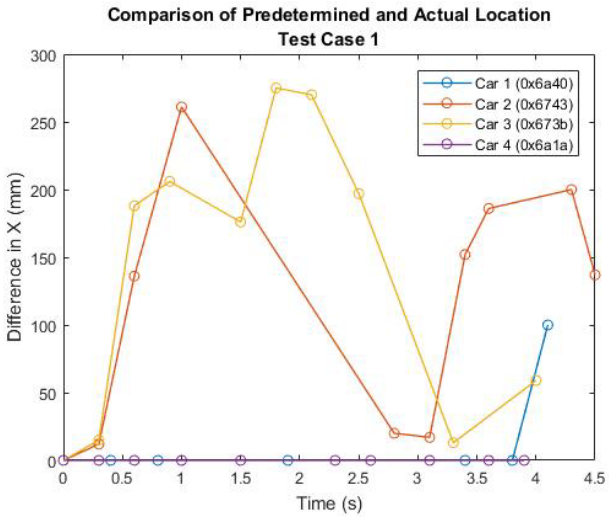


Figure 5: Comparison of Predetermined and Actual Locations Test Case 1a

Comparisons of their positions in the predetermined 2D grid and actual results were visualized in graphs in all test cases. The actual results are based on positions of remote-controlled cars acting as self-driving cars on the actual scaled-down track (Figure 1). The graphs show how much the actual locations differ from the predetermined locations of the cars. The difference between the predetermined and actual x and y coordinates in one test case is shown in Figures 5 and 6. The difference in x and y coordinates between the predetermined 2D grid location and the actual location is mostly negligible during each test's initial seconds. As the two cars accelerate, the actual cars' location falls slightly behind the predetermined location. However, the actual cars can also catch up back to the predetermined location, but with slight differences.

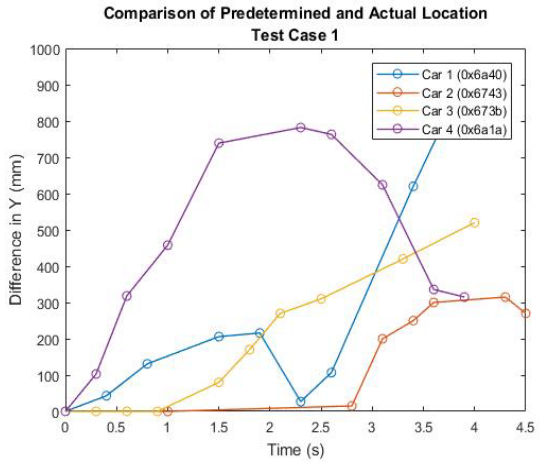


Figure 6: Comparison of Predetermined and Actual Locations Test Case 1b

One cause behind this type of deviation between the predetermined and actual results is due to the car's nature to not go directly straight despite not being issued any commands to turn, leading to it travel a greater distance than it would have otherwise traveled, creating sizeable deviations between its actual and simulated coordinates. It was also observed during testing that as the cars were used, the power banks' voltage level naturally also decreased, thereby making the cars' actual speeds more inconsistent compared to their simulated speeds, thus contributing further to the variation between the simulated and actual results.

3.4. Additional Cars

The hardware adaptation is limited to only four cars, but the simulation was designed to accept more than that. To add a new car to the program, select "Add New Car" in the authors' graphical user interface. The new car is then initialized based on the four cars' average characteristics since the new car does not have a hardware prototype. The system was able to simulate more cars than the physical hardware of four prototype cars, and the speeds of the cars were being adjusted like the procedure in the simulation of the prototype cars. Not only the predetermined 2D grid paths of the additional cars are generated, but the car commands are also sent using their Car IDs. Figure 7 shows one of these simulations of 8 cars in a zoomed crossroad.

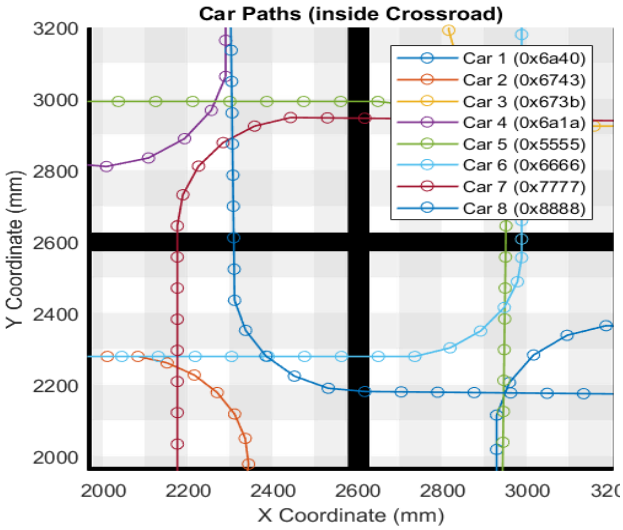


Figure 7: Eight Car Path Simulation - Zoomed in Crossroad

Reservations have an impact on the speed of the car. The eight-car scenarios resulted in more requests sent to the control system. Consequently, the increased number of cars caused the control system to issue more denials of reservations until the cars adjusted their speed accordingly. Figure 8 plots the reservation request results alongside the car's speed through the simulation's duration, while Figure 9 shows a car possessing an approved reservation alongside its speed versus time. When the car's request was approved, its speed is maintained until the car exits the crossroad.

The collections of reservation requests and status of the eight cars for all test cases were established. Here, the differences between each car's movement due to which gets approval first become apparent. Those cars that were given immediate approval, such as those who requested first, did not have to decrease their speeds. On the other hand, cars with several denials decremented several times before approval. Figure 10 shows the positive pulse of the reservation request that denotes approval lines up with its respective car beginning to maintain its speed. Likewise, the negative pulses that denote denial lines up with the car, decreasing their speeds. The approved reservation status alongside speed shows the window that the car must maintain its speed for the reservation duration. In Figure 11, the positive edge of the reservation marks the point where the speed is maintained. The

negative edge marks the point where each car's reservation is released; hence, the cars accelerate back. Between these two marks are the approved movements of the cars by the control system to avoid collisions.

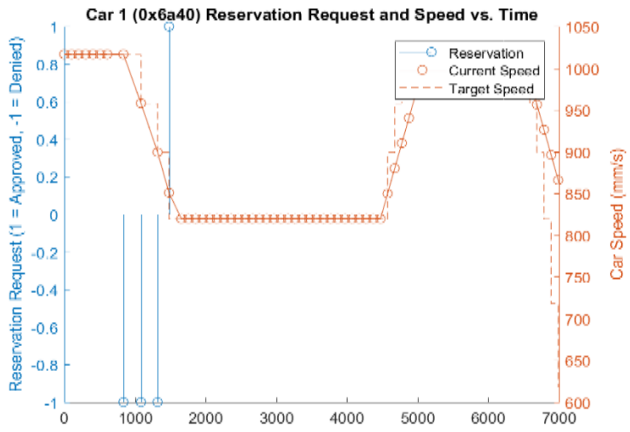


Figure 8: Reservation Requests Affecting Car's Speed

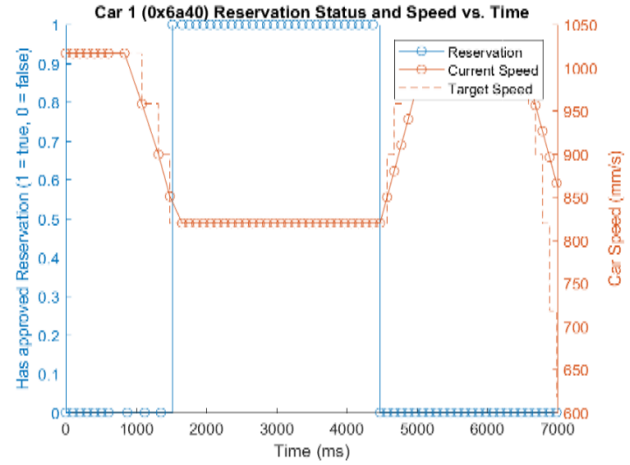


Figure 9: Reservation Status Alongside Car's Speed. The speed unit was based on the simulations.

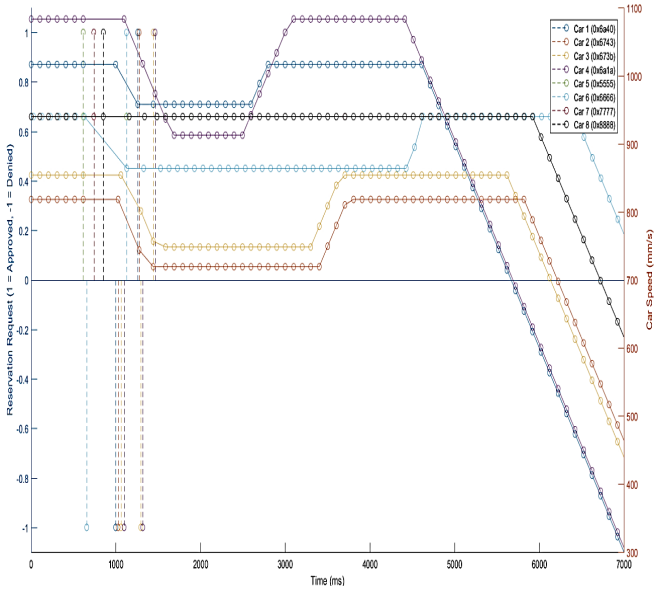


Figure 10: Effect of Reservation Requests on Car Speed. The speed unit was based on the simulations.

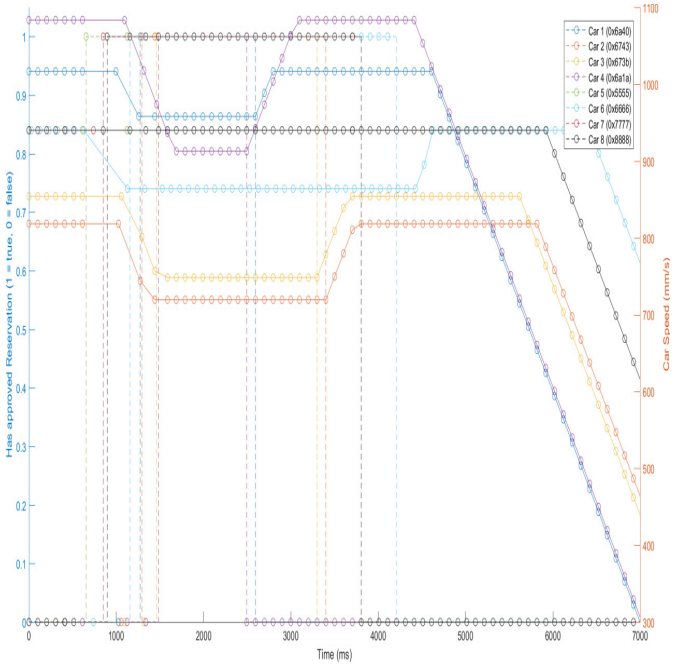


Figure 11: Having Approved Reservation and Car Speed. The speed unit was based on the simulations.

4. Conclusions and Recommendations

This work on a simulation design of collision avoidance of autonomous vehicles simultaneously crossing an intersection indicate that using the predetermined 2D grid can be an alternative localization scheme. Four prototype cars were modified with the specified modules and achieved the desired functionality of remotely controlling the control system's speed and steering. The control system was able to fulfill its function of controlling all the cars and negotiating the movement inside the crossroad per test case. The predetermined 2D grid was used as an alternative solution for determining the cars' locations increasing the accuracy of data and transmission. The result is a system that can control self-driving cars through a crossroad with minimal collisions. Furthermore, the control system using the predetermined 2D grid can perform simulation of more than four cars. Upon running the various test cases to verify the system with different car placements and maneuvers, it is concluded that the work was able to provide a fast method for vehicles to enter and exit a crossroad while avoiding slowdowns.

Recommendations include incorporating machine learning and cloud technology [5] in AVs and AIM, implementing simulation in different programming platforms, extending hardware implementation on large-scale vehicles, or 3D road maps [4]. It is also suggested to use uninterrupted, fast, and accessible communication technologies. Moreover, a useful feature would be more sensors that can keep track of attributes not necessarily related to motion. If a project were to implement a system like the predetermined 2D grid, it is recommended that the simulated cars' characteristics to be like the physical cars that it would eventually control. Expanding the model of the car's movement from linear acceleration to a more dynamic model would significantly improve the accuracy of the resultant predetermined coordinates, and consequently, more appropriate car commands. In addition, intelligent highway queue selectors

[10] can be explored for controlling data dissemination, which might result in coordinated traffic flow at crossroads.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

De La Salle University is acknowledged for supporting this work.

References

[1] E. A. Smith, "Statistics on Intersection Accidents," 2018. https://www.autoaccident.com/statistics-on-intersection-accidents.html
[2] R. Ewing, Traffic Calming State of the Practice. 1999. https://nacto.org/wp-content/uploads/2012/06/Ewing-Reid-1999.pdf
[3] N. Kazis, To Get Safer Streets, Traffic Lights and Stop Signs Aren't the Answer. 2011. https://nyc.streetsblog.org/2011/04/26/to-get-safer-streets-traffic-lights-and-stop-signs-arent-the-answer
[4] I. W.-H. Ho, S. C.-K. Chau, E. R. Magsino, and K. Jia, "Efficient 3D Road Map Data Exchange for Intelligent Vehicles in Vehicular Fog Networks," IEEE Transactions on Vehicular Technology, **69**(3), 3151–3165, 2020. doi: 10.1109/TVT.2019.2963346.
[5] K. F. Chu, E. R. Magsino, I. W.-H. Ho, and C.-K. Chau, "Index Coding of Point Cloud-Based Road Map Data for Autonomous Driving," in 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), 1–7, 2017. doi: 10.1109/VTCSpring.2017.8108280.
[6] SAE International and S. O.-R. A. V. S. COMMITTEE, "Surface Vehicle Information Report: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems," 2014.
[7] K. Dresner and P. Stone, "A Multiagent Approach to Autonomous Intersection Management," Journal of Artificial Intelligence Research, **31**, 591–656, 2008.
[8] T.-C. Au, S. Zhang, and P. Stone, "Autonomous Intersection Management for Semi-Autonomous Vehicles," Handbook of Transportation, 88–104, 2015.
[9] Vehicle Turning Paths Dimensions and Drawings. Dimensions.Guide [Online]. Available: https://www.dimensions.guide/collection/vehicle-turning-paths-radius
[10] E. R. Magsino and I. W. H. Ho, "An intelligent highway tollgate queue selector for improving server utilization and vehicle waiting time," in 2016 IEEE Region 10 Symposium (TENSYMP), 2016, 271–276, doi: 10.1109/TENCONSpring.2016.7519417.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 738-743 (2020)

www.astesj.com

Second International Virtual Conference on Multidisciplinary Research 2020

ASTESJ

ISSN: 2415-6698

Comparison of Support Vector Machine-Based Equalizer and Code-Aided Expectation Maximization on Fiber Optic Nonlinearity Compensation Using a Proposed BER Normalized by Power and Distance Index

Mark Renier M. Bailon[1], Lawrence Materum[1,2,*]

[1]*De La Salle University, 2401 Taft Ave., Malate, Manila, 0922, Philippines*

[2]*Tokyo City University, 1-28-1 Tamazutsumi, Setagaya, Tokyo, 158-8557, Japan*

A R T I C L E I N F O

Article history:
Received: 10 September, 2020
Accepted: 07 October, 2020
Online: 24 November, 2020

Keywords :
Coherent Fiber Optics
Nonlinearity Compensation
Support Vector Machine
Expectation Maximization

A B S T R A C T

Advances in optimizing optical fiber communications have been on the rise these recent years due to the increasing demand for larger data bandwidths and overall better efficiency. Coherent optics have focused on many kinds of research due to its ability to transport greater amounts of information, have better flexibility in network implementations, and support different baud rates and modulation techniques. These result in fiber-optic lines to provide faster speeds to end-users. Recent literature has looked into further developing digital signal processing techniques, while others have focused on fiber material optimization. Machine learning is another area of research that has garnered traction due to such demands. This survey discusses support vector machine (SVM) and code-aided expectation-maximization (CAEM) techniques on how they compensate for nonlinearity in coherent fiber optical communications. The study mainly focuses on how these techniques impact the performance of the transmissions where they are implemented and how they compensate for fiber optic nonlinearity through either the reduction of bit error rates (BERs), the improvements in the quality factor, or through a suggested index based on BER, power, and distance. Collating the results and based on a distinctive index, SVM is preferable in mid-range haul transmissions while CAEM for longer hauls.

1. Introduction

With the growing need to manage, transmit, process, and receive large amounts of data, optical fiber transmissions can find a niche in today's network communication age as it can address several network traffic issues. Compared to other communication methods such as radio wave propagation and other physical transmission media like twisted pair and coaxial cables, optical fiber transmission can deliver more data, operate more efficiently, occupy less space while having more capacity, and be less susceptible to interceptions. However, despite these advantages, optical communications have its fair share of disadvantages such as cost, complexity, and perhaps the most faced issue is phase sensitivity [1]. Over recent years, coherent optic fiber communications coupled with newly discovered digital signal processing techniques have improved and optimized data transmissions. One such stride was the shifting from single carrier multiplexing to coherent optical orthogonal frequency division multiplexing (CO-OFDM), which brought about advantages such as inter-symbol interference mitigation and higher bandwidth efficiency. However, OFDM's serious disadvantage is the higher peak-to-average power ratio (PAPR) that comes with it, which results in a phenomenon known as fiber optic nonlinear distortion [2]. Digital modulation techniques that are usually paired with OFDM, such as Amplitude shift keying (ASK) and Phase shift keying (PSK), and Quadrature amplitude modulation (QAM), are greatly affected by nonlinear distortions as these can significantly increase the BER of the system. Numerous studies and innovations such as those in [3]–[5] have risen to try to minimize the nonlinearity experienced in this multiplexing process such as varying or combining the digital signal processing (DSP) techniques applied, optimizing the material that coats the fiber optic core to reduce the birefringence and implementing neural networks in the transmitter/receiver or both to maximize the bandwidth delivered.

*Corresponding Author: Lawrence Materum, Email: lawrence.materum@dlsu.edu.ph

Just like any other transmission system, coherent optical fiber transmissions encounter their fair share of performance drops due to several types of induced noise. Amplified spontaneous emission (ASE) from inline amplifiers is one of these major contributors to linear noise. Laser phase noise from transmitters and the local oscillator also put a damper on system performance by a relatively significant margin. However, the largest contributor and perhaps the most significant cause for concern is the nonlinear phase noise (NLPN) caused by an interaction of the signal and the ASE mentioned above through the phenomenon known as fiber Kerr effect or Kerr induced nonlinearity [6].

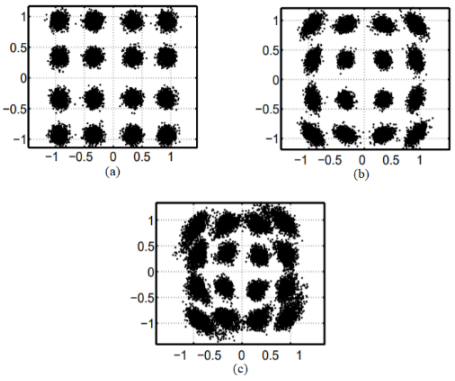


Figure 1: The effects on a 16-QAM system due to (a) Amplified Spontaneous Emission (b) Laser Phase Noise and (c) Nonlinear Phase Noise [7]

For transmissions like quadrature amplitude modulation (QAM), which relies heavily on signal amplitude and phase shifting in its transmission, noise can significantly impact its performance, specifically on its BER. A small amount of noise can incorrectly categorize the transmitted data. This error is especially apparent in optical transmissions because most utilizing the orthogonal frequency division multiplexing scheme produces a high peak-to-average power ratio resulting in noise that can increase the BER in a transmission medium.

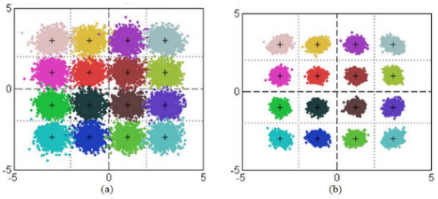


Figure 2: (a) 16-QAM constellation with high nonlinear phase noise; (b) with reduced noise [8]

Hence, solutions to addressing such nonlinearity are essential. In this work, two important techniques that work to that end analyze SVM and CAEM on how they affect optical fiber transmissions regarding their overall performance and fiber optic nonlinearity compensation. Performance comparison was made between the two using parameters of the different studies. In particular, the emphasis was on the use of coherent optical orthogonal frequency division multiplexing (CO-OFDM) or polarization division multiplexing using a 16-ary QAM (16-QAM) signal with varying fiber lengths and variables. The authors also propose a comparative index to fairly evaluate the two techniques based on the bit error rate (BER), power, and transmission distance. This paper is organized then as follows. Section 1 is followed by discussing the methods and a proposed comparative index in Section 2. Results and their discussions were done in Section 3, and recommendations given in Section 4.

2. Nonlinearity Compensation and Comparative Methodology

2.1. 16-QAM Least-Squares SVM Nonlinearity Equalizer

A 16-QAM CO-OFDM coupled with an SVM nonlinear equalizer is proposed in [2]. The optical fiber link is composed of multiple 100 km standard single-mode fiber (SSMF). Attenuation in the link is accounted for and compensated using Erbium-doped fiber amplifiers (EDFA). The ASE contributed by the EDFA is considered as white Gaussian noise. Inputs to the digital modulator are generated from a pseudo-random binary sequence (PRBS) module, which then undergoes the QAM modulation. An inverse fast-Fourier transform module is utilized to convert the time domain signal generated to an equivalent frequency domain. In this case, the pulses of the signal are kept ideal for simplifying the simulation. To maximize linear conversion between the OFDM signal and the optical field, the OFDM signal's in-phase and quadrature segments are used by a pair of Mach-Zehnder modulators (MZM). Both MZM operates via push-pull configuration and is configured to be biased at the minimum transmission point to remove the chirp phenomenon effectively. Once the optical signal reaches the receiver after traversing N-span amplifiers, it is converted into an electrical signal by a 90º photoreceiver. Noise due to the laser linewidth's imperfections is disregarded as the study aims to isolate fiber nonlinearities due to other noise [2]. The signal then undergoes the normal decryption process before being fed into the machine learning algorithm, after which it is fully demodulated, and the error rate is calculated.

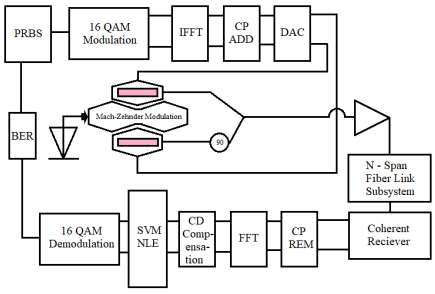


Figure 3: 16-QAM CO-OFDM SVM-NLE Diagram [2]

Support vector machines are powerful classifying tools yet can only be used as binary classifiers. Due to this restriction, this study's approach combines multiple two-class SVMs for a multi-class model to be used. Since the study uses 16-QAM as its modulation technique, the signal is divided into sixteen (16) individual clusters in a constellation wherein each cluster represents data in a unique binary form [9]. For a single two-class SVM classifier, N pairs of vectors (x_k, y_k), k=1,…,N where x_k and y_k are the input and output patterns, respectively undergo training to obtain the hyperplane. y_k is the labeling function where $y_k \in \{1, -1\}$. Through this training process, possible noise in the constellation is distributed efficiently and accurately. Based on the training data present, SVM aims to construct a classifier f(x)

$$f(x) = \text{sign}[(W^T \Phi(X_i) + b)] = \text{sign}\left[\sum_k^N \alpha_k y_k K(x, x_k) + b\right] \quad (1)$$

where w is the weight, α_k is the support vector, b is the bias term, and $\Phi(X_i)$ is the mapping function. The training process determines the weight, support vector, and bias terms used for the constructed classifier. For more complex data to be accurately separated, a mapping function is used to transform the training data x_k into a higher dimension. The SVM also utilizes a Kernel trick to help nonlinear decisions in mapping low complexity computations. The approach focused on this survey utilizes a radial basis function kernel in which only dot products are needed. $K(x_i,k_j) \equiv \exp(-y_{SVM}||x_i - x_j||^2))$, ySVM > 0, with y_{SVM} as the Kernel function.

Since there are more than two unique data sets to classify from in a 16-QAM constellation, the study employed a one versus rest rule, wherein a received data point would classify in a particular cluster if and only if it is accepted by that cluster and is rejected by the rest, if two or more clusters accept the data point then it is considered noise.

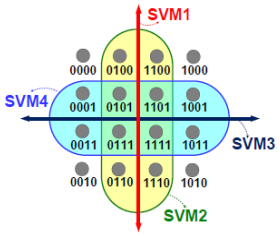


Figure 4: Implementation of a 4-level SVM to classify a 16-QAM constellation [10]

To further increase the SVM classification accuracy in noisier environments, the study opted to implement a least-square variant of the SVM studied in [11]. Least-Square SVM (LS-SVM) provides a more optimized solution using the following:

$$\min\left(\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i\right) \quad (2)$$

constrained by

$$y_i(w^T\phi(X_i) + b) \geq 1 - \xi_i \quad (3)$$

where ξ is termed as the slack variable which shows the error term which must satisfy the condition $\xi_i \geq 0$. C is known as the regularization parameter.

After the CD compensation process, both I and Q segments of the signal are fed into the LS-SVM in which the classifier is formed through a two-stage process of training and testing [2], [11].

1. Training
- Arrange label y_k, in-phase I and quadrature Q to format the SVM packet.
- Select the RBF Kernel function and scale I, Q to [0,1]
- Use cross-validation to determine the optimal C and y_{SVM} values.
- C and y_{SVM} values to train the SVM.

2. Testing
- Insert testing symbol.
- Compare predicted labels to transmitted symbols to determine and evaluate the BER.

2.2. Code-Aided Expectation Maximization

A wavelength-division multiplexing (WDM) and polarization division multiplexing (PolDM) system is considered in [12]. Nine simultaneously transmitting channels with the middle channel being the main focus of the study utilizes 16-QAM modulation with a 32 GBaud symbol rate. The optic signal goes through a 2640 km distance consisting of 33 spans of 80 km each. Dispersion-compensating fiber (DCF) is not utilized in the study; however, the same standard single-mode fiber (SSMF) is used for the fiber cable, and an erbium-doped fiber amplifier (EDFA) is employed to counteract fiber loss. The signal undergoes a chromatic dispersion compensation followed by a polarization demultiplexing at the receiver. A frequency estimation (FE) acts on the sampled signal to estimate the frequency with a margin of accuracy equal to or better than 4 MHz. A Viterbi algorithm compensates for any leftover frequency that utilizes an FIR filter length with phase averaging.

Due to noise correlating over time, traditional white noise assumption methods of demodulation are often suboptimal. The study in [12] partially compensates impairments done by both inter-channel and intra-channel nonlinear effects by exploiting the correlation of the phase noise through the use of CAEM. The phase noise correlation is dealt with by using a regularizer in the utility function of the algorithm. The CAEM describes as follows:

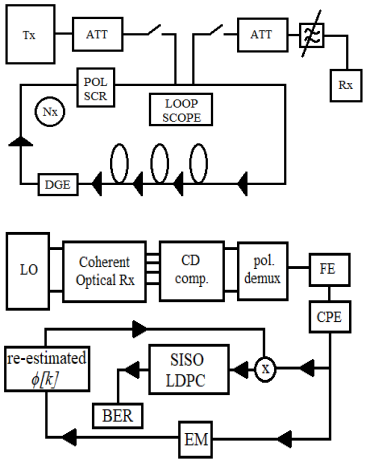


Figure 5: 16-QAM PDM CAEM Algorithm Diagram [12]

1. $r^i[k]$ will be set as the k th symbol after phase noise compensation in the ith iteration between the EM and forward error correction (FEC) process $\left(r^i[k] = y[k]e^{-j\hat{\phi}^i[k]}\right)$. $\hat{\phi}^0[k] = 0$ and $r^0[k] = y[k]$ are set as initial values.
2. The log-likelihood-ratio (LLR) is computed for each bit based on $r^i[k]$. Let s_l be the lth bit. The LLR is defined as follows if M-ary modulation is utilized.

$$\mathrm{LLR}(s_l) = \log\frac{p(s_l = 0 \mid r^i[k])}{p(s_l = 1 \mid r^i[k])} \quad (7)$$

$$l = (k-1)\log_2 M + 1, \ldots, k\log_2 M \quad (8)$$

It is assumed that the residual noise in $r^i[k]$ patterns itself in a circularly symmetric zero-mean white Gaussian distribution. Constellation points whose labels are $s_l = 0$ are defined in the set X_{l0} and likewise for X_{l1}.

3. An updated log-likelihood ratio $\mathrm{LLR}_o(s_l)$ is obtained by decoding a soft-input-soft-output (SISO) FEC based on the initial $\mathrm{LLR}(s_l)$. $\mathrm{LLR}_o(s_l)$ bits are then converted to probabilities. Assuming a constellation point, $a \in C$ is labeled as a logarithmic bit sequence $a_1, a_2, a_{3,\ldots,} a_{log_2 M}$ we obtain:

$$p\big(x[k] = a \big| y[k], \hat{\phi}^i[k]\big) = \prod_{m=1,\ldots,log_2 M} p\big(s_{(k-1)log_2 M+m} = a_m\big) \quad (9)$$

4. The utility function $E\big(\phi, \hat{\phi}^i\big)$ is then setup – Expectation Step.

$$E\big(\phi, \hat{\phi}^i\big) = \alpha\zeta(\phi) + (1-\alpha) * \sum_{k=1}^{N}\sum_{a \in C} p\big(x[k] = a \big| y[k], \hat{\phi}^i[k]\big) * \log p(y[k]|x[k] = a, \phi[k]) \quad (10)$$

where N is the total number of symbols, α is the BER optimization weight with $\alpha \in [0,1]$ The regularizer function $\zeta(\phi)$ is described as follows:

$$\zeta(\phi) = \sum_{k=3}^{N}((\ \phi[k] - w_1\phi[k-1] - w_2\phi[k-2]))^2 \quad (11)$$

where w_1 and w_2 are the noise AR analysis weights. The expression $p(y[k]|x[k] = a, \phi[k])$ is computed as follows:

$$p(y[k]|x[k] = a, \phi[k]) = \frac{1}{\sigma^2\sqrt{2\pi}}\exp\left(-\frac{\left|y[k] - ae^{j\phi[k]}\right|^2}{2\sigma^2}\right) \quad (12)$$

in which σ^2 is the noise variance that can be empirically computed with the aid of the training data.

5. The vector $\hat{\phi}^{i,1}$ is calculated assuming the below conditions - Maximization step:

$$\hat{\phi}^{i,j} = \arg\underbrace{max}_{\phi} E\big(\phi, \hat{\phi}^i\big) \quad (13)$$

where $\hat{\phi}^{i,j}$ is the estimated phase noise between FEC and EM in the ith iteration and the estimated noise between the E and M steps in the jth iteration. To numerically compute for $\hat{\phi}^{i,1}$ a gradient-ascent method is used.

6. The utility function is recomputed, however instead of using (step 3), the following is used.

$$p\big(x[k] = a \big| y[k], \hat{\phi}^{i,j}[k]\big) = \frac{1}{\sigma^2\sqrt{2\pi}}\exp\left(-\frac{\left|y[k] - ae^{j\hat{\phi}^{i,j}[k]}\right|^2}{2\sigma^2}\right) \quad (14)$$

7. Step 6 is repeated until $\hat{\phi}^{i,j}$ converges after which $\hat{\phi}^{i+1} = \hat{\phi}^{i,j}$ is set.
8. Steps 1 to 7 are repeated until $\hat{\phi}^i$ converges, and $\hat{\phi} = \hat{\phi}^i$ being the final phase noise is estimated.

2.3. Simulation Parameters

The following tables show the parameters and conditions of each approach.

Table 1: Least-Square SVM Parameters

Multiplexing	CO-OFDM
Distance	up to 1200 km
Amplifier	EDFA
Noise Figure	6 dB
Fiber	SSMF
Span Length	100 km
Attenuation Coeff	4.605 x 10^{-5} m^{-1} *
Dispersion Coeff	17 ps/(nm · km)
Nonlinear Coeff	-
EDFA Emission Factor	-
Symbol rate	40 Gbaud
Tx & LO Linewidth	-
Modulation	16-QAM
Sampling rate	-

Table 2: Code-Aided EM Parameters

Multiplexing	PDM
Distance	2640 km
Amplifier	EDFA
Noise Figure	-
Fiber	SSMF
Span Length	80 km
Attenuation Coeff	4.8354 x 10^{-5} m^{-1}
Dispersion Coeff	17.025 ps/(nm · km)
Nonlinear Coeff	1.3 $(W \cdot km)^{-1}$
EDFA Emission Factor	1.7741
Symbol rate	32 Gbaud
Tx & LO Linewidth	100 kHz
Modulation	16 QAM
Sampling rate	1 sample/symbol

3. Comparative Index, Results, and Discussions

The tables below show the results of each study's approach. Since comparison is to be done using the BER, findings with Q-

factor results were converted to their equivalent BER using the following [13]:

$$\text{BER} = \frac{1}{2}\text{erfc}\left(\frac{Q}{\sqrt{2}}\right) \approx \frac{1}{Q\sqrt{2\pi}}e^{-\frac{Q^2}{2}} \quad (15)$$

$$\text{Q(dB)} = 10\log_{10}(\text{Q}^2) = 20\log_{10}(\text{Q}) \quad (16)$$

where Q is the Q-factor and erfc () is the complementary error function. Since the (erfc) used in calculating BER is approximated, any BER values computed will be approximations. To ensure the calculations' high accuracy, BER values are displayed, having values up to seven decimal places.

3.1. Proposed Index

The authors proposed an index (17), which is a measure of how well each technique compares to others depending on the application. The index in Table 8 is calculated as follows:

$$M = \frac{\text{BER}}{\text{Power} \times \text{Distance}} \quad (17)$$

Table 3: Quality Factor (dB) for 1000 km and 1200 km fiber length at differing Launch Powers [2]

Launce Power	Distance (km)	
	1000	1200
-3 dBm	9.50	8.90
-4 dBm	10.20	9.55
-5 dBm	10.40	9.95
-6 dBm	10.30	9.90
-7 dBm	10.00	9.60
-8 dBm	9.50	9.10
-9 dBm	8.90	8.50

Table 4: BER for 1000 km and 1200 km fiber length at differing Launch Powers.

Launce Power	Distance (km)	
	1000	1200
-3 dBm	0.0015509	0.0029533
-4 dBm	0.0006563	0.0014644
-5 dBm	0.0005011	0.0009053
-6 dBm	0.0005743	0.0009635
-7 dBm	0.0008500	0.0013820
-8 dBm	0.0015509	0.0024036
-9 dBm	0.0029533	0.0043513

Table 5: BER at Launch Power = -6 dBm for different fiber lengths

Distance (km)	Launch Power	Q-factor (dB)	BER
400	-6 dBm	11.50	0.0000909
600	-6 dBm	11.25	0.0001389
800	-6 dBm	10.90	0.0002423
1000	-6 dBm	10.30	0.0005743
1200	-6 dBm	9.90	0.0009635

3.2. LS-SVM Results

The BER values in Table 3 are extrapolated from the graphical results in [2], while BER values in Tables 4 and 5 are calculated estimates from the data of [2] using (15) and (16).

3.3. CAEM Results

The Q-Factor values in Table 6 are extrapolated from the graphical results in [12]. BER values are calculated estimates using (15) and (16) based on the Q-Factor.

Table 6: Q-Factor (dB) and BER as a function of Launch Power at 2640 km fiber length

Launce Power	Distance (km)	Q-Factor (dB)	BER
-2 dBm	2640	4.6	0.0555459
-1 dBm	2640	8.5	0.0043514
0 dBm	2640	11.4	0.0001081
1 dBm	2640	13.1	0.0000033
2 dBm	2640	6.9	0.0155732

3.4. Comparative Results

Table 7 shows BER values attained from the different studies under highly similar conditions, thus making it possible to compare their results. LS-SVM can be compared under a -6 dBm launch power at 1200 km fiber length while CAEM has results at 0 dBm launch power with a 2640 km fiber length.

Table 7: BER values under similar attributes

Algorithm	Distance (km)	Fiber Link	Launch Power	BER
LS-SVM	1200	-	-6 dBm	0.0009635
CAEM	2640	-	0 dBm	0.0001081

Table 8: Results Based on the Proposed Index

Algorithm	BER	Power (Watts)	Distance (km)	Index
LS-SVM	0.0009635	0.0002512	1200	0.0031963
CAEM	0.0001081	0.0010000	2640	0.0000409

Low index values mean better overall performance for the algorithm. As shown in Table 8, the LS-SVM can be applied well in midrange haul and low complexity applications, whereas CAEM for long haul and high complexity optical networks. The index provided in (17) for measuring and comparing the performance in the studies [12], [2] is thus proposed for nonlinearity compensation performance comparisons. Nonetheless, it also recognized that multiple values should be considered and further simulated to get graphical representations of the results.

4. Recommendations

Based on the results, LS-SVM provides nonlinearity compensation in a CO-OFDM 16-QAM system, but for longer fiber lengths, CAEM provides a significantly lower BER value. These outcomes make CAEM a preferred choice when it comes to long haul optical transmissions. Complexity wise it is recommended to utilize LS-SVM. CAEM is preferred despite the higher complexity due to its significantly lower index. One potential future work is to verify the results in an experimental setup.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

De La Salle University is acknowledged for supporting this work.

References

[1] K. Kikuchi, "Fundamentals of Coherent Optical Fiber Communications," Journal of Lightwave Technology, **34**(1), 157–179, 2016. doi: 10.1109/JLT.2015.2463719.

[2] T. Nguyen, S. Mhatli, E. Giacoumidis, L. V. Compernolle, M. Wuilpart, and P. Mégret, "Fiber Nonlinearity Equalizer Based on Support Vector Classification for Coherent Optical OFDM," IEEE Photonics Journal, **8**(2),. 1–9, 2016, doi: 10.1109/JPHOT.2016.2528886.

[3] D. Semrau et al., "Achievable information rates estimates in optically amplified transmission systems using nonlinearity compensation and probabilistic shaping," Opt. Lett., **42**(1), 121–124, 2017. doi: 10.1364/OL.42.000121.

[4] D. Lavery, D. Ives, G. Liga, A. Alvarado, S. J. Savory, and P. Bayvel, "The benefit of split nonlinearity compensation for single channel optical fiber communications," in 2016 IEEE Photonics Conference (IPC), 2016, 799–802, 2016. doi: 10.1109/IPCon.2016.7831073.

[5] H. Hu et al., "Fiber nonlinearity compensation of an 8-channel WDM PDM-QPSK signal using multiple phase conjugations," in OFC, 1–3, 2014. doi: 10.1364/OFC.2014.M3C.2.

[6] G. P. Agrawal, "Nonlinear Fiber Optics," in Nonlinear Science at the Dawn of the 21st Century, Berlin, Heidelberg, 195–211, 2000.

[7] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Accumulation of nonlinear interference noise in fiber-optic systems," Opt. Express, **22**(12), 14199–14211, 2014. doi: 10.1364/OE.22.014199.

[8] T. Pfau, S. Hoffmann, and R. Noé, "Hardware-Efficient Coherent Digital Receiver Concept With Feedforward Carrier Recovery for $ M $-QAM Constellations," Journal of Lightwave Technology, **27**, 2009, doi: 10.1109/JLT.2008.2010511.

[9] E. Giacoumidis et al., "Kerr-induced nonlinearity reduction in coherent optical OFDM by low complexity support vector machine regression-based equalization," in 2016 Optical Fiber Communications Conference and Exhibition (OFC), 2016, 1–3, 2016.

[10] D. Wang et al., "Nonlinear decision boundary created by a machine learning-based classifier to mitigate nonlinear phase noise," in 2015 European Conference on Optical Communication (ECOC), 1–3, 2015. doi: 10.1109/ECOC.2015.7341753.

[11] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM transactions on intelligent systems and technology (TIST), **2**(3),. 1–27, 2011, doi: 10.1145/1961189.1961199.

[12] C. Pan, H. Bülow, W. Idler, L. Schmalen, and F. R. Kschischang, "Optical Nonlinear Phase Noise Compensation for 9x32 -Gbaud PolDM-16 QAM Transmission Using a Code-Aided Expectation-Maximization Algorithm," Journal of Lightwave Technology, **33**(17), 3679–3686, 2015, doi: 10.1109/JLT.2015.2451108.

[13] W. Freude et al., "Quality metrics for optical signals: Eye diagram, Q-factor, OSNR, EVM and BER," in 2012 14th International Conference on Transparent Optical Networks (ICTON), 1–4, 2012. doi: 10.1109/ICTON.2012.6254380.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 744-761 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Analysis of Long-term Equilibrium Relationship Between KRW, RMB, JPY Exchange Rates and International Financial Market Variables: Comparative Analysis of KRW, RMB, JPY

Moon-Kyum Kim*, Woong Ryeol Kim

Graduate School of Soongsil University, Seoul, 06978, Republic of Korea

ARTICLE INFO

Article history:
Received: 06 October, 2020
Accepted: 31 October, 2020
Online: 24 November, 2020

Keywords:
Currency
RMB
Yuan
CNY
CNH
KRW
JPY
NDF
VECM

ABSTRACT

This study used 15 VECM analysis models to analyze the relationship among exchange rates of South Korea, China and Japan; and between exchange rates of each country and international financial market variables. The analysis variables are the Won, Yuan, Yen spot exchange rates and international financial market variables. The results of the analysis are as follows: First, there was a long-term cointegration relationship between RMB, KRW, JPY and international financial market variables. Second, the analysis results of the VECM showed that explanatory power of Korea's offshore Won-Dollar accounts for 50% of the onshore Won-Dollar. The results of the analysis of the Yuan's VECM showed that the onshore Yuan (CNY) and offshore Yuan (CNH) exchange influence, but each has its own independent characteristics. Overall, the Won is more integrated with the international financial market than the Yuan. The Yen's relationship with variables in the international financial market was stronger than that of the Won and the Yuan. Third, offshore Won-Dollar, onshore KRW, and JPY had similar Granger causality relationship and impulse responses with international financial market variables. However, CNH and CNY indicated weaker than that of Won and Yen. The onshore Won-Dollar is shown to partially offset the shock from the onshore Yuan and offshore Yuan. The implications of this study are as follows: First, it is important to look at the exchange rate from the perspective of the international financial market. Second, Yuan investment and risk management are necessary considering the characteristics of the onshore and offshore Yuan which are interrelated but also distinctly unique markets. Third, it is necessary to manage the exchange rate position, taking into account the long-term equilibrium relationship among the Won, Yuan, and Yen currencies. Fourth, Korean companies should find a way to actively utilize the Won which shows the characteristics of partial internationalization.

1. Introduction

International credit rating agencies note changes in exchange rates in relation to country credit ratings. The exchange rate represents the credit risk of the currency country viewed in the financial market. In [1], the author argued that the credit rating announcement by the credit rating agency has a significant impact on the exchange rate and its influence varies depending on emerging markets or developed countries. In [2], the author argued that the depreciation of the exchange rate is a negative factor for the financial system of the country. In [3], the author argued that large differences in interest rates between local and foreign currencies would increase dollar loans with low interest rates. In [4], the author that entities in East Asian countries use foreign currency loans to reduce borrowing costs. They argued that company selectively use foreign currency loans in consideration of low foreign currency borrowing interest rate benefits and foreign exchange hedge costs arising from foreign currency borrowing. As they argument, the mid- to long-term exchange outlook, adjusted according to interest rates, expected currency values, macro-economic variables, etc., are important for the parties to the foreign exchange transaction. Therefore, it is important to understand the mid- to long-term equilibrium relationship between each currency and with variables in each currency and international financial markets variables. However, the exchange rate is generally highly volatile in the short term, as the following arguments suggest. In [5], the author argued that the short-term exchange rate determinant model by macroeconomic variables is almost

*Corresponding Author: Moon-Kyum Kim, Professor, Graduate School of Soongsil University. mgkim@ssu.ac.kr

indescribable. In [6], the author argued that the empirical analysis of advanced countries' currencies generally showed that the exchange rate was close to random walking, and that the won-dollar exchange rate was also random walking after the introduction of the floating exchange rate system in Korea. In [7], the author argued that the sensitivity of the trade-in between the local currency and the foreign currency was greater under the floating exchange rate system. He argued that the floating exchange rate system also limits the ability of monetary authorities to carry out their own policies. The following studies have shown that emerging economies' exchange rate-setting mechanisms and dollarization increase exchange rate volatility. In [8], the author studied the factors affecting the transactions of local and foreign currencies in emerging countries. He argued that the holding of the local currency or foreign currency depends on the level of interest rate on the foreign currency against the local currency and the expected depreciation of the local currency. Therefore, it was argued that the foreign currency trade was determined in accordance with reasonable expectations for the local currency and the foreign currency. In [9], the author argued that the dollarization generally increases in emerging economies. The dollarization is the use of the dollar (foreign currency) as a means of storing value instead of the local currency. The IMF defined the dollar phenomenon in emerging economies as the proportion of foreign currency deposits to broad cash. They used the dollarization level as the ratio of foreign currency deposits to M2. In [10], the author argued that the dollarization was a reasonable response from the investors to future uncertainties in emerging economies' currencies. The dollarization also affects macroeconomic policies, foreign exchange risk management and exchange rate systems in emerging economies. In [11], the author analyzed the Turkish currency using a VAR model. He argued that the dollarization is caused by instability in three macroeconomic variables: currency devaluation volatility, inflation volatility, and uncertainty forecast volatility. In [12], the author argued that the loss of confidence in economic management would reduce confidence in the local currency, thereby promoting the retention of foreign currencies. Bank plays a pivotal role in developing emerging economies' financial markets. Therefore, deepening the dollarization of banks' assets and liabilities increases foreign currency exchange risk. He argued that the dollarization is not significant in countries where the financial industry has grown to a considerable level, but in countries with small financial industries, the dollarization is relatively large. The above studies suggest that the medium- to long-term relationship of exchange rates is important for parties to foreign exchange transactions. The demand and supply of foreign exchange trading arise from the import-export transactions of enterprises and the hedging of foreign exchange positions held and the capital markets transactions invested in stocks and bonds. Thus, long-term exchange rate relationships between currencies and key variables in the international financial market are important for businesses and investors. South-Korea, China and Japan are the top 10 global import-export trading partners. In addition, China and Japan are Korea's largest export-import trading partners. The Korean won is a non-international currency, but the Chinese yuan and Japanese yen are SDR international currencies. As a non-international currency country, like most ASEAN countries, Korea has to do import-export transactions and capital market transactions with international currency countries. Therefore, it is an important issue for Korean import-export companies and pension fund investors to find out the long-term equilibrium relationship between the won dollar, the yuan, and the yen. Regarding exchange rate relationship between Korea, China and Japan. In [13], the author analyzed the non-international currency of the Won, the SDR international currency of the Yuan, and the Yen by applying CIP theory. The CIP theory is that there is an equilibrium relationship between the difference between the risk-free interest rates of the two currencies and the difference between the spot exchange rate and the forward exchange rate. Therefore, the country's foreign exchange market maintains equilibrium relationship under the CIP theory. In general, international currency is not regulated and there are no restrictions to trade, so the CIP theory is established. Therefore, it is in principle impossible to trade risk-free arbitrage transaction that arises from the collapse of the CIP equilibrium relationship. They argued that the Yen, which is an international currency, is impossible for risk-free arbitrage transaction when considering the transaction cost. However, they argued that the Yuan, which is in the process of internationalization, and the Won, a non-international currency, imply arbitrage opportunities through SWAP transactions. In [14], the author used the VECM (Vector Error Correction Model) model to analyze the relationship between the country CDS Premiums of Korea, China and Japan, the exchange rate and the variables of the international financial market. They argued that while the country CDS Premiums of Korea and Japan were more closely related to the variables in the international financial market, the Chinese CDS Premiums had a weak relationship. In [15], the author used the VECM model to analyze the relationship between the international financial market variables and the spot exchange rate of Latin America and the country CDS Premium. He argued that although variables in the international financial market affect the spot exchange rate of Latin America countries, their impact varies from country to country.

This study analyzes the long-term equilibrium between the Korean won (onshore KRW and offshore Won-Dollar, herein “NDF KRW” where NDF is non deliverable forward), the Chinese yuan (onshore CNY and offshore CNH), and the Japanese yen JPY currency. And analyze the long-term equilibrium relationship between these currency and international financial market variables. The analysis results compared the characteristics of a total of five currency markets by dividing them into won (KRW and NDF KRW) and yuan (CNY and CNH) and yen JPY. This study can be differentiated from the existent researches based on the following; First, we have identified the long-term equilibrium relationship between the exchange rates of Korea won, China yuan, and Japan yen from the perspective of the international financial market. Second, the degree of integration with the international financial market was compared by comparing and analyzing the non-international currency, the won, the yuan and the yen, the international currency. Third, it was found that although the onshore and offshore yuan, which is being promoted for internationalization, is related to each other, there is a unique market characteristic that is distinguished from each other. Therefore, we have provided suggestions that company and international investors need yuan trading and yuan position management in consideration of the onshore and offshore yuan characteristics. The composition of the study is as follows; Chapter 1 describes the purpose and background of the research. Chapter 2 describes prior research and the variables used in analysis. Chapter

3 described the research model and the research methodology. Chapter 4 basic statistics, unit root test, and co-integration test of the analysis data. Chapter 5 presents the analysis results of the VECM model. Chapter 6 presents summaries and conclusions.

2. Previous research

2.1. Dollarization

The dollarization is the use of the dollar (foreign currency) as a means of saving value instead of the local currency. However, there is no consensus as to the various hypotheses in previous studies about the dollar phenomenon and the cause of currency exchange trading. In [16], the author put together four types of research on the dollarization. First, the study that the dollarization increases when the policy and institutional financial system of the local currency finance is poor. In [17], the author and in [18], the author, and in [19], the author argued that financial markets are not naturally fostered, but are developed through the process of maintaining stable asset values through their own economic growth and continued policy and institutional legislation. In [20], the author and in [19], the author argued that the dollarization is caused by increased risk of default, expected currency devaluation and a lack of dollar liability mismatch information. They argued that market failures were caused by problems between financial industry instability and related legal systems, and the dollarization was also caused by poor policy results and financial market imperfections. Second, the study that devaluation of exchange rates causes capital outflow and aggravates effect on the financial statements of domestic companies. In [21], the author argued that, in the event of a crisis, a significant foreign debt repayment burden reduces aggregate demand and, therefore, further depreciation of the exchange rate occurs. They argued that this would worsen the company financial statements, making borrowing conditions more unfavorable. Third, the study of currency substitution in relation to the foreign currency hedging. If investors experience high inflation in assets held in their local currencies, they will replace their holdings with foreign currencies to maintain future asset values. In [22], the author outlined the theory of trade-in between two currencies. In [23], the author argued in empirical studies that inflation and currency exchange trading are strongly interrelated. In [24], the author argued that holding foreign currencies in emerging countries played a role in hedging macroeconomic risks in emerging country economies. He argued that if investors were concerned about inflation in their currencies, the demand for foreign currency holdings would continue to increase. Fourth, the study that portfolio investment manager will have foreign currency assets at an optimal rate to minimize the risk of their investment portfolios. In [25], the author used the Minimum Variance Portfolio (MVP) method to study with five Latin America countries. They confirmed that assets and liabilities of financial institutions are readjusted in accordance with changes in exchange rates to hedge the inflation. In [26], the author argued that the proportion of dollar assets in MVPs and the dollarization are positive correlations.

2.2. Internationalization of Yuan

In 2008, when the global financial crisis recurred, China proceeded the internationalization of the yuan, mainly in Hong Kong. The yuan was newly incorporated into the SDR international currency in October 2016. The internationalizing of the yuan was a major issue for non-international currency countries. It was because currency internationalization could be an alternative to addressing the exchange rate problems common to emerging Asian countries. However, it is necessary for companies and financial institutions in non-international currency countries that do business with China to understand the characteristics of the new international currency, the renminbi, to manage exchange rate risk. In [27], the author asserted four backgrounds for the internationalization of the yuan. First, currency internationalization brings about the development of the direct financial market, so it has the effect of increasing the efficiency of monetary policy. Second, internationalization of currency has a good function to reduce the difference in interest rates and exchange rates between currencies by enabling arbitrage trading when changes in monetary policy are expected. Third, it argued that currency internationalization could further expand the wealth effect of the currency by enticing investors in the international financial market. Fourth, it argued that currency internationalization could lead to an increase in yuan deposits abroad, thereby promoting the growth of the yuan's international financial market. They summed up the requirements for the internationalization of the yuan as follows: The increase in China's economic weight in the global economy, the convenient exchange function of the yuan, the development of financial market and the liberalization of the Chinese financial market, the development of the offshore yuan financial market and the expansion of the flexibility of the yuan exchange rate. In [28], the author and in [29], the author cited five common factors in currency internationalization: political, military power, economic scale, national financial soundness and financial market competitiveness. In [30], the author argued that the higher the proportion of GDP, which represents the size of the economy, and the higher the proportion of trade in the global economy, the higher the possibility of internationalization of currencies. Trade dependence and inflation, on the other hand, have a negative impact on currency internationalization. They also argued that the greater the share of the stock market, the higher the proportion of foreign bonds issued in the local currency, and the greater the financial opening, the greater the possibility of internationalization of currencies. In [31], the author studied the international demand for yuan as a means of value storage. He used the explanatory variables of existing research to analyze the yuan. He estimated that the demand for the yuan would be 1 to 2 percent of the global currency volume in 2015 and 3 to 8 percent in 2020. However, he said that if China actively liberalizes its capital market and opens up its financial markets, the yuan's demand may rise more than expected as the yuan's trading volume increases further in the international financial market. The previous study of internationalization of currency were summarized in Table 1 below.

Table 1: Pre-study of the internationalization of currency

		Description of variable			
[Dependent variable]		Macroeconomic variables	Currency market variables	Financial market variables	Others
[Requirements to be used as foreign	Chinn & Frankel	GDP portion Inflation rate	Exchange rate volatility	Foreign exchange	time lag

exchange reserves)			depreciation rate	transaction weight	
	Chen et al	GDP portion, trade size portion, inflation rate	Exchange rate volatility		time lag
	Li Daokui et al	GDP portion Inflation rate	Exchange rate volatility depreciation rate		
	Dae Won Oh	GDP portion, trade size portion, inflation rate	Exchange rate volatility	M2 / GDP Openness	
[Currency internationalization requirements]	Seok Hyeon, Sang Heon Lee	GDP portion, Trade portion, Trade dependency, Inflation		Stock Market Openness	Financial openness
[RMB internationalization requirements]	Haihong & Yongding Yu	China's Economy portion in the Global Market	Flexibility of exchange rate Convenient exchange	Financial liberalization Financial Market Development Offshore Market Development	

In [32], the author analyzed the impact of the onshore Chinese yuan (CNY) on the offshore yuan (CNH) market in Hong Kong. According to their analysis, the change in the implied interest rate of the forward exchange rate in China immediately affects the renminbi forward exchange rate market in Hong Kong. In [33], the author analyzed the effect of interest rate volatility transferred from the Chinese market to the Hong Kong offshore renminbi market through a variance decomposition analysis. They argued that volatility was transferred from the Chinese market to the Hong Kong offshore market at an average of 60.9%. However, the effect of transferring volatility from the Hong Kong offshore market to the Chinese local market was insignificant at 13.8%. In [34], the author analyzed the CNH and CNY markets from September 2010 to August 2013. They argued that CNY had a strong influence on CNH at the beginning of the analysis period, but argued that CNH's influence on CNY increased in the later period. Therefore, they argued that CNH Hong Kong's offshore yuan has predictive of CNY's Chinese onshore yuan. In [35], the author studied the difference in price between CNH offshore renminbi and CNY onshore currency. They argued that the yuan liquidity and risk aversion of international financial market investors were the main causes of the price difference between CNY and CNH and incurring arbitrage trading. They argued that an increase in the inflow of yuan from the CNY onshore market to the CNH Hong Kong offshore market narrowed the CNH-CNY price difference, and thus the yuan exchange rate was adjusted based on the available liquidity in the market.

2.3. Analysis of variable

We describe the variables used in the fifteen VECM analysis models. The first variable is the spot exchange rate. The exchange rate represents the credit risk of the currency country viewed in the financial market. In [2], the author argued that the depreciation of the exchange rate is a negative factor for the financial system of the country. In [36], the author argued that the depreciation of the exchange rate is a problem with the repayment of national debt. In [37], the author argued that a sharp depreciation of the exchange rate lowers the value of the assets denominated in the local currency, resulting in a lack of foreign currency liquidity, which increases the occurrence of a default risk. The second variable is the 10-year US treasury interest rate. In [38], the author argued that the 10-year U.S. government bond interest rate is the benchmark for the level of interest expected by investors in international financial markets. In [39], the author argued that the determinants of investment capital flowing into emerging economies are largely attributable to supply factors, such as the level of U.S. treasury bond interest rates and preference for safe assets in the international financial market. It argued that a 50bp drops in interest rates on 10-year U.S. treasury bonds would increase the total inflow of capital into emerging economies by 13 percent. The third variable is credit risk spread. In [40], the author argued that credit risk spreads represent investors' risk aversion and are explained by the spread difference between the 10-year U.S. treasury bond and credit risky bonds. In [41], the author argued that an increase in the hedge propensity of investors in the international financial market increases the credit risk spread, and that the credit risk spread is significant in explaining the EMBI index of emerging economies in 31 countries. Alsaka and ap Gwilym [1] argued that the announcement by the credit rating agency had a significant impact on exchange rate fluctuations. The fourth variable is the VIX index. VIX is Chicago Board Options Exchange Volatility Index and is a variability in S&P100 index option transactions. The index represents the level of hedge propensity among international financial market investors. The VIX index is to the extent that investors feel an event risk, and it is the risk compensation rate expected by investors who expect high returns in the international financial market. In [42], the author argued that the VIX index is an important explanatory variable that determines the rate of return on treasury bonds. In [43], the author argued that it is an important variable that explained the premium on government bonds in emerging countries. In [44], the author argued that VIX is an important variable that explains the CDS premium in Mexico, Turkey, and South Korea. In [38], the author argued that VIX is a factor that explains the premium on short-term government bonds. In [39], the author stated that the VIX index represents a preference for safe assets in the international financial market and that if the VIX index fall halves, total capital inflows from emerging economies would increase by 11%. The fifth variable is TED spread. TED spread is the difference between the 90-day yield on U.S. treasury bonds and the three-month interest rate on Euro-dollar Libor. TED spread represents the average level of risk felt by the counterparty in inter-bank money market transactions in the international financial market. Higher cross-bank counterparty risk burdens are considered to have increased TED spreads and increased hedging tendencies in the international financial market, worsening the funding situation. The change in TED spread implies changes in overall capital liquidity in the international financial market or changes in credit risk levels and bond rates. In [45], the author used the TED spread as a measure of the liquidity level of hedge funds. In [41], the author used TED spread as a measure of the liquidity of government-issued government bonds in the international financial market.

3. Research model

3.1. Research Analysis

In [46], the author analyzed the mid- to long-term relationship of the Turkish exchange rate using the VECM analysis model. He

constructed a nominal exchange rate and market exchange rate model and analyzed it using the VECM analysis method. They argued that the market exchange rate model has better predictability of the future mid-term exchange rate than a random walk with and without drift. In [15], the author also used the VECM analysis methodology to analyze the relationship between the Latin America spot exchange rate and the international financial market variables (the 10-year US treasury interest rate, the credit risk spread between the AAA credit rating and the BBB credit rating, VIX index, TED spread). This study also analyzed the Granger Causality, the impulse response, and the variance decomposition analysis using VECM analysis methodology that studied the mid- to long-term exchange rate relationship. The analysis variables are Korea won onshore (KRW) and offshore won, NDF KRW (KWD1m), Chinese onshore (CNY) and offshore yuan (CNH), Japanese yen (JPY), 10-year US treasury bond interest rate (T10Y), credit risk spread, VIX index, TED spread. The Chinese yuan is analyzed by dividing the Chinese onshore yuan (CNY) market, the Hong Kong offshore yuan (CNH) market, and the Seoul offshore yuan market (herein we refer as "CNK"). The Korean won is divided into the Seoul Foreign Exchange onshore (KRW) Market and the offshore NDF KRW market (KWD1M). The offshore won NDF KRW, has the shortest maturity of one month's forward rate. In [47], the author also analyzed the causal relationship between the spot exchange rate of the onshore won dollar and the offshore NDF KRW exchange rate (the same one-month forward rate). NDF KRW is a forward exchange contract in which only the difference between the forward exchange rate and the spot exchange rate is settled in dollars without having to exchange the principal of the contract at maturity. As of 2018, the Korea won's NDF market is worth about $9.5 billion. The won is non-international, but there is no limit to profit-taking arbitrage transactions in which investors take advantage of the price gap between the Seoul foreign exchange KRW market and the Seoul bond market and the offshore NDF KRW market. Therefore, it is assumed that the won has become substantially and effectively internationalized. The Japanese yen is International currency. Therefore, it analyzes the yen traded in international financial markets as there is no distinction between onshore and offshore markets. A total of fifteen VECM analysis models are used to verify the long-term cointegration relationship between variables and apply the VECM analysis methodology only if the long-term cointegration relationship is valid at statistical confidence level. The data analysis period is from December 1, 2014, when the yuan (CNK) began trading on the Seoul foreign exchange market to November 30 2017. The daily average trading volume of the Seoul offshore yuan (CNK) currency market is $1.5 billion to $2 billion, ranking third to fourth among the global offshore yuan market. There is no restriction on transactions because CNK is the same offshore yuan as the Hong Kong offshore yuan (CNH). It also means that the cross rate between the currencies can also affect the won-dollar KRW exchange rate. Therefore, the data analysis period was set for three years from December 1, 2014 to November 30, 2017 to compare fifteen analysis models.

3.2. Analysis Model

This study first tests the long-term equilibrium relationship between the analysis variables using the cointegration analysis method proposed by [48] the author. In general, time series data are non-stationary and therefore have a unit root. However, if there is a stable linear combination between these variables, it can be considered that there is a cointegration relationship between the variables, so that a long-term, stationary relationship exists. Therefore, regression analysis can be used when a cointegration relationship is recognized even among unstable time series data. When the cointegration relationship is confirmed, the relationship between the variables is analyzed using the VECM model, which includes the error correction term (ECT) in the VAR model. The VECM model can measure the effect of the time lag between endogenous variables and analyze the long-term equilibrium relation between variables as well as the short-term dynamic relations simultaneously. Nevertheless, since the VAR model is not constrained, the results may vary depending on the variable settings. Therefore, this study performed Granger causality analysis in addition. The impulse response analysis was used to analyze how external impacts were transferred between variables. The study also add variance decomposition analysis to analyze the magnitude of the influence of unexpected impacts on variables. It is necessary to verify the cointegration relationship between research variables in order to apply the VECM analysis methodology. Thus, this study conducted Johansson's cointegration analysis. The methods for verifying Johansen's cointegration vector are trace (λ trace) and maximum eigenvalue (λ max) tests. The test statistic is as follows:

$$\lambda_{trace} = -\ T\sum_{i=r+1}^{n} \ln\left(1-\hat{\lambda_i}\right) \quad (1)$$

$$\lambda_{max}(r, r+1) = -T\ln\left(1-\hat{\lambda_{r+1}}\right) \quad (2)$$

here, λ_i indicates the matrix eigenvalue and T is the observed number. The null hypothesis of λ trace is: "there is no cointegration"; if this is rejected, there exist at least r number of cointegration relations, and the variables become stable *I (0)* time series. Only after the cointegration verification is confirmed, the relationship between the endogenous variables is analyzed by applying the VECM analysis methodology. The cointegration test show the long-term equilibrium relationship between variables, but it does not suggest its direction. To overcome this problem, this study analysis the relationship between variables using VECM model. The use of the VECM model has the advantage of being able to identify both short- and long-term causal relationships, as it can identify not only the effects of the differential term of the independent variables on the dependent variables, but also the effect of the variation of the error correction term on the dependent variables. The VECM model is as follows:

$$\Delta[LKRW_t] = [\alpha_1] + \sum_{i=1}^{p} \alpha_i \Delta[LKRW_{t-i}]$$
$$+ \sum_{j=1}^{p} \alpha_j \Delta\left[LKRW1m_{t-j}\right] \ldots + [\theta_1][ECT_{t-1}] + \left[\varepsilon_{1,t}\right] \quad (3)$$

The extension of (3) of two variables into six variables shall be expressed as follow:

$$\begin{bmatrix} LKRW_t \\ LKWD1m_t \\ T10Y_t \\ CREDIT_t \\ LVIX_t \\ TED_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{bmatrix} + \sum_{i=1}^{p} \Delta \begin{bmatrix} \beta_{11i}\beta_{12i}\beta_{13i}\beta_{14i}\beta_{15i}\beta_{16i}\,\beta_{17i} \\ \beta_{21i}\beta_{22i}\beta_{23i}\beta_{24i}\beta_{25i}\beta_{26i}\,\beta_{27i} \\ \beta_{31i}\beta_{32i}\beta_{33i}\beta_{34i}\beta_{35i}\beta_{36i}\,\beta_{37i} \\ \beta_{41i}\beta_{42i}\beta_{43i}\beta_{44i}\beta_{45i}\beta_{46i}\,\beta_{47i} \\ \beta_{51i}\beta_{52i}\beta_{53i}\beta_{54i}\beta_{55i}\beta_{56i}\,\beta_{57i} \\ \beta_{61i}\beta_{62i}\beta_{63i}\beta_{64i}\beta_{65i}\beta_{66i}\,\beta_{67i} \end{bmatrix} \times \begin{bmatrix} LKRW_{t-j} \\ LKWD1m_{t-j} \\ T10Y_{t-j} \\ CREDIT_{t-j} \\ LVIX_{t-j} \\ TED_{t-j} \end{bmatrix} + \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \end{bmatrix} [ECT_{t-1}] + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \\ \varepsilon_{5t} \\ \varepsilon_{6t} \end{bmatrix} \quad (4)$$

where $\triangle$ is the difference coefficient, and $ECT\ t\text{-}1$ is the term describing the long-term equilibrium relation of the estimation coefficient as an error correction term, where $\alpha\ \beta\ \theta$ is the measurement coefficient value, p is the time difference and ε_t is the error term. The Granger causality test verifies the significance of the coefficient values measured in relation to the lagged variable *j* and the variable *i. If "H_{null} : j does not granger cause i*" is rejected, j shall become a Granger cause to i. Impulse response verification measures how quickly information is passed between variables when impacts are delivered to the VECM model with time lag. In other words, this study analyzed how the variables in the model react with time when a certain amount of impact is delivered. A variance decomposition analysis shows the relative importance of variables, as the estimation error of VECM shows the proportion that is explained by its own and other variables. Therefore, if there is no decomposition value of the variable due to the impact of the residual, this variable is considered as exogenous.

4. Analysis data

4.1. data

The analysis variables are summarized in Table 2. The analysis period is from December 1, 2014 to November 30, 2017, and is daily closing data. The closing price is determined by the accumulation of both new information for the day and trading activities between the parties. The data was obtained through Bloomberg or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Table 2: Analysis variables

	Market	Variables
	Global	US Treasury 10 Year
		Credit spread
		VIX index
		TED spread
Korea	onshore	KRW
	offshore	KWD1m
China	onshore	CNY
	offshore	CNH*, CNK**
Japan	Global	JPY

The exchange rate is the spot exchange rate. SPOT of the Chinese yuan, the Japanese yen, and the Korean won, against U.S. dollar. The Chinese yuan is analyzed by dividing it into onshore yuan (CNY) and offshore yuan in Hongkong (CNH). For the Korea won dollar, there is no offshore spot exchange rate. Therefore, the shortest term of one-month term in offshore market was used as a won-dollar offshore exchange rate (KWD1m). The analysis variables for the international financial market are the 10-year U.S. treasury bond interest rate, Credit risk spread, VIX index and TED spread, which are daily closing data. During the analysis period, there were no fundamental shift in the international financial market, such as the 1998 Asian financial market crisis or the 2008 international financial market crisis. Therefore, dummy variables were not considered. The analysis of basic statistics of yen, yuan, and won dollar is summarized in Table 3. The spot exchange rate and the VIX index were calculated as the difference between today's closing price taken on the log value and the previous day's closing price taken on the log value. The 10-year treasury interest rate, credit risk spread and TED spread are the difference between today's and the previous day's closing price of data.

Table 3: Raw data statistic

	onshore market		Global market							
	Korea	China	Korea	China		Japan	global market variable			
Variables	KRW	CNY	KWD1m	CNH	CNK	JPY	T10Y	CREDIT	LVIX	TED
Mean	1141.52	6.55	1142.34	6.56	174.38	114.17	2.09	1.70	14.72	34.53
Median	1136.95	6.58	1137.69	6.58	174.61	113.62	2.17	1.66	13.72	31.38
Maximum	1240.90	6.96	1246.00	6.98	189.53	125.63	2.63	2.30	40.74	68.93
Minimum	1068.60	6.15	1064.95	6.15	162.01	100.03	1.36	1.29	9.14	14.52
Std. Dev.	35.29009	0.24247	34.99482	0.23624	7.00772	6.40875	0.28092	0.23559	4.25084	10.46365
Observations	784	784	784	784	784	784	784	784	784	784
ADF	-2.1636	-1.6404	-2.1636	-1.7592	-1.5864	-1.5845	-2.1857	-0.3129	-4.5982***	-2.2357
PP	-2.2105	-1.6339	-2.2105	-1.7090	-1.3730	-1.5983	-2.0749	-0.6730	-4.4872***	-2.1803

* 10% significance level, ** 5% significance level, *** 1% significance level. KRW, CNY, and JPY are Korean, Chinese and Japanese spot exchange rates. CNH is the Chinese offshore yuan spot exchange rate, and KWD1m is the 1-month NDF KRW exchange rate. T10Y is the US treasury 10-year interest rate (%). CREDIT is the difference (%) of the spread between T10y and BBB credit rating bonds. VIX is the chicago options market index (%). TED is the difference between the libor 3-month interest rate and the T-bill rate (basis point)

Table 4: Stationary test

	onshore market		Global market							
	Korea	China	Korea	China		Japan	global market variable			
	KRW	CNY	KWD1m	CNH	CNK	JPY	T10Y	CREDIT	LVIX	TED
ADF	-2.1636	-1.6404	-2.1636	-1.7592	-1.5864	-1.5845	-2.1857	-0.3129	-4.59***	-2.2357
PP	-2.2105	-1.6339	-2.2105	-1.7090	-1.3730	-1.5983	-2.0749	-0.6730	-4.48***	-2.1803
ADF	-27.09***	-24.50***	-28.91***	-20.61***	-25.75***	-27.44***	-30.12***	-27.41***	-27.71***	-11.79***
PP	-27.07***	-24.43***	-29.03***	-24.56***	-25.73***	-27.44***	-30.13***	-42.64***	-30.09***	-27.92***

4.2. Data

The Stationary of the data was confirmed prior to the analysis. The study used ADF (Augmented Dickey Fuller) and PP (Phillips-Perron) unit root analysis method to verify the data. The analysis result of the yen, the yuan, and the won and international financial market variables are summarized in Table 4.

The first difference of the natural logarithm of spot exchange rate and VIX index data rejected the null hypothesis that "there is a unit root". Therefore, time series data satisfy the stationary condition. The first differenced data of 10-year US treasury interest rate (T10Y), Credit risk spread (CREDIT) and TED spread (TED) showed that the data satisfied the stationary without the unit root. Therefore, this study analyzed by using the log conversion for spot exchange rate, and VIX index data, and by using the level data for the US treasury 10-year interest rate, Credit risk spread, and TED spread.

4.3. Cointegration Test

The optimal time lag of the fifteen VECM model among the aforementioned six variables was determined prior to the cointegration test between the analytic variables. The optimal time lag was selected based on the time differences where AIC and SC values are minimized in the Akaike Information Criteria (AIC) or Schwarz Criteria (SC) analysis. The results of the time lag selection of the VECM (p) model are summarized in Table 6 below. The optimal time lag of the yen and Korea won is the time lag 3. The optimal time lag of the Chinese yuan is time lag 6. In the case of the yuan, the long-term equilibrium cointegration relationship was found to lag behind the yen and the won dollar. The time lag determined for each VECM model was applied to both the cointegration verification and the VECM analysis. The following includes the examination of whether there is at least one cointegration vector, which is a long-term equilibrium relationship between level variables of each time-series data, through cointegration analysis commonly used in financial time series data analysis. Table 5 summarizes the results of the number of the cointegration analysis of fifteen VECM analysis model between the analytic variables.

As summarized in Table 5, the null hypothesis of "no cointegration vector" was rejected at the 5% level for the all fifteen analysis model cases. Therefore, more than one cointegration relation between time series variables used in fifteen analysis model was revealed. The existence of the cointegration indicates that the individual variables are non-stationary, but when they are linked by a cointegration vector relation it indicates the stationary. Therefore, this study analyzed the lead-lag causality between variables using the VECM analysis model that includes the error correction term in the VAR model analysis.

Table 5: Cointegration vector

Currency		market sector (model No)		Cointegration vectors							
						dependent	independent variable				
KRW	on-shore KRW	[CNK] KWD1m, KRW, CNH, CNY, JPY	(1)		constant	LCNK	LKWD1m	LKRW	LCNH	LCNY	LJPY
				Coint EQ (1)	-4.49	1	0	0	-25.97316***	25.92838***	-0.118494
				Coint EQ (2)	-4.62	0	1	0	-26.39582***	25.37661***	-0.099479
				Coint EQ (3)	-4.55	0	0	1	-26.64894***	25.60423***	-0.104134
		[KRW] KWD1m, CNK, CNH, CNY, JPY	(2)		constant	LKRW	LKWD1m	LCNK	LCNH	LCNY	LJPY
				Coint EQ (1)	-4.55	1	0	0	-26.64894***	25.60423***	-0.104134
				Coint EQ (2)	-4.62	0	1	0	-26.39582***	25.37661***	-0.099479
				Coint EQ (3)	-4.49	0	0	1	-25.97316***	25.92838***	-0.118494
		[KRW] KWD1m	(3)		constant	LKRW	LKWD1m				
				Coint EQ (1)	0.11	1	-1.015389***				
		[KRW] KWD1m, T10Y, Credit, VIX, TED	(4)		constant	LKRW	LKWD1M	T10Y	CREDIT	LVIX	TED
				Coint EQ (1)	-8.52	1	0	0.081849	-0.895473***	1.074763***	-0.000894
				Coint EQ (2)	-8.47	0	1	0.079503	-0.872705***	1.044728***	-0.000904
	off-shore KWD1m	[KWD1m] KRW, CNK, CNH, CNY, JPY	(5)		constant	LKWD1m	LKRW	LCNK	LCNH	LCNY	LJPY
				Coint EQ (1)	-4.62	1	0	0	-26.39582***	25.37661***	-0.099479
				Coint EQ (2)	-4.55	0	1	0	-26.64894***	25.60423***	-0.104134
				Coint EQ (3)	-4.49	0	0	1	-25.97316***	25.92838***	-0.118494
		[KWD1m] KRW	(6)		constant	LKWD1m	LKRW				
				Coint EQ (1)	-0.11	1	-0.984844***				
		[KWD1m] KRW, T10Y, Credit, VIX, TED	(7)		constant	LKWD1m	LKRW	T10Y	CREDIT	LVIX	TED
				Coint EQ (1)	-9.44	1	0	0.153340	-1.253755***	1.587332***	-0.000418
				Coint EQ (2)	-9.52	0	1	0.158203	-1.290331***	1.636855***	-0.000389
	on-shore	[CNY] CNH, CNK,	(8)		constant	LCNY	LCNH	LCNK	LKWD1m	LKRW	LJPY
				Coint EQ (1)	-7.33	1	0	0	64.16184***	-63.55239***	0.235192

RMB	CNY	KWD1m, KRW, JPY		Coint EQ (2)	-6.87	0	1	0	61.64652***	-61.09849***	0.229879
				Coint EQ (3)	7.09	0	0	1	-62.4577***	60.8898***	-0.245945
		[CNY] CNH	(9)		constant	LCNY	LCNH				
				Coint EQ (1)	0.06	1	-1.030136***				
		[CNY} CNH, T10Y, Credit, VIX, TED	(10)		constant	LCNY	LCNH	T10Y	CREDIT	LVIX	TED
				Coint EQ (1)	-12.80	1	0	0.773481	-4.477629***	6.285092***	0.005925
				Coint EQ (2)	-12.96	0	1	0.786428	-4.551355***	6.38005***	0.006126
	off-shore CNH	[CNH] CNY, CNK	(11)		constant	LCNH	LCNY	LCNK			
				Coint EQ (1)	0.23	1	-1.004402***	-0.04332***			
		[CNH] CNY, CNK KWD1m, KRW, JPY	(12)		constant	LCNH	LCNY	LCNK	LKWD1M	LKRW	LJPY
				Coint EQ (1)	-6.87	1	0	0	61.64652***	-61.09849***	0.229879
				Coint EQ (2)	-7.33	0	1	0	64.16184***	-63.55239***	0.235192
				Coint EQ (3)	7.09	0	0	1	-62.4577***	60.8898***	-0.245945
		[CNH] CNY	(13)		constant	LCNH	LCNY				
				Coint EQ (1)	-0.06	1	-0.969678***				
		[CNH] CNY, T10Y, Credit VIX, TED	(14)		constant	LCNH	LCNY	T10Y	CREDIT	LVIX	TED
				Coint EQ (1)	-5.27	1	0	0.194119	-1.489189***	2.081043*	-0.000288
				Coint EQ (2)	-5.25	0	1	0.192055	-1.473122***	2.066396*	-0.000370
JPY	[JPY] T10Y, Credit, VIX, TED		(15)		constant	JPY		T10Y	CREDIT	LVIX	TED
				Coint EQ (1)	-2.87	1		-0.22638***	0.524198***	-0.882197***	0.002019

5. Analysis Result

5.1. VECM Analysis

Table 6 shows the relationship between the rate of change of the analytic variables by applying the time lags to the fifteen VECM model using the long-term cointegrating equilibrium equations. The fifteen VECM analysis models' error correction term, ECT values are expressed as Coint EQ (1). In the VECM model analysis, if there are two cointegrations, it is expressed as Coint EQ (1) and (2). In the VECM analysis model, the ECT, error correction term must be negative (-) and statistically significant at the same time. Table 6 analysis results showed that 12 of the total 15 analysis models had negative (-) error correction terms and were statistically significant at the same time. However, in analysis models 5, 6 and 12, the error correction term was negative (-), but it was not significant.

* Korea won dollar and Seoul offshore yuan

We found four onshore won-dollar models, three offshore won-dollar models, and a total of seven VECM analysis models. The following Table 6 shows that all four VECM models of the onshore won-dollar were significant. The long-term equilibrium relationship explanation power of the VECM model was found to be around 50 percent to 68 percent. However, the explanation power of the three offshore won-dollar VECM models was about 3 percent to 6 percent. The error correction term, Coint EQ of the five VECM analysis models was negative (-) and statistically significant. However, analysis models No. 5 and No. 6 showed that the error correction term were negative (-), but were not statistically significant. The VECM Model No. 1, CNK, shows a long-term equilibrium relationship between the won-dollar and the yuan and the yen. Since CNK is an offshore yuan traded in the Seoul foreign exchange market, there are no restrictions on its transactions. Therefore, it has been shown that the Seoul offshore yuan (CNK) has a long-term equilibrium relationship with onshore yuan (CNY) and offshore yuan (CNH) and Japanese yen (JPY) either. These results suggest that although the yuan's liquidity in the Seoul foreign exchange market is relatively insufficient compared to Hong Kong. Therefore, it secures insufficient liquidity through cross-rate transactions between currencies of KRW↔USD↔CNH, CNY, and JPY. The amount of liquidity of a currency used in transactions turned out to be the influence factor of these currencies. It gives impact on the Seoul offshore yuan (CNK) exchange rate either. VECM model No. 2, onshore won-dollar (KRW) exchange rate, indicated that there was a long-term equilibrium relationship with the offshore won-dollar exchange rate, the yuan and the yen. This result implied that the won-dollar exchange rate received impact by the yuan and yen flowing through the export/import trade and inter-bank money markets transactions between Korea, China, and Japan. VECM Model No. 4 is an analysis of the relationship between the onshore won-dollar (KRW) and the variables in the international financial market. The analysis results indicated that the won-dollar exchange rate has a long-term equilibrium relationship with variables in the international financial market. VECM Analysis Models No. 1, No. 2, No. 3 and 4 show that the error correction factor of the offshore won-dollar exchange rate (KWD1m) is greater than that of the other variables. VECM model No. 3 analyzed the relationship between two variables, the onshore won-dollar exchange rate (KRW) and the offshore won-dollar (KWD1m). According to the analysis of the long-term equilibrium relationship between the two variables, the offshore won-dollar exchange rate explains the onshore won-dollar exchange rate about 50%. These results are similar to [47] the author's argument. They studied the Granger causal relationship between the KRW exchange rate and the offshore NDF KRW market. They argued that the offshore one-month forward exchange market was strongly causative in the onshore spot exchange market and that the volatility shock was exchanged between the two markets. In [34] the author analyzed the relationship between offshore yuan (CNH) and onshore yuan (CNY) for the period from September 2010 to August 2013. They argued that the CNH Hong Kong yuan has a predictive power for the CNY yuan in China. The Korean won has not yet been internationalized, but like the yuan, the onshore won-dollar exchange rate has been consistently and strongly influenced by the offshore market won-dollar exchange rate. In addition, the offshore won-dollar exchange rate, KWD1m is found to have a long-term equilibrium relationship with four variables in the international financial market, and as a result, the Korean won-dollar exchange rate (KRW) is affected by variables in the international financial market either. The results suggest that Korea onshore won-dollar exchange market and the Korea won-

Table 6: VECM analysis

	KRW							Yuan							JPY
[dependent variable]	[onshore KRW]				[offshore KWD1m]			[onshore CNY]			[off-shore CNH]				
	[CNK]	[KRW]	[KRW]	[KRW]	[KWD1m]	[KWD1m]	[KWD1m]	[CNY]	[CNY]	[CNY]	[CNH]	[CNH]	[CNH]	[CNH]	[JPY]
variable 1	LKWD1m	LKRW1m	LKWD1m	LKRW1m	LKRW	LKRW	LKRW	LCNH	LCNH	LCNH	LCNY	LCNY	LCNY	LCNY	
variable 2	LKRW	LCNK		T10Y	LCNK		T10Y	LCNK		T10Y	LCNK	LCNK		T10Y	T10Y
variable 3	LCNH	LCNH		CREDIT	LCNH		CREDIT	LKRW1m		CREDIT		LKWD1m		CREDIT	CREDIT
variable 4	LCNY	LCNY		LVIX	LCNY		LVIX	LKRW		LVIX		LKRW		LVIX	LVIX
variable 5	LJPY	LJPY		TED	LJPY		TED	LJPY		TED		LJPY		TED	TED
Coint EQ (1)	-0.43***	-0.584***	-0.60***	-0.617***	*-0.1303*	*-0.0646*	-0.1747**	-0.056**	-0.035*	-0.058**	-0.064**	*-0.0455*	-0.051*	-0.063*	-0.002*
Coint EQ (2)	-0.1237	0.2444***		0.16879**	*0.6307****	*0.6114*	0.6353***	0.1095*		0.05109		*0.0261*		0.065***	
Coint EQ (3)	-0.0416	0.0911**			*0.3452****			-0.191***				*-0.235****			
time lag	3	3	2	3	*3*	*2*	3	3	4	3	4	*3*	6	2	3
adjusted $R^{\wedge 2}$	68.48%	50.94%	49.97%	51.29%	*3.21%*	*0.66%*	6.47%	7.92%	7.34%	8.63%	4.78%	*5.25%*	5.46%	4.93%	9.42%
model number	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)

government bond market and the inter-bank money market are substantially and effectively open, so this increase the link to the international financial market. In [49] the author and in [50] the announcement argued that certain currency shocks are transferred through the interbank money market. It also argued that the impact from the US dollar currency affects the interbank money markets in Hong Kong (HKD), Australia (AUD), Japan (JPY), and Korea (KRW). However, they argued that the impact of the US dollar currency impact on China (Yuan), Thailand (THB) and Malaysia (RHB) is limited. The fact that the Korean won dollar exchange rate is strongly related to the international financial market is in line with [49] the author's argument and with [50] announcement. The VECM model relationships model No. 5 and model No. 6 were not significant. The analysis shows that offshore won-dollar exchange rates are more affected by international financial market factors than by regional factors between Korea, China and Japan in Northeast Asia.

*China Yuan

We found three onshore analysis model and four offshore analysis models, and a total of seven VECM analysis models. The three VECM models of the onshore yuan (CNY) were shown to be statistically significant. The three long-term equilibrium relationship VECM models of the onshore yuan had the explanatory power of between 7% and 8%. The explanatory power of the four VECM models of the Hong Kong offshore yuan (CNH) was between 4% and 5%. Although the error correction term of the six VECM analysis models was significant, analysis model 12 showed that the error correction term was not significant. The VECM model No. 8 Chinese onshore yuan (CNY) was found to have a long-term equilibrium relationship with the onshore won dollar (KRW) and the yen (JPY). The analysis results of VECM Model No. 10 and No. 14 show that onshore and offshore yuan have long-term equilibrium relationship with international financial market variables. The VECM model No. 11 Hong Kong offshore yuan (CNH) has a long-term equilibrium relationship with the Chinese onshore yuan (CNY) and the Korean offshore yuan (CNK). The unusual was that, VECM analysis models No. 9 and 13, show that China's long-term equilibrium relationship between onshore and offshore yuan is weaker than Korea's onshore and offshore won-dollar relations. In the previous model No. 3 Korean Won-dollar VECM analysis model, offshore Won-dollar's explanation power on the onshore KRW was about 50%, at 1% significant level. However, the explanatory power of the VECM No. 9 and No. 13 analytical models analyzed the Chinese yuan was 7.3% and 5.4%, respectively, and it was found to be significant at the 10% level. It is unusual for the long-term equilibrium relationship between onshore and offshore yuan markets to be weaker than the non-international currency of Korea won. This analysis result is interpreted by the fact that the yuan onshore-offshore trade related cross border transactions are liberalized, but capital market transactions are limited. Overall, the yuan's long-term equilibrium relationship with other currencies was weaker than that of Korea won. And these results are in line with [49] the author's argument that the impact of the shock from the US dollar currency on Chinese yuan is limited.

*Japan Yen

VECM Analysis Model No. 15 of Table 6 is the result of an analysis of the long-term equilibrium relationship between the Japanese yen and variables in the international financial market. The analysis indicated that the yen exchange rate is more related to the variables in the international financial market than that of the won and the yuan. This result suggest that the yen is international currency and the Japanese bond markets are fully open to international financial market investors, so, the greater association with the variables of the international financial market than the non-international currency, the Korea won and the yuan, which is on the way of internationalized. Analysis models 1, 2, and 8 show that the onshore won dollar (KRW) and the onshore yuan (CNY) have a long-term equilibrium relationship with the yen (JPY). This result show that since most of the demand and supply of foreign exchange transactions occurred through import-export transactions and were supplied to the foreign exchange markets in Korea and China, so the currency market result in long term equilibrium relationship. However, looking at analysis models Nos. 5, 6, and 12, it was found that the offshore won (KWD1m) and the offshore yuan (CNH) had no long-term equilibrium relationship with the yen (JPY). This result is interpreted as that, in the offshore international financial market, since investors trade currencies in consideration of the country risk of the currency, a long-term equilibrium relationship between the offshore won and the offshore yuan is not established with the international currency, the yen.

5.2. Granger Causality Analysis

The VECM model has no restriction on the analytical model and may cause a fictitious spurious regression analysis problem. As such, this study further conducted a widely used in time series

analysis, Granger causality test. The null hypothesis of Granger causality test is that "the offshore won, KWD1m has no effect on the spot exchange rate of the onshore won, KRW", and the direction of the causal relationship in Table 7 was written as "KWD1m→KRW". If the null hypothesis is rejected, the offshore won shall have an impact on onshore KRW exchange rates. Table 7 summarizes the results of the Granger causality analysis between the exchange rates of Japanese yen, Chinese renminbi, and Korean won, and between the exchange rates of the three currencies of Korea, China, and Japan and variables in the international financial market.

Table 7: Granger Causality analysis of won, yuan, yen and international financial market variables

Currency		market sector (model No)		Granger Causality relation			Time lag (1)	Time lag (2)	Time lag (3)	Time lag (4)	Time lag (5)
							Chi-square	Chi-square	Chi-square	Chi-square	Chi-square
KRW	on-shore KRW	[CNK] KWD1m, KRW, CNH, CNY, JPY	(1)	KWD1m	→	CNK	14.9383	56.2589	214.79	165.636	135.221
				CNH	→	CNK	*na*	*na*	*na*	2.94628**	2.39355**
				CNY	→	CNK	3.33834*	*na*	*na*	2.05321*	*na*
				KRW	→	CNK	*na*	77.2157***	58.6483***	45.4459***	36.4794***
				JPY	→	CNK	6.40453**	3.18475**	3.45293**	3.06304**	2.74506**
				CNK	→	*KWD1m*	*na*	*na*	*na*	*na*	*na*
				KRW	→	*KWD1m*	*na*	*na*	*na*	*na*	*na*
				CNH	→	*KWD1m*	*na*	*na*	*na*	*na*	*na*
				CNY	→	*KWD1m*	*na*	*na*	*na*	*na*	*na*
				JPY	→	*KWD1m*	*na*	*na*	*na*	*na*	*na*
				KWD1m	→	KRW	732.949***	379.449***	254.111***	189.808***	152.381***
				CNK	→	KRW	*na*		2.49953*	2.63884**	2.13671*
				CNH	→	KRW	*na*	16.3348***	13.3646***	10.049***	9.04209***
				CNY	→	KRW	*na*	12.1681***	10.988***	8.23484***	6.57248***
				JPY	→	KRW	*na*	5.78389***	6.56095***	5.00083***	4.69926***
				CNY	→	CNH	4.22086**	6.22332***	3.6091**	2.61292**	2.23464**
				CNK	→	*CNH*	*na*	*na*	*na*	*na*	*na*
				LKWD1m	→	*CNH*	*na*	*na*	*na*	*na*	*na*
				KRW	→	*CNH*	*na*	*na*	*na*	*na*	*na*
				JPY	→	*CNH*	*na*	*na*	*na*	*na*	*na*
				CNK	→	*CNY*	*na*	*na*	*na*	*na*	*na*
				LKWD1m	→	*CNY*	*na*	*na*	*na*	*na*	*na*
				KRW	→	*CNY*	*na*	*na*	*na*	*na*	*na*
				CNH	→	CNY	13.5758***	14.8769***	11.4008**	8.35465***	6.67746***
				JPY	→	*CNY*	*na*	*na*	*na*	*na*	*na*
				KWD1m	→	JPY	2.99698*	*na*	*na*	*na*	*na*
				KRW	→	JPY	3.07157***	*na*	*na*	2.07901*	2.3439**
				CNK	→	*JPY*	*na*	*na*	*na*	*na*	*na*
				CNH	→	*JPY*	*na*	*na*	*na*	*na*	*na*
				CNY	→	*JPY*	*na*	*na*	*na*	*na*	*na*
			(2)	*[KRW] KWD1m, CNK, CNH, CNY, JPY*							
			(3)	*[KRW] KWD1m*							
		[KRW] KWD1m, T10Y, Credit, VIX, TED	(4)	T10Y	→	KRW	*na*	*na*	4.50778***	3.50753***	3.53952***
				Credit	→	KRW	3.33958**	*na*	*na*	*na*	*na*
				VIX	→	KRW	5.16845**	4.82868***	5.68474***	4.39125***	3.57154***
				T10Y	→	KWD1m	*na*	2.86013*	2.92025**	3.17319**	3.53952***
				Credit	→	KWD1m	4.01793**	*na*	*na*	*na*	*na*
				TED	→	KWD1m	*na*	4.19474**	2.73643**	3.00317**	2.4705**
				KWD1m	→	T10Y	*na*	3.16114**	2.02383*	2.48292**	1.8993*
	off-shore KWD1m		(5)	*[KWD1m] KRW, CNK, CNH, CNY, JPY*							
			(6)	*[KWD1m] KRW*							
			(7)	*[KWD1M] KRW, T10Y, Credit, VIX, TED*							
RMB	on-shore CNY		(8)	*[CNY] CNH, CNK, KWD1m, KRW, JPY*							
			(9)	*[CNY] CNH*							
		[CNY] CNH, T10Y, Credit, VIX, TED	(10)	TED	→	CNY	3.57194**	3.87336**	2.81656**	2.17497*	1.95067*
				CNY	→	Credit	3.76118*	2.47433*	*na*	*na*	*na*
				CNY	→	VIX	5.01105**	3.03027**	*na*	*na*	*na*
				CNY	→	TED	*na*	*na*	*na*	*na*	1.86105*
	off-shore CNH		(11)	*[CNH] CNY, CNK,*							
			(12)	*[CNH] CNY, CNK, KWD1m, KRW, JPY*							
			(13)	*[CNH] CNY*							
		[CNH] CNY, T10Y, Credit VIX, TED	(14)	TED	→	CNH	*na*	3.64619**	2.49048*	*na*	*na*
				CNH	→	Credit	3.52004*	3.13441**	3.03439**	2.40199**	2.44063**
				CNH	→	VIX	3.4716*	2.36669*	*na*	*na*	*na*
JPY	[JPY] T10Y, Credit, VIX, TED		(15)	T10Y	→	JPY	*na*	6.90707***	5.24848**	4.46118***	3.72127***
				Credit	→	JPY	*na*	14.591***	13.7139***	10.8005***	8.83545***
				VIX	→	JPY	*na*	3.52556**	2.28807*	*na*	*na*
				TED	→	JPY	3.00953*	6.51429***	4.74023***	4.46049***	4.40631***
				JPY	→	T10Y	*na*	*na*	*na*	*na*	*na*

	JPY	→	Credit	7.08832***	10.0327***	6.64887***	4.78381***	4.11865***
	JPY	→	VIX	6.24512**	4.23995**	2.92458**	2.29145*	1.9398*
	JPY	→	TED	*na*	3.43594**	2.43067*	*na*	2.01376*

* (KWD1m → KRW) the direction of the arrow is the direction of the granger causal relationship. 'na' is no the granger causal relationship does not exist at the significance level. The Granger causal relationship analysis results of analysis models 2, 3, 5 to 9, 11 to 13 are overlapped with model 1

*Korea won dollar and Seoul offshore yuan

Models No. 1 through No. 4 in Table 7 summarizes the variables affecting the Granger causal relationship between the Korean won (KRW) and the Seoul offshore yuan (CNH). The causality analysis among the variables showed that offshore won (KWD1m) was the biggest cause-and-effectiveness for Korean onshore won (KRW). After KWD1m, the onshore won exchange rate was shown to be causal in the order of Hong Kong's offshore yuan (CNH), China's onshore yuan (CNY) and Japan's yen (JPY) at 1% significant level. These results suggest that the offshore won-dollar exchange rate had the greatest impact on the onshore won-dollar exchange rate. And it is in line with the findings of [34] the author who claimed that the influence of the Chinese offshore yuan on the onshore yuan has increased. The unusual was that the Chinese yuan causes more on the onshore won dollar exchange rate than the yen. These results are interpreted as reflecting the largest export-import trade volume of Korea with China. Model 4 analyzed the Granger causal relationship between the won dollar and international financial market variables. The analysis results show that 10-year US treasury bonds, Credit risk spread, and VIX index causes Korean onshore won dollars (KRW) at 1%, 5%, and 1% significance levels, respectively. Models No. 5 through 7 analyzed and summarized the variables affecting causality in offshore won dollars. According to the Granger causal analysis among the variables, the offshore won-dollar exchange rate (KWD1m) has no causal relationship with the onshore dollar, the Hong Kong Overseas Yuan (CNH), the Chinese onshore yuan (CNY) and yen (JPY). Similar to this analysis, the analysis results of VECM model No. 5 showed that offshore won has no long-term equilibrium with the yuan and yen. In [51] the author argued that the depreciation of the exchange rate is an increase in the sovereign credit risk. The analysis results of this paper and in [51] the author reconfirm that when international financial market investors trade offshore won (KWD1m), they trade based on the expected rate of return versus risk considering the country-specific factor only. In addition, offshore dollars are freely traded in overseas markets without being restricted by Korea foreign exchange transaction regulations, and the difference in price at maturity is settled in US dollars. Therefore, the absence of a liquidity problem in offshore won have affected the analysis results. This analysis result is reconfirmed in the Granger causality relationship analysis result of the model 7 variables. International financial market variables such as US 10-year treasury bonds, Credit risk spread, and TED spread were found to cause offshore won dollars (KWD1m) at 1%, 5%, and 5% significant level, respectively. The results of the Granger causality relationship analysis of Models 5 and 7 show that the offshore won-dollar exchange rate has no causal relationship with the yuan and yen but has a causal relationship with international financial market variables. These results provide policy implications that the analysis of future won-dollar exchange rate fluctuations should be approached from the perspective of global investment portfolio investor in international market.

*China Yuan

Models 8 to 10 in Table 7 summarize variables that have a Granger causal effect on the Chinese onshore yuan (CNY). Since the variables used in the Granger causal relationship analysis are the same as those of models 1 to 4, we need to refer the analysis result of model 1. The analysis results of this study showed that the Hong Kong offshore yuan, CNH was causal to the Chinese onshore yuan (CNY) at a significance level of 1%. However, it was found that the Seoul offshore yuan (CNK), the offshore won dollar (KWD1m), the onshore won dollar (KRW), and the Japanese yen (JPY) do not cause the Chinese onshore yuan (CNY). Model No. 10, the analysis of Granger causality between the yuan and the variables in the international financial market showed that only the TED spread Granger causal the yuan at a 5% significance level. However, 10-year US treasury interest rate, the Credit risk spread, the VIX index did not cause onshore yuan. Models 11-14 summarize the variables that have a Granger causal effect on the Hong Kong offshore yuan (CNH). Since the variables used in the Granger causal relationship analysis are the same as those of models 1 to 4, we refer to the ones summarized in model 1. According to the results of Granger causal relationship analysis between variables, it was found that the Chinese onshore yuan (CNY) caused the Hong Kong offshore yuan (CNH) at a 1% significance level. In the model 14, the analysis of the Granger causal relationship between the offshore yuan and the variables of the international financial market variables, it was found that only TED spread caused Hong Kong offshore yuan (CNH) at a 5% significance level. However, similar to the onshore yuan (CNY), the 10-year US treasury interest rate, Credit risk spread, VIX index, did not appear to be Granger causal to the Hong Kong offshore yuan (CNH). These results show that since dollar liquidity required for currency exchange transactions is secured through the interbank money market, the TED spread, which represents the level of liquidity in the interbank money market, has an impact. In the case of Korean won, the offshore won dollar market caused onshore won, but not in the opposite case. In the case of the yuan, however, it was found that the onshore-offshore yuan exchanged its causal relationship. In [35] the author argued that yuan liquidity and risk aversion of investors in international financial markets are the main reasons for creating price differences and generating profit-taking arbitrage transaction between CNY and CNH. They argued that increased inflow of yuan from the onshore CNY market to the Hong Kong CNH offshore market would reduce the difference in CNH-CNY prices, so that the yuan exchange rate would be adjusted based on the onshore CNY, which has more liquidity. According to [50] the statement, China began supplying yuan liquidity to Hong Kong's offshore markets in cooperation with HKMA on November 10, 2014. Therefore, the onshore yuan's Granger causal of offshore yuan is showed. Offshore yuan have been influenced by the ample liquidity of the onshore yuan market. In [34] the author analyzed CNH and CNY markets for the period from September 2010 to August 2013. They argued that CNH's influence on CNY has increased, and argued that CNH Hong Kong offshore yuan has

predictive power over CNY China onshore yuan. The onshore yuan market is limited to investors in the international financial market, while the Hong Kong currency market has no trading limitation. Therefore, as argued by [34] the author, it has been shown that the price-discovery effect of the Hong Kong offshore yuan market may have affected the onshore yuan market. The yuan (CNY and CNH), the SDR currency, is the same yuan, but the onshore yuan and the offshore yuan exchange each other's independent causal relationship and each has a distinct characteristic. These analysis results provide important implications for investment and risk-hedge transactions for companies that use the yuan to trade in import-exports and for investors in the international financial market.

*Japan Yen

Model 15 of Table 7 summarizes the Granger causal relationship of international financial market variables to the Japanese yen. The analysis results showed that all four variables (the 10-year US treasury bond, credit risk spread, VIX index, and TED spread) were responsible for the change in the yen's exchange rate at 1%, 1%, 5%, and 1%, significant level respectively. The aforementioned analysis of the Korean won dollar showed that three variables in the international financial market affect the won-dollar exchange rate. In yuan case, only TED spread affects the yuan's exchange rate. It was expected that the yen, the international currency, would be the strongest in the strength of Granger causality relationship among the variables in the international financial market, Chinese the yuan that is in the process of internationalization, would be the next and the won-dollar, the non-international currency would be the weakest. However, the analysis result showed that the non-international currency, the won, has a stronger Granger causal relationship with variables in the international financial market than that of Chinese the yuan. In [14] the author analyzed the effects of international financial market variables and currency exchange rates on the country CDS premium. They argued that the Granger causal relationship of the international financial market variables and the won-dollar exchange rate that affect the Korea country CDS premium is stronger than that of the China country CDS premium. Therefore, when a shock occurs in the international financial market, the shock could be dispersed and absorbed into the Korea CDS premium and the onshore and offshore won-dollar currency market. However, for China the shock could be concentrated in the CDS premium only. The analysis results that the Korean won dollar has a stronger Granger causal relationship with international financial market variables than the Chinese yuan is in line with their arguments.

5.3. Impulse Response Analysis

The Granger causality test informs the direction of the causal relationship between variables. However, the Granger causality test cannot determine whether the direction of the causal relationship is positive (+) or negative (-). Therefore, the impulse response analysis was performed to measure the impact of a specific variable on other variables. The Impulse response results are summarized in Table 8 for the results of the 1 to 10th and 20th, and 30th days.

Table 8: Impulse response analysis of won, yuan, yen and international financial market variables

Curr-ency		model number		Response of dependents variable to one standard deviation of innovations						
				days	depen-dent	independent variables				
KRW	on-shore KRW	[CNK] KWD 1m, KRW CNH CNY JPY	(1)		LCNK	LKRW1m	LKRW	LCNH	LCNY	LJPY
				1 day	0.00261	0.00000	0.00000	0.00000	0.00000	0.00000
				2 day	0.00110	0.00163	0.00062	(0.00087)	(0.00004)	(0.00002)
				3 day	0.00074	0.00412	0.00031	(0.00175)	(0.00052)	(0.00010)
				4 day	0.00034	0.00399	0.00024	(0.00207)	(0.00077)	0.00039
				5 day	0.00020	0.00400	0.00022	(0.00196)	(0.00055)	0.00062
				6 day	0.00002	0.00399	0.00054	(0.00184)	(0.00067)	0.00072
				7 day	(0.00005)	0.00394	0.00081	(0.00176)	(0.00094)	0.00073
				8 day	(0.00011)	0.00399	0.00087	(0.00161)	(0.00107)	0.00071
				9 day	(0.00015)	0.00408	0.00084	(0.00147)	(0.00117)	0.00068
				10 day	(0.00017)	0.00414	0.00080	(0.00138)	(0.00127)	0.00067
				20 day	0.00014	0.00434	0.00060	(0.00061)	(0.00182)	0.00056
				30 day	0.00033	0.00437	0.00051	(0.00032)	(0.00201)	0.00049
		[KRW] KWD 1m, CNK CNH CNY JPY	(2)		LKRW	LKRW1m	LCNK	LCNH	LCNY	LJPY
				1 day	0.00372	0.00000	0.00000	0.00000	0.00000	0.00000
				2 day	0.00236	0.00292	0.00006	0.00003	(0.00014)	(0.00007)
				3 day	0.00201	0.00468	0.00009	(0.00001)	(0.00003)	0.00009
				4 day	0.00195	0.00471	(0.00031)	(0.00010)	(0.00001)	0.00027
				5 day	0.00213	0.00479	(0.00022)	(0.00013)	(0.00004)	0.00053
				6 day	0.00227	0.00455	(0.00029)	0.00004	(0.00014)	0.00064
				7 day	0.00240	0.00453	(0.00032)	0.00022	(0.00027)	0.00059
				8 day	0.00245	0.00457	(0.00036)	0.00035	(0.00042)	0.00055
				9 day	0.00244	0.00461	(0.00034)	0.00047	(0.00052)	0.00052
				10 day	0.00241	0.00466	(0.00031)	0.00057	(0.00059)	0.00050
				20 day	0.00236	0.00481	0.00007	0.00125	(0.00106)	0.00038
				30 day	0.00233	0.00484	0.00025	0.00149	(0.00121)	0.00032
		[KRW] KWD 1m	(3)		LKRW	LKRW1m				
				1 day	0.00372	0.00000				
				2 day	0.00235	0.00292				
				3 day	0.00203	0.00474				
				4 day	0.00197	0.00485				
				5 day	0.00194	0.00482				
				6 day	0.00194	0.00481				
				7 day	0.00195	0.00481				

Curr-ency		model number		Response of dependents variable to one standard deviation of innovations						
				days	depen-dent	independent variables				
RMB	on-shore CNY	[CNY] CNH CNK KWD 1m, KRW JPY	(8)		LCNY	LCNH	LCNK	LKRW1M	LKRW	LJPY
				1 day	0.00182	0.00000	0.00000	0.00000	0.00000	0.00000
				2 day	0.00210	0.00038	(0.00006)	(0.00005)	0.00000	(0.00004)
				3 day	0.00195	0.00044	(0.00003)	(0.00008)	0.00007	(0.00003)
				4 day	0.00206	0.00041	(0.00006)	(0.00011)	0.00014	(0.00004)
				5 day	0.00209	0.00043	(0.00004)	(0.00009)	0.00011	(0.00005)
				6 day	0.00206	0.00045	(0.00004)	(0.00009)	0.00008	(0.00005)
				7 day	0.00208	0.00046	(0.00004)	(0.00009)	0.00008	(0.00005)
				8 day	0.00209	0.00048	(0.00004)	(0.00008)	0.00008	(0.00005)
				9 day	0.00209	0.00049	(0.00004)	(0.00008)	0.00007	(0.00005)
				10 day	0.00209	0.00050	(0.00003)	(0.00009)	0.00007	(0.00005)
				20 day	0.00211	0.00056	(0.00001)	(0.00011)	0.00006	(0.00006)
				30 day	0.00212	0.00058	0.00001	(0.00011)	0.00005	(0.00007)
		[CNY] CNH	(9)		LCNY	LCNH				
				1 day	0.00182	0.00000				
				2 day	0.00209	0.00038				
				3 day	0.00194	0.00044				
				4 day	0.00205	0.00041				
				5 day	0.00209	0.00044				
				6 day	0.00208	0.00045				
				7 day	0.00208	0.00046				
				8 day	0.00209	0.00048				
				9 day	0.00210	0.00050				
				10 day	0.00210	0.00051				
				20 day	0.00212	0.00061				
				30 day	0.00213	0.00065				
		[CNY] CNH T10Y Credit VIX TED	(10)		LCNY	LCNH	T10Y	CREDIT	LVIX	TED
				1 day	0.00182	0.00000	0.00000	0.00000	0.00000	0.00000
				2 day	0.00208	0.00038	0.00009	(0.00005)	0.00005	(0.00011)
				3 day	0.00193	0.00045	0.00012	0.00001	(0.00001)	(0.00017)
				4 day	0.00204	0.00043	0.00010	0.00000	(0.00001)	(0.00019)
				5 day	0.00206	0.00047	0.00011	0.00000	(0.00001)	(0.00015)
				6 day	0.00203	0.00051	0.00013	0.00000	(0.00003)	(0.00014)
				7 day	0.00205	0.00054	0.00013	0.00000	(0.00006)	(0.00014)

				8 day	0.00194	0.00481				
				9 day	0.00194	0.00482				
				10 day	0.00194	0.00482				
				20 day	0.00194	0.00482				
				30 day	0.00194	0.00482				
		[KRW] KWD1m T10Y Credit VIX TED	(4)		LKRW	LKRW1m	T10Y	CREDIT	LVIX	TED
				1 day	0.00371	0.00000	0.00000	0.00000	0.00000	0.00000
				2 day	0.00233	0.00284	(0.00000)	0.00003	(0.00004)	(0.00030)
				3 day	0.00201	0.00461	0.00055	0.00036	0.00037	(0.00049)
				4 day	0.00192	0.00460	0.00073	0.00052	0.00047	(0.00040)
				5 day	0.00203	0.00460	0.00112	0.00070	0.00038	(0.00071)
				6 day	0.00215	0.00446	0.00136	0.00063	0.00037	(0.00090)
				7 day	0.00216	0.00431	0.00137	0.00055	0.00032	(0.00093)
				8 day	0.00217	0.00428	0.00138	0.00054	0.00024	(0.00089)
				9 day	0.00217	0.00423	0.00137	0.00053	0.00016	(0.00086)
				10 day	0.00216	0.00418	0.00136	0.00052	0.00008	(0.00084)
				20 day	0.00212	0.00411	0.00142	0.00048	(0.00026)	(0.00083)
				30 day	0.00211	0.00409	0.00144	0.00046	(0.00036)	(0.00083)
	off-shore KWD1m		(5)	na	LKWD1m	LKRW	LCNK	LCNH	LCNY	LJPY
			(6)	na	LKWD1m	LKRW				
		[KWD1m] KRW T10Y Credit VIX TED	(7)		LKWD1m	LKRW	T10Y	CREDIT	LVIX	TED
				1 day	0.00584	0.00000	0.00000	0.00000	0.00000	0.00000
				2 day	0.00546	0.00003	0.00062	0.00037	0.00048	(0.00062)
				3 day	0.00509	0.00017	0.00101	0.00081	0.00056	(0.00062)
				4 day	0.00505	0.00047	0.00134	0.00082	0.00057	(0.00092)
				5 day	0.00473	0.00061	0.00141	0.00056	0.00054	(0.00091)
				6 day	0.00462	0.00067	0.00134	0.00058	0.00044	(0.00085)
				7 day	0.00460	0.00067	0.00131	0.00056	0.00033	(0.00079)
				8 day	0.00454	0.00067	0.00128	0.00055	0.00023	(0.00076)
				9 day	0.00452	0.00067	0.00129	0.00054	0.00016	(0.00074)
				10 day	0.00451	0.00067	0.00130	0.00054	0.00009	(0.00075)
				20 day	0.00445	0.00065	0.00137	0.00050	(0.00024)	(0.00075)
				30 day	0.00442	0.00065	0.00140	0.00049	(0.00034)	(0.00075)
JPY		[JPY] T10Y Credit VIX TED	(15)		JPY		T10Y	CREDIT	LVIX	TED
				1 day	0.00597		0.00000	0.00000	0.00000	0.00000
				2 day	0.00593		0.00087	(0.00131)	(0.00036)	(0.00071)
				3 day	0.00583		0.00115	(0.00156)	(0.00027)	(0.00098)
				4 day	0.00589		0.00137	(0.00139)	(0.00011)	(0.00136)
				5 day	0.00580		0.00124	(0.00143)	(0.00003)	(0.00134)
				6 day	0.00578		0.00121	(0.00144)	0.00009	(0.00130)
				7 day	0.00575		0.00124	(0.00141)	0.00021	(0.00127)
				8 day	0.00571		0.00122	(0.00140)	0.00032	(0.00128)
				9 day	0.00568		0.00122	(0.00138)	0.00042	(0.00129)
				10 day	0.00566		0.00122	(0.00136)	0.00050	(0.00130)
				20 day	0.00552		0.00121	(0.00129)	0.00093	(0.00131)
				30 day	0.00549		0.00121	(0.00127)	0.00104	(0.00131)

				8 day	0.00206	0.00057	0.00013	0.00000	(0.00007)	(0.00014)
				9 day	0.00206	0.00060	0.00014	0.00000	(0.00009)	(0.00014)
				10 day	0.00207	0.00062	0.00014	0.00000	(0.00010)	(0.00013)
				20 day	0.00210	0.00073	0.00017	(0.00000)	(0.00020)	(0.00012)
				30 day	0.00210	0.00075	0.00017	(0.00001)	(0.00023)	(0.00011)
		[CNH] CNY CNK	(11)		LCNH	LCNY	LCNK			
				1 day	0.00261	0.00000	0.00000			
				2 day	0.00293	0.00025	(0.00008)			
				3 day	0.00269	0.00032	(0.00014)			
				4 day	0.00248	0.00040	(0.00004)			
				5 day	0.00244	0.00049	(0.00016)			
				6 day	0.00247	0.00049	(0.00017)			
				7 day	0.00245	0.00050	(0.00016)			
				8 day	0.00243	0.00055	(0.00013)			
				9 day	0.00240	0.00056	(0.00011)			
				10 day	0.00237	0.00059	(0.00011)			
				20 day	0.00219	0.00074	(0.00004)			
				30 day	0.00213	0.00079	(0.00001)			
off-shore CNH			(12)	na	LCNH	LCNY	LCNK	LKRW1M	LKRW	LJPY
		[CNH] CNY	(13)		LCNH	LCNY				
				1 day	0.00260	0.00000				
				2 day	0.00291	0.00028				
				3 day	0.00266	0.00035				
				4 day	0.00245	0.00046				
				5 day	0.00242	0.00059				
				6 day	0.00235	0.00080				
				7 day	0.00235	0.00069				
				8 day	0.00237	0.00062				
				9 day	0.00241	0.00064				
				10 day	0.00240	0.00064				
				20 day	0.00231	0.00070				
				30 day	0.00227	0.00072				
		[CNH] CNY T10Y Credit VIX TED	(14)		LCNH	LCNY	T10Y	CREDIT	LVIX	TED
				1 day	0.00260	0.00000	0.00000	0.00000	0.00000	0.00000
				2 day	0.00293	0.00026	0.00002	(0.00001)	0.00006	(0.00019)
				3 day	0.00268	0.00028	0.00004	0.00012	0.00003	(0.00025)
				4 day	0.00255	0.00027	0.00007	0.00009	(0.00003)	(0.00025)
				5 day	0.00251	0.00033	0.00008	0.00007	(0.00007)	(0.00024)
				6 day	0.00248	0.00037	0.00008	0.00008	(0.00010)	(0.00024)
				7 day	0.00244	0.00040	0.00008	0.00008	(0.00012)	(0.00024)
				8 day	0.00240	0.00044	0.00009	0.00008	(0.00014)	(0.00024)
				9 day	0.00236	0.00047	0.00009	0.00008	(0.00016)	(0.00024)
				10 day	0.00233	0.00049	0.00009	0.00008	(0.00017)	(0.00025)
				20 day	0.00211	0.00067	0.00010	0.00008	(0.00021)	(0.00026)
				30 day	0.00203	0.00072	0.00010	0.00008	(0.00020)	(0.00026)

* a negative number in parentheses

*Korea won dollar and Seoul offshore yuan

Table 8 summarizes the impulse response results of 15 analysis models. The analysis results of this study showed that Model No. 1, the Seoul offshore yuan (CNK) reacts positively to the shock from the offshore dollar. It responds positively to shocks from the onshore dollar and yen. However, it reacts negatively to the shock of the yuan offshore and onshore in China. It was found that the Seoul offshore yuan currency (CNK) is most strongly affected by the impulse shock from the offshore won-dollars (KWD1m). It was found that the influence of the offshore won dollar on the Seoul offshore yuan (CNK) was greater than that of the CNH, and CNY. These results are found to be due to the fact that the lack of yuan liquidity in the Seoul offshore yuan market compared to the Hong Kong and Chinese yuan markets. It was found that the Seoul offshore yuan (CNK) market financed insufficient yuan liquidity through offshore won dollar cross rate arbitrage transaction. In addition, the impact from onshore won (KRW) and offshore won (KWD1m), Chinese onshore yuan (CNY), Hong Kong offshore yuan (CNH), and yen (JPY) to the Seoul offshore yuan (CNK) was mostly reflected within 1 to 6 time lag. This result is interpreted as the Seoul offshore yuan (CNK) exchange rate being adjusted over time through arbitrage trading between the foreign exchange markets. Model No. 2 onshore won dollar (KRW) was found to react positively to the shock from offshore won dollars. It reacts positively to shocks from the Hong Kong offshore yuan and yen. However, it reacts negatively from the shock of the Chinese onshore yuan (CNY). The onshore won dollar (KRW) has a positive response to the shock from the offshore won dollar (KWD1m) and the yen and remains constant. The onshore won dollar showed adjustment response to the impact of the offshore yuan, but a constant negative (-) response to the impact from the onshore yuan. The result is interpreted as the fact that it takes time for the onshore yuan to be delivered to the Seoul offshore yuan market through import-export transactions and limited capital market transactions. This can also be seen in the VECM analysis model No. 9 and No. 13 in Table 6 which has optimal analysis time lag was 4 and 6 days. It is also unusual for the onshore won-dollar, KRW to have different impulse reactions of the onshore yuan and offshore yuan. The onshore won-dollar showed an adjustment reaction that reverse the positive sign to negative to the impact from the Hong Kong offshore yuan (CNH). These results are interpreted that since onshore won-dollar markets and Hong Kong offshore yuan markets have no restrictions on foreign exchange trading in general, so the two markets adjust to the external shocks with cross rate arbitrage transactions. However, onshore won-dollar continuously responded negatively to the shock from the Chinese onshore yuan. The reason for such results can be interpreted that because foreign exchange transactions

between the Seoul and China are only export-import related transactions are liberalized but capital market transactions are still limited, so the adjustment responses to the shocks from the outside are different. In [49] the author argued that the impact of the shock from the US dollar currency on Chinese yuan was limited. The onshore won-dollar's negative reaction to the shock of the onshore yuan is in the same line as his argument. However, the direction of onshore and offshore yuan shock transfer to the onshore Korean won is partly opposite. This is interesting in that the impact of the yuan on the onshore won-dollar exchange rate is partially offset by offshore yuan. And these results can be interpreted that the different characteristics of the onshore and the offshore yuan contributed in part to the stabilization of the onshore won-dollar exchange rate. Models 4 and 7 onshore and offshore won dollars responded almost identically to shocks from international financial market variables. The onshore and offshore won dollar responded positively to the impact of the 10-year US treasury bond and credit risk spread. In the VIX index shock, it reacted to positive (+) and then converted to negative (-), and the shock disappeared. It responded continuously negative to the impact from the TED spread. Overall, the Korean onshore and offshore won dollar and the Japanese yen had similar shock reactions to the impact from the international financial market. However, the onshore and offshore yuan showed relatively weaker responses to shocks from variables in the international financial market than that of the won and yen. In [39] the research report argues that the determinants of investment capital inflow to emerging countries are largely influenced by factors on the supplier side, such as the level of US treasury bond rates and preference for safe assets in the international financial market. They argued that a 50bp fall in the 10-year US treasury bond yield would increase the total capital inflow to emerging countries by 13%. Korean onshore and offshore won dollars and Chinese onshore and offshore yuan responded positively to the shock arising from 10-year US treasury bonds. Rising risk free interest rates on 10-year U.S. treasury bonds increase demand for the U.S. dollar and lead to capital outflows from emerging markets, which in turn could weaken emerging economies' currencies. This study also confirmed the results consistent with the IMF research. In [39] the research report argued that the VIX index represents the preference for safe assets in international financial markets, and that if the VIX index falls in half, the total capital inflow of emerging countries increases by 11%. However, the onshore and offshore won dollars were slightly different from the IMF argument. The impact from the VIX index was weaker than US 10-year treasury and credit spread. The reason for such results can be interpreted that, because the impact from the VIX index can be adjusted through the NDF KWD1m transaction and the offshore won dollar (KWD1m) readjusts the onshore won dollar (KRW). This result is also explained as the volume of offshore won-dollar, NDF transactions in the international financial market accounts for about 20% of the total global NDF transactions. The won dollar, the yuan, and the yen all reacted in the same negative direction to the shock from the TED spread. Changes in the TED spread implies changes in the overall liquidity level in the international financial market, or changes in credit risk levels and bond yields. In [45] the author used the TED spread as a measure of hedge funds' liquidity level. In [49] the author argued that the impact of a particular currency is spill over through the interbank money market. As he claimed, this study also reconfirmed that the shock generated by TED spread, a variable in the international financial market, was similarly transmitted to Korea won, China yuan and Japan yen. The reason for such results can be interpreted that because banks in Korea, Japan, and China, which are counterparties in money market transactions, do not have a large difference in financing capacity in the international interbank market, and therefore all respond similarly to the shock from the TED Spread change.

*China Yuan

Table 8 summarizes the analysis results of Model No. 8, Chinese onshore yuan (CNY). CNY was found to react negatively to shocks from offshore won dollars (KWD1m), Seoul offshore yuan (CNK) and yen (JPY). On the other hand, Chinese onshore yuan responded positively to shocks from Hong Kong offshore yuan and onshore won dollar. The Chinese onshore and offshore yuan reacts positive (+) to the shocks generated from each other because the two markets are closely interconnected. The Chinese onshore yuan (CNY) reacted positively to the shock from the Korean onshore won (KRW). This phenomenon is attributed to the fact that the 250 billion dollars import-export transactions volume between South Korea and China. Analysis of models 10 and 14, onshore and offshore yuan, and international financial market variables, showed that the onshore and offshore yuan responded similarly to shocks from international financial market variables. The unusual was that the yuan is the SDR international currency, but the response to the impact from international financial market variables was relatively weak compared to the non-international currencies Korean won and international currency Japanese yen. This phenomenon once again support [49] the author's argument that the impact of the US dollar currency on Chinese yuan is limited.

*Japan Yen

The analysis result of model No. 15, yen, JPY summarized in Table 8 shows that JPY consistently responds to the shocks from international financial markets. The yen responded positively to shocks from 10-year US treasury bonds and credit risk spreads, and negatively (-) to shocks from the VIX index and TED spreads.

5.4. Variance Decomposition Analysis

Variance decomposition analysis is to measure the variance of the prediction error of the VECM analysis model and divide the weight for each variable. Table 9 summarized analysis result.

Table 9: Variance decomposition analysis of won, yuan, yen and international financial market variables

Curr-ency	model number		Variance decompositions of dependent variables								
			days	S. E	depen-dent	independent variables					
on-sho	[CNK]	(1)	10 day	S. E	LCNK	LKRW1m	LKRW	LCNH	LCNY	LJPY	
				0.0134	4.8881	73.8031	2.0152	13.946	3.6724	1.6745	

Curr-ency	model number		Variance decompositions of dependent variables								
			days	S. E	depen-dent	independent variables					
on-sho	[CNY]	(8)	10 day	S. E	LCNY	LCNH	LCNK	LKRW1m	LKRW	LJPY	
				0.0066	95.4783	4.1412	0.0407	0.1492	0.1486	0.0420	

KRW	re KRW	[KRW]	(2)	20 day	0.0202	2.1805	77.7151	2.0285	8.3040	8.1100	1.6619
				30 day	0.0254	1.4867	78.7531	1.7485	5.5444	11.006	1.4613
					S. E	LKRW	LKWD1m	LCNK	LCNH	LCNY	LJPY
				10 day	0.0157	24.6279	73.6158	0.2814	0.3046	0.3747	0.7956
				20 day	0.0234	21.2813	74.2448	0.1611	1.9899	1.6258	0.6969
				30 day	0.0295	19.7039	73.5180	0.1423	3.5161	2.5471	0.5725
		[KRW]	(3)		S. E	LKRW	LKWD1m				
				10 day	0.0156	20.5450	79.4550				
				20 day	0.0227	17.1122	82.8878				
				30 day	0.0280	16.0438	83.9562				
		[KRW]	(4)		S. E	LKRW	LKRW1m	T10Y	CREDIT	LVIX	TED
				10 day	0.0154	22.8658	68.8931	4.8228	1.0169	0.3475	2.0539
				20 day	0.0220	20.6905	69.2345	6.4061	0.9974	0.2285	2.4430
				30 day	0.0270	19.9320	69.1256	7.0975	0.9654	0.2990	2.5806
	off-shore		(5)			LKWD1m	LKRW	LCNK	LCNH	LCNY	LJPY
			(6)			LKWD1m	LKRW				
		[KWD 1m]	(7)		S. E	LKWD1m	LKRW	T10Y	CREDIT	LVIX	TED
				10 day	0.0164	89.9893	1.0681	5.0957	1.2291	0.5770	2.0409
				20 day	0.0223	88.4382	1.4175	6.3739	1.1847	0.3566	2.2291
				30 day	0.0270	87.6089	1.5451	7.0189	1.1424	0.3758	2.3090
JPY	[JPY] T10Y Credit VIX TED		(15)		S. E	JPY		T10Y	CREDIT	LVIX	TED
				10 day	0.0195	88.1914		3.4020	4.6873	0.2073	3.5121
				20 day	0.0273	86.5664		3.7050	4.7074	0.9522	4.0690
				30 day	0.0333	85.6938		3.8121	4.6436	1.5557	4.2948

RMB	re CNY	[CNY]	(9)	20 day	0.0095	94.4583	5.1621	0.0231	0.1807	0.1165	0.0592
				30 day	0.0118	93.8791	5.7306	0.0152	0.2071	0.0970	0.0711
					S. E	LCNY	LCNH				
				10 day	0.0066	95.7357	4.2643				
				20 day	0.0096	94.4312	5.5688				
				30 day	0.0119	93.5308	6.4692				
		[CNY]	(10)		S. E	LCNY	LCNH	T10Y	CREDIT	LVIX	TED
				10 day	0.0066	93.7528	5.4277	0.3049	0.0065	0.0673	0.4408
				20 day	0.0096	91.2001	7.7013	0.4060	0.0032	0.3224	0.3671
				30 day	0.0120	89.8291	8.8248	0.4644	0.0023	0.5514	0.3279
	off-shore CNH	[CNH]	(11)		S. E	LCNH	LCNY	LCNK			
				10 day	0.0081	96.7287	3.0479	0.2234			
				20 day	0.0110	94.3654	5.4780	0.1566			
				30 day	0.0132	92.6297	7.2585	0.1118			
			(12)			LCNH	LCNY	LCNK	LKRW1M	LKRW	LJPY
		[CNH]	(13)		S. E	LCNH	LCNY				
				10 day	0.0081	95.2963	4.7037				
				20 day	0.0112	93.8654	6.1346				
				30 day	0.0135	93.0042	6.9958				
		[CNH]	(14)		S. E	LCNH	LCNY	T10Y	CREDIT	LVIX	TED
				10 day	0.0081	96.9695	1.9251	0.0760	0.0863	0.1644	0.7787
				20 day	0.0109	94.2397	4.1311	0.1265	0.0981	0.4395	0.9650
				30 day	0.0129	92.2013	5.8894	0.1510	0.1067	0.5628	1.0888

*Korea won dollar and Seoul offshore yuan

In Table 9, it was found that the offshore won dollar (KWD1m) accounts for more than 70% of the variance of the onshore won dollar (KRW) and the Seoul offshore yuan (CNK). In [52] the author argued that the volatility of the offshore yuan had a great influence on the exchange rate of the onshore yuan. The results of this study also showed similar results to their argument. The analysis results of this study showed that the onshore won dollar exchange rate is greatly affected by the offshore won dollar. The reason for such results can be interpreted that, First, the won dollar is effectively and practically internationalized in general. Korean won, transactions between onshore and offshore markets are processed through the cross rates arbitrage transactions. Various SWAP transactions using interest rate differential between the Seoul foreign exchange market and the international financial market support this finding. Therefore, the Korea won may be considered as partly internationalized currency. Second, offshore won-dollar (KWD1m) transactions, which have absolute influence on the onshore won-dollar (KRW), have no liquidity problems. This is because offshore NDF won-dollar payments are settled in U.S. dollars at the maturity, so they are not subject to KRW liquidity constraints. Third, it is because investors in the international financial market actively utilize the offshore won-dollar market. This phenomenon is reaffirmed that about 20 percent of the world's NDF currency market transactions are DNF Korea won- dollar. Fourth, Korea is the top 10 global exporters and importers. Therefore, the demand for won-dollar transactions related to import-export settlement and capital market transactions are sizeable in onshore and offshore markets. The analysis results of this study showed that the variance decomposition analysis results between the onshore-offshore won dollar and the variables of the international financial market (analysis models 4 and 7) are similar. In addition, the analysis of variance decomposition of the Korean won, the Japanese yen and the international financial market variables show similarity either. This phenomenon is attributed to the fact that the Korea won has both the characteristics of emerging market currencies and international currencies either.

*China Yuan

The results of analysis of models No. 8-14 in Table 9 show that the onshore-offshore yuan variable explains 92%-95% of the total variance of yuan itself. Compared to the won dollar and yen, the onshore-offshore yuan showed relatively weak explanatory power of other currencies and international financial markets variables. The results of this analysis also reaffirm of [49] the author's argument that the impact of the US dollar is limited in the yuan.

*Japan Yen

The results of the analysis of variance decomposition of the Japanese yen (JPY) of model No. 15 summarized in Table 9 are similar to the results of analysis of Korean onshore and offshore won dollars in analysis models No. 4 and No. 7. The explanatory power of the variables in the international financial market for the won dollar and the yen were similar.

6. Summary and Conclusion

This study analyzed the relationship between renminbi (onshore and offshore), won dollar (onshore and offshore), yen spot exchange rate, and international financial market variables using the VECM model. The analysis period is 3 years from December 1, 2014 to November 30, 2017, and daily closing price data were used. Analysis variables are US dollar-based Chinese yuan (onshore and offshore), Korean won dollar (onshore and offshore), Japanese yen spot exchange rate, and international financial market variables. During the analysis, there were no unusual situations in the international financial market, such as the 1998 Asian financial market crisis or the 2008 global financial market crisis. Therefore, dummy variables were not considered. We used Bloomberg data. The results of the analysis are as follows: First, there was a long-term cointegration equilibrium relationship between yuan (onshore and offshore), won dollar (onshore and offshore), yen spot exchange rate, and international

financial market variables. Second, In the results of VECM analysis of won dollar, yen and yuan, it was found that there is a long-term equilibrium relationship between the Seoul offshore yuan (CNK), Hong Kong offshore yuan (CNH) and the China onshore yuan, (CNY). The result is interpreted as the yuan's liquidity in Seoul's offshore foreign exchange market is relatively insufficient compared to Hong Kong's, and the lack of the liquidity is financed through cross rate transactions between currencies of KRW↔USD↔CNH, CNY. The onshore won dollar (KRW) has a long-term equilibrium relationship in the VECM model from the relationship between the offshore won dollar, the onshore and offshore yuan, and the yen. The explanatory power of offshore won dollars (KWD1m) to onshore won dollars (KRW) using the VECM model was found to be about 50%. And it was found that the offshore won dollar exchange rate has a long-term equilibrium relationship with four variables in the international financial market. As a result, Korea's onshore won-dollar exchange rate was found to be largely influenced by international financial market variables. China's Onshore yuan (CNY) is found to have a long-term equilibrium relationship with the Hong Kong currency (CNH), the onshore and offshore won dollar, and the yen. CNY has also a long-term equilibrium relationship with four variables in the international financial market. Hong Kong's offshore yuan (CNH) has a long-term equilibrium relationship between China's onshore yuan (CNY) and Korea's offshore yuan (CNK). There is also a long-term equilibrium relationship with four variables in the international financial market. However, Hong Kong's offshore yuan has not had a long-term equilibrium with the onshore and offshore won and the yen. The unusual was that although there was a long-term equilibrium relationship between the onshore and offshore yuan, the relationship was weaker than that of the onshore and offshore won-dollar. Overall, the yuan has a weaker long-term equilibrium relationship with other currencies than that of the won dollar. This phenomenon can be interpreted as the yuan is a currency in the progress of internationalization, and still foreign exchange transactions are limited. Commercial export-import cross border yuan transactions are liberalized but capital market transactions are still limited. VECM model analysis of onshore won dollar and onshore yuan have a long-term equilibrium relationship with yen. The reason for such results can be interpreted that the supply of the won dollar, yuan, and yen currency positions are exchanged in the foreign exchange markets of Korea, China and Japan through export-import transactions between Korea, China and Japan. However, the offshore won dollar (NDF, KWD1m) and offshore yuan (CNH) does not show long-term equilibrium relationship with the yen (JPY). This result is interpreted to be because investors in the international financial market, which do not have transaction restrictions, take into account country specific factors only when they trade in currency market. The results of the analysis of the yen VECM model showed that the yen is more correlated with international financial market variables than that of the won dollar and yuan, as expected. Third, the results of the Granger causality relationship between the won dollar, yen, yuan and international financial market variables. It was found that the offshore won dollar (KWD1m) was the biggest causative (Granger causality) to the Korean onshore won dollar (KRW). Subsequently, in the order of the Hong Kong offshore yuan (CNH), the Chinese onshore yuan (CNY) and the Japanese yen (JPY). It was found that the Chinese yuan has slightly stronger Granger causality the Korean won than that of the Japanese yen. The offshore won dollar exchange rate (KWD1m) showed no Granger causal relationship from the onshore won dollar (KRW), Hong Kong offshore yuan (CNH), Chinese onshore yuan (CNY), and Japanese yen (JPY). However, the international financial market variables such as 10-year US treasury bonds, Credit risk spread, and TED spread were found to Granger cause offshore won (KWD1m). Chinese onshore yuan (CNY) was found to Granger cause Hong Kong offshore yuan (CNH) at a 1% significance level. And the reverse case was also significant at the 1% level. As for yuan, onshore and offshore yuan were found to exchange Granger causality relationship. However, it was found that international financial market variables do not Granger cause onshore and offshore yuan. In addition, onshore yuan (CNY) and offshore yuan (CNH) each independently showed a Granger causal relationship with the won dollar and the yen. Therefore, it was found that onshore and offshore yuan have distinct characteristics. This phenomenon is interpreted as the fact that although the yuan is an internationalized currency, only export-import commercial transactions are liberalized but capital market transactions are still limited. The Japanese yen (JPY) has a stronger Granger causal relationship than that of the won dollar and yuan. It was found that three out of four variables in the international financial market have a Granger causal effect on the Korean won dollar exchange rate. However, it was found that only one TED spread has a Granger causal effect on the yuan exchange rate. It was found that international financial market variables had a stronger Granger causal relationship on the non-international currency, the Korean won dollar, than that of the SDR international currency, the Chinese yuan. Fourth, the results of Impulse response analysis. The onshore won dollar (KRW) and the Seoul offshore yuan (CNK) were found to be most strongly affected by the impact from the offshore won dollar (KWD1m). The onshore won dollar (KRW) consistently reacts negatively to the onshore yuan (CNY) shock, but reverse response of positive (+) to negative (-) with different time lag of the offshore yuan (CNH) shock. Therefore, some of the onshore and offshore yuan shocks delivered to the onshore won-dollar (KRW) exchange rate were found to be mutually offset. Overall, Korean onshore (KRW) and offshore won dollars (KWD1m) and Japanese yen (JPY) showed similar impulse responses to shocks from international financial markets. However, the impulse response of the onshore (CNY) and offshore yuan (CNH) to international financial market variables was weaker than that of the won dollar and yen. China's onshore and offshore yuan responded positively to the mutual shock. These results are interpreted as the two markets are closely related to each other as China supplies yuan liquidity to the Hong Kong offshore (CNH) market. Fifth, the result of Variance decomposition analysis. The offshore won dollar (KWD1m) was found to account for more than 70% of the variance of Korean onshore won dollar (KRW) and the Seoul offshore yuan (CNK). International financial market variables have similar degrees of explanation power for the variance of Korean onshore and offshore won and Japanese yen. On the other hand, international financial market variables showed weaker explanatory power for Chinese onshore and offshore yuan compared to Korean won and Japanese yen. These results suggest that the Korea won dollar is closely related to international financial market.

The implications of this study are as follows: First, from the perspective of the international financial market, we compared and

analyzed the won dollar (onshore and offshore), yuan (onshore and offshore), and yen. As a result of analyzing the long-term equilibrium relationship with the VECM model, it was found that the offshore won-dollar exchange rate and the international financial market variables explained more than 50% of the onshore won-dollar exchange rate. Therefore, it was revealed that it is important to monitor the international financial market for changes in the onshore won-dollar exchange rate. Second, overall, the Korean won dollar was found to have greater integration with the international financial market than the Chinese yuan, the SDR international currency. Therefore, it is necessary to consider the characteristics of the won, which is practically and partially similar to the international currency. It also revealed that a new perspective on the Korea won-dollar exchange rate is needed from the perspective of global portfolio investors. Third, the SDR currency, the yuan, is closely related to the onshore yuan (CNY) and the offshore yuan (CNH), but at the same time, each has its own unique market characteristics. Therefore, it was revealed that companies and investors that trade in yuan need to manage currency risks in consideration of the characteristics of the onshore and offshore yuan market. Fourth, it was revealed that for the development of the Korean financial market, it is necessary to prepare a way to actively utilize the won, which has already partially internationalized, in areas where possible, such as SME import-export transaction, investment of sovereign wealth funds. Under the modern monetary economy, where integration with global financial markets is deepening and the role of finance is important, the Korean won will need a partial and managed internationalization process.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

I would like to thank you Graduate School of Software, Open Source lab, Soongsil University.

References

[1] R. Alsakka, O. ap Gwilym, "Foreign exchange market reactions to sovereign credit news," Journal of International Money and Finance, 2012, doi:10.1016/j.jimonfin.2012.01.007.
[2] D.J. Jüttner, D. Chung, W. Leung, "Emerging market bond returns - An investor perspective," Journal of Multinational Financial Management, 2006, doi:10.1016/j.mulfin.2005.05.005.
[3] H.S. Basso, O. Calvo-Gonzalez, M. Jurgilas, "Financial dollarization: The role of foreign-owned banks and interest rates," Journal of Banking and Finance, 2011, doi:10.1016/j.jbankfin.2010.11.018.
[4] M. Brown, S. Ongena, P. Yeşin, "Foreign currency borrowing by small firms in the transition economies," Journal of Financial Intermediation, 2011, doi:10.1016/j.jfi.2010.12.001.
[5] R.A. Meese, K. Rogoff, "Empirical exchange rate models of the seventies. Do they fit out of sample?," Journal of International Economics, 1983, doi:10.1016/0022-1996(83)90017-X.
[6] R. Sangdai, "Analysis of exchange rate characteristics before and after the introduction of floating exchange rates," Bank of Korea, 3–24, 2000.
[7] G. Ortiz, Dollarization in Mexico: Causes and consequences, 1983.
[8] P. Agenor, M.S. Khan, "Foreign currency deposits and the demand for money in developing countries," **50**, 101–118, 1996.
[9] C.M. Reinhart, E. Carmen, K.S. Rogoff, M.A. Savastano, "Addicted to Dollars *," **50**, 1–50, 2014.
[10] M. Lebre de Freitas, "The dynamics of inflation and currency substitution in a small open economy," Journal of International Money and Finance, 2004, doi:10.1016/j.jimonfin.2003.10.003.
[11] K. Metin-Özcan, V. Us, "Dedollarization in Turkey after decades of dollarization: A myth or reality?," Physica A: Statistical Mechanics and Its Applications, 2007, doi:10.1016/j.physa.2007.06.018.
[12] I. Asel, "Financial Sector Development and Dollarization in the Economies of Central Asia," Unpublished Paper (London: European Bank for …, 2010.
[13] Woong Ryeol Kim, "Analysis of RMB Covered Interest Rate Parity and its Implication," The Academic Society of Global Business Administration, **16**(3), 129–159, 2019.
[14] W.R. Kim, M.K. Kim, "Empirical Analysis of RMB Price Discovery Effects between China Sovereign CDS spread, offshore RMB and Global Financial Market Variables: Comparative analysis of China, Japan and South Korea," in Proceedings - 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2019, 2019, doi:10.1109/SNPD.2019.8935719.
[15] A.T. Wang, S.Y. Yang, N.T. Yang, "Information transmission between sovereign debt CDS and other financial factors - The case of Latin America," North American Journal of Economics and Finance, 2013, doi:10.1016/j.najef.2013.02.023.
[16] I. Marcelin, I. Mathur, "Financial sector development and dollarization in emerging economies," International Review of Financial Analysis, 2016, doi:10.1016/j.irfa.2016.04.010.
[17] I. Marcelin, I. Mathur, "Financial development, institutions and banks," International Review of Financial Analysis, 2014, doi:10.1016/j.irfa.2013.09.003.
[18] I. Marcelin, I. Mathur, "Privatization, financial development, property rights and growth," Journal of Banking and Finance, 2015, doi:10.1016/j.jbankfin.2014.03.034.
[19] I. Mathur, I. Marcelin, "Institutional failure or market failure?," Journal of Banking and Finance, 2015, doi:10.1016/j.jbankfin.2014.12.018.
[20] E.L. Yeyati, "Financial dollarization: Evaluating the consequences," Economic Policy, 2006, doi:10.1111/j.1468-0327.2006.00154.x.
[21] A. Korinek, "Excessive Dollar Borrowing in Emerging Markets: Balance Sheet Effects and Macroeconomic Externalities," SSRN Electronic Journal, 2011, doi:10.2139/ssrn.967524.
[22] P.E. Guidotti, C.A. Rodriguez, "Dollarization in Latin America: Gresham's Law in Reverse?," Staff Papers - International Monetary Fund, 1992, doi:10.2307/3867472.
[23] A. Kokenyne, J. Ley, R. Veyrune, "Dedollarization," IMF Working Papers, **10**(188), 1, 2010, doi:10.5089/9781455202225.001.
[24] A. Berg, E. Borensztein, Full Dollarization - The Pros and Cons, Internatiol Monetary Fund, 2000.
[25] A. Ize, E.L. Yeyati, "Financial dollarization," Journal of International Economics, 2003, doi:10.1016/S0022-1996(02)00017-X.
[26] P. Honohan, "The retreat of deposit dollarization," International Finance, 2008, doi:10.1111/j.1468-2362.2008.01223.x.
[27] International Monetary Institute, "The internationalization of the Renminbi," The Internationalization of the Renminbi, **9781315884**, 1–221, 2013, doi:10.4324/9781315884967.
[28] J. a Frankel, "NO SINGLE CURRENCY REGIME IS RIGHT FOR ALL COUNTRIES OR AT ALL TIMES," Time, 1999.
[29] G. Michalopoulos, Global Divergence in Trade, Money and Policy-the internationalization of the euro: trend, challenges and risks, 2006.
[30] S. Hyun, S. Lee, "On the Determinants of Currency Internationalization: Implications for Korean Won Internationalization (in Korean)," Economic Analysis, **19**(1), 43–75, 2013.
[31] C. won Park, Essays on International Monetary System Issues, Hanyang University, 2012.
[32] D.W.Y. Leung, J. Fu, "Interactions between CNY and CNH Money and Forward Exchange Markets," SSRN Electronic Journal, 2014, doi:10.2139/ssrn.2448569.
[33] F.X. Diebold, K. Yilmaz, "Measuring financial asset return and volatility spillovers, with application to global equity markets," Economic Journal, 2009, doi:10.1111/j.1468-0297.2008.02208.x.
[34] Y.W. Cheung, D. Rime, "The offshore renminbi exchange rate: Microstructure and links to the onshore market," Journal of International Money and Finance, 2014, doi:10.1016/j.jimonfin.2014.05.012.
[35] M. Funke, C. Shu, X. Cheng, S. Eraslan, "Assessing the CNH-CNY pricing differential: Role of fundamentals, contagion and policy," Journal of International Money and Finance, 2015, doi:10.1016/j.jimonfin.2015.07.008.
[36] M.D. Bordo, C.M. Meissner, M.D. Weidenmier, "Identifying the effects of an exchange rate depreciation on country risk: Evidence from a natural experiment," Journal of International Money and Finance, 2009, doi:10.1016/j.jimonfin.2008.10.004.

[37] M. Schneider, A. Tornell, "Balance sheet effects, bailout guarantees and financial crises," Review of Economic Studies, 2004, doi:10.1111/j.1467-937X.2004.00308.x.
[38] I. Petrova, M.G. Papaioannou, D. Bellas, "Determinants of Emerging Market Sovereign Bond Spreads: Fundamentals Vs Financial Stress," IMF Working Papers, 2010, doi:10.5089/9781455210886.001.
[39] IMF, Recent Experiences in Managing Capital Inflows—Cross-Cutting Themes and Possible Policy Framework, 2011, doi:10.5089/9781498339353.007.
[40] M.A. Segoviano Basurto, C. Caceres, V. Guzzo, "Sovereign Spreads: Global Risk Aversion, Contagion or Fundamentals?," IMF Working Papers, 2010, doi:10.5089/9781455200795.001.
[41] J. Hilscher, Y. Nosbusch, "Determinants of sovereign risk: Macroeconomic fundamentals and the pricing of sovereign debt," Review of Finance, 2010, doi:10.1093/rof/rfq005.
[42] C. Moser, "The Impact of Political Risk on Sovereign Bond Spreads - Evidence from Latin America," Proceedings of the German Development Economics Conference, Göttingen 2007 / Verein Für Socialpolitik, Research Committee Development Economics, 2007.
[43] L.E. Kodres, K. Hartelius, K. Kashiwase, "Emerging Market Spread Compression: Is it Real or is it Liquidity?," IMF Working Papers, 2008, doi:10.5089/9781451868722.001.
[44] J. Pan, K.J. Singleton, "Default and recovery implicit in the term structure of sovereign CDS spreads," Journal of Finance, 2008, doi:10.1111/j.1540-6261.2008.01399.x.
[45] M. Teo, "The liquidity risk of liquid hedge funds," Journal of Financial Economics, 2011, doi:10.1016/j.jfineco.2010.11.003.
[46] E. Warshaw, "Parallel Currency Markets and the Monetary Exchange Rate Model: A VECM Application to Turkey Over 1987–1998," Eastern European Economics, 2016, doi:10.1080/00128775.2016.1194215.
[47] L.S. Lee Jaeha, "Causal relationship between the KRW onshore spot market and the offshore NDF KRW spot market," Korea Financial Management Association, **17**(2), 211–227, 2000.
[48] S. Johansen, "Statistical analysis of cointegration vectors," Journal of Economic Dynamics and Control, 1988, doi:10.1016/0165-1889(88)90041-3.
[49] L.F. and I. Yu, A STUDY ON THE TRANSMISSION OF MONEY MARKET TENSIONS IN EMEAP ECONOMIES DURING THE CREDIT CRISIS OF 2007 – 2008, 2009.
[50] HKMA, Introduction of Intraday Repo under the Renminbi Liquidity Facility, HKMA, 2014.
[51] G. Ferrucci, Empirical Determinants of Emerging Market Economies' Sovereign Bond Spreads, 2003, doi:10.2139/ssrn.597422.
[52] Y.W. Cheung, C.H. Hui, A. Tsang, "The RMB central parity formation mechanism: August 2015 to December 2016," Journal of International Money and Finance, 2018, doi:10.1016/j.jimonfin.2018.04.011.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 762-767 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Optimization of Multi-user Face Identification Systems in Big Data Environments

Majdouline Meddad[*,1], Chouaib Moujahdi[2], Mounia Mikram[1,3], Mohammed Rziza[1]

[1]*LRIT Laboratory, Associated Unit to CNRST (URAC 29) Rabat IT Center, Faculty of Sciences, Mohammed V University in Rabat, 10010, Morocco*

[2]*Scientific Institute of Rabat, Mohammed V University in Rabat, 10010, Morocco*

[3]*School of Information Sciences in Rabat, 10010, Morocco*

ARTICLE INFO

Article history:
Received: 29 August, 2020
Accepted: 14 November, 2020
Online: 24 November, 2020

Keywords:
Big Data
Biometric Distribution
Face Identification
Convolutional neural network

ABSTRACT

Computer vision offers several strategies that permit computers to comprehend the substance of inputted data to extract the relevant highlights features. That gives the possibility to develop several successful recognition systems like face identification. One of the enormous difficulties these days is the way to have a prompt identification face in a multi-client identification system. We propose in this paper, an optimization of some face identification systems in big data environments that manages the cost of an appropriate identification time while holding a good performance of the system. The used systems are based on: CNN with Inception V3, PCA, LDA and LBPH. Our experimental results indicate that the performance can be preserved while reducing considerably the running time. LBPH with a Radius equal to 2 and 8 neighbors gives the best accuracy of that is equal to 95,71% with an identification time of 4.44 seconds.

1 Introduction

Computer vision offers several techniques for machines to understand the content of images like human vision. To understand the content of images, machines need some image processing steps to be able to extract features from the images for several applications such as face identification, object recognition and object tracking.

The substantial idea of a biometric system is to utilize acquisitions of behavioral and/or physiological characteristics of human like face, fingerprint, signature and voice for recognition/verification. Currently, several face identification systems use large-scale databases that contain a big number of subjects such as celabA Dataset that contains 10177 identities and 202599 images. Large-scale databases present a big open challenge to create an operational biometric identification that offers an acceptable identification time. The Scalability [1] of such tremendous systems makes the challenge harder to be solved. This work is concerned with the creation and optimization of such biometric identification systems.

This paper is an extension of a previous work presented during 2019 in the 1st International Conference on Smart Systems and Data Science (ICSSD) [2], where we have tested and built several identification systems (i.e., LBPH [3],[4], PCA [5], LDA [5] and deep learning with inception V3 [6]) that use a big database with some hard conditions.

The remainder of the paper is structured as follows. The related works are presented in Section 2. Section 3 presents the different used face identification systems with all optimization steps. Our experimental results are introduced in Section 4 Our conclusion and perspectives are given in Section 5.

2 Related Work

Identification time, efficiency and security are three crucial issues from which large-scale identification systems suffer. Due to these problems. Several published kinds of research deal with such problems. On one hand, some published approaches tackle security issues whom we can cite [7],[8] that implemented Biometric anti-spoofing technique by using the trait randomization technique.

On the other hand, published approaches can be part of two significant classes: learning distribution approach (first class) and Processing distribution approaches (second class). The principal idea of the second class is sharing tasks to several workers in the cluster to handle the issue of treatment large amounts of data in a few

[*]Corresponding Author: Majdouline Meddad, Email: majdouline_meddad@um5.ac.ma

times by employing the big data technologies in the cloud environment for parallel-distribution treatment. Well known, deep learning models are hefty in computing when it needs more iterations to get the best performant model which needs to use an approach of the first class with big data technologies to distribute the learning process.

For the first class, the crucial idea presented in [9] is merging the utility of the neural network algorithm with the MapReduce technique to face the extensive data problem in handwritten characters. The authors in [10] proposed for the handwriting recognition digit by firstly applying an elastic distortion to the entered data then Implemented a computing method with distribution to gain the time cost by mapping the dataset into several machines of the cluster. In [11], the author proposed some parallelization methods which are MRBPNN 1, 2 and 3) based on a MapReduce technique to face exhaustive scenarios in the conditions of the size of data and neurons.

For the second class, [12] displays a MapReduce technique of the process of converting frames of video to grayscale images as a preprocessing technique to extract features from converted images. In [13], the author apply a distributed medical application in two different techniques of distribution such as MapReduce and Sun Grid Engine regarding running time. In [14], the author introduced an approach for storing and processing satellite images by a remote sensing processing tools OTB with the MapReduce technique of Big Data technologies. In [15], the author proposed to use the MapReduce for parallel model execution to extract the similarity of the tumor image. We trust that the principal drawback of this class is hard to manage the cluster to ensure continuity of processing in a distributed platform and Hadoop technology requires a significant memory.

3 Face Identification In Big Data Environment

The conception of a biometric system carries four principal parts as shown in FIG.1.

- The sensor part (Grey Box) that obtains the biometric information to remove a computerized portrayal during enlistment, identification or verification.
- The characteristic extraction part (Green Box) that empowers decreasing the dimensionality of the representation, emitted by the sensor part, by extricating the significant capabilities. This part can possess a sub-part to test the norm of gained biometric information.
- The database part (Red Box) that comprises the biometric formats of the enlisted subjects inside the framework.
- The decision part (Orange Box) that contrasts the test format and the enlisted formats of the database part requires a decision with regards to the sort of the framework.

The correspondence between these parts becomes tedious inside of having huge database for identification systems. To fix this difficulty, we will utilize some technologies for parallelizing these correspondence tasks inside various slaves. Indeed, current datasets are quickly expanding in terms of size and multifaceted nature, also, there is an importunate need to create solutions to extract pertinent data by utilizing methods based on statistics.

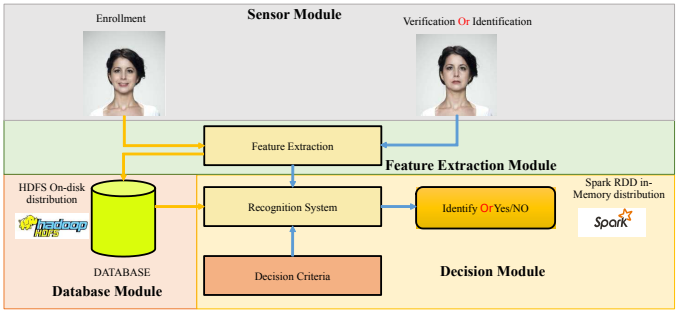


Figure 1: The general conception of a biometric system [16]

An effective biometric framework might be a framework that primarily controls the capacity of the database to be changed in terms of size or scale. they are numerous big data frameworks that are utilized for overseeing a huge data, as Apache Spark framework [17] and Hadoop [18] that run dataflow diagrams over the slaves of a cluster.

In the data-parallel computing process, a client program is arranged into an execution plan graph (EPG), This EPG is the central information structure utilized for distributing tasks, control distributed jobs in the slaves and adapting to non-critical failure. Once a job is running, EPG is unaltered at runtime aside from some restricted adjustments. This makes it hard to utilize dynamic optimization procedures that could significantly improve the disseminated execution dependent on runtime data. Optimus [19] is a framework for progressively rewriting an EPG at runtime.

MillWheel [20] is a system for constructing a data processing with a high size of data in a few time where the user describes the code and the directed computation graph for specific nodes then the framework oversees determined state and recording continuously the flow with non-critical failure in stream processing.

For supporting embedded devices, Sonora [21] is a framework that enhance the stream abstraction for the accompanying reasons. To begin with, streams effectively bind together numerous data activities in cloud and Mobile.

Apache Spark, Naiad [22] and Hadoop run in less calculation time when there is enough memory inside the group of machines to convey the working arrangement of the calculation. Spark keeps up MapReduce's linear adaptability and adaptation to non-critical failure, it is a lot quicker, a lot simpler to program, the distributed data of the scope of computationally heavy tasks, including graphic process, interactive queries, flows and Machine Learning (ML).

In this paper, the jobs of a huge face identification system are parallelized by utilizing Hadoop[23] with Apache Spark introduced by Directed Acyclic Graph (DAG) in figure 3. Every Spark application includes a driver program that completes the user's function and carries out distributes operations on a group of the machine (cluster). The main reflection that Spark furnishes could be resilient distributes dataset (RDD), which could be an assortment of parceled items on a cluster that will run in equal. Users can even request that

Spark remain RDD in memory, permitting it to proficiently reutilize operation in parallel. when a node is failed RDD can automatically be recovered.

3.1 Distributed tasks of LBPH, PCA and LDA in Apache Spark

The utilized identification system is predicated on [3]. However, we have made some changes, to support it inside the Apache Spark, following the stages beneath (see additionally figure 2).

We firstly make first RDD (RDD1) by loading all test images such as Binary Files from HDFS then we split it into sixty parts. Secondly, we apply a transformation of the RDD1 to make RDD2 by transforming the color image to grayscale. Thirdly, we create RDD3 by applying transforms of loaded images in RDD 1 to extract real labels. Fourthly, we made RDD 4 by extracting the content of previous RDDs (2 and 3) by applying a collective action. Finally, the identity of each image is parallelized and compared with the right label extracted in RDD3.

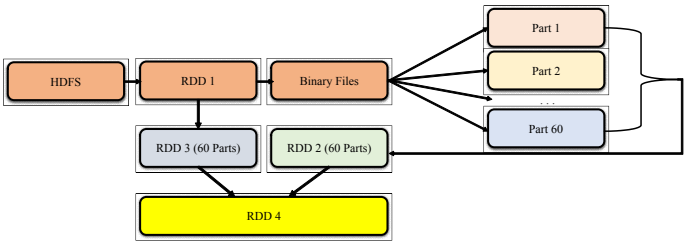


Figure 2: The process of parallelization in Apache Spark

To diminish the running time within the identification stage using the normal PCA, LBPH and LDA identification systems. Initially, the first Apache Spark RDD was created by recovering the test set in 60 parts as binary files from the Hadoop HDFS, and afterward, every thread of our computer changes every picture of every part to grayscale to create the second RDD of grayscale images. Then a third RDD is created by extracting the right label of each test image to collect their identity.

Secondly, we extract the PCA, LDA and LBPH characteristic vectors and compare the vector of the test with those of the learning set.

Thirdly, we predict the identity of the image of the test with the minimal distance overall comparison of the images from the learning set. We used HDFS in the database module to store images and Apache Spark for characteristic extraction and decision parts.

3.2 Pre-trained CNN model, LBPH, PCA and LDA for Multi-user identification system

In this sub-section, we have utilized the Transfer learning procedure where a model produced for tackling an issue is reutilized for a subsequent issue in a timesaving way by starting from patterns that have been learned in the first task, in this manner you raise antecedent learning and keep away from starting from scratch.

We have utilized the Inception V3 as an extractor of pertinent characteristics. It is utilized besides to dispose of the last completely connected layer, at that point, we train the rest of the model as a learning set element extractor for the new dataset. Inception V3 extract 2048 values for every picture. We call these highlights CNN codes. Whenever we have removed the codes for whole pictures. We have prepared a totally connected layer with a Softmax classifier for our system.

A while later, to claim less period, due to the mechanism of spark's build-in broadcast the model is firstly distributed to the cores of our machine in [24]. At that point, this model is used on each core and utilized to identify the identity of an entered image. For this work, the models of a huge identification face system are distributed by utilizing just Apache Spark technology [23].

In our work, every worker takes an image to identify by turning the facial identification system based on PCA, LDA or LBPH to predict the identity of a client by calculating the similarity value with all over the training set as shown in figure 3(a).

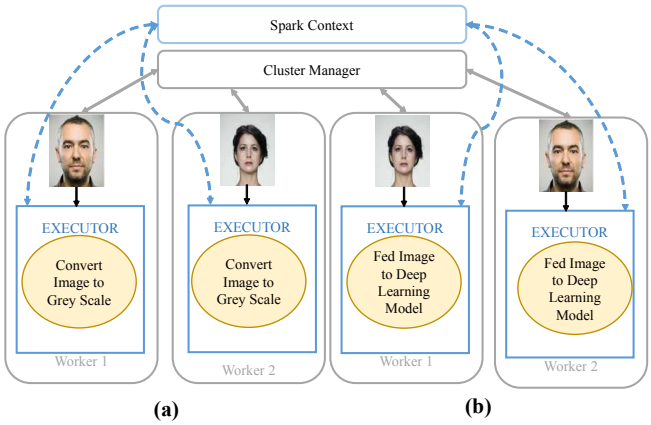


Figure 3: Multi-user identification system: (a) for PCA, LDA and LBPH, (b) for Deep Learning Model

Firstly, the image from RGB (color) to grayscale is converted, afterwards, Every worker compute the eigenvectors, LBP histogram and fisherFace vectors then every worker of the cluster identifies the identity of the tested image from saved PCA, LDA or LBPH models which contain all computed values as shown in figure 3(a) and shared to all cores of the machine.

We employed in this paper a pre-trained CNN Model (Inception V3). Firstly, we have only trained the classification part of the Inception V3 model due to the computational cost of learning to create a model that can identify the person from his/her face image which we need to reutilize the model trained from the Imagenet database to our chosen dataset, we have trained the classification part to induce the correct weights and bias of the fully connected layer (FC) for identification face system, thereafter, we fed the test image into the model to urge the identity, this is often the task of every worker within the cluster as illustrated in figure 3(b) where every worker takes a test image from Hdfs then fed it into the model to extract characteristics that contain 2048 values of every vector then pass it through classification part (Fully connected layer trained to create the facial identification system) then introduce the identity as the last step.

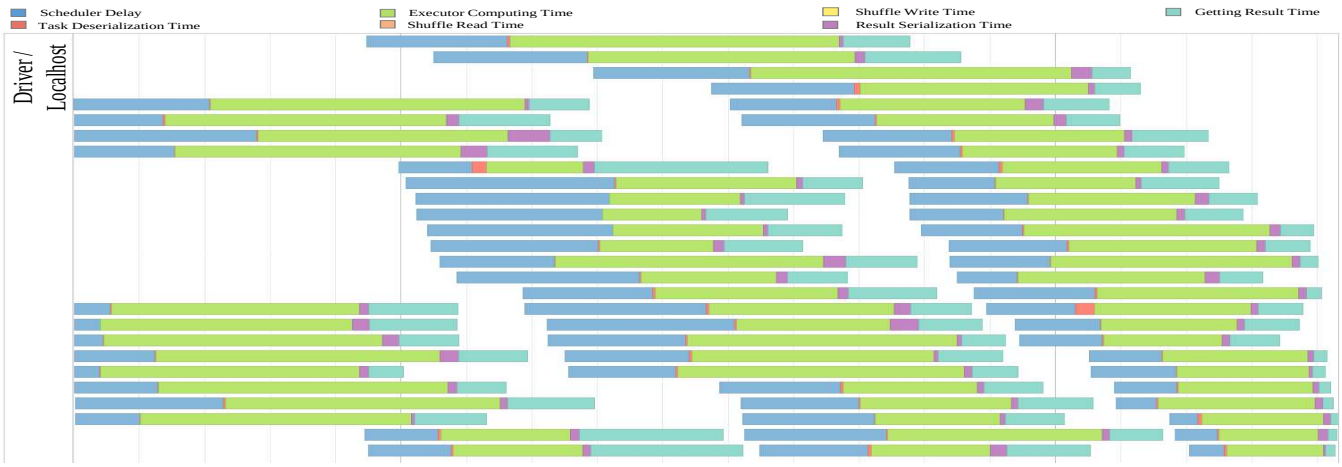


Figure 4: The run time using 6 cores (12 threads) of various tasks

4 Experimental Results

In this section, we evaluate the optimization of various identification systems : the FisherFace [5], LBPH [3][4], , FisherFace [5] and Inception V3 [6] that involve 48 layers can distinguish between 1000 items (e.g., Flower, Fish, Bird, etc) in terms of identification time and the precision. For the evaluation, we have got used the Extended Yale Faces B database [25].

4.1 Database Description and Evaluation

We applied 17 changes (e.g., Rotation, Blur, Noise, etc.) for every subject of the database utilized in this paper [26] to make a bigger database and also to make the identification task more harder.
This database involves 28 subjects with 16128 images with 9 poses and in 64 lighting conditions, In our paper, we are just engaged by 26 subjects. Consequently, the database consists of 107730 images in total. We have suggested splitting the database into three sets as described in figure 5.

4.2 Results and discussion

In the aim to reduce the running time, we run our system using Apache Spark technology by making a collection of RDDs. Every thread of the computer core runs a part of partitioned RDD in 60 parts as illustrated in the figure 4. Each job could be a part of the step, as an example, the primary job contains 60 tasks, in as much as our machine involve 12 threads, 12 tasks are going to be run in parallel (see figure 4). figure 6 illustrates a comparison between various face identification systems that utilize EigenFace, LBPH, FisherFace, and CNN model in Apache Spark [24].

The output of the last convolutional layer of the CNN model is called cnn code extracted from the entered images (Trained images) during the primary part of the learning phase, this stage takes 4 hours and 19 minutes by using Apache Spark broadcasting to calculate all codes. Running the identical stage of the learning step it takes 14 hours without utilizing the Apache Spark broadcasting, is an off-the-cuff term of a feature vector that plays a vital part of identification, which is a representation of a picture in less number of values that will be used for classification.

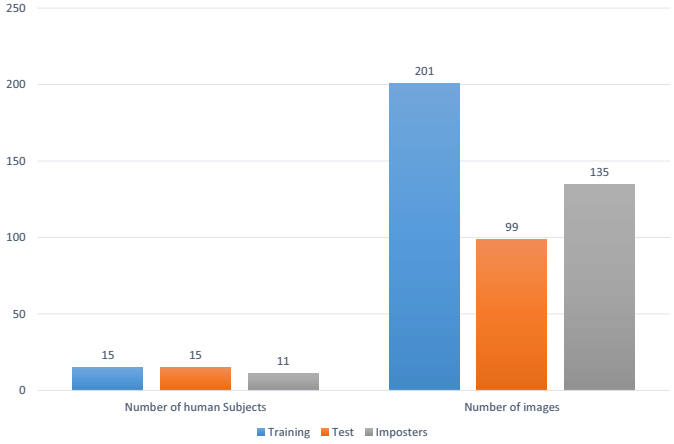


Figure 5: Database Distribution

This final layer was created to give a collection of values that is fitting to the classifier to separate among users. This implies that it needs to be a meaningful and good representation of the images because it should have as many relevant features for the classifier to create an accurate decision by just using a few values.

During the learning phase, every picture is reutilized various times during the learning phase and every code takes an all-inclusive time to be calculated. To quicken the job we conserve these codes on our machine to not recalculate them consistently. On the off chance that we re-execute the framework once more without feed in images to the model.

When the codes are all extracted in the first step, the specific development of the last layer of the neural network model starts. Many series of outputs will be during this step that can show the exactness of the validation and hence the precision of the learning part.

The precision of the learning shows how the model learned from the pre-owned pictures to distinguish between identities. Validation exactness presents the coordinating rate utilizing an arbitrarily

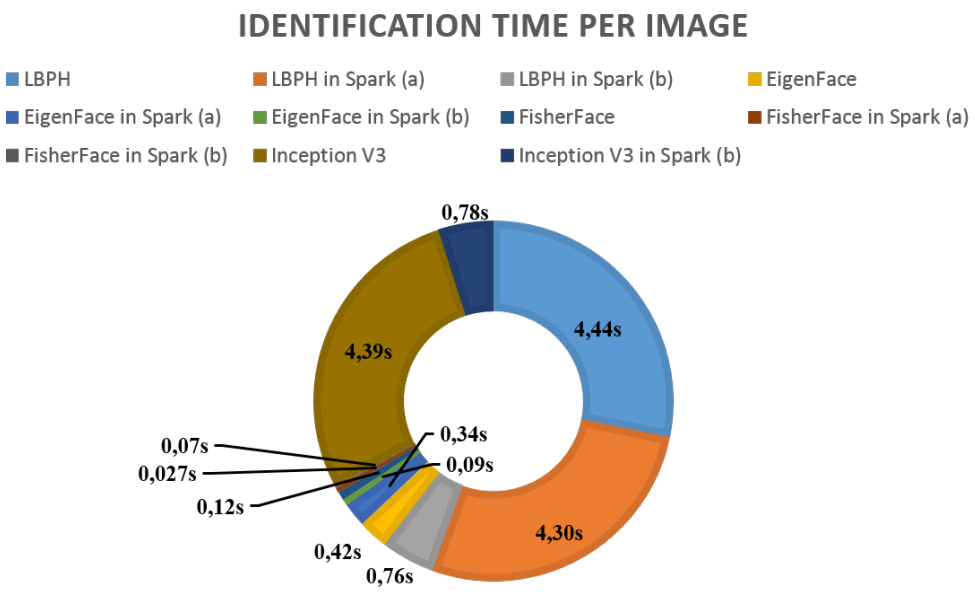


Figure 6: Overall identification time Comparison between The Different Tested Identification Systems with/without distribution

chosen group of pictures from a remarkable learning set. That validation exactness depends on the work of an assortment of data that is not contained inside the learning set if the precision of the learning is high. However, the validation set precision stays low, this suggests that the model is remembering specific highlights inside the learning set pictures. In each validation or learning step, we select 100 arbitrary pictures of the entire validation/learning set, we find its codes inside the cache and brings them into a definitive layer to get identities.

These predictions are contrasted with the specific labels to optimize the classification part and update the bias and weights by utilizing some optimizers like Adam Optimizer and Gradient Descent.

By a learning rate of 0.01 with 250000 iterations using the Gradient Descent Algorithm and Adam Optimizer we respectively get 91, 47% and 91.66%. By a learning rate of 0.1 with 500000 iterations using the Gradient Descent Algorithm and Adam Optimizer we respectively get 91.82% and 92.31%.

By a learning rate of 0.05 with 250000 iterations using the Gradient Descent Algorithm and Adam Optimizer we respectively get 92, 30% and 90.84%. By a learning rate of 0.1 with 500000 iterations using the Gradient Descent Algorithm and Adam Optimizer we respectively get 92.24% and 91.45%.

By a learning rate of 0.001 with 250000 iterations using the Gradient Descent Algorithm and Adam Optimizer we respectively get 87.01% and 91.85%. By a learning rate of 0.1 with 500000 iterations using the Gradient Descent Algorithm and Adam Optimizer we respectively get 88.73% and 92.10%.

We set in the learning phase the proportion of learning set at 80% and we preserved 10% for validation and test. In every iteration we have entered 100 images, after every 10 iterations, we have got evaluated the model using 100 images of the validation set that are chosen arbitrarily. after finishing the learning phase, we fed all the test set through the model to know the precision and how the way the model can classify the inputted faces. After being aware of the good precision of the model in the last part of the learning phase we run our test set for every learning rate and variety of iterations as shown in figure 6. Globally, the precision value is expanded by expanding the number of iterations. If the learning rate is tiny, the precision value is expanded smoothly which makes the learning becomes more reliable, however, the optimization will take longer because steps towards the minimum of the loss function are very small. If the learning rate is high, the precision is expanded rapidly and that we will have the overshoots of the minimum which makes aggravate the loss.

Our main aim is to increase the identification time per image of the test while getting the good performance of the models. This purpose is reached by running our system in Apache Spark (see figure 6). As shown in figure 6 the quick system is FisherFace followed by the EigenFace, then the CNN model, and at last the LBPH.

To advance the exactness of systems, we have made some changes to a few parameters of the LBPH algorithm, such as changing the number of neighbors and the Radius value. For Inception V3, we have made changes to a few parameters simply like the number of iterations, the chosen optimizer algorithm, and subsequently the learning rate value.

For the LBPH algorithm with a Radius equal to 1 with 8, 6 and 4 neighbors, we respectively get 94.05%, 93,74% and 91.91% an Accuracy of identification.

For the LBPH algorithm with a Radius equal to 2 and 8,6 and 4 neighbors, we respectively get 95.71%, 95.43% and 94.02% an Accuracy of identification.

All systems are executed in a server that contains 6 cores with 12 threads and 4.70 GHz Turbo frequency with 12Go of the storage memory.

5 Conclusion and Perspectives

In this paper, we introduced an optimization of multi-user identification face systems in Big Data environments. Hadoop and Apache Spark are two technologies accustomed to parallel tasks and make multi-user systems. This optimization has been examined and evaluated, as far as accuracy and identification time using various classical identification face systems. Our experimental results show that the performance is preserved while offering an affordable identification time for oversized scale systems. In our future work, we have plan to create a multi-view identification face system using a Siamese Convolutional network to tackle the issue of having a few facial acquisitions.

Acknowledgment The financial support of the "Centre National pour la Recherche Scientifique et Technique" CNRST, Morocco, is acknowledged by Majdouline Meddad.

References

[1] K. Janocha, W. M. Czarnecki, "On loss functions for deep neural networks in classification," Schedae Informaticae, 2016, doi:10.4467/20838476SI.16.004.6185.

[2] M. Meddad, C. Moujahdi, M. Mikram, M. Rziza, "A Multi-User Face Identification system with distributed tasks in a Big Data Environment," in ICSSD 2019 - International Conference on Smart Systems and Data Science, 2019, doi:10.1109/ICSSD47982.2019.9003112.

[3] T. Ahonen, A. Hadid, M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, doi:10.1109/TPAMI.2006.244.

[4] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, "Local binary patterns and its application to facial image analysis: A survey," IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, 2011, doi:10.1109/TSMCC.2011.2118750.

[5] A. Khan, H. Farooq, "Principal Component Analysis-Linear Discriminant Analysis Feature Extractor for Pattern Recognition," **8**(6), 267–270, 2012.

[6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, doi:10.1109/CVPR.2016.308.

[7] K. Okereafor, C. Onime, O. Osuagwu, "Enhancing Biometric Liveness Detection Using Trait Randomization Technique," in 2017 UKSim-AMSS 19th International Conference on Computer Modelling Simulation (UKSim), 28–33, 2017, doi:10.1109/UKSim.2017.44.

[8] K. OKEREAFOR, O. Osuagwu, C. Onime, "Biometric Anti-spoofing Technique Using Randomized 3D Multi-Modal Traits," 2018, doi:10.6084/m9.figshare.12479165.v1.

[9] K. Sun, X. Wei, G. Jia, R. Wang, R. Li, D. C. Oct, "Large-scale Artificial Neural Network : MapReduce-based Deep Learning," .

[10] N. Basit, Y. Zhang, H. Wu, H. Liu, J. Bin, Y. He, A. M. Hendawi, "MapReduce-based deep learning with handwritten digit recognition case study," in Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016, 2016, doi:10.1109/BigData.2016.7840783.

[11] Y. Liu, J. Yang, Y. Huang, L. Xu, S. Li, M. Qi, "MapReduce Based Parallel Neural Networks in Enabling Large Scale Machine Learning," Computational Intelligence and Neuroscience, 2015, doi:10.1155/2015/297672.

[12] M. Yamamoto, K. Kaneko, "Parallel Image Database Processing With Mapreduce and Performance," International Journal of Electronic Commerce Studies, 2012, doi:10.7903/ijecs.1092.

[13] T. Ben-Nun, T. Hoefler, "Demystifying parallel and distributed deep learning: An in-depth concurrency analysis," ACM Computing Surveys, 2019, doi:10.1145/3320060.

[14] I. Chebbi, B. W, F. I. R, "Improvement of Satellite Image Classification : Approach Based on Hadoop / Map Reduce," in 2nd International Conference on Advanced Technologies for Signal and Image Processing - ATSIP'2016 March 21-24, 2016, Monastir, Tunisia, 31–34, 2016.

[15] D. D. Manojbhai, K. K. Pradipkumar, R. Rajamenakshi, "Big image analysis for identifying tumor pattern similarities," in Proceedings of 2016 International Conference on Advanced Communication Control and Computing Technologies, ICACCCT 2016, 2017, doi:10.1109/ICACCCT.2016.7831596.

[16] A. K. Jain, P. Flynn, A. A. Ross, Handbook of Biometrics, Springer Publishing Company, Incorporated, 1st edition, 2010.

[17] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, I. Stoica, "Spark: Cluster computing with working sets," in 2nd USENIX Workshop on Hot Topics in Cloud Computing, HotCloud 2010, 2010.

[18] S. G. Manikandan, S. Ravi, "Big data analysis using apache hadoop," in 2014 International Conference on IT Convergence and Security, ICITCS 2014, 2014, doi:10.1109/ICITCS.2014.7021746.

[19] Q. Ke, M. Isard, Y. Yu, "Optimus: A Dynamic Rewriting Framework for Data-Parallel Execution Plans," in Proceedings of the 8th ACM European Conference on Computer Systems, EuroSys '13, 15–28, Association for Computing Machinery, New York, NY, USA, 2013, doi:10.1145/2465351.2465354.

[20] T. Akidau, A. Balikov, K. Bekiroğlu, S. Chernyak, J. Haberman, R. Lax, S. McVeety, D. Mills, P. Nordstrom, S. Whittle, "MillWheel: Fault-Tolerant Stream Processing at Internet Scale," Proc. VLDB Endow., **6**(11), 1033–1044, 2013, doi:10.14778/2536222.2536229.

[21] X. Chen, I. Beschastnikh, L. Zhuang, F. Yang, Z. Qian, L. Zhou, G. Shen, J. Shen, "Sonora: A Platform for Continuous Mobile-Cloud Computing," Technical report, 2012.

[22] D. G. Murray, F. McSherry, R. Isaacs, M. Isard, P. Barham, M. Abadi, "Naiad: A timely dataflow system," in SOSP 2013 - Proceedings of the 24th ACM Symposium on Operating Systems Principles, 2013, doi:10.1145/2517349.2522738.

[23] A. V. Hazarika, G. Jagadeesh Sai Raghu Ram, E. Jain, "Performance comparision of Hadoop and spark engine," in Proceedings of the International Conference on IoT in Social, Mobile, Analytics and Cloud, I-SMAC 2017, 2017, doi:10.1109/I-SMAC.2017.8058263.

[24] T. Hunter, "Deep Learning with Apache Spark and TensorFlow," 2016.

[25] D. K. K.C. Lee , J. Ho, "Extended Yale Face B," 2016.

[26] A. S. Georghiades, P. N. Belhumeur, D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, doi:10.1109/34.927464.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 768-780 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Surge Protection Device for Ex Application

Teik Hua Kuan*, Kuew Wai Chew, Kein Huat Chua

Department of Electrical and Electronic Engineering, Universiti Tunku Abdul Rahman, Sungai Long 43000, Malaysia

ARTICLE INFO

Article history:
Received: 12 July, 2020
Accepted: 11 November, 2020
Online: 24 November, 2020

Keywords:
Lightning
Surge Protection Device
Hazardous area
Ex area
Explosion Protection Technique
Overvoltage Transient

ABSTRACT

A modern process plant is controlled, operated and safeguarded by the sophisticated distributed control system or a programmable logic controller and an emergency shutdown system. This process plant's operation could be seriously affected by lightning, electrical switching, and power outrage. Lightning propagation can travel kilometers away, which has the risk of disturbing the control system's operation. The process plant's control system could damage when it is subjected to surges induced by lightning, power outrage, or heavy load switching. To protect equipment damage from these surges surge protection device is required. A number of the process plant uses flammable material for its production or during the manufacturing processes, explosive gas or vapor is released created an explosive atmosphere known as hazardous (Ex) area. Surge protection device mounted in the Ex area or used in the Ex control loop needs to be carefully selected to protect the equipment and at the same time, it does not become a source of ignition. This paper provides practical consideration and comprehensive detail in revealing the science and engineering behind Ex application on surge protection devices. The technical presentation is limited to equipment protection; structural protection against lightning will not be discussed here.

1. Introduction

This journal article is an extension of the original conference paper entitled, "Practical approach on surge protection device for Ex application," presented in 2019 1st. International Conference on Electrical, Control and Instrumentation Engineering (ICECIE), Kuala Lumpur, Malaysia [1].

Lightning is a natural phenomenon that can be seen in volcanic eruptions, pyrocumulonimbus clouds, heavy snowstorms, hurricanes and thunderstorms [2]. Lightning-induced voltage surges can enter into the electrical system via the capacitive, inductive and resistive coupling [3]. According to the report from NFPA [4], in the United States alone, from 2007 to 2011, an estimated 22,600 fire cases were caused by lightning, with a $451 million loss in direct property damage per year. The report also cited several explosion incidents caused by the lightning surge. One of them is a coal mine (Ex area) explosion in January 2006 in West Virginia, which claimed 12 lives. A protective circuit like a surge protection device or SPD is used to protect the electrical apparatus and its accessories from the damage created by transient surge activity induced by the lightning and surges generated by switching heavy reactive load or power supply interruption and prevent a fire. SPD has to be carefully selected and with proper installation for its intended function; else, instead of preventing equipment failure, it could become the source of fire or explosion as reported by [5–7].

In the original paper, the following are either not evaluated, missing or too brief as the reader is expected to have the background knowledge to understand the subject. In the present work, the critical area left out by the original paper is being addressed and enhanced. The rectification and enhancements are highlighted below:

- The reasons why SPD is needed for the modern electronics system for the application in a safe, hazardous area and critical control loop application.
- Why is the selection criteria on SPD for safe area application not sufficient for hazardous area application?
- Presentation on the lightning protection zone (LPZ) and the SPD types are too brief on the original paper. Therefore it is decided to separate them into two sub-sections with enhancement and with an application example. Definition and a summary of the key points are also being addressed and tabulated.

*Corresponding Author: Teik Hua Kuan; Adress: Universiti Tunku Abdul Rahman, Sungai Long 43000, Malaysia; Contact No: +60123173650; Email: thkuan@1utar.my or thkuan@klaesb.com

- Application on lightning protection level's (LPL) concept; what needs to be done before implementing LPL and the efficiency level for various LPLs are added.
- The description of the maximum continuous operating voltage has been revised to reflect the practical application.
- The illustration on the concept of the "Temperature Classification" is enhanced with a practical example.
- Elements required to be present for an explosion to occur and the concept of various explosion protection techniques being implemented are added.
- Using "Spark Test Apparatus" to obtain the minimum ignition curve for inductive, capacitive and resistive load at the field. The application of the minimum ignition curves is also added.
- Requirement on the hot permit to carry out maintenance and typical application example is added for flameproof, Exd protection.
- Hot surface investigation and typical applications example for increased safety, Exe protection is added.
- Missing entity safety parameter description on the cable capacitance and inductance for intrinsic safety, Exi protection technique is rectified.
- Surge withstanding capability (SWC) consideration is added as it is missing in the original paper.
- Why are two SPDs needed for a control loop? Why one of the SPD will have the Temperature Classification, the other one is not necessary?

The subject of explosion protection for industrial automation applications (non-military) is a specialist field; hardly any higher education institution conducting the courses. Hence, most of the knowledge on various explosion protection technique is coming from the industry. There are very few industry players who produce multiple Ex-rated products and conduct multiple Ex-product research; most of the manufacturer only specialist in their respective areas of expertise. For example, field device manufacturers (field devices like pressure transmitter, I/P converter, control valve, solenoid valve, etc.) like Honeywell, Emerson or Yokogawa only produce Ex-rated field devices, not other devices like Ex-rated junction box or cable gland. Therefore it is difficult for the industrial user to familiarize themselves with various explosion protection techniques for all the control components. This paper helps the industrial user to implement surge protection in the Ex area systematically. It investigates and evaluates the selection criteria and SPD's application requirement for Ex area application with various explosion protection techniques. Safety calculation and proper installation are also presented. It details down all factors involved in implementing the solutions. Two application examples are quoted in detail.

The paper starts the discussion on the need for the SPD in the modern electronics system in the safe area, hazardous area and the critical control loop application; The working principle of the SPD then moves on to the selection criteria for Ex applications and the common explosion protection technique that could be applied together with SPD. Three explosion protection techniques are selected for evaluation and discussion in detail on its application, namely the application of SPD using Exi (intrinsically safe), Exd (flameproof) and Exe (increase safety) protection method in the common explosion protection technique section. The paper then provides two working examples of how the selection and application of SPD using Exi and Exd methods. Safety parameter calculation, distance requirement, temperature consideration, etc., are presented in the examples. The paper ends with a summary of the key area of consideration and criteria in selecting the SPD for Ex application.

2. Why Is Surge Protection Important In The Modern Electronics System?

Modern equipment demands for more application features, faster speed, smaller physical dimensions and lower power consumption. This makes the circuit designer squeeze more transistors into a smaller IC chip, using the material to operate at a lower voltage and lower power consumption. The progress in developing faster, denser and lower power consumption IC Chip makes the modern electronics circuit more vulnerable to the induce transient surges than the early components that use vacuum tube and discrete transistor. Hence, transient surge protection is becoming more important in the modern electronics system. Figure 1 shows the increase of equipment failure rate from the early vacuum tube technology to the modern IC chip caused by transient surges.

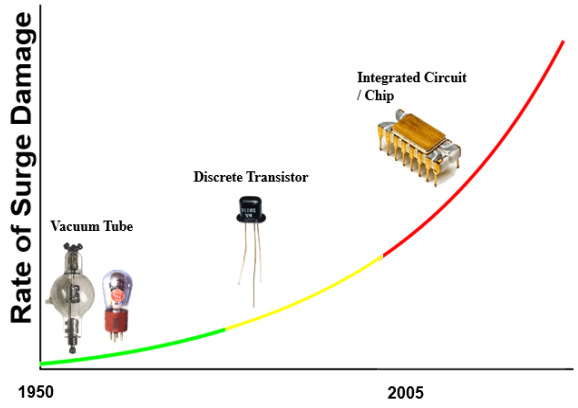


Figure 1: Increased Rate of Equipment Failure [8].

On the other hand, modern electronic systems also rely a lot more high-performance electronic components to execute the critical task that cannot afford to fail. Failure will lead to plant shutdown and production losses, which will cost millions of dollars. For example, emergency shutdown systems (ESD), depending on the process and the safety requirement, can be designed as a standalone, redundant system or triple redundant system to cope with equipment failures. A redundant design plus a logic solver algorithm allows the ESD to detect the real loop failure and decide when to shut down the plant using the voting system. Once the decision is made, the plant is then shut down according to the pre-determined sequence to avoid plant accident or explosion. The redundancy is often achieved by duplicating or triplicating the field sensor, input-output cards, control processors and even the actuators. Almost all the field sensors, except for the wireless system, are connected to ESD by cables, potentially entry

routes for the transient surges. Therefore, SPDs are included in the system design to improve the resilience of such systems from the effect of transients.

When a process plant uses flammable material for its production or during the manufacturing processes, explosive gas or vapor is released created an explosive atmosphere known as hazardous (Ex) area. SPD mounted in this Ex area needs to be carefully selected to protect the equipment from transient surges and at the same time, it does not act as a source of ignition. SPD requirements for the hazardous (Ex) location are often more stringent than in the non-hazardous (non-Ex) locations.

When sizing up the SPD for equipment protection for safe area application, generally, the let-through voltage (should be below the SWC [9–11] of the device to be protected), kA rating, connection method (either series or parallel) of the SPD need to be considered. However, when the application is involved with the Ex area, this is not sufficient as the SPD could become the source to initiate an explosion. The additional consideration for the SPD to use in the Ex area includes the SPD's mounting location, the applicable explosion protection technique, equipment surface temperature, and equipment heat dissipation.

3. Working Principle of SPD

SPD protects the electrical installation, consisting of electrical equipment, consumer unit, wiring and accessories, from electrical power surges known as transient overvoltage. The lightning activity, electricity is interrupted and immediately re-established or the switching of a heavy reactive load could instigate the power surge. Figure 2 illustrates the working principle of a 4-20mA signal line SPD. When there is no transient surge, the SPD does nothing; it is transparent to the circuitry. Whenever a surge appears, the SPD will be triggered instantly. It clamps the SPD's output voltage, depending on the model used, to a pre-determined level and diverts all the excessive current to the ground via a protective earthing terminal.

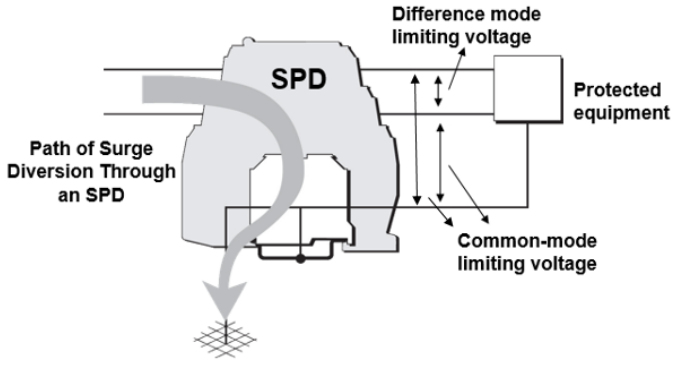


Figure 2: Working Principle of a signal SPD (redrawn from [1] with additional information).

4. Selection Criteria on SPD for Ex Area Applications

When selecting an SPD for use in the Ex area, all the selection considerations and criteria on SPD for the safe area application are applied. Additionally, all other factors that could cause the SPD to trigger an explosion and other ignition sources should be considered as well.

Typically, SPDs are mounted in a panel at the control room to protect the systems and its component against surge. If there is a requirement to protect the field device from surge, additional field mounted SPDs will be required. In other words, the process control loop will require two SPDs. One SPD is mounted as close as possible to the field device (to protect the field device), another one should be mount as close as possible to the controller (to protect the controller) at the control cabinet.

4.1. Types of SPDs

IEC 61643-11:2012 and NFPA 79 standard defined three different types of surge protection devices (SPDs). They are Type 1, Type 2 and Type 3, respectively. The SPD types provide information on where an SPD may be installed, as shown in Figure 3. The energy discharged from the surge is brought down in various stages, starting with the more robust SPD (Type 1) and finer protection Type 2 and Type 3 SPDs. This protection arrangement and co-ordination are represented with the Lightning Protection Zoning (LPZ), which divides the structure based on the lightning strike effects.

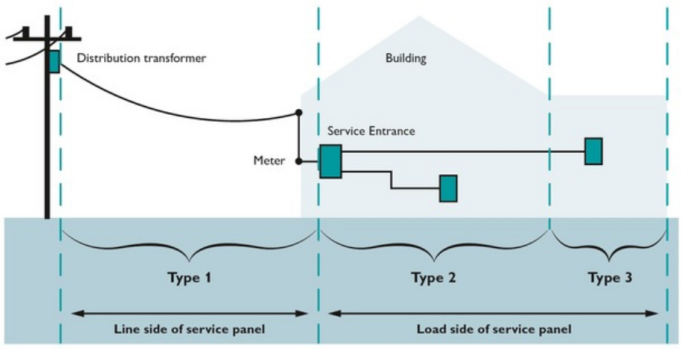


Figure 3: Types of SPD and Its Mounting Location. [12].

Type 1 SPD: Provide lightning protection against transient overvoltage instigates from direct lightning strokes where the lightning discharge voltage could spread from one of the structures earth conductor to the other network conductors. Type 1 SPD is characterized by a 10/350μs surge current waveform per IEC 61643-11:2012

Type 2 SPD: It is the primary protection system for low voltage electrical installations. Protect against transient overvoltage caused by switching or indirect lightning strokes. Typically, it is installed in the electrical distribution board (DB) to prevent transient surges in the electrical installations and protect the loads. Type 2 SPD is characterized by a 8/20μs surge current waveform per IEC 61643-11:2012

Type 3 SPD: Type 3 SPD is aimed to provide local protection on sensitive and critical loads. This SPD has a low discharge capacity. It is only installed as a supplement to Type 2 SPD and it is installed as near as possible to the sensitive loads. They are widely used as hard-wired devices and generally combined with Type 2 SPDs for fixed installations. In many cases, they are built-in and incorporated into a surge-protected socket outlet, surge-protected portable socket-outlets for telecommunication lines and data

protection. Type 3 SPD is characterized by a combination of voltage waveform (1.2/50 µs) and current waveform (8/20 µs).

Table 1 provides a summary of Types of SPD, its test waveform and the current rating, as well as the typical technology used and the Lightning Protection Zone (LPZ). LPZ will be discussed in the following section.

Table 1: Types of SPD, Test Waveform, Applicable Technology and LPZ [13].

	LPZ 0/1	**LPZ 1/2**	**LPZ 2/3**
Typical SPD Installation Point	Service Entrance (e.g. Main distribution board or telecom NTP)	Sub-distribution board or telecom PBX frame	Terminal Equipment (e.g. socket outlet)
Type of SPD	**1**	**2**	**3**
Surge Test Waveform	10/350 µS current	8/20 µS current	Combination 8/20 µS current and 1.2/50 µS voltage
Typical peak test current (per mode)	25KA	40KA	3KA
Typical Technology (Power Line)	Spark Gap	Metal Oxide Varistor (MOV)	Metal Oxide Varistor (MOV)

4.2. Lightning Protection Zone, LPZ

IEC 62305-4:2010 describes the LPZ concept. LPZ identifies and divides a structural area according to the risk level exposed to the Lightning Electromagnetic Impulse (LEMP). It coordinates the type of SPD's kA current rating requirement within the zone. The concept of LPZ taken from IEC 62305-4: 2010 is as shown in Figure 4. ANSI/IEEE C62.41 also has a similar concept, as shown in Figure 5.

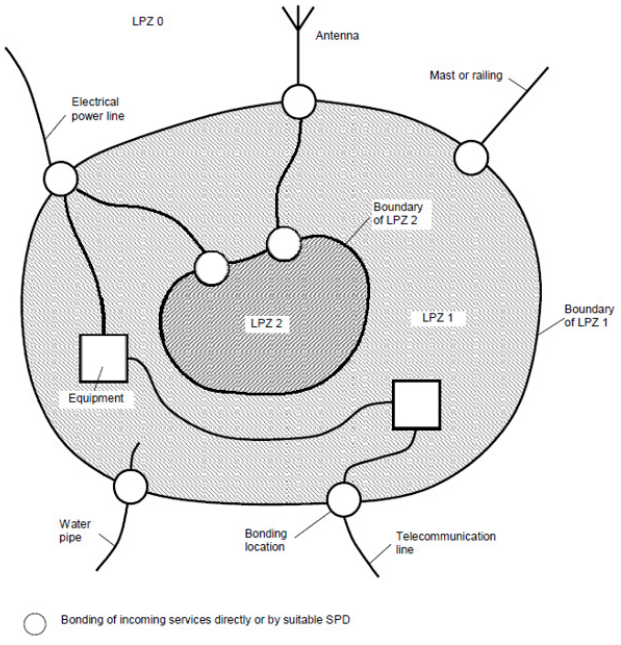


Figure 4: The Principle of LPZ - Division into Different LPZ per IEC 62305 part 4 [14].

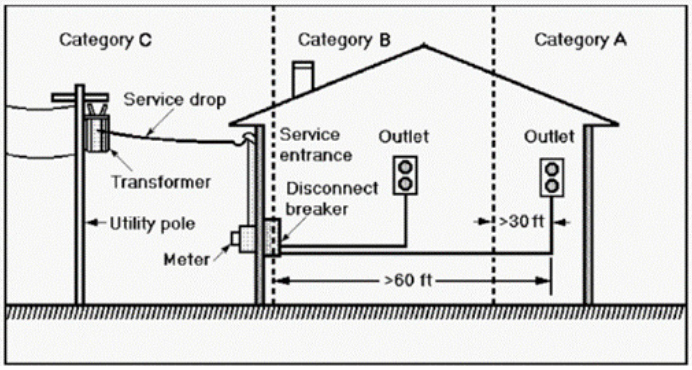


Figure 5: Surge Protective Category as per ANSI/IEEE C62.41 [15].

The LPZs are defined as follows:

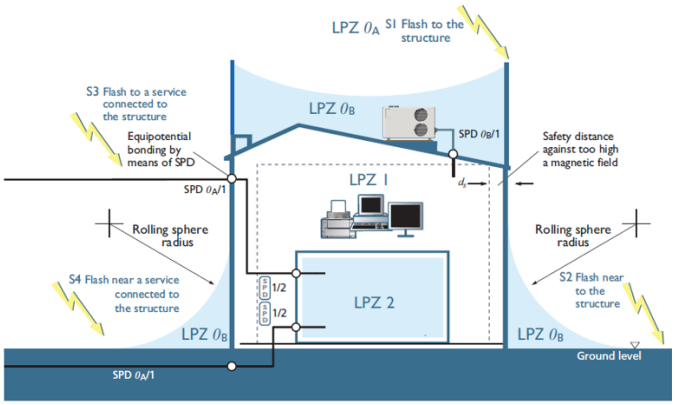
- **LPZ 0_A**: It is an open zone and not protected by the external lightning protection system (LPS). The equipment installed is directly exposed to the direct atmospheric discharges. Hence, the device installed in this zone must be able to handle the entire surge current generated and it is exposed to the whole magnetic induction.
- **LPZ 0_B**: This zone also refers to the open zone but is protected by the external lightning protection system, such as a lightning rod. This method provides direct lightning strikes protection but suffers from total exposure to the magnetic field.
- **LPZ 1**: It is the zone inside a building. At this zone, there is a partial lightning current transmitted from external surges. However, the degree of induced currents is less than zone LPZ 0_A and LPZ 0_B. Between LPZ 0_B and LPZ 1, Type 1 SPD should be installed to protect the downstream equipment.
- **LPZ 2, LPZ n**: This zone refers to an area further inside the building where low surges may occur. The shielding and SPDs, both at the limits of the different zones and protecting the local equipment, allowing further reduction of the induced current.

The LPZs could be view as two categories, as shown in Figure 6, namely two external zones, which are LPZ 0_A and LPZ 0_B; and usually two internal zones, which are LPZ 1 and LPZ 2. If necessary, further zoning can be introduced to further reduce the electromagnetic field and lightning current.

4.3. Lightning Protection Levels, LPL

According to the IEC 62305 standard, LPL can be classified into four levels, namely Level I, Level II, Level III and Level IV. Each level has a set of parameters to predefined the maximum and minimum lightning surge current. Refer to Figure 7, it can be seen that the object or device to be protected will be assigned to one of the four LPL levels before a lightning protection system is implemented. Efficiency in LPL I is the highest at 99% and LPL IV is the lowest at 84%. The maximum lightning surge current set by LPL is used to design and the selection of the SPD. According to the IEC 62305, any lightning surge current parameters exceeding LPL 1 (i.e., 200kA) will be excluded from lightning protection. The chances and probability of occurrence outside the range are less than 2 %. The cost of erecting a lightning protection system, for example, the necessary protection angles, distances from loops, protection profiles and distances from arrestors for

LPL I is much more than that of LPL IV. IEC 62305 Part 1 provides the detail on the LPL probability of occurrence.



LPZ O_A Direct flash, full lightning current, full magnetic field
LPZ O_B No direct flash, but partial lightning or induced current, full magnetic field
LPZ 1 No direct flash, but partial lightning or induced current, damped magnetic field
LPZ 2 No direct flash, or partial lightning, but induced currents, further damped magnetic field
Protected volumes inside LPZ 1 and LPZ 2 must respect safety distances d_s

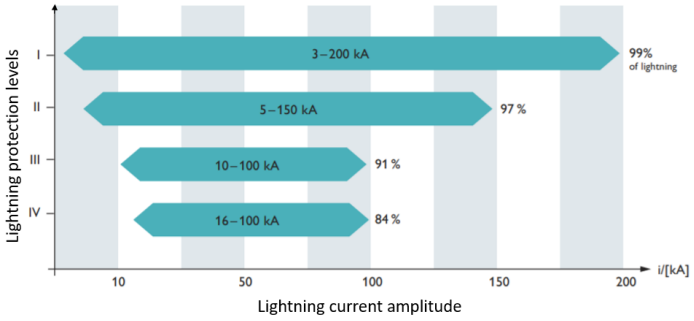


Figure 7: Lightning Protection Levels per IEC 62305-1 with Corresponding Minimum and Maximum Lightning Current Amplitude (redrawn from [16] with additional information).

4.4. Parameter of SPD

Common SPD specifications recognized by IEC 61643 are described below. They are used to select suitable SPD.

- **Nominal Voltage, U_N**
  It is the standard operating supply voltage of the apparatus or system circuit based on the use envisaged for the SPDs.
- **Maximum Continuous Operating Voltage, U_C**
  The maximum voltage that can be permanently applied to the SPD without damaging and triggering it. U_C rating of the SPD is selected such that its triggering voltage is slightly higher than the nominal operating voltage (U_N) of the equipment to be protected. This is to avoid SPD been accidentally triggered during normal operating.
- **Voltage Protection Level, U_P**
  U_p is known as the "Let-Through" voltage or the clamped voltage in the industry. U_P defines the maximum output voltage that could appear at the SPD terminal when tested with a specific waveform per IEC 61643. Generally, U_p will be the maximum output terminal voltage of the SPD when subjected to transient surge. In practice, the SPD's U_p rating should be chosen so that its value is lower than the maximum SWC voltage of the apparatus to be protected.
- **Nominal Load Current, I_L**
  I_L is defined as the maximum continuous RMS load current that is allowed to pass through the SPD without triggering and damaging the SPD, including the capability to withstand the current continuous thermal overload. This term is only applicable when the SPD is connected in series. Since a parallel connection is capable of handling an unlimited load current; hence, this term is not relevant to SPD connected in parallel.
- **Impulse Current, I_{imp}**
  It is the peak current with 10/350μs waveform, direct strike, flowing through the SPD to test mainly for Type 1 SPD.
- **Nominal Discharge Current, I_n**
  Peak current with 8/20μs waveform flowing through the SPD for the test of Type 2 SPD.

4.5. Implementation of SPD in the Ex Area

All guidelines highlighted in sections 4.1 to 4.4 are applicable when applying and installing SPD in hazardous areas. Additional factors described below to ensure the SPD does not become a source of the explosion also need to be considered.

4.5.1 Hazardous Area Classification

Classification of the hazardous area is a risk management tool that handles potentially flammable material to support explosion safety verification. Hazardous areas are defined by the occurrence of the type of hazards present, such as gas, vapor or dust; the probability of that hazard may be present; flammable concentrations on the lower explosion limits (LELs) and upper explosion limits of the explosive substance (UEL) as shown in Table 2. Hence, hazardous area classification is a systematic way to review, analyze and classify plant safety according to the explosive substance. This allows an optimum selection of the electrical apparatus or equipment installed in a hazardous atmosphere. ATEX and IEC standards classify the Ex area into Industry Group, Gas Group, Temperature Classification and Zone. The hazardous area's electrical apparatus and equipment selection is based on the hazardous area classification defined by the ATEX / IEC standard.

Industry Groups:

- **Group I** is the mining industry, where equipment is used in the explosive environment that contains methane gases or vapors of its equivalent hazard, as shown in Table 3.
- **Group II** is the surface industry, as shown in Table 3. It is then further divided into three subgroups, that is Group IIA, Group IIB and Group IIC. Group IIA for the environment containing propane gases or vapors of its equivalent hazard; Group IIB for the environment containing ethylene gases or vapors of its equivalent hazard and Group IIC for the environment containing hydrogen gases or vapors of its equivalent hazard.
- Group III is also for the surface industry but with dust and fibers hazards. They are again subdivided into Group IIIA, IIIB and IIIC according to the nature of the explosive environment, as shown in Table 4.

Table 3 provides the information on Industry Group I and II; and the minimum ignition energy (MIE). MIE is the

minimum energy required to ignite the easiest explosive gas-air mixture.

Table 2: Flammable Properties [17].

Combustible Gas LEL's and UEL's			
		LEL	UEL
Acetylene	C2H2	2.5%	100%
Benzene	C6H6	1.2%	8.0%
Methane	CH4	5.0%	15.0%
Ethane	C2H6	3.0%	15.5%
Ethyl Alcohol	CH2H5OH	3.3%	19.0%
Ethyl Ether	(C2H5)2O	1.7%	36.0%
Ethylene	C2H4	2.7%	36.0%
Hexane	C6H14	1.1%	7.5%
Hydrogen	H2	4.0%	75.6%

Table 3: Industry Group I and II for Mining and Surface Industry [17].

Industry	Gas Group	Substance	MIE	Selection of equipment
Mining Industry Group I	I	Methane	200µJ	Equipment only suitable for mines unless it is additional approved for surface: Marked I/IIC
Surface Industry Group II	IIA	Propane	180 µJ and higher	II; IIA; IIB and IIC certified equipment may be used in an IIA area.
	IIB	Ethylene	60 µJ -180 µJ	II; IIB and IIC certified equipment may be used in an IIB area.
	IIC	Hydrogen	20 µJ – 60 µJ	Only II & IIC certified equipment is allowed to be used in an IIC area.

Dusts and Fibres (Group III) Surface	IIIA	Combustible flying's
	IIIB	Non-conductive dust
	IIIC	Conductive dust

Temperature Classification

Temperature Classification is also known as ignition temperature, auto-ignition temperature (AIT) or T class. It defines the device's maximum surface temperature under a normal or abnormal condition with respect to the ambient temperature of 20°C to 40°C (unless otherwise stated) [18]. Depending on the types of explosion protection techniques used or its combination, for example, Exd, Exe or Exi, the temperature corresponds to the device's external surface's maximum temperature or the maximum temperature of the device's internal surface (for example, Exi device putting into Exd junction box). For a combination of explosion protection techniques used (for example, Exdei – Exd, Exe and Exi are used in the device), the final temperature classification of the device needs to be determined. All flammable materials will have an AIT, the temperature at which ignition will occur without an external source. Apparatus must be selected with its temperature classification below the AIT. This will ensure the Ex environment's apparatus is not exposed to the gas-air mixture with a temperature exceeding its AIT. If several different flammable materials are present within the same area, the material that gives the lowest AIT dictates the device's overall applicable temperature class. The temperature classification only applicable to Industrial Group II (surface industry), as shown in Figure 8. Temperature classes do not apply to Group I (mining) applications. Mining equipment has either a rigid 150°C or 450°C limits [18]. Be noted that the device's temperature class does not represent the operating temperature range of the device, but rather the maximum allowable surface temperature with respect to a default ambient temperature of 20°C to 40°C with the presence of flammable material.

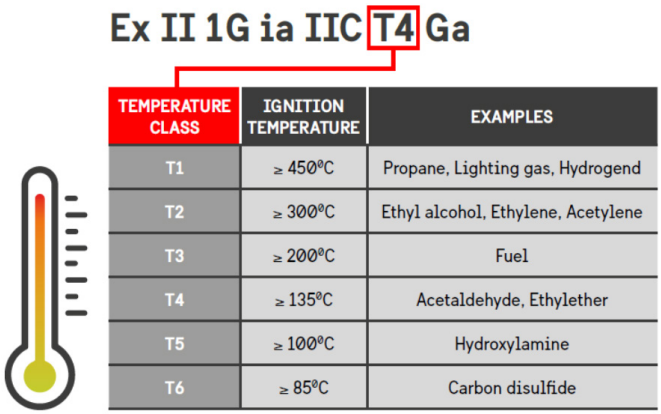


Figure 8: Equipment temperature classification [19].

According to the ATEX directives, all the hazardous area installation apparatus, including SPD, must be marked with its temperature class and assigned an appropriate temperature classification according to the nature of hazard that might be present to maintain its integrity and prevent unexpected explosion due to AIT. For instant, if the hazard present is butane gas, its AIT is 372°C, then all apparatus used in that environment cannot generate a surface temperature exceeding 372°C. In this case, the apparatus rated T2 (300°C) to T6 (85˚C) will meet the requirement. This also means that for this case, all the apparatus used in the butane gas environment cannot have a surface temperature equal to or greater than 372˚C under fault or regular operation. Any equipment that generates 372˚C or more surface temperature could increase the possibility of an explosion.

Table 5: Zone Classification [20].

Zones and Equipment Categories				
Zones		Typical Definitions of Zones (for guidance only)	ATEX Minimum Permissible Equipment Category for Zone	Minimum Equipment Integrity Requirements
Gases and Vapours zones	Dusts			
0	20	Explosive atmosphere is present continuously, for long periods or frequently	1	Equipment must be safe under normal operation, expected and rare malfunction
1	21	Explosive atmosphere is likely to occur under normal operation, occasionally	2	Equipment must be safe under normal operation, expected malfunction
2	22	Explosive atmosphere may occur under abnormal operation and persist for a short period only	3	Equipment must be safe under normal operation.
The higher the probability of an explosive atmosphere occurring and persisting, the lower the allowable potential risk of ignition and the higher the safety integrity requirements of the equipment to be installed. The relationship between zones and equipment categories can be varied following a complete risk assessment				

Zoning Classification

- Zone Classification: Classified concerning the probability of a potentially explosive environment being present in a 3-dimensional region or space and the duration for which it is likely to exist. The available explosion protection methods also depend a lot on the zone classification assessment and its

assignment. The classification of the hazardous area is shown in Table 5.

- This rule of thumb derived from API RP 505 and IP 15 stated if the hazard is continuously present for more than 1000 hours in a year is classified as Zone 0. If the hazard present between 10 to 1 000 hours in a year is classified as Zone 1. If the hazard presents less than 10 hours in a year is classified as Zone 2.

Equipment Marking

- Equipment Marking: All certified apparatus used in the Ex area must carry mandatory marking. The standard uniformly defines the marking requirement. The marking on all equipment intended for use in a potentially explosive environment should provide all necessary and essential information for safe operation and easy identification at the site. Figure 9 and Figure 10 provide a self-explanation and sample of the manufacturer marking according to ATEX Directive 94/9/EC and the IECEx certification scheme, respectively.

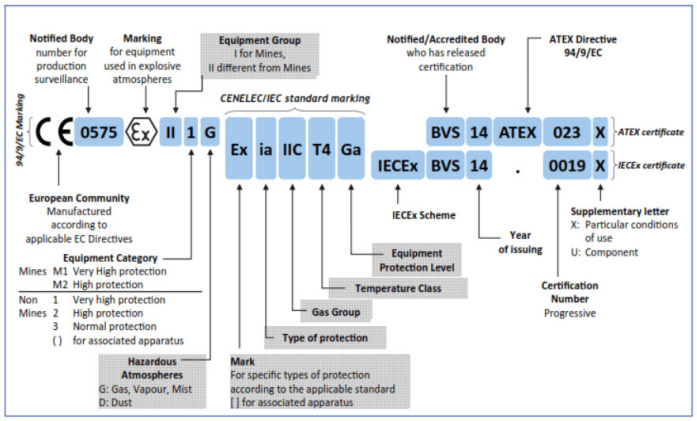
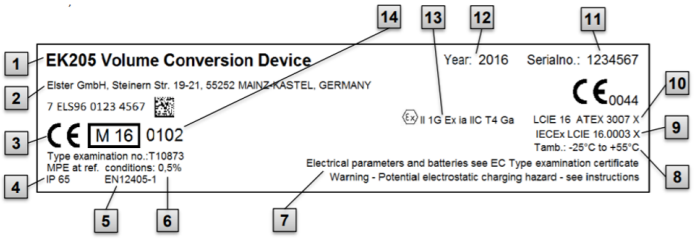


Figure 9: Summary of Apparatus Marking According to ATEX Directive 94/9/EC and IECEx Certification Scheme [21].



1	Type designation	8	Ambient temperature range
2	Manufacturer and address	9	IECEx - EC type examination certificate no.[5]
3	CE marking	10	ATEX - EC type examination certificate no.
4	IP protection class data	11	Serial number
5	Reference to EN 12405-1	12	Year of construction
6	Measurement accuracy data	13	Ex marking
7	Notices	14	Metrology marking

Figure 10: Typical Manufacturer Marking Tag [22].

5. SPD in Common Explosion Protection Technique

For an explosion to occur, three elements must be present at the same time, namely the hazard (flammable or explosive substance), supporter (air or oxygen) and the source of ignition (electrical spark or hot spot). All the explosion protection technique is to remove one or more of this element or contain the explosion so that it will not propagate to a bigger area or outside the enclosure. There are many explosion protection techniques used by the industry. However, only three common techniques and two practical examples based on Exi, intrinsic safety and Exd, flameproof explosion protection technique will be discussed here. Table 6 summarizes the various explosion protection techniques used in the industry, Ex code and its applicable zoning.

Table 6: The IEC Methods of Protection and Permitted Zone of use [23].

Method of Protection	Ex code	Permitted in:		
		Zone 0	Zone 1	Zone 2
Encapsulation	ma	✓	✓	✓
	mb		✓	✓
Oil immersion	o		✓	✓
Powder filling	q		✓	✓
Pressurisation	p		✓	✓
Flameproof	da	✓	✓	✓
	db		✓	✓
	dc			✓
Intrinsic safety	ia	✓	✓	✓
	ib		✓	✓
	ic			✓
Increased safety	e		✓	✓
Type of protection "n"	nA			✓
	nR			✓
	nC			✓

5.1. Intrinsic Safety, Exi

- Permitted to use in Zone 0, 1 and 2.
- Could carry out maintenance repair work without the need to turn off the power supply.
- The use of "Uncertified Simple Apparatus" is allowed when connected through a safety barrier.
- Typical Application with Exi protection is for instrumentation signal, control circuit and other low power apparatus like CCTV.

Exi is a low-energy signaling technique. It brings down the electrical energy level for the Ex area apparatus used below the ignition curve. In this concept, any electrical spark, thermal or hot spot effect in any part of the circuitry intended for use in Ex areas do not have sufficient energy to cause an explosion. Hence, the operation for live maintenance repair work can be carried out and is allowed. The energy levels available for signaling are small but useable and more than adequate for most instrumentation systems.

The minimum ignition energy is obtained experimentally using "SPARK TEST APPARATUS" [24]. The experiments have produced a general agreed "IGNITION CURVES". These are the designer's curves with suitable safety factors for safety parameters calculation (section 6.1 provides the detail on safety parameters calculation). Three curves are resulting from the experiments; they are Resistive, Capacitive and Inductive Ignition Curve, as shown in Figure 11 to 13, respectively.

The resistive ignition curve (Figure 11) provides the entity safety parameter on the maximum open-circuit voltage denotes as Voc and the maximum short circuit current denotes as Isc allowed to go into the hazardous area without causing an explosion for different Gas Group. Any combination of the voltage and current

touching or above the curve could cause an explosion. A value below the curve will not have sufficient energy to cause an explosion. Equations (1) and (2) in section 6.1 provide the detail on the application.

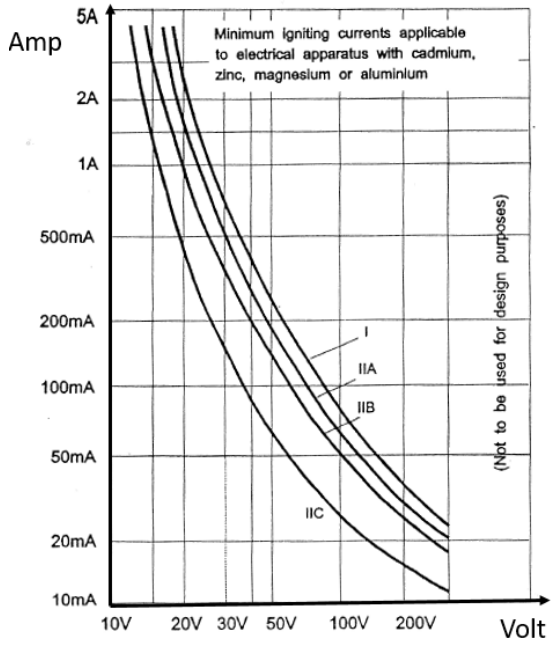


Figure 11: Resistive Ignition Curve [24].

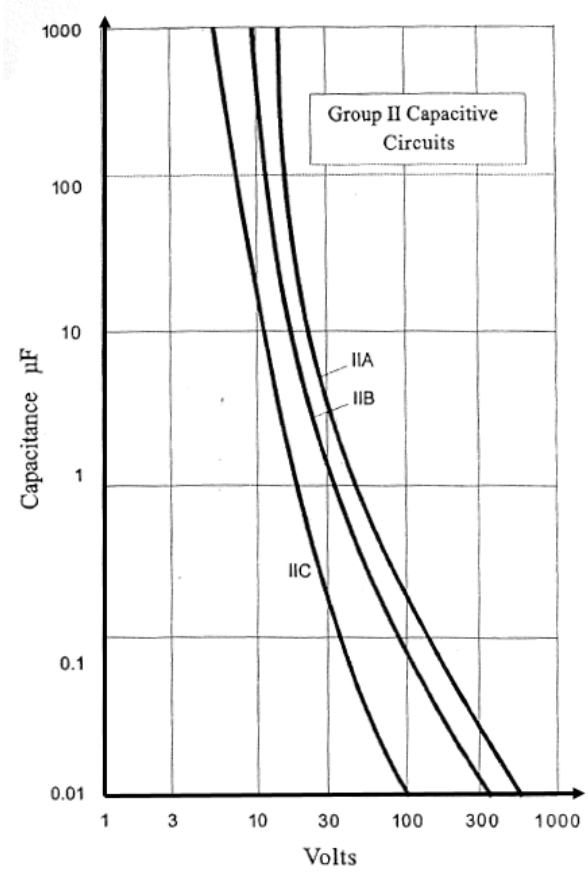


Figure 12: Capacitive Ignition Curve [24].

Capacitive ignition curve (Figure 12) provides the entity safety parameter on the maximum external capacitance (Ca) at the applied voltage allowed to be connected in the hazardous area for different Gas Groups. If the external capacitance (Ca) at the applied voltage is below the curve, it will not have sufficient energy to cause an explosion. Equation (3) in section 6.1 provides the detail on the application.

Inductive ignition curve (Figure 13) provides the entity safety parameter on the maximum external inductance (La) at the applied voltage allowed to be connected in the hazardous area for different Gas Groups. If the external inductance (La) at the applied voltage is below the curve, it will not have sufficient energy to cause an explosion. Equation (4) in section 6.1 provides the detail on the application.

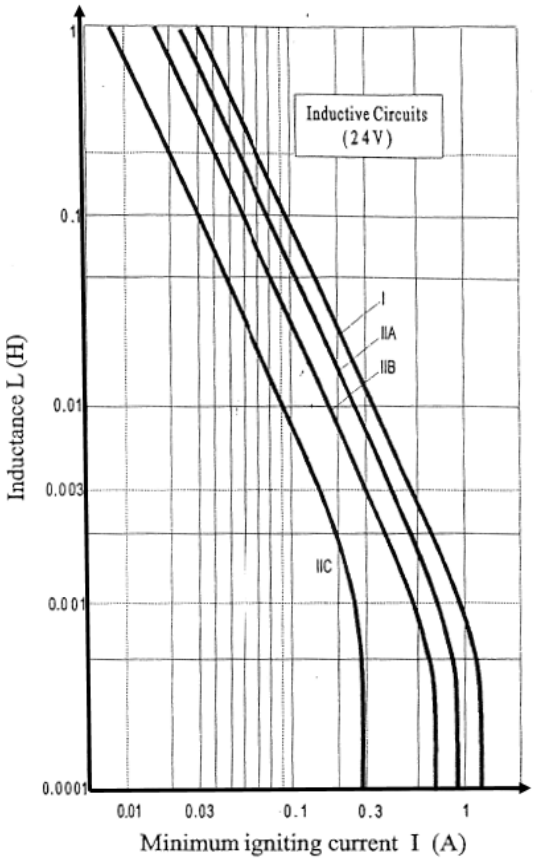


Figure 13: Inductive Ignition Curve [24].

The IEC 60079-11 standard defines three categories of Exi protection. They are "Exia", "Exib" and "Exic" with different fault tolerance levels, as summarized in Table 7. Only Exia category is permitted to apply in Zone 0.

With the introduction of intrinsically safe technique, the uses of “Uncertified Simple Apparatus” are allowed as defined by IEC 60079-25 in a hazardous area without additional certification when the simple apparatus is connected to the field through a certified safety barrier. IEC 60079-25 defines “Simple Apparatus” as an apparatus that does not generate or store energy more than 1,5V, 100mA and 25mW, such as mechanical switches and resistors, thermocouples and PT100. The simple apparatus also required to pass the insulation test at 500 Vrms (one-minute duration) to be used in the Ex area. The application of Simple Apparatus in the Ex area significantly brings down the cost of implementing intrinsically safe systems.

Table 7: Fault Tolerance on Exia, Exib and Exic [23].

Level of protection	Zone of use	Fault Count
Ex ia	Zone 0	Safe with 2 countable faults
Ex ib	Zone 1	Safe with 1 countable fault
Ex ic	Zone 2	Safe with no countable faults (ie safe in normal operation)

5.2. Flameproof, Exd

- Permitted to use in Zone 1 and 2 only.
- The system needs to power down to carry out maintenance work. In other words, no live maintenance is allowed.
- Hot permit is required to carry out maintenance work.
- Wiring enters the Exd enclosure is via a flameproof cable gland or flameproof barrier.
- Typical applications will be rotating machines, lighting, switches (start/stop button) and junction boxes.

The flameproof protection technique relies on housing design to ensure an explosion or ignition inside the enclosure will not propagate outside the enclosure. It is achieved using different flame path designs. There are three possible flame path designs, as shown in Figure 14. The flame path construction plays a significant role. It creates a longer flame traveling distance so that the hazard gas generated during the internal explosion has sufficient time to cool down below the permitted AIT. The enclosure design also must be capable of withstanding the explosion's pressure and, at the same time, preventing the hazard gas from entering the enclosure and contact with the protected equipment inside the enclosure.

Figure 14: Types of Joint

The Exd enclosure usually is supplied in a complete assembly. The manufacturer or its approved agent does all the internal components assembly and wiring work. Both the manufacturer and the approved agent must have a certified facility and trained person to carry out the work. An approved test authority certifies the assembled system as a single entity as system approval (see section 6). Only a single system certificate will be issued. The test procedure for the approval include:

- Ventilation – Consider the available free air volume to ensure sufficient air circulation and ventilation after all the electrical components are fitted and wired.
- Temperature Classification – Consideration of the temperature rises to comply with the temperature classification and define the temperature class;
- Creepage and Clearance: Check all creepage and clearance distance requirements (see Figure 15 on the difference between creepage and clearance distance).
- Pressure Piling – Consideration of the rise in air pressure resulting from the internal enclosure explosion by considering the worst-case scenario on the gas-air mixture.

Future component replacement during maintenance should use back the initially designed component specification. If the replaced component has a larger or smaller dimension than the initially installed part, it will affect the enclosure's internal geometry and ventilation. When a larger than the original component is replaced, pressure piling could happen. If a smaller sized component is replaced, more free air volume will be generated. Meanwhile, the heat conductivity and heat dissipation characteristics of an object are affected by the geometry, ventilation and the material used. This will significantly affect the temperature classification rating. The initial temperature rating of the enclosure could be void.

Drilling and tapping the hole on the enclosure for fitting the cable gland, breather and conduit by a third-party technician is strictly prohibited or else the certification is considered void. The installation of these auxiliary components can only be carried out by the enclosure's manufacturer or certified installer. The type of threads, thread pitch and the clearance tolerance for a particular junction box must be compatible with the cable glands, breather or conduit as flame paths exist at these points.

To improve the ingress protection (IP rating), for example, for weatherproofing applications, non-hardening grease material or non-setting grease or compounds could be applied to the joint. Installation of gaskets to increase the ingress protection is not allowed if they are not part of the original design. It will affect the internal enclosure pressure release efficiency during the explosion. The gasket can only be replaced. Consult the manufacturer and check the installation manual on the type of gasket that could be used.

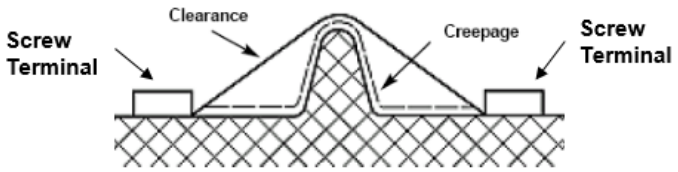


Figure 15: Different between Creepage and Clearance Distance

5.3. Increased Safety, Exe

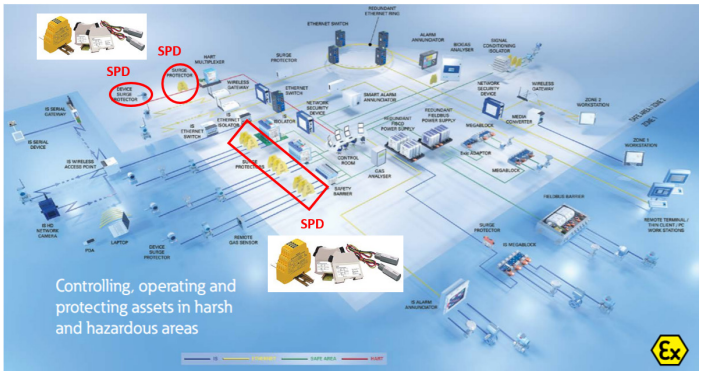
- Permitted to use in Zone 1 and 2 only.
- Required to turn off the power before carrying out maintenance work. In other words, live maintenance is not allowed.
- Minimum IP rating is IP54.
- No hot surface.
- Good mechanical strength (detail refer to EN50014).
- Typical applications will be in lighting, junction boxes and terminal blocks.

Increased Safety, Exe is an explosion protection technique applied to the hazardous area installation that ensures no hot spot or no excessive temperatures and no sparking occurrence from hazardous area electrical equipment and accessories. Equipment and the design that normally cause sparks are excluded from use within this method of explosion protection.

Exe relies on housing construction to withstand mechanical impact with a minimum of IP54 rating. The design of Exe ensures that under any operating mode, the apparatus does not contain normally arcing, sparking devices or hot surfaces that could cause ignition. This protection is realized by using high integrity insulation material. The designer also needs to consider the temperature de-rating factor of the insulation materials and enhanced terminal creepage and clearance distance (made wider than is generally the case in the general industry). The terminal design must be generously dimensioned for the intended cable connection and ensure that conductors are securely fastened to prevent arcing (loose connection can cause arcing). When designing the Exe protection method, a possible spark created by friction should be taken into consideration as well.

6. Method to Protect Assets in Hazardous Area

As shown in Figure 16, the mounting locations of the SPDs in a typical hazardous process plant.



6.1. Application Example on SPD in Exi Control Loop

- Permitted to use in Zone 0, 1 and 2.
- Maintenance work could be carried out with the power on. In other words, live maintenance is allowed
- No further certification is required when using "Uncertified Simple Apparatus" in combination with a certified safety barrier

There are two concepts of Exi applicable to intrinsically safe systems. They are “System Approval” and “Entity Concept”. The meaning of system approval indicates that a single certificate is issued by the approving authority such as BASEEFA or TUV, including the SPD, safety barrier and the field device, as shown in Figure 17 as the components inside the dotted line box. Uncertified control equipment, for example, DCS, PLC or ESD are installing before the safety barrier in the control room; they are not included in the system certificate. In fact, they do not require certification because the energy transmitted to the hazardous area in the field is already clamped down by the safety barrier when there is any fault that happened. Although there is no certification requirement for the uncertified control equipment, they are governed by the installation code, such as the requirement on cable segregation for Ex and Non-Ex cable; creepage and clearance distance for the terminal block; color code and so on. System certificate can only be issued after comprehensive testing and assessment by both the approval authority and the equipment manufacturer. Installation guidelines must be followed strictly according to the certificate/manual's instruction and wiring diagram provided. In the case where the fault happened, all replacement parts will have to be replaced with the exact original model used else; the certification will no longer be valid. This is not so flexible and, in some cases, will post difficulty when any of the parts are obsolete. When the exact replacement part is not available, then replacement on an entirely new set of the system will be required; otherwise, the newly installed component needs to be re-certified.

The most significant advantage of a system certification is to ensure that it is a proven and tested system. The approving authority and equipment manufacturer will be fully responsible for explosion protection in terms of safety. However, this led to the absence of flexibility in changing the component when it is discontinued. The responsibility of the installer is to follow strictly the installation instruction provided by the certificate. As for the entity concept, all intrinsically safe loop components, as shown in Figure 17 as the dotted line box, need to be individually tested and certified to use in Ex area. A separate certificate will be issued individually with all the safety parameters and other special clauses if any. It is the responsibility of the user to match the safety parameter, which will be described later. The utilization of the entity concept provides more flexibility in terms of parts selection, repair and equipment replacement. Regardless of the manufacturer, any product could be used from the safety perspective as long as the safety parameters are satisfied. However, in practice, the device can only be used if they fulfilled both safety and operational parameters. The operational requirement, such as it must have enough voltage and current to drive the field device.

Safety Parameter on Entity Concept

The entity safety parameters are described below:

$$Vmax \geq Voc, \quad (1)$$

Where Vmax is the maximum voltage specified by the field device manufacturer (tested by approval authority) that could be applied safely to the terminal of an intrinsically safe field device, Voc denotes the maximum open-circuit voltage (tested by approval authority) that could appear at the safety barrier’s output terminal when the fault occurs.

$$Imax \geq Isc, \quad (2)$$

Imax is the maximum input current specified by the field device manufacturer (tested by approval authority) that could be applied safely to the terminal of an intrinsically safe field device. Isc denotes the maximum output short-circuit current (tested by approval authority) that could be generated from the safety barrier's output terminal when a fault occurs.

$$Ci + Ccable \leq Ca, \quad (3)$$

Ci is the intrinsically safe field device's internal unprotected capacitance (tested by approval authority). Ccable denotes the cable capacitance.

Ca is the maximum external capacitance (individually tested by approval authority), including all devices at the field that could be connected safely to the hazardous terminal of the safety barrier.

$$Li + Lcable \leq La, \quad (4)$$

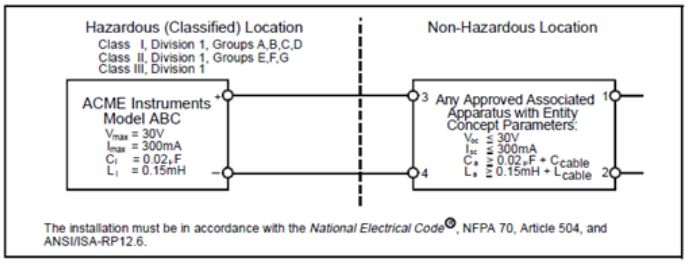
Li is the intrinsically safe field device's internal unprotected inductance (tested by approval authority). Lcable denotes the cable inductance.

La is the maximum external inductance (individually tested by approval authority), including all devices at the field that could be safely connected to the terminal of a safety barrier.

The value for Vmax, Voc, Imax, Isc, Ci, Ca, Li and La could be obtained from the individual intrinsically safe certificate issued by the approval authority. It is usually supplied together with the product, where Ccable and Lcable can be obtained from the cable manufacturer's datasheet.

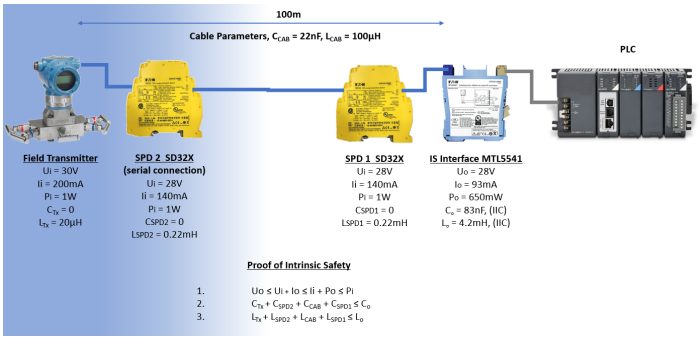
Only the system that fulfills all four mentioned equations, i.e., equations (1), (2), (3) and (4) could be used and safely installed according to the entity concept. However, the user also must take into consideration the operational requirements; for example, it should have sufficient voltage and current to drive the field equipment. Only systems that fulfill both operational and safety requirements will be implemented.

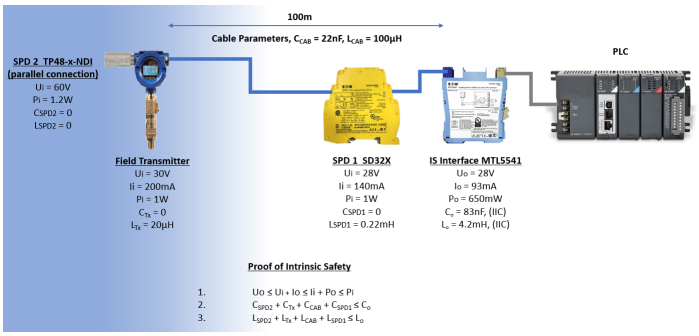
Entity parameters may be found on the control drawing supplied by the equipment manufacturer, along with other pertinent information regarding proper connections, conditions of use, etc. Figure 18 is an example of a control drawing for Exi equipment assessed based on the entity concept.



When an SPD is mounted in the safe area where no hazard substance is present, for example, in the control room, temperature classification and the type of hazard classification such as hazardous gas are not applicable; therefore, it will not be considered here. However, for field mounted SPD where hazard substance is present (Ex area); temperature classification and hazard classification need to be considered.

Figure 19 shows two SPDs connected in series, while Figure 20 shows one of the SPD, field mounted SPD connected in parallel and the other SPD (mounted in a safe area) is connected in series. For the case where all SPDs are mounted in the safe area, then there is no specific requirement for temperature and hazardous area classification. However, temperature and hazardous area classification must be taken into consideration if SPDs are to be mounted in a hazardous area. The hazardous area classification information of the individual Ex rated device can be found in the certifying authority's intrinsically safe certificate. Since all the SPDs in Figure 19 and Figure 20 are part of the Intrinsically Safe loop; therefore, it is a mandatory requirement to perform the entity safety parameter checking (equation (1) to (4)) to ensure its operational safety.





6.2. Application Example on SPD in Exd Control Loop

- Permitted to use in zone 1 and 2 only.
- Need to turn off the power before maintenance work is carried out. In other words, no live maintenance is allowed.
- Hot permit is required to perform any maintenance job.

SPD used in the Ex area is required to mount inside the Exd enclosure. The following information is required to determine the suitable size of Exd enclosure:

- To determine the enclosure's overall size, all the dimension on the individual component to be mounted inside the Exd enclosure is necessary. The system design for heat dissipation and ventilation on top of component size needs to put into the calculation as well.

- Temperature profile or T classification, heat or power dissipation information for the individual component is needed to identify the overall temperature classification of the Exd enclosure. In this case, if the temperature, heat or power dissipation information is not available, the system will need to perform temperature profiling to determine the maximum surface temperature by powered on for an extended period of time using a temperature probe. A bigger system may be required to power on for multiple hours until the temperature rise reaches a stable point where the measured temperature achieves equilibrium, stable and saturated.
- The manufacturer needs to know the type of hazard that could be present, for example, propane, ethylene or others, to determine the kind of joints and its design, either flanged joint, threaded joint or spigot joint. However, in most cases, it is the other way around; the manufacturer states their design is suitable for what kind of hazard. The user's choice is based on the hazard present in their application.
- Hazardous area mounting location or zoning: This is to determine the appropriate type of explosion protection technique. The flameproof (Exd) protection method can only be permitted to apply to Zone 1 or Zone 2 only.

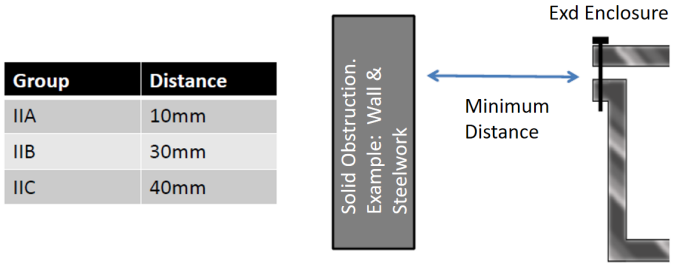


Figure 21: Obstruction of Flame Path (redrawn from [1] with additional information

Figure 22: Sample of the Exd Enclosure

During installation, it is crucial to ensure that there is no blockage such as a wall, brackets, conduit, weather guards or other electrical equipment near the opening of the flame path joint, as shown in Figure 21. The presence of such obstruction could reduce the efficiency of the flame path in the event of an internal explosion. Therefore, the minimum distance requirement between the flame path opening and the obstruction must be satisfied. Figure 21 provides the minimum distance requirement as recommended by IEC 60079-14.

Sample of the Exd enclosures fitted with various Ex and Non-Ex components are shown in Figure 22. It is also a common industrial practice that all the internal assembly and wiring work inside the Exd junction box is carried by the manufacturer or a certified center or an approved agent to ensure compliance.

7. Summary

When selecting SPD for hazardous area application, two criteria involve:

- Equipment protection

The parameters to be considered to protect the equipment from surge:

i. Types of SPD: SPD installation location correspondence to discharge surge energy. This is different from hazardous area classification. It is applied to the installation location for both Ex and non-Ex area.
ii. LPZ and LPL to determine the kA rating of the SPD.
iii. Let-through voltage (Voltage Protection Level, U_P): SPD should be selected where the let-through voltage is below the SWC of the device to be protected.
iv. Bandwidth: only require when it is used for the communication line.
v. Maximum continuous operating voltage, U_C: The U_C rating of the SPD should be selected slightly higher than the U_N (nominal operating voltage) of the equipment to be protected so that during normal operation, the SPD is transparent to the system.
vi. Nominal load current, I_L: For a series connection, the load current needs to be determined so that the SPD's current passing will not damage the SPD. The SPD must be able to withstand the continuous thermal current load. This only applies when the SPD is connected in series whereas SPD connected in parallel is able to handle unlimited load current.
vii. Series voltage drop: This mainly applicable for signal SPD where a series connection is applied. After adding the SPD, the user needs to ensure the voltage drop does not affect the circuit operation.

- Explosion protection

These are the factors that are nothing to do with the equipment protection from surge but rather to ensure the SPD itself does not become the source of ignition.

i. Industry Group: provides information on the type of industry (example: mining or surface industry) and the type hazard substance such as vapor, gas or dust that may be present, and its minimum ignition energy.
ii. Temperature Classification: provides information on the maximum surface temperature of the equipment at fault condition. The temperature classification of the equipment mounted in the Ex area should fall below AIT of the hazardous substances present. For the process loop application, only the SPD mounted at the field will have the temperature classification (Ex are). The other SPD mounted in the control room (non-Ex area) will not have the temperature classification.

iii. Zoning Classification provides information on the probability of a particular hazardous substance being present. The applicable type of explosion protection method depends on the Zoning information and the type of hazard present.

iv. Two approval concepts for safety evaluation could be adopted, entity or system approval. Entity Approval is a technique to combine individually approved apparatus into an intrinsically safe circuit. Two apparatus from different manufacture could be interconnected together as long as it could fulfill the safety calculation spell out by equation (1), (2), (3) and (4), no need for the two apparatus being tested in combination. System Approval is a technique to verify the intrinsic safety of a circuit in which all of the interconnected devices are examined all together as a complete system.

Conflict of Interest

No conflict of interest.

References

[1] T.H. Kuan, K.W. Chew, K.H. Chua, "Practical approach on surge protection device for Ex application," in 2019 IEEE International Conference on Electrical, Control and Instrumentation Engineering, ICECIE 2019 - Proceedings, 2019, doi:10.1109/ICECIE47765.2019.8974781.

[2] J. Kain, Severe Weather 101: Lightning FAQ, NOAA National Severe Storms Laboratory, Nov. 2020.

[3] MTL, Lightning surge protection for electronic equipment-a practical guide, Rev G, Measurement Tecnology Limited, Luton, 2016.

[4] M. Ahrens, LIGHTNING FIRES AND LIGHTNING STRIKES, National Fire Protection Association Fire Analysis and Research Division LIGHTNING FIRES AND LIGHTNING STRIKES, Quincy, 2013.

[5] A. Davis, Surge protector sparks fires instead of preventing them, homeowners say, Click2Houston.Com, 2015.

[6] K.F. Company, Kimberton Fire Company - Chester County, PA, Firehouse Solutions, 2020.

[7] P.˘ Stroch, "ScienceDirect Do not underestimate danger of explosion; Even dust can destroy equipment and kill," Perspectives in Science, **7**, 312–316, 2016, doi:10.1016/j.pisc.2015.11.048.

[8] MTL, Working Principle of Surge Protection Device, Luton, 2019.

[9] F. Martzloff, Surge Withstand Capability of Various Devices, Schenectady NY, 2019.

[10] F. Kloer, G. McDonald, N. Desai, TRICON v10 Nuclear Qualification Project: SURGE WITHSTAND TEST REPORT, Boxborough, Massachusetts, 2008.

[11] I.A. Metwally, A. Gastli, M. Al-Sheikh, "Withstand capability tests of transient voltage surge suppressors," Electric Power Systems Research, **77**, 859–864, 2007, doi:10.1016/j.epsr.2006.07.011.

[12] PHOENIX CONTACT | Information on NFPA 79, 2018 Editi, PHOENIX CONTACT GmbH & Co. KG, Blomberg, Germany, 2018.

[13] Guide to Surge Protection Devices (SPDs): selection , application and theory, BEAMA Ltd, London, 2018.

[14] IEC, INTERNATIONAL STANDARD Protection against lightning-Part 4: Electrical and electronic systems within structures, 2.0, INTERNATIONAL ELECTROTECHNICAL COMMISSION, 2010.

[15] N. Sclater, J.E. Traister, Surge protective categories defined by ANSI/IEEE C62.41, Electrical Engineering Portal, Nov. 2020.

[16] The basics of surge protection, PHOENIX CONTACT GmbH & Co. KG, Blomberg, Germany, 2016.

[17] C.P. Yelland, M.I. Lythe, Selection of explosion protected equipment for hazardous locations, EE Publishers, 2016.

[18] Temperature Classification (T class) - what it means, Control and Instrumentation.Com, Nov. 2020.

[19] IS MY TORCH CORRECTLY CERTIFIED BY ATEX?," LEARN ALL ABOUT ATEX DIRECTIVE 2014/34/EU, 2017, Peli Products Germany, Nov. 2020.

[20] I. Harrison, "HAZARDOUS AREA CLASSIFICATION," When is hazardous area classification needed?, Ex-Consulting, Nov. 2020.

[21] G. T., Understanding Hazardous Area Classification, Instrumentation Tools, Nov. 2020.

[22] Elster, Operating manual Volume conversion device EK205, Mainz-Kastel, Germany, 2016.

[23] P. Saward, Ex ic Intrinsic safety's new protection level, Luton, UK, 2016.

[24] T.H. Kuan, "INTRISIC SAFETY MADE EASY Title: INTRINSIC SAFETY MADE EASY," in Instrumentation & Control Society, ICS, Kuala Lumpur, 2006.

[25] MTL, Controlling, operating and protecting assets in harsh and hazardous areas MTL scope of supply, Measurement Technology Ltd, Luton, UK, 2017.

[26] T. Report, Intrinsically Safe System Assessment Using the Entity Concept, North Carolina, 1995.

[27] MTL, "Entity Concept Calculation," MTL Training Slide, Luton, UK, 2020.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 781-787 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Optimal Irrigation Strategy using Economic Model Predictive Control

Luisella Balbis*

Department of Computer Engineering, University of Bahrain, Kingdom of Bahrain

ARTICLE INFO

Article history:
Received: 05 August, 2020
Accepted: 17 November, 2020
Online: 24 November, 2020

Keywords:
Soil Moisture
System Modelling
Optimization

ABSTRACT

In many countries, irrigation water is one of the major contributors to water scarcity. In the present study, a novel optimized irrigation system which minimizes water consumption in irrigation is presented. The system is based on a predictive control algorithm, which foresees the water need of the crop, and regulates the time and amount of irrigation to maintain the soil moisture around an optimal level, while taking into account system constraints. The predictive feature of the algorithm requires a model of the soil moisture, which is obtained from the actual meteorological data of the Kingdom of Bahrain. The optimization problem is formulated as an Economic Model Predictive Control (EMPC) problem and implemented using MATLAB. The simulation experiments show that the novel system yields a reduction of water consumption around 8% and 16% compared with the PID and On-off controllers, respectively, while maintaining an optimal soil moisture level.

1. Introduction

Reducing water consumption is a fundamental constituent of water management, particularly in water-scarce areas. In many countries, irrigation water is one of the major contributors to water scarcity, and reducing irrigation water while increasing water efficiency is a way to overcome this issue. This work, which is an extension of the paper originally presented in 2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO) [1], looks at different practices used to reduce the use of irrigation water and proposes the use of an advanced optimization technique as possible solution to the problem.

Different mulching practices, irrigation techniques and irrigation strategies can reduce water consumption in irrigation. Mulching refers to the practice of covering the soil with protective material to reduce evapotranspiration. Irrigation techniques (dripping, furrow, sprinkler etc.) refer to the way irrigation water is applied, which also affects evapotranspiration. Irrigation strategies refer to the timing and the amount to irrigation water provided to the crop. Irrigation scheduling, which refers to when and how much to irrigate, is a fundamental concept associated with irrigation strategies [2], [3]. Although mulching practices, irrigation techniques and irrigation strategies are all important tools, in [2] it is found that for optimal results, the irrigation strategy should be improved first, followed by mulching practices and the irrigation technique. Therefore, in the present study we will focus on irrigation strategies and, in particular, on water-efficient irrigation technology.

Technological developments include automated irrigation systems, that irrigate at set times and for a certain amount of time. Although these automated techniques present advantages in terms of increased water efficiency and reduced farmers'intervention, the performances of these open loop control systems in terms of water consumption are not yet optimal. The main reason of their unsatisfactory performance is that they can cause under- or over-watering, since weather and crop conditions are not taken into account. This of course can lead to spoiled crop and water wastage.

Greater expectation is placed on technologies which include sensors and closed-loop watering systems [4]-[6]. The simplest approach uses sensors in an on/off control scheme, where irrigation is turned on or off depending on the reading of a moisture sensor above or below set values.

More complex control strategies, which often are model-based, have been described in different papers [7]-[10]. In general, model-based control strategies require models of the soil moisture dynamics including variables such as climatological data, crop water needs, water saturation etc. [7], [8]. Among such advanced control strategies is Model Predictive Control (MPC), as proposed in the research described in [9] and [10]. The models presented in [9] and [10] are based on the water balance equation, which requires measuring or estimating the initial soil moisture, rainfall,

*Corresponding Author: Luisella Balbis, lbalbis@uob.edu.bh

water capacity of the soil and evapotranspiration (ET). From these measurements, the water balance equation gives an estimation of the soil moisture. In these papers, the MPC algorithm is formulated as an optimization problem, which reduces set point tracking errors.

Similarly to the work presented in [9] and [10], our work describes the development of a soil moisture dynamic model and its application in a predictive control setting. However, the predictive control strategy here proposed is based on Economic Model Predictive Control (EMPC), which aims at maximizing the economic performance of the irrigation system rather than reducing tracking errors. Constraints in the optimization problem are used to guarantee that optimal soil moisture is obtained. Moreover, in the present study the soil moisture model is based on atmospheric conditions of the Kingdom of Bahrain [11].

The remainder of this paper is organized as follows. Section 2 describes the derivation of the soil moisture model based on the water balance equation and the measurement of meteorological data of the Kingdom of Bahrain. Section 3 describe the principles of MPC and presents EMPC as a special case of the MPC policy. In Section 4 the outcomes of employing the EMCP are shown and comparisons are drawn with On-off and PID automated irrigation systems. Finally, conclusion and recommendations are included in Section 5.

2. System Identification

This section presents the grey-box modeling of the soil moisture dynamics using difference equations. Grey-box modeling is a method for identifying the system's dynamics that uses measurement obtained from experimental data combined with physical knowledge of the system. For parameters estimation, state space models are often used, where the state, input and output of the system are related by difference equations.

2.1. Physical Modelling

2.1.1. Water-balance equation

The model developed here is based on the water-balance equation represented graphically in Figure 1 [9], [10]:

$$M(k+1) = M(k) + IR(k) + RAIN(k) + CR(k) - ET_c(k) - DP(k) - RO(k) \quad (1)$$

where k is the sampling time, M is the moisture of the soil, IR is the irrigation water, $RAIN$ is the precipitation, CR is the capillary rise, ET_c is the crop evapotranspiration, DP is the deep percolation and RO is the water runoff.

Equation (1) expresses the variation in soil moisture as the difference between the water influx (capillary rise, precipitation and irrigation) and the water losses (evaporation, transpiration, water runoff and deep percolation) [10].

Geomorphological and meteorological data of the Kingdom of Bahrain show a generally flat land and very scarce precipitations. Therefore, it is reasonable to assume that water runoff, precipitation, and capillary rise are negligible. Under these assumptions, (1) can be reduced to:

$$M(k+1) = M(k) + IR(k) - ET_c(k) - DP(k) \quad (2)$$

Equation (2) reduces the dependency of soil moisture to three variables: irrigation, evapotranspiration and deep percolation.

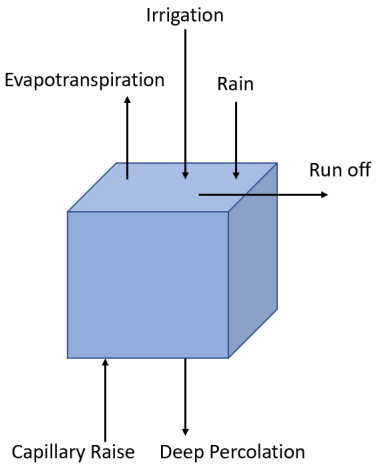


Figure 1: Water Balance Model

2.2. Evapotranspiration Estimation

Evapotranspiration is defined as the movement of water from the soil and the crop into the atmosphere. Many factors influence evapotranspiration rate, such as crop type and growth stage, weather factors (air temperature, wind, humidity etc.), environmental conditions (e.g. soil salinity) and land management (use of fertilizers and pesticides, etc.) [12].

Different approaches exist to estimate evapotranspiration [13]. In the present study, the FAO Blaney-Criddle estimation approach has been used. Despite being a simpler method compared to others, it has been proven to be satisfactory under a variety of weather conditions [14], [15].

Using this method, the actual evapotranspiration ET_c is estimated for a particular type of crop C as:

$$ET_c = K_c \cdot ET_o \quad (3)$$

In (3), the crop coefficient K_c depends on the crop type and the growth stage of the crop. ET_o is the estimation of the evapotranspiration from the reference crop with hypothetical qualities (actively growing green grass of 8–15 cm height). In the present study, ET_o is assumed to be affected by climatic parameters, according to the following empirical FAO Blaney – Criddle equation:

$$ET_o = p \cdot (0.457 \cdot T_{mean} + 8.128) \quad (4)$$

where T_{mean} is the average daily temperature [°C] and p is the average annual percentage of daytime.

2.3. Deep Percolation

Deep percolation (DP) is a hydrologic process, where water penetrates below the roots of the crop. The quantity of percolation during irrigation depends on the characteristics of the soil as well as the soil humidity.

In the present study, DP is assumed to be directly proportional to M with proportionality constant K_p: when M increases, DP

increases due to differential pressure [16]. This is expressed by the following equation:

$$DP(k) = K_p \cdot M(k) \tag{5}$$

In (5), K_p is a constant that depends on the type of soil (sandy, clay, silt, loamy, etc)

Substituting (3) and (5) in (2) we get a difference equation which leads to a model of the water balance dynamic in the form:

$$M(k+1) = K_s \cdot M(k) + IR(k) - K_c ET_o(k) \tag{6}$$

where K_s=1-K_p.

2.4. Model Structure

Equation (6) gives a mathematical model which contains two parameters, K_s and K_c. In our work, these parameters are unknown, meaning that the value of crop coefficient and the type of soil used in our experiment are unknown. Therefore, a gray-box system identification process, which uses a the first principles model of the system and data coming from direct system measurements, was used to estimate the values of the unknown parameters.

The linear difference equation (6) can be re-written in the form of a discrete state space model as:

$$M(k+1) = [K_s]M(k) + [1]IR(k) + [-K_c]ET_0(k) \tag{7}$$

$$M(k) = [1]M(k) \tag{8}$$

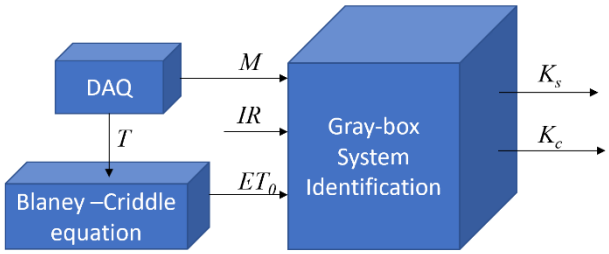
where $M(k)$, $IR(k)$, $ET_0(k)$ are the state vector, the input signal, and the disturbance of the system.

The state-space model (7) and (8) presents two parameters, K_s and K_c, that need to be estimated from measured data.

2.5. Parameters Estimation

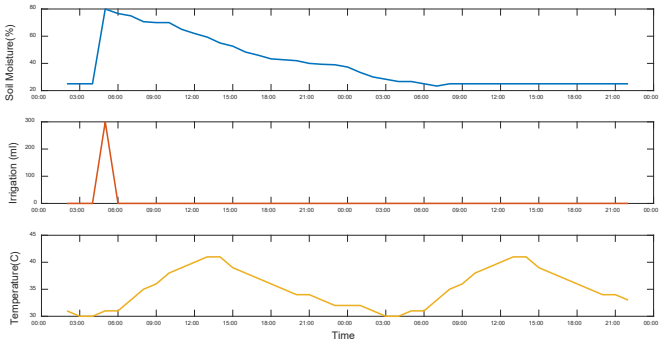
Parameter estimation was carried out using Matlab System Identification Toolbox R2018a, which provides a tool for deriving mathematical models of dynamic systems given measured input-output data. The toolbox was used to estimate the values of K_s and K_c, through a gray-box model estimation process.

Figure 2 shows the process of model identification.



A set of measurements (see Figure 3), called identification data, was collected using a data acquisition system implemented on an Arduino Uno microcontroller board.

The experimental setup consists of a temperature sensor LM35, and a soil moisture sensor FC-28. The temperature sensor outputs a voltage which is directly proportional to the instantaneous temperature. The soil moisture sensor measures the volumetric soil moisture as a function of electric conductivity in the soil. In our experiment, the position of the soil moisture sensor was fixed at 10 cm depth. The sensor was calibrated by taking two measures; one with the soil entirely wet and one with the soil entirely dry. The moisture was expressed in percentage, with the values of the measurements mapped from 0 (dry soil) to 100 (soaked soil).



For our experiments, the soil moisture M and the temperature T near the moisture sensor were synchronously measured every hour over a period of two days. The irrigation IR was modeled as a pulse signal, since it provides a great amount of water but for a limited time [10].

The Blaney-Cradle equation block in Figure 2 estimates ET_o using (4). In our experiment, the value of the variable p in (4) was substituted by the actual percentage of daytime for the time of the year in which the experiment was conducted. In the same way, the variable T_{mean} was substituted by the actual temperature T measured at time k.

2.6. Model Validation

To validate the model obtained in the previous section the output of the simulated model was compared with measured data.

For this purpose, another set of input data was measured and then inputted to the derived model of the system. The input data set used in the validation is shown in Figure 4.

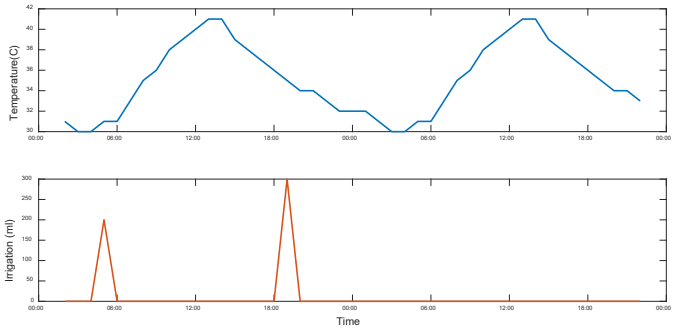


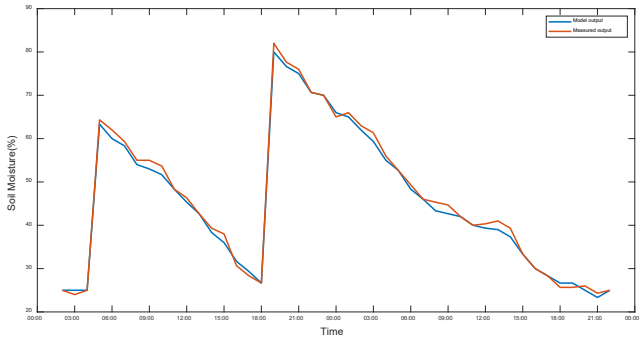
Figure 4: Validation Data

Figure 5 shows the difference between the output of the model and the measured soil moisture. The plots show a good agreement between simulated and measured outputs, with negligible time delay between the irrigation event and the change in soil moisture dynamics.

A further comparison between simulated and measured outputs to verify the reliability of the identified model was done calculating the Mean Absolute Error (MAE) as follows:

$$MAE = \frac{\sum_{k=1}^{T} |\widehat{M}_k - M_k|}{T} \quad (9)$$

where $\widehat{M}_k - M_k$ is the difference between the predicted and the measured output value and T=48 is the total number of samples. The obtained *MAE* value of 2% is considered an indication of a valid model.



Measured and Simulated Soil Moisture

3. Economic MPC

3.1. EMPC Basic Concepts

The term Predictive Control refers to a control design method characterized by [17]:

- use of a model to forecast the system output at future time instants (prediction horizon);
- computation of an input control sequence that optimizes a given objective function;
- receding horizon strategy, where the input control sequence is computed by taking into account the predicted outputs, but only the first control signal of the sequence is applied to the system.

The different algorithms which belongs to the Predictive Control group present differences in terms of the chosen system models (impulse, step, state-space, CARIMA, fuzzy models, etc.), disturbance (constant, decaying, filtered white noise etc.) and in the cost-function to be minimised.

A general MPC mathematical formulation is given by the following optimization problem [18]:

$$J = min \sum_{i=0}^{H_c-1} f\left(x_{k+i|k}, u_{k+i|k}\right) \quad (10)$$

where the variables $x \in R^n$ and $u \in R^m$ denote the state and control action at the instant k and H_c denotes the control horizon. The notation $k+i|k$, with $k \in Z$, indicates the estimation of the future variables based on the measurements at current time instant k.

Traditional MPC controllers minimize a quadratic cost function which tracks output and manipulated variable references.

The general prediction model can be expressed by the discrete equation:

$$x_{k+i+1|k} = g(x_{k+i|k}, u_{k+i|k}) \quad (11)$$

where the predicted states depend on the future states and control actions from instant k to instant $k+H_p$.

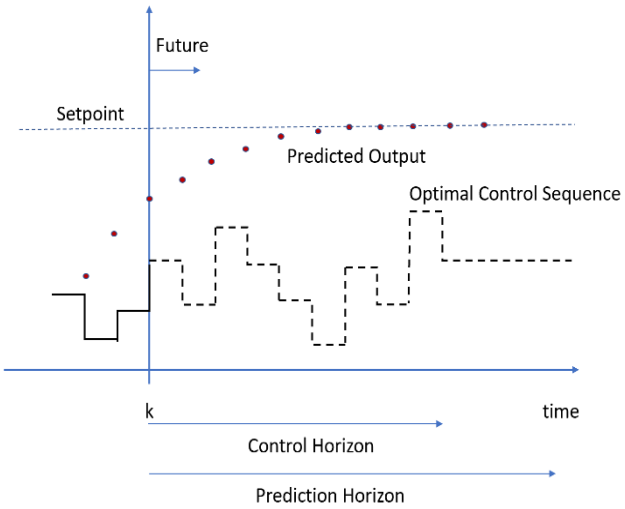
The minimization problem is usually subject to a set of constraints which have to be satisfied over the predictive horizon. These constraints are due to physical or operational limitations of the real system and they take the general form:

$$h(x_{k+i|k}, u_{k+i|k}) \leq 0 \quad (12)$$

At each time instant, MPC measures or estimates current state x_k and predicts the system state evolution over the prediction horizon H_p using the current state and the model of the system.

The optimal control sequence over the control horizon H_c is found solving the minimization problem (10). Only the first control action is applied to the system, and the remaining results are discarded.

In absence of disturbances and plant-model mismatch, the optimization problem could be solved as an open-loop problem, and the input sequence found at k=0 could be applied to the system for all $k \geqslant 0$. In a more realistic scenario, the behavior of the real system differs from the predicted model output. To cater for this mismatch, the optimal control problem is solved again at the next sampling time. Using the state measurement/estimation at time k+1, the cycle of prediction and optimization is repeated, moving the horizon forward (Figure 6).



Receding Horizon

In recent years, the EMPC approach has been proposed with successful applications in various fields [19].

Contrary to traditional MPC, EMPC is employed to maximize profitability, rather than minimize tracking errors. This can be done by defining a cost function that, directly or indirectly, measures the process economics [20]. The formulation of cost function depends on the particular applications, such as the operating profit, production rates, product selectivity, and product yield. In EMPC the cost function, as well as the plant dynamics and the constraints can be linear or nonlinear. Therefore, generic performance cost functions and nonlinear optimization algorithms are used to solve the minimization problem. The structure of the basic EMPC algorithm is shown in Figure 7.

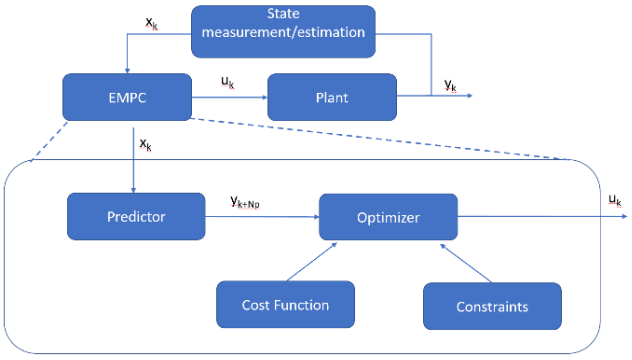


Figure 7: Basic EMPC Structure

3.2. EMPC Problem Formulation

In our work, we utilize an EMPC approach to minimize water consumption while maintaining the soil moisture close to an optimum value. We consider a linear economic stage cost function; additionally, both plant dynamics and constraints are linear. Because of this, our formulation of the EMPC problem is linear. The fundamental steps of the devised EMPC algorithm are the following:

- At the time k, measure $M(k)$ and compute the optimal control sequence $\overline{IR}(k)$ by solving a stochastic optimization problem over the control horizon N_c.
- Apply only the first computed value $IR(k/k)$ as input to the system.
- At the time k+1, measure $M(k+1)$ and repeat the optimization.

At each time instant k, a vector $\overline{M}(k)$ of the moisture predicted values is computed, using the model given by (7) and (8):

$$\overline{M}(k) = [M(k+1/k) \quad \dots \quad M(k+N_p/k)] \tag{13}$$

The stochastic optimization problem becomes deterministic if we assume that ET_o does not change over the prediction horizon N_p., i.e.:

$$\overline{ET_0}(k) = [ET_0(k) \quad \dots \quad ET_0(k)] \tag{14}$$

where $ET_o(k)$ is calculated using (4) and the current measured temperature $T(k)$.

Let's define the future control signals as the vector:

$$\overline{IR}(k) = [IR(k/k) \quad \dots \quad IR(k+N_c-1/k)] \tag{15}$$

Combining (13), (14) and (15) with the model (7) and (8) we obtain:

$$\overline{M}(k) = \Theta M(k) + \Phi \overline{IR}(k) + \Psi \overline{ET_0}(k)) \tag{16}$$

where:

$$\Theta = \begin{bmatrix} K_s \\ \vdots \\ K_s^{Np} \end{bmatrix} \tag{17}$$

$$\Phi = \begin{bmatrix} -K_c & 0 & \dots & 0 \\ -K_c K_s & -K_c & 0 & 0 \\ \vdots & \vdots & -K_c & 0 \\ -K_c^{Nc} K_s & -K_c^{Nc-1} K_s & \dots & -K_c \end{bmatrix} \tag{18}$$

$$\Psi = \begin{bmatrix} -K_c & 0 & \dots & 0 \\ -K_c K_s & -K_c & 0 & 0 \\ \vdots & \vdots & -K_c & 0 \\ -K_c^{Np} K_s & -K_c^{Np-1} K_s & \dots & -K_c \end{bmatrix} \tag{19}$$

The purpose of the designed control system is to obtain optimal economic performance. The following stage cost is assumed to reflect the nominal economic cost of irrigating the system over one sampling period, that is:

$$L(k) = IR(k) \tag{20}$$

The cost function does not consider the cost of operating the valves or the cost of pumping water, since their minimization is not considered an objective in the present study.

The resulting optimization problem adopted in our EMPC controller formulation is:

$$\min_{IR(k),v(k)} \sum_{i=0}^{N_c-1} IR(k+i/k) + \sum_{i=0}^{N_p} \rho v(k+i) \tag{21}$$

In addition to the stage cost, the objective function contains the terms $\rho v(k+i)$ which represent soft economic constraints.

The minimization problem (21) is subject to:

$$\overline{M}(k) = \Theta M(k) + \Phi \overline{IR}(k) + \Psi \overline{ET_0}(k)) \tag{22}$$

$$\overline{M}(k) \le \overline{M}_{opt}(k) + \overline{V}(k) \tag{23}$$

$$\overline{M}(k) \ge \overline{M}_{opt}(k) - \overline{V}(k) \tag{24}$$

where $\overline{M}_{opt}(k), \overline{V}(k)$ are vectors of size N_p defined as follows:

$$\overline{M}_{opt}(k) = [M_{opt} \quad \dots \quad M_{opt}] \tag{25}$$

$$\overline{V}(k) = [v(k/k) \quad \dots \quad v(k+N_p/k)] \tag{26}$$

The value M_{opt} represents the soil moisture that should be kept constant for optimal operating conditions. The variable $v(k)$ is a slack variable that is introduced to allow for small variation of the soil moisture $M(k)$ around the optimal value. That is, we nominally want to satisfy $M(k)$=M_{opt}, and when these constraints are violated, an additional cost $\rho v(k+i)$ is paid, weighted by the penalty parameter ρ. The choice of the parameter ρ is critical; when it is chosen too small, the value of the soil moisture might be far from the optimal one; when it is chosen too big, we might incur into infeasibility issues.

4. Simulation Results

The performances of PID controller, On-off controller and the proposed EMPC are compared. Not surprisingly, the behavior of the three control techniques is very different, as it can be observed in Figure 8 and Figure 9.

The On-off controller uses the measurement from a moisture sensor to switch the irrigation pump on or off based on whether the measurement is below or above maximum and minimum limits. The output from the device is either 100% on or off, with no middle

state. Figure 8 shows the outcome using the On-off controller. In this example, the soil moisture minimum and maximum levels are fixed at 50% and 70%. As expected, the output presents a variation of the moisture level (overshoot) around the optimal value of 65%. Reducing the thresholds could help improving the underwatering issue, but it would imply a higher switching of the actuators and a persistent overwatering problem.

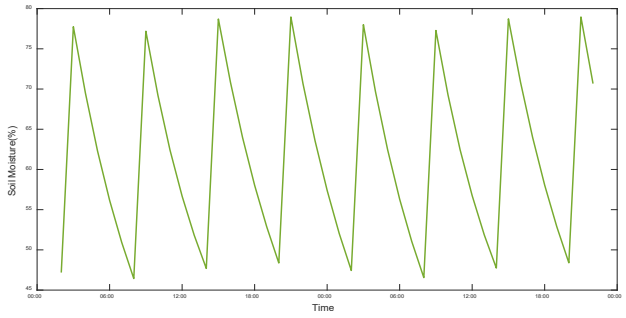


Figure 8: Soil Moisture with On-Off Controller

In Figure 9 the outcomes of the PID and EMPC approaches are shown. It can be noticed that the EMPC maintains the value of the soil moisture around the optimal value better that the PID controller in presence of disturbances, since it considers the effect of $ET_o(k)$ on the moisture level over the prediction horizon.

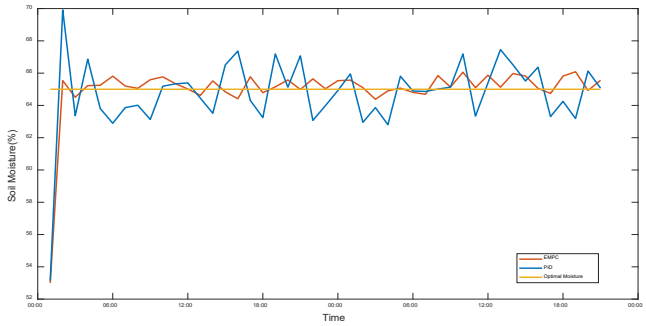


Figure 9: Soil Moisture Dynamics with EMP and PID Controllers

The three irrigation systems are also compared in terms of accumulated error, defined as:

$$CE = \sum_{k=1}^{T} \left| M_{opt_k} - M_k \right| \tag{27}$$

where the error is the difference between the optimal soil moisture and the simulated soil moisture. The cumulative error gives an indication of the ability of the controller to keep the soil moisture close to the optimal value. As it can be seen in Figure 10, the EMPC has smaller cumulative error compared to the other irrigation methods.

The total water consumption over the 48 hours period is around 2400 ml for the On-off controller, 2160 ml for the PID based system and 2000 for the EMPC. Therefore, the EMPC yields a reduction of water consumption around 8% and 16% compared with the PID and On-off controllers, respectively, while keeping the soil moisture within optimal values.

The simulations demonstrate that the EMPC scheme applied to an irrigation system provides benefits compared to conventional irrigation methods, since it maintains the moisture level near the nominal optimum, while providing economic benefit.

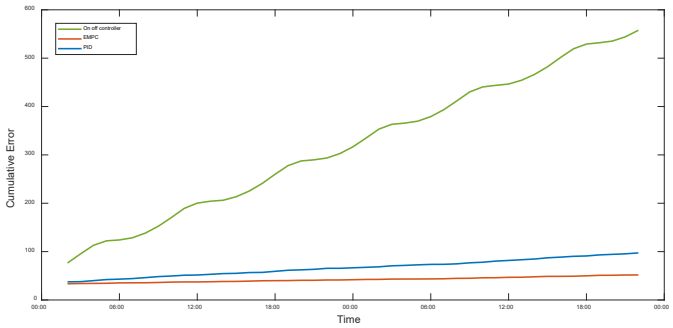


Figure 10: Cumulative Error for On-Off Controller, PID and EMPC

5. Conclusions

This work presents the process of developing an optimal controller regulating the time and amount of irrigation to maintain the soil moisture around an optimal level, while taking into account system constraints. The proposed control strategy, based on EMPC techniques, was implemented in MATLAB and results were compared with traditional irrigation methods based on PID and On-Off controllers. The results suggest that EMPC applied to irrigation systems can lead to increased efficiency and reduced water consumption.

Further improvement of this work would require a more accurate model of the system. A current limitation is given by the model developed, which can be improved by collecting sensors measurements over a sparse area and under different operating conditions, and using more complicated approaches to estimate ETo such as the FAO Penman Montheith technique.

Furthermore, developments on this work could explore the implementation of the proposed EMPC algorithm on hardware and evaluate the feasibility of such strategy in real time.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] L. Balbis, "Economic Model Predictive Control for Irrigation Systems," in 2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO), Manama, Bahrain, 2019. https://doi.org/10.1109/ICMSAO.2019.8880332.

[2] A. D. Chukalla, M. S. Krol, A. Y. Hoekstra, "Marginal cost curves for water footprint reduction in irrigated agriculture: guiding a cost-effective reduction of crop water consumption to a permit or benchmark level" Hydrol. Earth Syst. Sci., **21**(7), 3507–3524, 2017. https://doi.org/10.5194/hess-21-3507-2017, 2017.

[3] A. D. Chukalla, M. S. Krol, A. Y. Hoekstra, "Green and blue water footprint reduction in irrigated agriculture: effect of irrigation techniques, irrigation strategies and mulching" Hydrol. Earth Syst. Sci., **19**(12), 4877–4891, 2015. https://doi.org/10.5194/hess-19-4877- 2015.

[4] R. Aqeel-ur, A.Z. Abbasi, S. N. Islam, Z. A. Shaikh, "A Review of Wireless Sensors and Networks Applications in Agriculture" Computer Standards & Interfaces, **36**, 263-270, 2014. https://doi.org/10.1016/j.csi.2011.03.004.

[5] P. Alagupandi, R. Ramesh and S. Gayathri, "Smart irrigation system for outdoor environment using Tiny OS," in 2014 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC), Chennai, India, 2014. https://doi.org/10.1109/ICCPEIC.2014.6915348.

[6] M.W. van Iersel, M. Chappell, J. D. Lea-Cox, "Sensors for Improved in Efficiency of Irrigation in Greenhouse and Nursery Production",

HortTechnology, **23**(6), 735-746, 2013. https://doi.org/10.21273/HORTTECH.23.6.735
[7] S. Kulkarni, "Innovative Technologies for Water Saving in Irrigated Agriculture" International Journal of Water Resources and Arid Environments, **1**(3), 226-231, 2011. ISSN 2079-7079
[8] J. B. Million, T. H. Yeager, J. P. Albano, "Evapotranspiration-based irrigation scheduling for container-grown Viburnum odoratissimum" HortScience, **45**, 1741-1746, 2010. https://doi.org/10.21273/HORTSCI.45.11.1741
[9] S. K. Saleem, D. K. Delgoda, S. K. Ooi, K. B. Dassanayake, L. Liu, M. N. Halgamuge, H. Malano, " Model Predictive Control for Real-Time Irrigation Scheduling", in 2013 4th IFAC Conference on Modelling and Control in Agriculture, Horticulture and Post Harvest Industry, Espoo, Finland, 2013. https://doi.org/10.3182/20130828-2-SF-3019.00062.
[10] C. Lozoya et al., "Model Predictive Control for Closed-Loop Irrigation", in 2014 19th World Congress, Cape Town, South Africa, 2014. https://doi.org/10.3182/20140824-6-ZA-1003.02067.
[11] L. Balbis and A. Jassim, "Dynamic Model of Soil Moisture for Smart Irrigation Systems," in 2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Sakhier, Bahrain, 2018. https://doi: 10.1109/3ICT.2018.8855748.
[12] M. Todorovic, Crop evapotranspiration, in book: Water Encyclopedia: Surface and Agricultural Water, John Wiley & Sons Publisher, 2005.
[13] R.G. Allen, L.S. Pereira, D. Raes, M. Smith, "Crop evapotranspiration: guidelines for computing crop water requirements" in FAO Irrigation and Drainage, FAO, Rome, Italy ,1998.
[14] H. R. Fooladmand, "Evaluation of Blaney-Criddle equation for estimating evapotranspiration in south of Iran" in African Journal of Agricultural Research. **6**(13), 3103-3109, 2011.
[15] O. E. Mohawesh, "Spatio-temporal calibration of Blaney-Criddle equation in arid and semiarid environment" in Water Resour. Manage., **24**(10), 2187-2201, 2010. https://doi.org/10.1007/s11269-009-9546-7
[16] S.K. Ooi, I. Mareels, N. Cooley, G. Dunn, G. Thoms, "A Systems Engineering Approach to Viticulture On-Farm Irrigation" in 2018 17th World Congress. Seoul, Korea, 2008. https://doi.org/10.3182/20080706-5-KR-1001.01618
[17] E.F. Camacho, C. Bordons Introduction to Model Predictive Control, in book: Model Predictive Control. Advanced Textbooks in Control and Signal Processing, Springer, 2007.
[18] J. M. Maciejowski, Predictive Control with Constraints, Prentice Hall, 2002.
[19] T. Tran, K. Ling, J. M. Maciejowski, "Economic Model Predictive Control - A Review", in 2014 31st International Symposium on Automation and Robotics in Construction and Mining, ISARC, Sydney, Australia, 2014. https://doi.org/10.3182/20080706-5-KR-1001.01618.
[20] J. B. Rawlings, D. Angeli, C. N. Bates, "Fundamentals of economic model predictive control" in 2012 51st IEEE Conference on Decision and Control (CDC), Maui, Hawaii, USA, 2012. https://doi.org/10.1109/CDC.2012.6425822.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 788-800 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Dependency Head Annotation for Myanmar Dependency Treebank

Hnin Thu Zar Aye*, Win Pa Pa

Natural Language Processing Lab, University of Computer Studies, Yangon, Yangon, 11411, Myanmar

ARTICLE INFO

Article history:
Received: 22 September, 2020
Accepted: 17 November, 2020
Online: 24 November, 2020

Keywords:
Dependency head
Universal Dependencies
Treebank
Annotation schemes

ABSTRACT

Complete manual annotation of dependency treebank needs resources like annotators and annotation tools and takes long time and has high possibility of inconsistent annotations for free word order languages such as Myanmar. This paper describes a dependency head annotation scheme with Universal part-of-speech and Universal Dependencies for Myanmar dependency treebank. Currently 22,810 sentences and 680,218 tokens were annotated from three corpora for Myanmar dependency treebank. Some language specific issues are also described with examples. Raw syntactic structures were annotated automatically by UDPipe according to the Universal Dependencies based on Universal-part-of-speech tag scheme. Then unsupervised annotated dependency head structures have been manually updated in post processing. To be reliable and speedy post process with reduced errors for manual updating, selected sentences were added to the training data after being updated. After that the model has been retrained and the remaining sentences were parsed by UDPipe. Post processing was repeated until all sentences were updated. Some specifications of dependency annotation schemes in sentences encountered in post processing are presented with examples. For parsing performance of annotated data, cross validation tests and parsing experiments were performed. Moreover, annotated treebank data have also been evaluated by CoNLL 2017 evaluation script for parsing performance. Results of parsing experiments and evaluation are also reported by unlabeled and labeled attachment scores and demonstrated that the proposed method is a suitable way for building Myanmar dependency trees. Moreover, syntax structures of treebank are also analyzed and syntax information is also presented. This dependency head annotation for dependency treebank is the first work for Myanmar language as far as we know.

1. Introduction

A treebank annotated with syntax or dependency structure is an essential and important resource for natural language processing systems in any language. Treebank annotated dependency structures is called dependency treebank. While phrasal constituents and syntax rules could not provide a direct role in sentences, syntactic dependency information of a sentence can describe directed grammatical relations between words. Moreover, dependency grammar is also able to deal with morphologically rich and relatively free word order languages. Dependency treebank is also critical resource in any language to develop natural language processing applications [1].

Many dependency treebanks have been constructed manually for many languages such as Czech, German, Danish, French, Portuguese, Estonian, Russian, Dutch, Danish, Turkish, Basque, Italian, English [2], Norwegian [3], Finnish [4], Romanian [5], Ancient Greek and Latin [6], Vietnamese [7],

Annotating dependency syntactic information in sentences is still a hard task for Myanmar having free word order nature. Moreover, currently there is still low resource for syntactic information for Myanmar language.

The Myanmar grammar is different from other languages of ASEAN countries such as Thailand, Vietnam, Malaysia and these languages already have treebank resources. Myanmar has been similar structures with the other SOV order languages such as Japanese, Chinese, and Korean and also a head final language. According to these properties, for Myanmar, a dependency-based head finalization has been proposed for statistical machine translation (SMT) in [8]. Although the proposed method was being able to improve a baseline SMT result without requiring parallel

*Corresponding Author: Hnin Thu Zar Aye, hninthuzaraye@ucsy.edu.mm

training information, it depends on dependency parser of source-side language to achieve higher performance than unsupervised baseline result [8]. Having similar syntactic structures of Japanese and Myanmar, a parsing approach has been proposed by applying SMT by using Japanese as pivot in [9]. The proposed method performed well with satisfying results due to the good performance of Japanese parser [9]. Although Myanmar dependency syntax structures were applied competently for SMT [8] and parsing [9] without annotated Myanmar corpus, there was dependence on intermediate language dependency parser like English and Japanese as limitation

Progress of unsupervised dependency parsing researches is increasing with the shared tasks of the tenth conference on computational natural language learning (CoNLL-X) in recent years [10]. A Universal part-of-speech tag (U-POS) set has been proposed as standard for research in unsupervised induction of syntactic structure [11]. Universal Dependencies (UD) is a project developing cross-linguistically consistent treebank annotation for many languages [12]. Currently, there are over 100 treebanks of more than 70 languages available in the UD inventory. Treebanks have derived UD format from existing formats in many languages like Korean [13]. UDPipe has been proposed to easily perform basic natural language processing tasks from tokenization to parsing in CoNLL-U format, the revised version of CoNLL-X format, for treebanks of UD without requiring any other external data [12]. UDPipe has been applied in dependency treebank building [14, 15].

Currently, Asian language treebank (ALT) project has developed a Myanmar syntax treebank with a parallel corpus by annotators using web-based tool [16]. Building treebank needs annotators and applied tools for processes of tree building. Therefore, it becomes hard for low resource language like Myanmar.

Having limitation of related works [8, 9], no resource for dependency parsing and information of Myanmar, improvement of unsupervised parsing researches [10], and simple trainable facility of UDPipe with CoNLL-U format, as our motivation, we annotated a corpus by applying U-POS tags and unsupervised dependency parsing by UDPipe of UD project to get raw syntax information for Myanmar [17]. Then, as future work of unsupervised annotation, manual post processing is carried out on the unsupervised parsed results with reference dependency structures to build dependency treebank in order to apply in parsing Myanmar sentences by deep learning approaches.

This paper presents dependency head annotation to build Myanmar dependency treebank by U-POS tag sets and UD 2.0 guidelines by updating in post processing on unsupervised annotated corpus. Building Myanmar dependency treebank contains two main parts. The first is automatic annotating by unsupervised dependency parsing by UDPipe to get raw universal dependency syntactic information [17]. The second is manual post processing by dependency structures for correct dependency heads. During post processing, UDPipe is applied in a bootstrapping manner to be consistent updating with reduced post processing time..

The organization structure of paper is as follow: Section 2 briefly describes nature of Myanmar language and sentences. Section 3 presents corpus information and overview of annotation scheme. Section 4 describes corpus pre-processing tasks before annotation. Section 5 presents Myanmar language specific tags and their related U-POS tags and mapping scheme between language POS and U-POS tags. Section 6 describes post checking and updating unsupervised dependency heads in sentences with examples. Section 7 reports parsing experiments with post processed data. Section 8 discusses about parsing and evaluation results of treebank. Section 9 concludes all sections and presents future work.

2. Myanmar Language

Myanmar (Burmese) is a member of the Lolo-Burmese grouping of the Sino-Tibetan language belonging to the Southern Burmish branch of the Tibeto-Burman languages. It is an official language in Myanmar. It is also the first language of the Bamar people, the principal ethnic group and related ethnic groups of the country, and a second language of ethnic minorities in Myanmar. Myanmar (Burmese) is a tonal, pitch-register, and syllable-timed language, largely monosyllabic and analytic, with a subject–object–verb word order. It is morphological rich and agglutinative language.

2.1. Nature of Myanmar Sentences

Myanmar sentences can be written with formal or colloquial style. Two types of Myanmar sentence construction are simple and complex sentence types. Simple sentence has only one nominal phrase, action maker or subject, and one verb phrase. Complex sentence has two or more clauses or simple sentences, joined with conjunction, or post positional markers, or particles to modify the followed part which might be phrase or clause of main sentence.

Sentence / (Type)	ကိုယ် တွင် ညို သော်လည်း ၊ မျက်နှာ နှင့် လက် များ မှာ မည်း နေ သည် ။ (Complex sentence)													
Clauses / (Type)	***ကိုယ်***တွင် ***ညို*** သော်လည်း ၊ (Dependent clause)					***မျက်နှာ*** နှင့် ***လက်*** များ မှာ ***မည်း***နေ သည် ။ (Independent root clause)								
Phrases/ (Type)	***ကိုယ်*** တွင် (Noun)		***ညို*** သော်လည်း ၊ **(Adjective)**			***မျက်နှာ*** နှင့် ***လက်*** များ မှာ (Noun)					***မည်း***နေ သည် ။ **(Adjective/Root)**			
Words	***ကိုယ်***	တွင်	***ညို***	သော်လည်း	၊	***မျက်နှာ***	နှင့်	***လက်*** များ		မှာ	***မည်း***	နေ	သည်	။
Morphenes	***ကိုယ်***	တွင်	***ညို***	သော်လည်း	၊	***မျက်နှာ***	နှင့်	***လက်***	များ	မှာ	***မည်း***	နေ	သည်	။
Morpheme translation	body	at	brown	although	,	face	and	leg	-s	-	black	-	-	.
Sentence translation	Although body is brown, face and legs are black .													

Table 1: Suffixes of example sentence from Figure 1

Types	Nominal marker (post positional marker)	Plural marker (particle)	Verb suffix (particle)	Verb marker (post positional marker)
Morpheme Words	“တွင်”,“များ”, “မှာ”	“များ”	“နေ”	“သည်”

Table 2: Composition of Myanmar Dependency Treebank

Corpus Source	Domain	Sentences	Tokens	Average Length	Remark
myPOS	Wikipedia	11,010	239,534	21.75	manually annotated
Web News	Websites’ News	1,800	66,710	37.06	manually annotated
ALT	Wiki news	10,000	373,974	37.39	Automatic annotated
	Total	22,810	680,218		

Myanmar noun phrases are usually ended with different types of nominal markers, postpositional markers (PPMs), to identify their roles in sentences such as subject, object but most colloquial style noun phrases are written without these markers. Myanmar sentences might have one or more phrases or clauses. Some subordinate clauses are usually used to represent more detail meaning for modified parts and also placed before modified ones. Having these nested parts, defining correct dependency between sub and main parts of the sentence takes more time. Figure 1 illustrates an example sentence structure.

There are four phrases in sentence of Figure 1. In Myanmar language, adjective ended with verb maker suffix is verb phrase. In the example sentence, sentence ended verb phrase is composed of adjective with post positional verb marker Four suffix types of morphemes of the example sentence are described in Table 1. Bold and italic words are main content words of each phrase. Main root phrase of a sentence or clause is final verb phrase of that sentence. Dependent sentence or clause modifies the independent root sentence. Therefore, main root of example sentence in Figure 1 is the last right most verb phrase.

Moreover, most formal sentences are usually ended with verb or verb phrase. However, some sentences may not be ended with verbs because of the Myanmar sentence writing style in which there are presence of hiding verbs hidden for actions of being or having or living or coming actions. Moreover, Myanmar sentences can be written by many forms according to nature of free word order language, and some special cases of sentence construction of Myanmar grammar. Moreover, emphasized noun phrases can be put at the beginning of the sentence according to writer’s idea.

These conditions can also be complicated and time-consuming issues to define correct dependency heads and relations for phrases in sentences. Besides above nested cases, there arises one issue in dependency tree building.

3. Corpus Annotation

This section presents corpus statistic and overall architecture of dependency head annotation.

3.1. Corpus Information

Currently two corpora have been annotated manually by dependency head information and one corpus has been annotated in unsupervised way for Myanmar treebank. The myPOS corpus consists of 11,010 sentences written by formal and colloquial format from the Myanmar Wikipedia including various areas such as economics, history, news, politics and philosophy [18]. The sentence writing style of myPOS corpus is very similar to current mostly used standard formal and colloquial style. Therefore, we selected first this corpus to annotate to get similar syntactic structures of current written styles of Myanmar sentences and annotated them as first standard sentences for other corpora. Web News corpus contains 1,800 sentences written by current modern writing styles in Myanmar news websites. ALT corpus contains 10,000 translated sentences from Wiki news. The statistical information of current Myanmar dependency treebank is presented in Table 2.

3.2. Overview Structure of Annotation

Using difference corpora in treebank, word segmentation and POS tagging style of each corpus is different. Being morphological rich and agglutinative language, most words are segmented to provide morphological level syntax information in this work. Therefore, pre-processing is needed to carry out to be the same word segmentation and POS-tagged scheme among different corpora. U-POS tags, CoNLL-U shared task format, shared Japanese model, and UDPipe of UD project were used to get raw universal dependencies in this work.

The corpora were transformed to CoNLL-U format by adding U-POS tags in pre-processing. After transformation, they were annotated by unsupervised dependency parsing by using shared Japanese model in UD project [17] because of similar conditions in grammar structure of Myanmar and Japanese and dependency structures and UD of Japanese [19,20]. Then unsupervised annotated dependency heads were manually post checked by human annotator to be more correct dependency head nodes. One annotator was done manual post checking on automatic annotated results by learning Myanmar language grammar books and books written by linguistic experts of Myanmar language. Therefore, dependency head annotation of Myanmar dependency treebank has two steps: initial corpus pre-processing and post processing on unsupervised annotated results of pre-processing.

In post processing, to be consistent and fast checking and updating, selected updated data were repeatedly trained with all other not updated data in corpus by UDPipe. After checking and updating manually selected 2,000 sentences, they were added to the training data, Then the training model was retrained to parse the remaining sentences in corpus by the updated model. Parsed

results were updated manually and trained again until all sentences were post processed in corpus. The overall architecture of annotation is illustrated in Figure 2. Detail explanation with examples will be presented in Section 4 and Section 6.

4. Pre-processing

This section presents overview of corpus pre-processing step. Before dependency structures annotation, it is needed to check and update word segmentation and POS tagging of used corpora because some corpora might have different segmentation and POS tagging styles based on their original created purposes. Using unique word segmentation and POS tagging can be easy to transform dependency corpus. It is also better and easier having unique word segmentation and POS tagging than different formats to transform dependency corpus.

To have a consistent Myanmar POS tag scheme, a new general POS tag scheme has been defined. It will be explained in following sub section. Moreover, U-POS tags were also added to the CoNLL-U format in order to get raw universal dependencies syntax structures.

In Myanmar sentences, some words might also have different POS tag forms for the same word. Tagging correctly for each content word in sentences is important for correct dependency head. One example of these issues is presented in Table 3. The quality of word segmentation and POS tagging can mainly impact providing correct dependency information of sentences in corpus. Therefore, rechecking word segmentation and POS tagging of used corpus before annotation is very important.

Then, the correct POS tagged corpus was transformed CoNLL-U format by adding related U-POS tags for most language POS tags by manual python script as UD 2.0 format. In adding U-POS tags by python script, it is needed to check again manually U-POS tags of Myanmar conjunctions because it is also needed to be correct form of two conjunction tags of U-POS as described in Table 5 according to the content words of sentence.

5. Part-of-speech Tagging Scheme

There are ten main POS tags in Myanmar language such as noun, pronoun, adjective, verb, adverb, post-positional marker, particles, conjunctions, interjection and punctuation [21]. For easy mapping to U-POS tags, 16 POS tags have been defined for language specific tags which are presented in Table 4 with examples.

Most tags of these were already defined in ALT [6] except proper noun (PRPN) and text number (TNUM). In the previous work of this, these two tags were defined as noun, “N” [17]. In most Myanmar corpora, proper nouns and text numbers are tagged as noun, but they are already defined as proper noun and number in U-POS tag set. Therefore, these two tags were also added in Myanmar language tag set in order to be fast and easy mapping between Myanmar POS tags and U-POS tags.

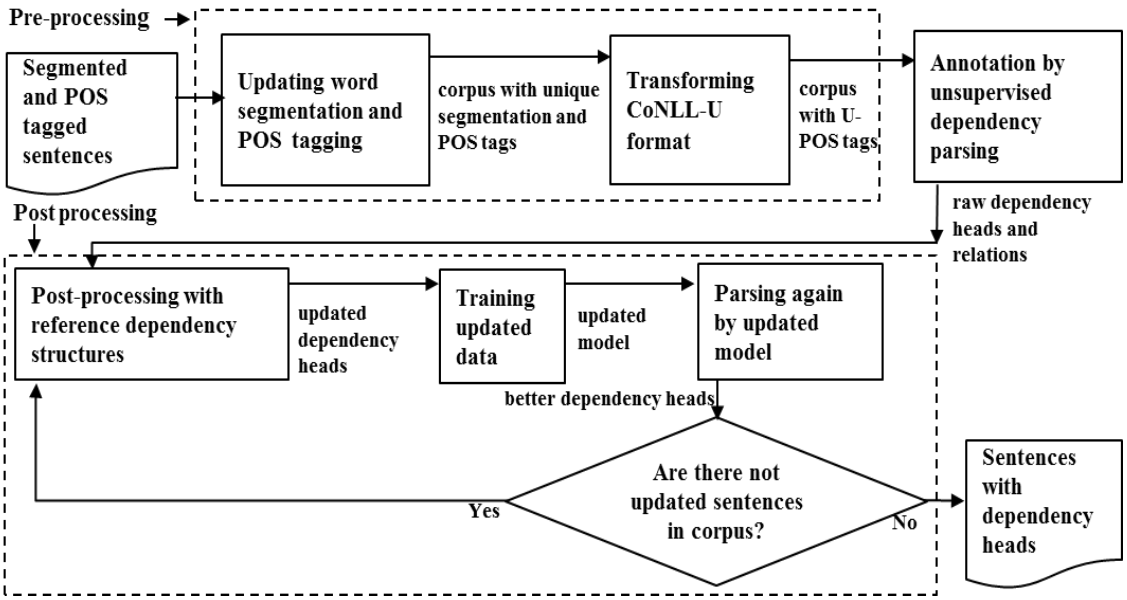


Figure 2: Overview architecture of dependency head annotation

Table 3: Example POS tagging for same word

Words	Correct Tags	Meanings	Examples in phrases
နေ	verb	live	Myanmar : ရန်ကုန် မှာ နေ Words in English: Yangon in live - Translation : Live in Yangon.
နေ	particle	describing continuous action	Myanmar : လေ့လာ နေ သည် Words in English: study - - Translation : is studying

Table 4: Myanmar POS tags for language specific tag set

POS Tag	Description	Example words
N	Noun	သတင်းစာ (Newspaper), ခန်းမ (Hall), ကုမ္ပဏီ (Company)
PRPN	Proper Noun	စက်တင်ဘာ (September), အာရှ (Asia)
NUM	Number	ဝ (0), ၁(1), ၂(2), ၃(3), ၄(4)
TNUM	Number text letter	သုည (zero) , တစ် (one) , နှစ် (two), သုံး (three), လေး(four)
FOR	Foreign word	Carsh, Federal, process, Bit
ABB	Abbreviation	ညွှန်/ချုပ် (Director-General)
PRON	Pronoun	ကျွန်တော်/ ကျွန်မ/ ကျွန်ုပ် (I), ထို (that), မည်သူ (who)
ADJ	Adjective	ကျယ်ဝန်း (wide), ယဉ်ကျေး (polite), မြင့်မား (high), လေးလံ (heavy)
ADV	Adverb	မကြာခဏ (frequently), အလွန် (very), လောလောဆယ် (currently)
V	Verb	စား (eat), သွား (go), ရေးသား (write), ဖြစ် (be), ရှိ (has/have), တည်ရှိ (exist)
CONJ	Conjunction	နှင့် (and), သို့မဟုတ် (or), ၍/သောကြောင့် (because)
PART	Particle	များ/တို့/တွေ (plural marker), ခန့် (about), ခဲ့ (past marker), နိုင်(can), ကြ (plural action marker)
PPM	Post positional marker	သည်/က/မှာ (nominal marker for subject), ၌ (at) , ၏ (of) , ဖြင့် (by)
PUNC	Punctuation	၊ , ။ , “ , (or -LRB-,) or -RRB- , “, ”
SB	Symbol	% , $
INT	Interjection	အိုး (Oh), အမလေး (Oh my god!)

Table 5: Mapping scheme between Universal POS and Myanmar language specific POS tags

U-POS Tag	Description [Examples]	Myanmar Language POS
NOUN	Noun	N
		FOR
PROPN	Proper Noun	PRPN
		ABB
NUM	Number	NUM
		TNUM
PRON	Pronoun	PRON
ADJ	Adjective	ADJ
ADV	Adverb	ADV
VERB	Verb	V
CCONJ	Coordinating Conjunction [နှင့် (and), သို့မဟုတ် (or)]	CONJ
SCONJ	Subordinating Conjunction [၍/ သောကြောင့် (because), နှင့်တစ်ပြိုင်နက် (as soon as), ပါက (if)]	
PART	Particle	PART
ADP	Adposition	PPM
PUNCT	Punctuation	PUNCT
SYM	Symbol	SB
INTJ	Interjection	INT

Table 6: Mapping scheme between Universal POS and Myanmar language specific POS tags

U-POS Tag	Description	myPOS	ALT	Web News	Total
NOUN	Noun	55,590	71,246	18,237	145,073
PRON	Pronoun	2,680	7,864	331	10,875
PROPN	Proper Noun	12,377	23,168	3,375	38,920
NUM	Number	6,261	11,240	2,259	19,760
ADJ	Adjective	7,108	7,776	1,393	16,277
ADV	Adverb	2,868	4,686	852	8,406
VERB	Verb	34,046	62,876	11,057	107,979
CCONJ	Coordinating Conjunction	4,624	6,368	1,584	12,576
SCONJ	Subordinating Conjunction	6,587	7,493	1,452	15,532
PART	Particle	52,413	78,268	13,294	143,975
ADP	Adposition	38,838	64,624	8,674	112,136
PUNCT	Punctuation	15,845	28,255	4,199	48,299
SYM	Symbol	199	116	3	318
INTJ	Interjection	98	0	0	98

All conjunctions in Myanmar sentences were tagged as CONJ which were mapped to one of two conjunctions of U-POS tags: CCONJ and SCONJ, coordinating and subordinating conjunction respectively. Mapping scheme between U-POS tags and Myanmar POS tags can be seen in Table 5. Total frequencies of U-POS tags in each corpus of treebank are listed in Table 6.

6. Post-processing

Annotation method and types of annotation schemes are important in dependency treebank construction. The annotation was based on Universal Dependencies. In the UD annotation scheme, dependency relations are expressed between words and main content words attached leaf words such as function words in sentences [22]. Main parts of syntactic structures are head nodes and relation links called as dependency relation labels between words.

Generally Myanmar nouns, adjectives, and verbs are formally written with suffixes such as post positional markers or particles. Currently, dependency relation labels are automatic unsupervised annotated results without post processing in treebank. Only dependency head information was post processed.

Overview of the linking structures of dependency head node words between dependent words in most occurred phrases in sentences had been presented in our previous work [17]. In that work, only construction of dependency link arc connections between heads and dependents words had been presented. Therefore, the arcs directed to the heads from dependents.

After the previous work, unsupervised dependency annotation results have been post processed to update dependency head links based on the dependency structures described in our previous work. Updating dependency relation labels needs more time because of few annotators and issues of sentence writing styles. Therefore, dependency relations labels are still automatic unsupervised annotated labels after post processing. To describe full referenced dependency information, in this paper, dependency relation labels between head node words and dependent words will be presented with unsupervised annotated labels in sample sentences according to the viewpoint of the language typology.

6.1. Proper Noun and Possessive Phrase

The specific unique names of common nouns are called as proper nouns as in other languages. Myanmar proper nouns are usually found in before or after common nouns [21]. In Figure 3, two proper nouns, following two common nouns can be seen in example sentence. In that sentence, example possessive case can also be seen. Possessive case of a noun or pronoun can be written by a post positional suffix marker, “ ၏”, to show possession of following right noun by left proper noun of it.

6.2. Compound Noun

Myanmar compound nouns might have two or more words and POS tags of these words can be nouns, verbs, adjectives or adverbs [21]. An example compound noun in sentence of Figure 4 contains two consecutive nouns and means “memorial stamp”. In this case, the left noun modifies the right one.

6.3. Numeral Phrase

Most numeral noun phrases can be written by three general formats which are described in Table 7 with their meanings. Counting amount can be written by digit number or text number [21]. Sample numeral phrases in sentence can be seen in Figure 5.

6.4. Adjective Phrase

Adjectives modify nouns and are usually placed before or after noun in Myanmar sentences. Myanmar adjectives can be written as simple or simple adjective with suffix particles or transformed adjective [21]. Suffixes of adjective or example adjective phrases are presented in Table 8. Sample dependency of adjective phrase in sentence can be seen in Figure 6.

6.5. Adverb Phrase and Verb Phrase

Adverbs can be written as simple or transformed adverb with the suffix particle, “ စွာ”, followed by verb, or adjective, or adverb.

Example adverb phrases can be seen in Table 9. Example dependency of adverb phrase in sentence can be seen in Figure 7.

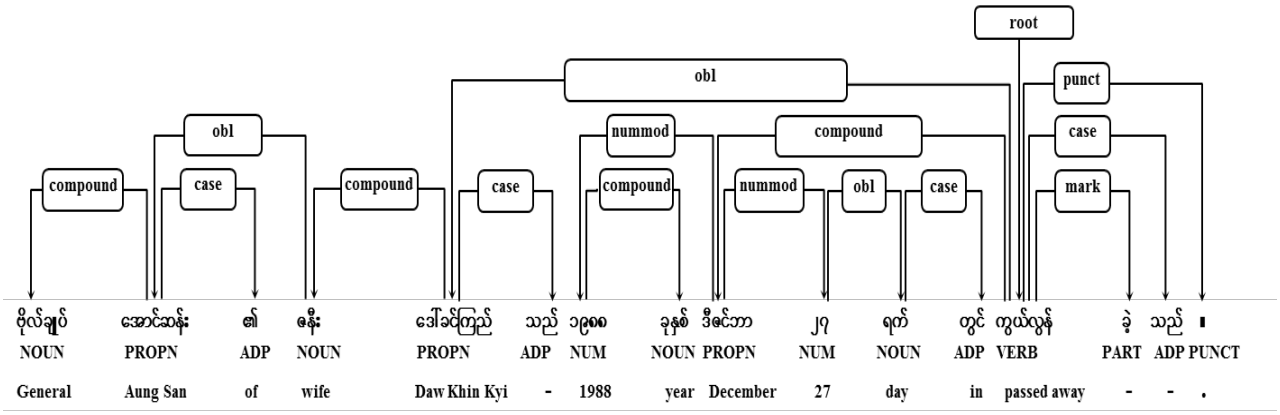


Figure 3: Proper noun and possessive case examples in sentence

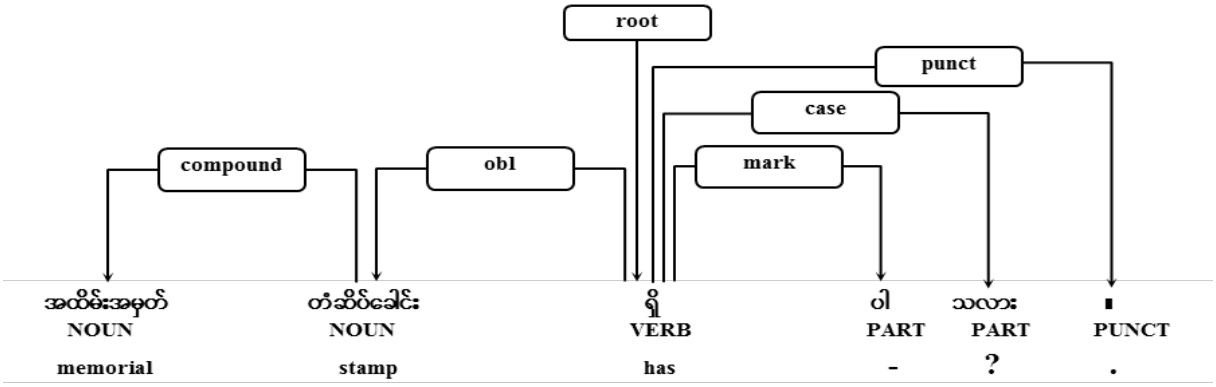


Figure 4: Compound noun example in sentence

Table 7: Numeral phrase formats

Numeral Phrase Forms	Example	Meaning	Remarks
N NUM/TNUM PART/N	မှတ်ချက် ၂/နှစ် ချက် comment 2/two - (Glossary)	2/two comments	
N N (noun affixed particle, "အ", to form common nouns) NUM/TNUM	မှတ်ချက် အချက် နှစ်ဆယ် comment fact twenty (Glossary)	twenty comments	if counted amount is exact numbers of multiple of ten, hundred, thousand, etc.
N N(transformed or common noun) NUM/TNUM PART/N	မှတ်ချက် အချက် နှစ်ဆယ် comment fact twenty two - (Glossary)	twenty two comments	

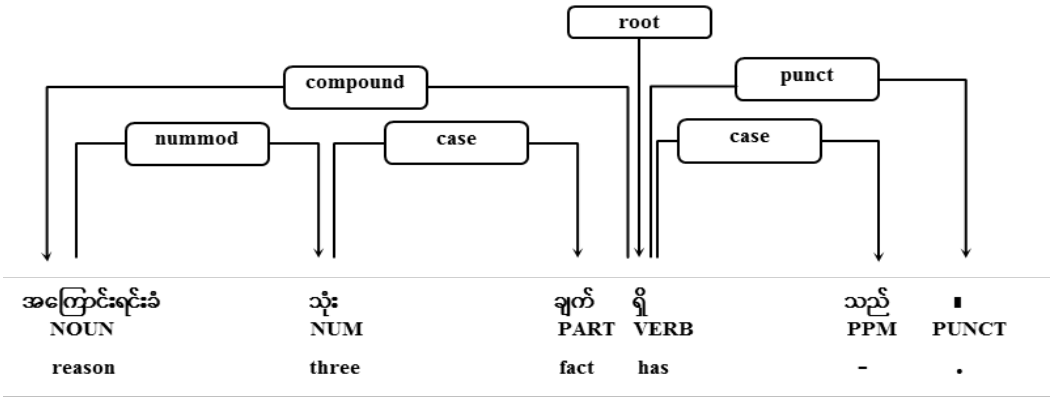


Figure 5: Numeral phrase example in sentence

Table 8: Suffixes of adjective or adjective phrases and examples

Particles used to suffix or transform adjective or adjective phrase		သော, သည့်, မည့်, တဲ့
Example Myanmar adjective	***Translations***	***Remarks***
ကြီး သော အိမ် **big** - house	big house	Simple adjective followed suffix
အိမ် **ကြီး** house **big**	big house	Simple adjective
အစိုးရ ပေး သည့် အိမ် government provide - house	house provided by government	Transformed adjective

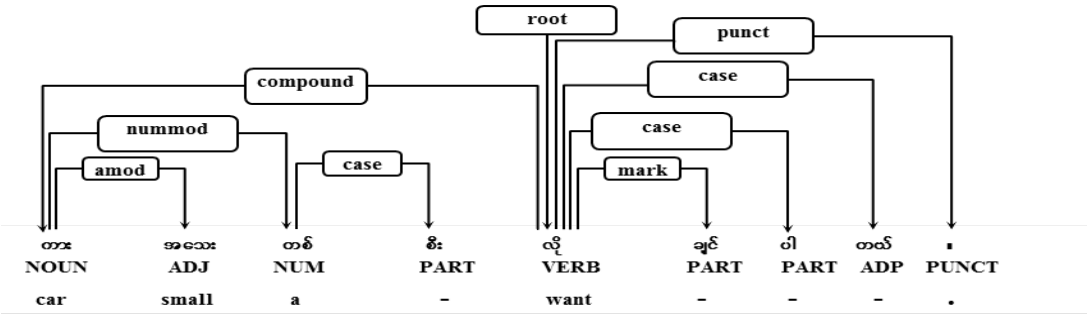


Figure 6: Adjective phrase example in sentence

Table 9: Example of adverb forms

Example Myanmar adverb phrases	**Meanings**	**Remarks**
မြန်မြန်(quickly)	quickly	Simple adverb
မြန်မြန်သွက်သွက် (quickly)	quickly	Simple adverb
လျင်မြန် စွာ quick -	quickly	Transformed adverb with suffix particle

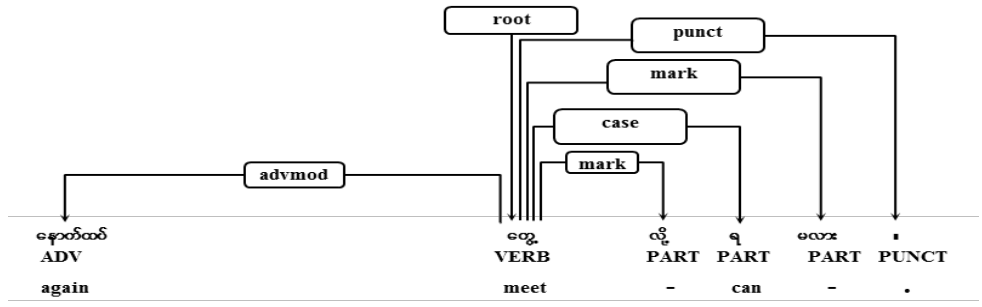


Figure 7: Adverb and verb phrase example in sentence

Table 10: Suffixes of verbs

Suffixes to verb	**Description of usage and example**	**Suffix Type**
သည်, ၏, ပြီ, မည်, မယ်, တယ်	to form verb by showing tense Examples: သွား သည် (go) သွား မယ် (will go)	post positional markers
မှာ, ပါ, စမ်း	to express giving order or answering action Examples: လုပ် ပါ (giving order or requesting “Do” action) မောင်း မှာ (answering or forecasting “drive” action based on the meaning of content words in sentence)	particles

Verb phrases in sentences are usually formed by one of the verb ended post positional markers or particles [21] as presented in Table 10 .

In addition, zero or more particles can be followed by verbs before verb ended markers to express the full state of action. In example sentence of Figure 7, the verb phrase, " တွေ့လို့ရမလား", means "can meet" and "တွေ့" is main verb and the suffix sequence with three particles, "လို့ ရ မလား", means asking politely for the requested action. That main verb phrase is also modified by left adverb phrase, " နောက်ထပ်". As a result, the verb, "တွေ့", is root of that sentence. The whole sentence means "Can meet again?",

6.6. Conjunctions

In Myanmar language, conjunctions are used to combine not only clauses or simple sentences but also related words or phrases. Moreover, simple sentences can be connected by suffixes: post positional markers, " က, မှာ, ကို", or particles "ဟု , သော , သည့် , မည့်, တဲ့" to form noun or adjective clause in sentence. If two or more simple sentences are connected by post positional markers or particles or conjunctions, combined sentence becomes complex sentence and clauses represent the roles as subjects or objects or adverbs. Most clauses ended by conjunctions are usually adverb modifiers in main sentence. Therefore, the role of dependent clauses can be divided into three types: noun clause, adjective clause, and adverb clause based on their roles in sentences. If combined clauses or sentences contains the same subject, subject can be omitted in dependent clause or independent clause or in both. Similarly, object noun can also be omitted in dependent clause or independent clause if it is placed in one [21, 23].

Some conjunctions described in Table 11 are used to connect words, phrases to connect two sentences to give coordinated extra meaning [21, 23].

Table 11: Example conjunctions

Conjunctions	**Usage**
နှင့် , လည်းကောင်း…လည်းကောင်း, ရော…ပါ, ရော… ရော, သို့မဟုတ်, ဖြစ်စေ…ဖြစ်စေ, ဖြစ်ဖြစ်…ဖြစ်ဖြစ်, သော်လည်းကောင်း…သော်လည်းကောင်း, မှတစ်ပါး, ပြီး	to connect words, phrases in sentence for extra meaning
ထို့ပြင်, ထို့အပြင်, ၎င်းပြင်, သည့်ပြင်	to give connection between prior sentence and next by giving coordinated extra meaning.

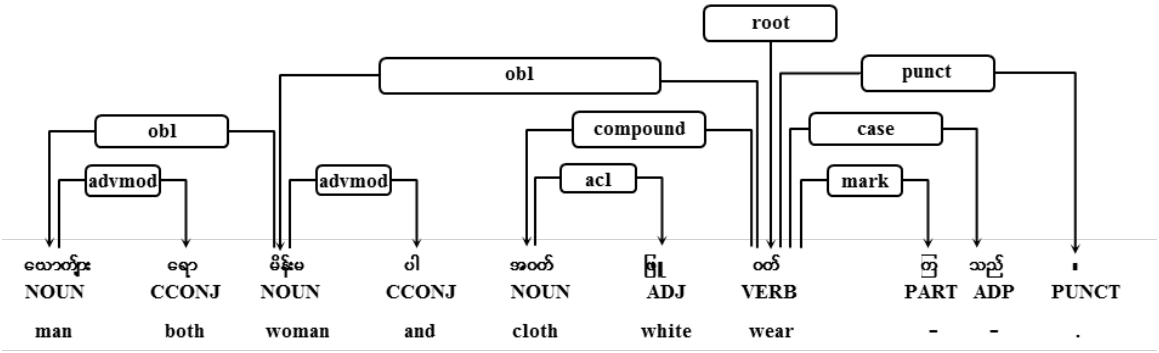


Figure 8: Coordinated phrase example in sentence

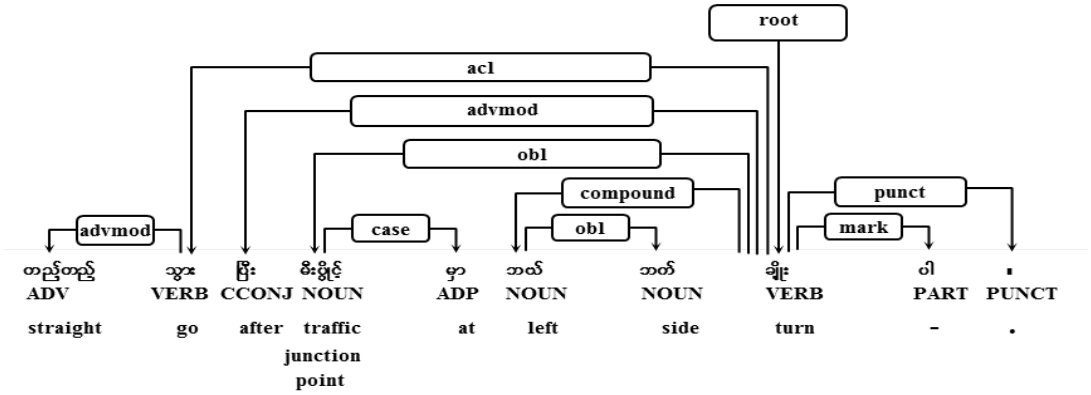


Figure 9: Coordinated clauses example in sentence

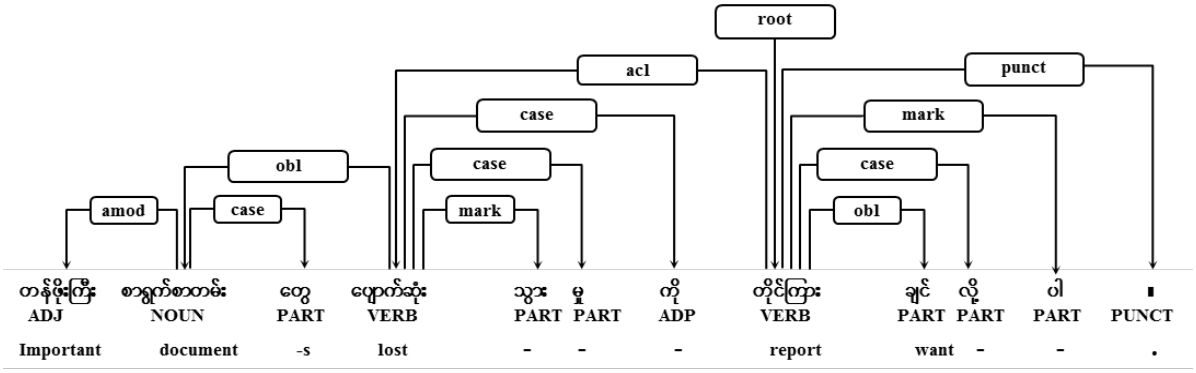


Figure 10: Complex sentence with noun clause

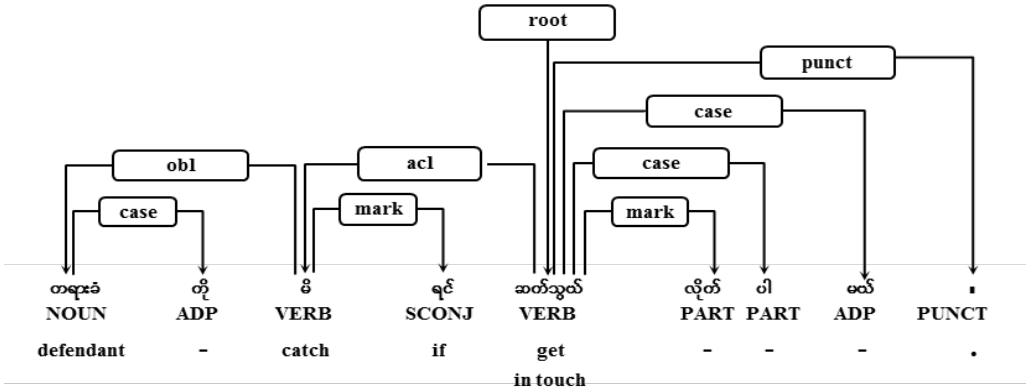


Figure 11: Complex sentence joined with subordinating conjunction

Table 12: Average parsing precision scores of each corpus

Corpus	Sentences		Average Scores	
	Train	Test	UAS	LAS
myPOS	10,010	1,000	89.06	86.67
Web News	1,620	180	87.03	84.83
ALT	9,000	1,000	91.40	90.24

Sample coordinated phrase with conjunctions in sentence can be seen in Figure 8. Moreover, complex sentence combined two simple sentences by coordinating conjunction can also be seen in Figure 9. The left side of conjunction is dependent sentence or clause to give first action and the right side of conjunction is independent clause to give final action as shown in Figure 9. Subjects of clauses are omitted and main sentence means "After go straight, turn left at the traffic junction point.". Therefore, the conjunction is tagged as "CCONJ", coordinating conjunction, for U-POS tag.

Some Myanmar conjunctions such as "လျှင်/ရင်/ပါက, သောကြောင့်, နှင့်တစ်ပြိုင်နက်, သကဲ့သို့, စေရန်, သောအခါ, သော်လည်း,.." , "if, because, as soon as, as/like, in order to, when, although,.." , are frequently used to connect clauses to provide the required meaning for main clauses [21, 23]. Therefore, these types of conjunctions are tagged as "SCONJ", subordinating conjunction, for U-POS tag.

Example complex sentence including noun clause as an object role in sentence can be seen in Figure 10. Main complex sentence means "Lost of important documents is wanted to be reported." The PPM, " ကို", is used to connect the left clause to represent as object noun for the action of root verb, "တိုင်ကြား", of main sentence.

In addition, example complex sentence joined with subordinating conjunction type, " ရင်" (if) , can be seen in Figure 11. Subjects are omitted in both dependent clause and main independent clause. It means that "If defendant is caught, I will get in touch".

7. Parsing Experiment

The main purpose of building treebank is to use in dependency parsing. This section presents parsing experiments executed for performance of treebank. UDPipe is an open-source trainable pipeline processing tool to perform segmentation, POS tagging, lemmatization and dependency parsing without any other external

data for multiple languages. And it is also available under open-source Mozilla Public Licence (MPL) and provides bindings for C++, Python, Perl, Java and C#. UDPipe 1.2 has been used in our parsing experiments.

We divided 90% and 10% of total sentences from each corpus of treebank for training and test data to evaluate each corpus by 10-fold cross validation test. We arranged alternate order of test sentences range and all the rest for train data in each corpus and split them into equally-sized parts for each validation test by python script. We calculated average score of all validation tests to report as total average score of cross validation test for each corpus.

8. Results and Evaluation

In this section, the statistical results of parsing experiments to evaluate parsing performance of treebank and evaluation results will be described. For each corpus performance of treebank, average parsing precision scores are calculated from the total results of cross validation tests in each corpus and they are listed in Table 12. Average precision scores measured by unlabeled attachment score (UAS) and label attachment score (LAS) of each corpus are over 89% and 86%, over 87% and 84 %, and 91% and 90% in myPOS, Web News, and ALT respectively .

Each corpus of treebank was evaluated by the CoNLL 2017 UD parsing evaluation script. The evaluation results will be described in following sub section. In addition, syntax structures of the referenced dependency types were also analyzed by manual python script which counts types of dependency structures being countable types in each sentence. The analyzed results will be described in next sub sections.

8.1. Evaluation

Corpora contained in treebank were evaluated by the CoNLL 2017 UD parsing evaluation script, "conll17_ud_eval.py" [24], to know their parsing accuracy. Gold standard file and system output file are input to the evaluation script to evaluate the data. The parsing model was trained with all current data of treebank to generate system output of each corpus. The current accuracies comparing standard annotated data with system output parsing result of each corpus can be seen in Figure 12.

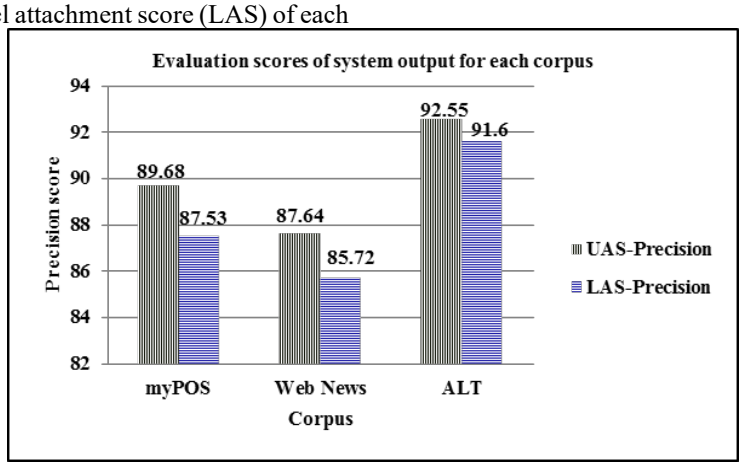


Figure 12: Precision scores of treebank data

Table 13: Information of sentence lengths in treebank

Sentence Type	myPOS	ALT	Web News	Total sentences of each range	Remark
Simple short	2,253	13	74	2,340	<=10 words
Short	6,463	3,858	732	11,053	>=11 and <=30 words
Normal range	1,929	4,398	568	6,895	>=31 and <=50 words
Long	349	1,687	401	2,437	>=51 and <=100 words
Very long	16	44	25	85	>100 words

Table 14: Phrases and clauses structures in treebank

Phrase Types	myPOS	ALT	Web News	Total
Noun	27,083	48,466	6,316	81,865
Proper Noun	10,514	17,750	3,013	31,277
Numeral Noun	5,790	10,852	2,125	18,767
Compound Noun	21,096	25,706	9,537	56,339
Adjective	9,535	11,993	1,581	23,109
Adverb	3,027	4,580	837	8,444
Verb	16,593	19,136	3,584	39,313
Phrases with Conjunctions	2,830	3,534	921	7,285
Clauses	7,027	9,837	1,950	18,814

Precision scores of UAS and LAS in system outputs of myPOS and Web News data are over 89% and 87%, and 87% and 85% respectively in comparing manually updated standard data. System output precision scores of UAS and LAS for ALT data is over 92% and 91% in comparing with automatic unsupervised annotated standard data.

As sentences of ALT corpus are longer than myPOS and Web News corpus, precision scores of unsupervised annotated ALT is more than manual updated standard data of myPOS and Web News corpus. However, system outputs of myPOS and Web News corpus are better and more similar to reference dependency structures than system outputs of ALT in manual random checking on system output trees.

Although currently Myanmar ALT has been developed manually, there is still no dependency resource for Myanmar. Constructing treebank manually is an error-prone and slow process. The high precision scores of cross validation tests and evaluations by post processed model illustrate that the proposed method can produce efficient precision scores for dependency parsing. With consistent and faster annotation, the proposed method will provide fast dependency tree building for Myanmar which has free word order and issues mentioned in Section 2

The length of sentences is also one main part of treebank characteristics for having various types of syntax and syntactic structures. Therefore, the ranges of sentences in treebank were also analyzed by classifying five levels by python script and resulted sentence ranges are listed in Table 13.

8.2. Syntax Analysis

The syntax information is one of the most important characteristics of treebank to know syntax status. It is difficult to count all syntax information of natural language sentences by program script exactly because of the issues of phrases and many sentence construction types discussed in Section 2. However, overview syntax structures of formal sentences from each corpus of treebank could be extracted by python script based on sequence order of words and POS tags information of phrases and clauses written by formal literature written style. The program counts the related phrase and clause types from word and POS tag sequences of sentences annotated as example dependency structure types described in Section 6. However, some informal phrases and clauses not ended with the formal related post positional markers or particles, could not be counted by the program. Overall countable syntax structures are listed in Table 14.

9. Conclusion

This paper has presented annotating dependency heads based on Universal Dependencies framework for Myanmar dependency treebank. Myanmar POS tags and U-POS tags of treebank, issues of tagging, and mapping scheme between two tag sets have also been presented. In addition, dependency head annotation schemes have also been described with sample sentences in line with grammatical point of views. This work is first for Myanmar to the best of our knowledge. Parsing experiments have also been executed for performance of treebank and results have also been presented. To conclude, contribution is first dependency head annotation for building Myanmar dependency treebank and can provide useful information for direct dependency parsing for Myanmar sentences by Myanmar model in future.

As future work, firstly we would intend to add more annotated sentences from same and different domains such as News articles data and Myanmar grammar books data to treebank because correct syntactic forms are able to provide faster annotation and better useful syntax information for Myanmar dependency treebank. The next work is to update unsupervised dependency labels according to the standard of Universal Dependencies based on Myanmar grammatical point of views.

References

[1] N. Green, “Dependency Parsing”, In WDS 2011 Proceedings of Contributed Papers, WDS 2011, 137–142, 2011, doi:10.1007/1-4020-4889-0_3..

[2] T. Kakkonen, “Dependency treebanks: methods, annotation schemes and tools”, In Proceedings of the 15th NODALIDA conference, 94–104, 2005, doi: arXiv:cs/0610124

[3] P. E. Solberg, “Building gold-standard treebanks for Norwegian”, In Proceedings of the 19th Nordic Conference of Computational Linguistics, 2013 (NODALIDA 2013), Linköping University Electronic Press, 459-464, 2013.

[4] K. Haverinen, J. Nyblom, T. Viljanen, V. Laippala, S. Kohonen, A. Missilä, S. Ojala, T. Salakoski, and F. Ginter, “Building the essential resources for Finnish: the Turku Dependency Treebank”, Language Resources & Evaluation, **48**(3), 493-531, 2014, doi: 10.1007/s10579-013-9244-1.

[5] C. MĂRĂNDUC, and C. A. PEREZ “A Romanian Dependency Treebank”, In International Journal of Computational Linguistics and Applications, **6**(2), 83-103,2015.

[6] D. Bamman, and G. Crane, "The Ancient Greek and Latin Dependency Treebanks," In Language Technology for Cultural Heritage 2011, Springer, 79-98, 2011, doi:10.1007/978-3-642-20227-8_5.

[7] K. Nguyen, “BKTreebank: Building a Vietnamese Dependency Treebank”, In Proceedings of the 11th International Conference on Language Resources and Evaluation 2018 (LREC-2018),European Languages Resources Association (ELRA), 2164-2168, 2018.

[8] C. Ding, Y. K. Thu, M. Utiyama, A. M. Finch, and E. Sumita, “Empirical Dependency-Based Head Finalization for Statistical Chinese-, English-, and French-to-Myanmar (Burmese) Machine Translation”, in Proceedings of the 11th International Workshop on Spoken Language Translation, 184-191, 2014.

[9] C. Ding, Y. K. Thu, M. Utiyama, and E. Sumita, “Parsing Myanmar (Burmese) by Using Japanese as a Pivot”, Proceedings of 14th International Conference on Computer Applications, 158-162, 2016

[10] D. Mareˇcek, “Twelve Years of Unsupervised Dependency Parsing”, ITAT 2016 Proceedings, CEUR Workshop Proceedings , **1649**, 56–62, 2016..

[11] S. Petrov, D. Das, and R. McDonald, “A Universal Part-of-Speech Tagset”, In Proceedings of the 8th International Conference on Language Resources and Evaluation, 2089-2096, 2012.

[12] M. Straka, J. Hajic, and J. Strakov ˇ a´, “UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing,” In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), 4290–4297, 2016.

[13] J. Chun, N. Han, J. D. Hwang, and J. D. Choi, “Building Universal Dependency Treebanks in Korean”, In Proceedings of the 11th International Conference on Language Resources and Evaluation 2018 (LREC-2018). European Languages Resources Association (ELRA), 2194-2202, 2018.

[14] E. Badmaeva, and F. M. Tyers, “A Dependency Treebank for Buryat”, In 15th Internationtal Workshop on Treebanks and Linguistic Theories (TLT15), 1-12, 2017.

[15] T. Rama, and S. Vajjala, “A Dependency Treebank for Telgu”, In Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories (TLT16), 119–128, 2018.

[16] Y. Kyaw Thu,, W. Pa Pa, M. Utiyama, A. Finch, and E. Sumita, “Introducing the Asian Language Treebank (ALT)”, In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16) 2016 May , 1574-1578, 2016. DOI: 10.1109/ICSDA.2016.7918974

[17] H. T. Z. Aye, W. P. Pa, and Y. K. Thu, “Unsupervised Dependency Corpus Annotation For Myanmar Language”, In Proceedings of the 2018 Oriental COCOSDA-International Conference on Speech Database and Assessments, IEEE, 78-83, 2018, doi:10.1109/ICSDA.2018.8693009

[18] K. W. W. Htike, Y. K. Thu, Z. Zhang, W. P. Pa, Y. Sagisaka, and N. Iwahashi, "Comparison of Six POS Tagging Methods on 10K Sentences Myanmar Language (Burmese) POS Tagged Corpus", In Proceedings of the CICLING 2017, April 2017..

[19] S. Mori, H. Ogura, and T. Sasada, "A Japanese word dependency corpus", Proceedings of the 9th International Conference on Language Resources and Evaluation 2014 May 26, 753–758, 2014.

[20] T. Tanaka, Y. Miyao, M. Asahara, S. Uematsu, H. Kanayama, S. Mori, and Y. Matsumoto, "Universal Dependencies for Japanese", In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16) , 1651-1658, May 2016.

[21] Department of the Myanmar Language Committee, Myanmar Grammar, Ministry of Education, The Republic of the Union of Myanmar, 2013.

[22] P. Osenova, and K. Simov, "Treebanks, Linguistic Theories and Applications Multilingual Treebanks: the Universal Dependencies Case", In 30th European Summer School in Logic, Language and Information, 2018.

[23] U Thaung Lwin, Advanced New Method Myanmar Grammar) Second Edition, New Method Book Store (နည်းသစ်စာအုပ်တိုက်), 2015.

[24] A. Kutuzov, "Dependency Parsing (Project A) ", In INF5830, Fall 2017,University of Oslo, 2017

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 801-810 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

A Model-Driven Approach for Reconfigurable Systems Development

Ismail Ktata[*,1,2], Naoufel Kharroubi[1]

[1]*Computer Science Department, Khurma University College, Taif University, Taif, 21944, Kingdom of Saudi Arabia*

[2]*CES laboratory LR11ES49, National Engineering School, University of Sfax, Sfax, 3038, Tunisia*

ARTICLE INFO

Article history:
Received: 27 September, 2020
Accepted: 18 November, 2020
Online: 24 November, 2020

Keywords:
Model-Driven Engineering
Ontology
Semantic Rules
SWRL Reasoner
Reconfigurable Architecture
Scheduling

ABSTRACT

Reconfigurable systems are considered as promising technology that enables the design of more flexible and dynamic applications. However, actually existent design flows are either low-level (so complex) or they lack support for automatic synthesis. In this paper, we present an ontology-based modeling approach for reconfigurable systems. Our approach is based on model-driven engineering process and addresses dynamic features on application-level enabling early exploration of the execution behavior of the system. The model is characterized by a logical and syntactical description conform with application domain knowledge and respect a number of metamodel constraints. These elements are semantically presented by an ontology language. We successfully implemented a system model with several tasks and resources and made scheduling test for the application graph.

1. Introduction

The system designers have been relying two choices for computing system implementing, varying largely in terms of performance and flexibility [1]. One is considered for the general-purpose systems built to accommodate flexibility problem. The compromise, in this case, is the system performance because of the absence of hardware resources dedicated for specific application. The second is carried out to develop an optimized hardware circuit dedicated and adapted to meet the demands of a specific application. Such system will be at issue following a minor need for change, because a costly redesign will be carried out in order to adapt new system parameters.

The present work introduces a new paradigm based on configurable computing that can solve performance problems as well as flexibility problems, and hardware design. We offer the performance of dedicated circuits by changing hardware configurations. The reprogram ability of the computing hardware explains that new architecture needs to emulate different computer architectures [2]. Mapping software operation is facilitated with getting the ability to reconfigure its connections. A number of manufacturers introduce FPGAs characterized by partial and dynamic reconfiguration abilities [3]. When a part of FPGA logic resources and interconnections is reconfigured while the remainder device continues operating, the partially reconfigurable FPGA is. Dynamic reconfiguration deals with logic and interconnections reconfiguration in FPGA from an application to the next, that is what we call dynamic reconfiguration or run-time reconfiguration (RTR) [4]. In such situation, without adding external physical (hardware) resources, in a static or dynamic state at run time, when the system can change/adapt its basic computational structure to suit the changing requirements of a program, this is referred to as a Configurable Computing System (CCS) [5]. Our current CCS realizations has a general-purpose microprocessor augmented set with an FPGA. Using reconfigurable logic as a resource shared by multiple applications, a hardware operating system (HwOS) facilitates system setting operations to overcome the problems due to real time considerations. Real time multi-task systems' requirement is tasks scheduling to complete their execution depending on their deadlines [6]. Tasks known periodic in advance are scheduled off-line. Instead, irregular and unknown tasks in advance are scheduled on-line. That makes specific constraints concerning computing resource recovery, preemption and interruption time, tasks' migration between different executing resources, limited memory and power use, hardware costs, etc. [7]. To overcome such complexities, high-level development techniques are required. That researching methods dealing with modeling and simulation of complex and heterogeneous systems are called model driven engineering approaches.

[*]Corresponding Author: I. KTATA, i.ktata@tu.edu.sa, ismail.ktata@enis.rnu.tn

Model-Driven Engineering (MDE) methods tackle with the specific constraints of a specific domain and execute model checking that may identify many errors and limit failure risks at an early period of the system's life cycle [8]. In this context, the present paper focuses on the use of an ontology model [9] in order to describe the dynamic behaviors of the components of a system within the schedulability analysis and verification context [10].

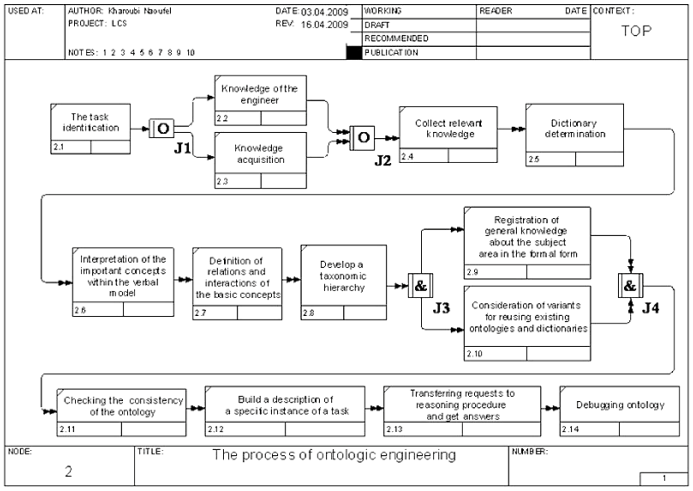


Figure 1: Process of Ontology Engineering

The proposed ontology is based on component behavior representation (Figure 1). Motivation behind our proposal is:

- System abstraction at a high level expressed in an ontology language enabling a prior evaluation and analysis of the impacts of possible online changes and reconfigurations.
- As a consequence of different multiple tools use, probable incoherencies may appear between design model and safety (normal) model.
- For architectural design, there is a heterogeneity in textual description, semantic representations, and syntax of the existing modeling languages.
- The need of reusable knowledge repository.
- Generating a safe executing model to be implemented.

For safety assessment, we elaborate, an ontology-based model description [11], [12]. We have several error behavior items that are stored in the ontology. Such reusable elements are a result of nominal models mixed with failure modes: each object is executable design system specification components under nominal behavior models. We transformed into formal safety models the obtained extended system models. A series of constraints (such as temporal and precedence constraints) are then added to the model in order to make it eligible to be analyzed and simulated by different tools, and so be reliable to making right decisions. All constraints are mapped onto an OWL domain ontology [13] written with the Protégé editor (Figure 2) [14]. Incorporated OWL description checked, the inference rules logic reasoned, and the ontology are able to detect mainly semantically inconsistent parts [15] as well as lack of model elements.

All errors interpreting cost are avoided: these errors encountered in the extended system model formal representation, are hardly to conceive. During the work, proposed architecture design is formulated in the Architecture Analysis and Design Language (AADL) [16], [17]. As a result, precise execution semantics are expressed for modeling software systems and their target hardware architecture. Safety models are generated in the AltaRica formal language [18]. We have opted for AADL and AltaRica languages, but we might as well have opted for other similar formalisms. The AADL language is adapted to describe functional and software architectures of embedded systems. The AltaRica language is an event-centric formal language suited for safety functioning modeling. Its translation is implemented in the form of model transformation. It would also have been possible to use other software such as ATL for example. There are series of analysis tools processed this language [19]-[21]. The proposed approach can be applied to validation contexts.

This work takes advantage of all previously cited domains: it is related to the scheduling problem in reconfigurable systems with a high degree of dynamicity. To overcome such complexity, we propose to work at a high abstraction level using an MDE approach combined with ontology description. The considered ontology language is OWL which gives a formal description of the database and the knowledge base and a semantic hierarchy of the whole system domain. In addition, after implementing the application constraints as rules, the ontology reasoning engine enhances the knowledge/data bases with new inferred axioms and data. All this is in order to have a perfect scheduling decision that takes into account all parameters/conditions and is flexible and dynamic according to the rule changes that may take place.

Related works are reviewed in Section 2. The domain ontology and the model-driven engineering process is presented in the Section 3. Section 4 is for the model transformation process and in the Section 5 we deal with the experimental case study. Then we conclude with a summary of the proposed process and we comment on its potential impact.

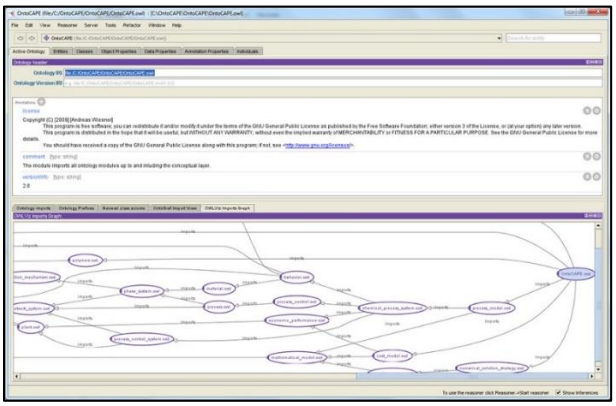


Figure 2: Protégé Framework Interface

2. Related Works

Reconfigurable embedded systems are very complex and include different Hw/Sw components. This led researchers to address new high-level design approaches to abstract that complexity, to facilitate development and to give more degree of flexibility when making changes and fixing errors. In this way, most research works on two principle modeling approaches: Architecture Analysis & Design Language and MARTE

(Modeling and Analysis of Real-Time Embedded systems) [22]. Those two methods consider only predefined reconfiguration models, so they cannot respond to the dynamicity of reconfigurable systems. MDE methods cope with that and permit to transform the model to another model until reaching a target one which responds better to a specific architecture and application domain. Moreover, the ontology is considered as a language that gives a formal description of a variety of knowledge (e.g. properties, features, relations) and different components (e.g. data, entities) forming a specific domain. Ontology languages are used in many domains, such as system engineering, semantic web application, logistics, artificial intelligence, system architecture, etc. It combines different techniques (e.g. semantic search, model matching) and formulates rules in order to tackle with systems complexity and domain conflicts. In [23] authors focus on e-health applications and propose a description for the whole software based on model-driven-architecture (MDA) approach. They exploit all steps of that approach, but especially the first stage called Computation-Independent-Model, with the integration of domain ontologies to represent particular information and all structures and relations existing in the system. In [24] authors apply the MDE approach on semantic web applications and provide an output model that generates annotated user interfaces and reduces some repetitive processes that may occur on the web pages.

To our knowledge, works related to model transformation based on an ontology language are not abundant. The main work related to that subject is [25]. Even in that paper, the purpose was to enhance cross-organizational modelling by adopting automatic generation then evolution of transformed model, a concern which is out of the scope of this paper. Two works on AADL transformation to AADL Altarica were made [26]: firstly, a model transformation [27], [28] based on system hardware architecture from the AADL and selected model reused from a project to another. Secondly, transformations were enriched with failure propagations written in AADL Error [29] and derived from AADL code. Transformation operation can be done if analysis and component relationships are defined. The important difference of our ontology-based model can be resumed in the semantic connection between the AADL and Altarica languages which permits to overcome the difference between them in term of syntax and scope [30], [31].

The focus of this work is to transform traditional representations of reconfigurable systems and make them more significant and related to their application domain. This led us to implement MDE approach combining the ontology languages (describing domain constraints in semantic rules) and applied them to solve scheduling problem.

3. Ontology Domain and the Model-Driven Approach

We propose an engineering process to model a dynamic application mapped on configurable computing system. The process is based on (1) modelisation of the application to be created, (2) analysis and verification of the rules relating different models, and (3) test of the schedulability of the application from the resulting model. We assume that a CCS is computer-based system consisting of hardware and software components, integrated in a physical environment, and requiring different timing/ressource requirements on the final outcome.

3.1. Application Modeling Approach

We aim in this work to exploit domain knowledge in design models. We want to formalize the system domain design, its annotations and its knowledge in a single model. To do this, we followed a four-step approach (shown in Figure 3):

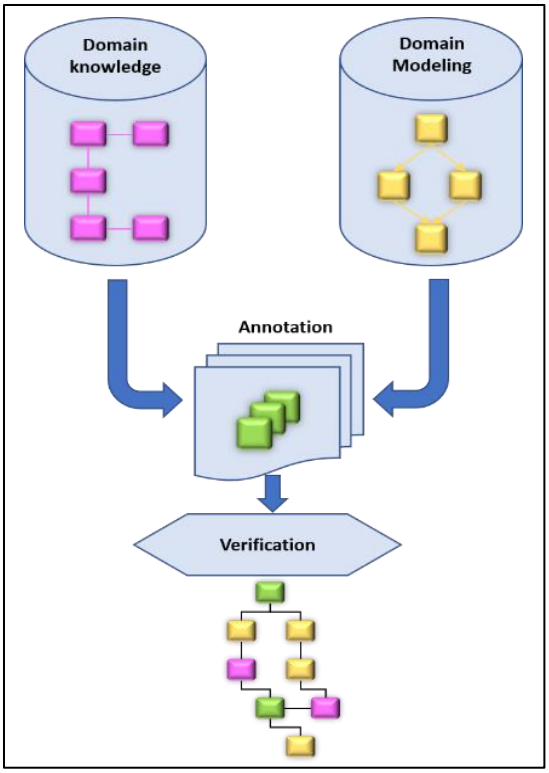


Figure 3: Proposed Application Modeling Approach

- The first step is to formalize and explicitly define the information (knowledge) of the domain of application in a formal ontology (entities, relationships, constraints, etc.). The properties of the ontology thus designed are domain-linked and completely independent of any context of use. In many cases, this ontology can be constructed from pre-existing standard ontologies.
- The second phase concerns the definition of design models. This means the formal definition of the properties corresponding to a given specification of the application domain.
- The third step, called annotation, is to define specific relationships between the different entities of the two previous models (design and ontology). Different relationships are available, they have their specific properties and describe specific rules related to the application domain.
- A verification step is required in any approach. This consists of validating the context of design models and domain properties expressed in the ontology, as well as the logic of the rules defined in the annotation step.

At the end, the final designed model has the advantage of well describing the domain knowledge with more details and properties derived from the ontology and rules generated during the annotation.

3.2. Scheduling Model

As mentioned above, we consider highly dynamic applications with soft real-time constraints. With such applications, missing a deadline does not cause catastrophic consequences, but only a performance degradation (quality of service parameter). In order to schedule these applications, we need a description that exhibits dependencies between the tasks and their own characteristics.

Each task is defined as a class in the ontology model. It represents a computational activity that needs to be performed according to a set of constraints (called slots in the ontology model). Each task i, for $i = \{1, 2, \ldots, k\}$ is defined by:

- ξ represents a set of active tasks T_i forming the application. Considered tasks could be sporadic and aperiodic.
- a_i : stand for the arrival time of task T_i.
- C_i : stand for the maximum computation (execution) time of task T_i.
- c_i : stand for the computation time of task T_i, i.e., the remaining worst case execution time WCET is needed for a computing elements (processor or logic circuit), at the current time, to complete the execution of task T_i without interruption.
- d_i : stand for the absolute deadline of task T_i.
- D_i : stand for the relative deadline of task T_i.
- S_i : stand for the first start time of task T_i.
- s_i : stand for the last start time of task T_i.
- f_i : stand for the estimated finishing time of task T_i.
- L_i : stand for the laxity of task T_i.
- R_i : stand for the remaining time of task T_i.

Baruah *et al.* [32] present a necessary and sufficient test for synchronous tasks with pseudo-polynomial complexity. For that reason, we present the following equations defining relations among the parameters defined above:

$$d_i = a_i + D_i \tag{1}$$

$$L_i = d_i - a_i - C_i \tag{2}$$

$$R_i = d_i - f_i \tag{3}$$

$$f_1 = t + c_1;\ f_i = f_{i-1} + c_i \ \forall\, i > 1 \tag{4}$$

$$R_1 = d_1 - t - c_1 \tag{5}$$

$$R_i = R_{i-1} + (d_i - d_{i-1}) - c_i \tag{6}$$

For any other task T_i, with $i > 1$,

$$f_i = f_{i-1} + c_i \tag{7}$$

and, by equation (3), we have:

$$R_i = d_i - f_i = d_i - f_{i-1} - c_i = d_i - (d_{i-1} - R_{i-1}) - c_i = R_{i-1} + (d_i - d_{i-1}) - c_i \tag{8}$$

The information above present the major inputs of the task class ontology. In our model, we define the scheduling ontology basic classes as follows:

- Basic scheduling elements (including tasks, resources and activities) and relation between them.
- Basic required components and constraints (including hard/soft constraints, temporal restrictions and resource limitations).
- Ontology of instances: each class has one or more related subclasses.

The Protégé editor provides the definitions for basic object types and properties such as application graph (Figure 4), tasks set properties (Figure 5), relations, numbers, etc. Figure 6 shows the data properties of a created task refereed to its temporal features in the ontology framework. Figure 7 is related to processor creation, where processor is either a software processor or a CLB for hardware task. Figure 8 represents slots where defined the properties of a task, their status and needed executing resources.

We mean by resources all hardware components that are responsible for the execution of tasks. Resources could be a processor (for software tasks) or a discrete number of logic circuit called CLB (configurable logic blocks) for hardware tasks in an FPGA architecture. Resources are defined in separate classes and are characterized by their execution time and/or number of CLB.

Rules and constraints are modelled as distinctive class. The class constraint has the same definition for both hard and soft constraints, and is applied to tasks or resources. The hard constraints are the rules that cannot be violated under any conditions, while the soft ones have to be satisfied by the completion time of the scheduler. For example, if we consider the temporal constraints in the EDF (Earliest Deadline First) scheduler [33], the semantic rule would be:

$$\begin{aligned} &\text{Task}\,(?T_i) \wedge \text{has\_D}_i\,(?T_i, ?d_i) \\ &\rightarrow \text{sqwrl:select}\,(?T_i) \wedge \text{sqwrl:orderBy}\,(?d_i)\,, \end{aligned} \tag{9}$$

as EDF is a preemptive algorithm which assigns the highest priority to be executed to the task T_i that has the lowest deadline D_i.

The both soft and hard constraints are applied to a task as well as a resource through the class schedule, which helps to satisfy schedulability conditions of the application execution.

3.3. Model Implementation

The Ontology-based model engineering architecture (Figure 9) is based on the Architecture Analysis & Design Language (AADL) and describes how the proposed ontology organizes the error models and the components into structural and functional interdependent hierarchies. AADL is considered as a textual

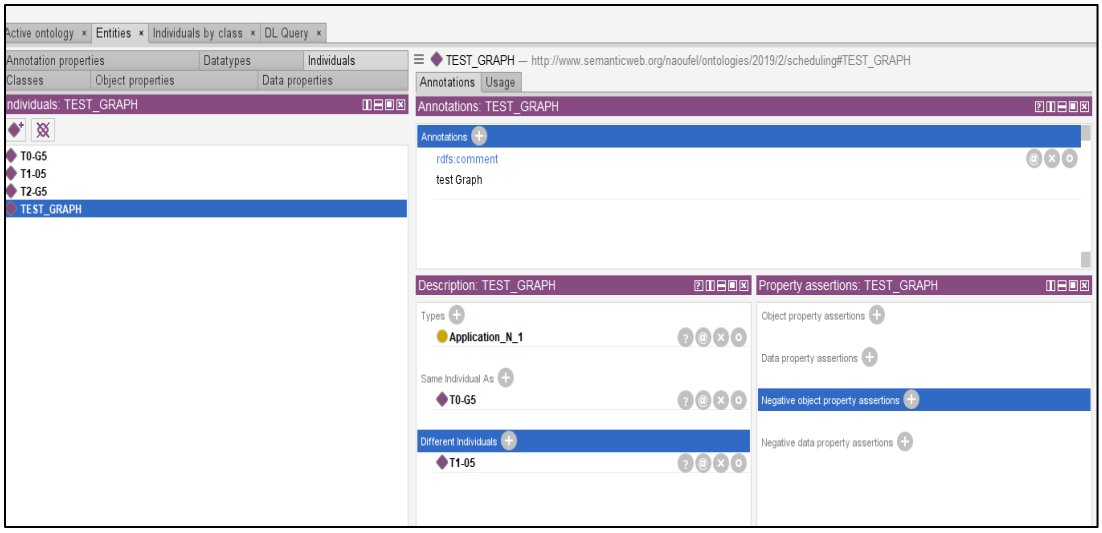


Figure 4: Application Graph Description

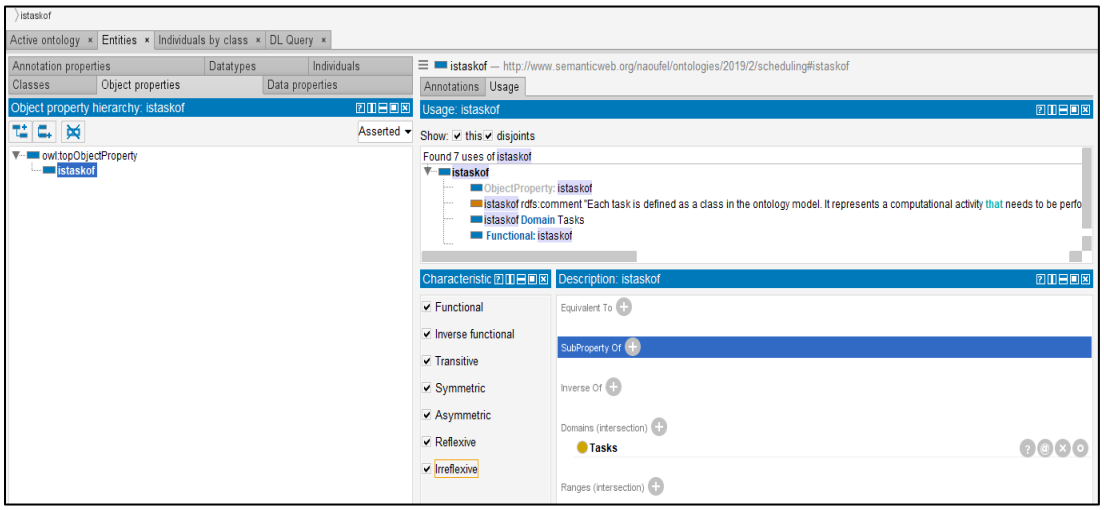


Figure 5: Task Class Properties

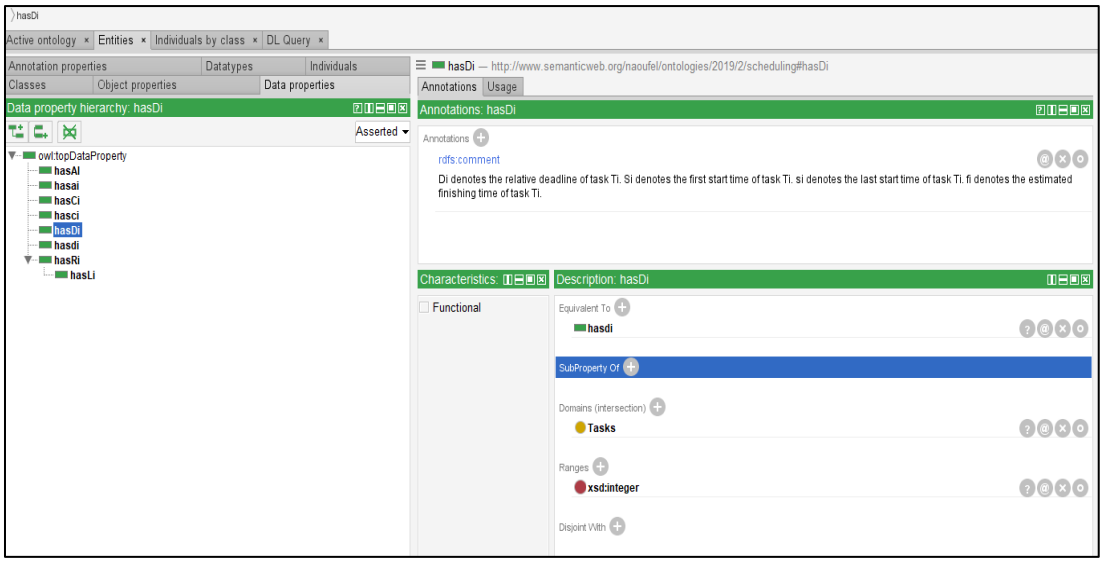


Figure 6: Task Creation With its Temporal Features

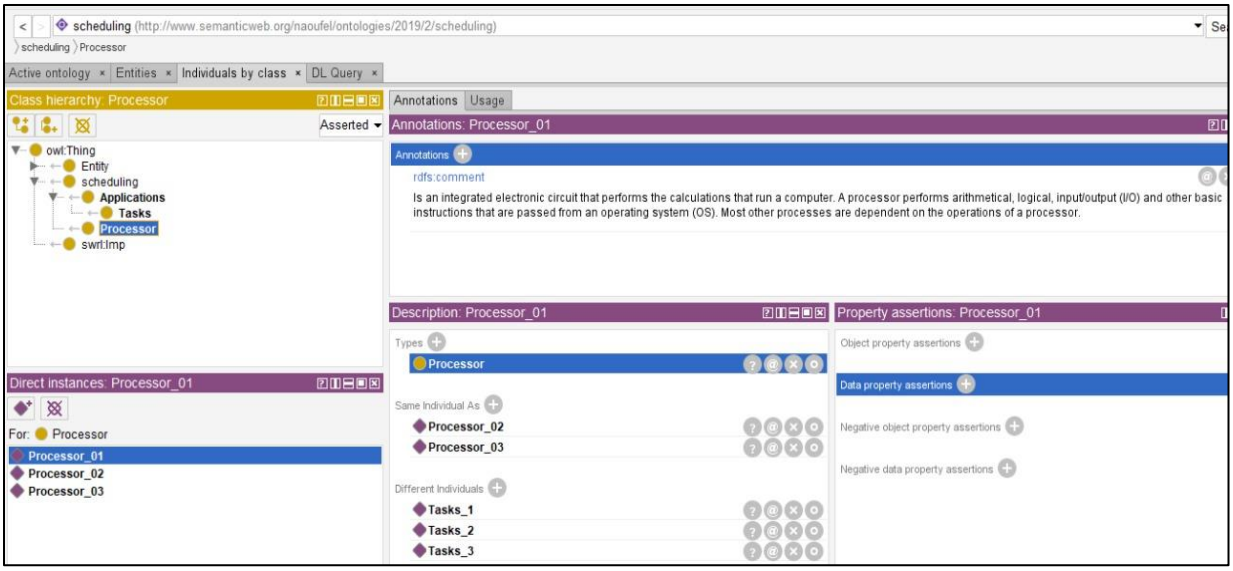


Figure 7: Processors Class Creation

language and graphical one too. It is used to specify the software (SW) and hardware (HW) architecture for the needed critical level of real time systems [34].

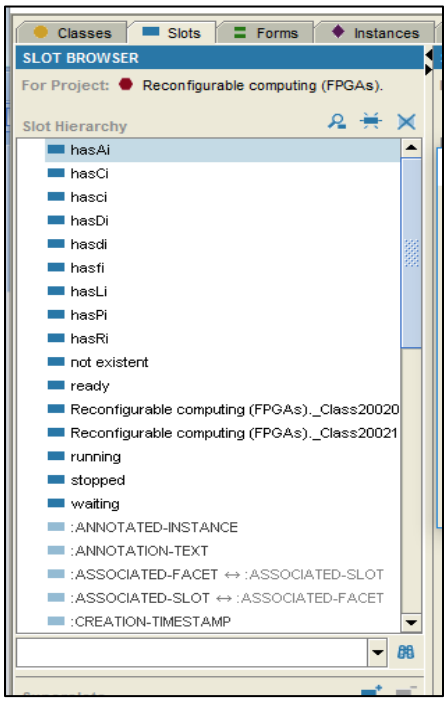


Figure 8: Task Properties, Status and Resources

In AADL language, each system is defined by its components. Each component identifies a number of elements of the actual system architecture. In our case, systems based on reconfigurable architecture are composed of a hardware part and software one. The AADL language permits to define software components (process, thread, data), as well as hardware components (device, processor, memory, etc.). Moreover, this design language defines precise legality rules that control the different component assemblies, and this in both static and dynamic (on-execution) way. To describe the communication between components, AADL defines the connectors and ports. A port performs a particular task in the context of the connector that connects components. All instances of connectors (including their ports and roles), compositions of components and components define the implementation operation. To enable linking flows to internal states, a model describes what initial states may evolve and it includes events, states, transitions to perform various analyses with different constraints on the model.

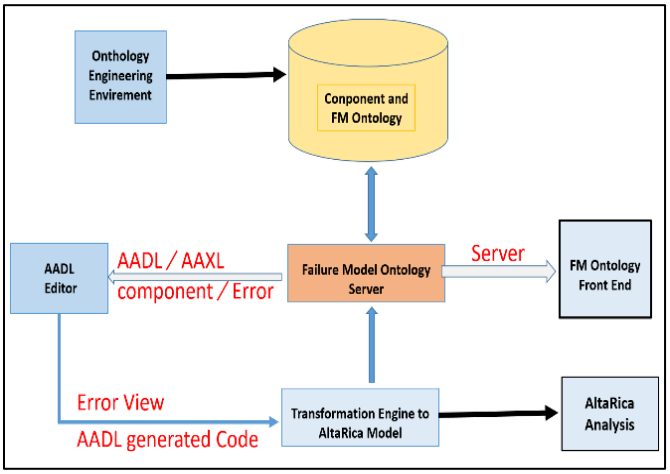


Figure 9: Model Engineering Process Based on Ontology

The objective of the design of the ontology is to prove the correlation of constructs related to the AADL specification, the retrieval, the storage operation's permission, and to offer the desirable reasoning capabilities among them. We specified the ontology [33] in the OWL DL fragment of OWL and we applied Pellet DL reasoner combined with the Protégé ontology editor, and this to customize inference rules' determination.

The proposed ontology is mainly based on three aspects [36]. Firstly, specify all necessary constructs for AADL core that permit to model components (e.g. data) with their corresponding characteristics and so facilitate ontology to add or extract new components. Secondly, for each AADL error, ontology offers the needed concepts to represent error models in terms of states and events, as hardware errors, computational problems and memory

exceptions, and thirdly, the ontology offers the connections and subcomponents (constructs) for component implementation. To achieve this task, we use ontology-based representation of component to detect rules custom inference [37]. The system detects three inconsistencies types making certain formal analysis models invalid: error of incomplete or missing transition; conflict transition that may occur when more than one event is triggered simultaneously; and finally, error in the state of a component due to an error in the failure scheduling scenario. The rules related to inconsistent semantics permit new inconsistent levels to be added. Meta-modeling design approach followed the ontology and determines instantiations and implementation steps. We represent the AADL component as an ontology class. The error models are presented as the ontology hierarchy subclass. The proposed model-driven engineering process enforces rules and constraints on the associations between the error models and the components' constraints. It permits to select and resume the failure modes from the error model hierarchy, then associate them with the nominal mode. The ontology model is then checked for possible component inconsistencies and transform the extended architecture model to the safety one and finally analyze the safety model with tools that provide model checking and simulation.

4. Model Transformation Process

For discovering structurally equivalent constructs, existing modeling experience is considered a basic step for model transformation, and the first set of rules is dedicated to the transformation of AADL components into AltaRica nodes. The transformation rules are therefore driven by the mapping of knowledge with the constructs used through the underlying domain ontology. The benefit of the transformation from AADL models to AltaRica is to expand the set of safety evaluation tools for AADL. Hence, all system components (tasks, data, CLBs, processors, memory, etc.) are transformed to AltaRica nodes while conserving their same original features.

Then, AltaRica state statements are generated based on the AADL components' properties. We manage a set of rules which concern the components' error models. The AADL error, states and transitions are transformed to corresponding AltaRica which are filled with assignments found in the matched AADL. Another set of rules focuses on failure propagation. Since no support failure capabilities are made, a transition declaration creates additional component variants and other transformation rules set related to the used architecture design process and the associated AADL editor is made. The problem is that the editor tool follows a control-flow based approach, which make a semantic gap compared to the data-flow approach of the AltaRica specification.

An AADL component can be of two kinds from a safety point of view: either its properties are filled or are not filled *{lost}*. In the following, we will focus on the generic transformation. The names of the variables vary according to the connection ports of the component and the formulas of the component assertion depend on the data flow path. Indeed, any component of an AADL system:

$C = \{D^{in}, D^{out}, K, Ty\}$, where D refers to data dependencies related to the task data flow, K refers to the knowledge related to properties of safe operation of the component, and Ty refers to the type of the component,

is translated into an AltaRica node:

$N = \{ F, S, D^{in}, D^{out}, \Sigma, \sigma, I\}$, where F is a field of finite values of the variables, S is for state, Σ is a set of events, σ is an affirmation function $S \times D^{in} \rightarrow D^{out}$, and I refers to initial conditions,

such as:

- the state variable s takes its *{correct}* value if the component is working normally otherwise its value is equal to the name of the failure mode *{lost}* declared in AADL ;
- the dependency connections D^{in} and D^{out} form the AADL (and also the AltaRica) component interface;
- a failure transition is produced by events $Evts \subset \Sigma$ leading to a failure mode;
- for each outgoing flow variable $Out_j \in D^{out}$, we consider the set $\{In_i \mid (In_i, Out_j) \in D^{in} \times D^{out} \}$ of incoming flow variables on which the Out_j variable depends. Then, if we consider only the loss of a component, the statement associated with Out_j is the following Out_j = *{if s = correct and* In_i *= correct then correct, else lost)*;
- finally, we consider that, initially, components states are correct *I(s) = {correct}*.

Considering an AADL system $\boldsymbol{\psi} = \{D^{in}, D^{out}, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n, A, R\}$, $\boldsymbol{\psi}$ is transformed into an AltaRica node $N = \{N_0, N_1, \ldots, N_n, V\}$ with:

- each subsystem $\boldsymbol{\psi}_i$ is transformed into an AltaRica N_i node;
- N_0 has for dependencies variables D^{in} and D^{out} ;
- R refers to connections between components and is equivalent to the σ function;
- the synchronization vector V is given by the allocation relation A: if a task $\boldsymbol{\psi}_i$ is allocated to a processor $\boldsymbol{\psi}_j$ then $(evt_i, evt_j) \in V$ for $evt \in \Sigma$.

5. Experimental Case Study

We choose as a case study a 3D image synthesis application. It is a complex application with some flexibilities related to the computation time of some tasks and their executing occurrence. The application class hierarchy is implemented in OWL Protégé editor shown in Figure 10. The input of the application graph is a set of the coordinates of the different polygons' summits that represent the 3D object. All coordinates are defined relative to a local space where the 3D object is located. Those coordinates are manipulated by different arithmetic functions (tasks) that create the animation, such as : Loading, Scaling, Adding, Rotating, Translating, etc. Tasks are implemented with their temporal features, and precedence relation according to the application graph. Rules are implemented by Semantic Web Rule Language (SWRL) interface (Figure 11) and combined with the scheduling ontology in order to improve the domain knowledge. Examples of some temporal rules are defined in equations (1) to (8), other rules related to allocation are like in (10) to indicate that processor can hold only one task to execute.

$$Tasks(?T) \wedge hasLi(?T, ?L) \wedge Processor(?P) \wedge hasProcessor(?T, ?P) \rightarrow sqwrl:select(?T) \wedge sqwrl:orderBy(?L) \wedge sqwrl:groupBy(?P) \quad (10)$$

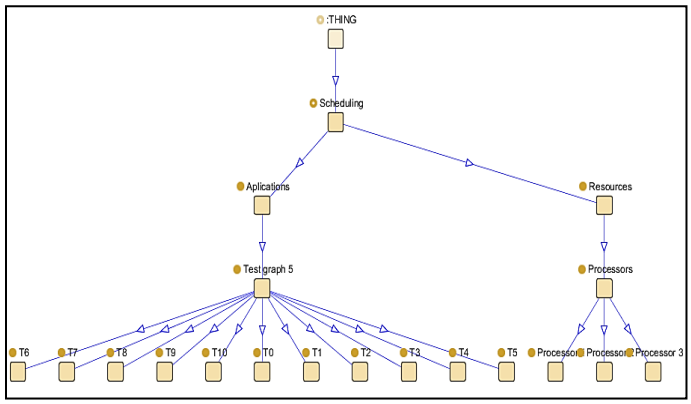


Figure 10: Class Hierarchy in Scheduling Ontology

After creating the SWRL rules, the reasoner component is able to infer the user query based on the ontology knowledge and predefined rules. To type queries, we tested two methods:

- The first method is simple and use existing information in the knowledge base without the need to any inferences. We use such method to verify existing properties in our ontology knowledge. For example, if we want to verify which processor is available for accepting a task execution, or the remaining computation time of some tasks.
- The second query method triggers the rule-based reasoner for the inference process with the knowledge which may lead to enhance and enrich the knowledge base. For example, if we consider tasks migration between processors or CLBs, and we want to decide migration of task T executing on processor1 when processor1 is overloaded. Hence, the rule-based reasoner has to infer the possibility of such migration based on the status of each processor and check that whether this change (so this query) is feasible or not.

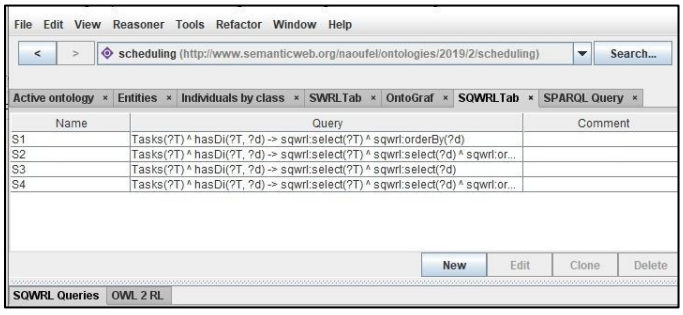


Figure 11: Example of SWRL Rules

Inference results are resulting from SPARQL queries. The SPARQL service permits to check different information as the status of tasks and it provides feedback concerning events and decision failures. The tested query results are correct, and the rule-based reasoner is useful for the scheduling problem. Based on that formalized temporal/resource rules combined with the semantic reasoner and inference rules, the ontology model gives the resulting possible tasks scheduling (Figure 12).

Some examples of successfully implemented scheduling rules (EDF algorithm (9), Least Laxity First LLF algorithm (11), Rate Monotonic algorithm RM (12):

$$Tasks(?T) \wedge hasLi(?T, ?L) \rightarrow sqwrl:select(?T) \wedge sqwrl:orderBy(?L) \quad (11)$$

$$Tasks(?T) \wedge hasCi(?T, ?C) \rightarrow sqwrl:select(?T) \wedge sqwrl:orderBy(?C) \quad (12)$$

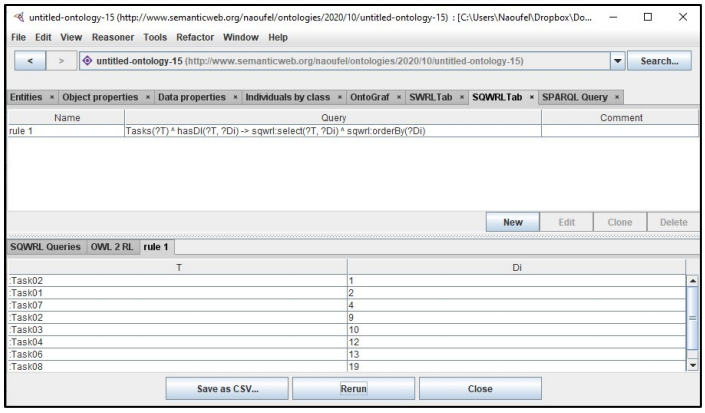


Figure 12: SQWRL Rules Result

By using AADL, nominal models were designed and combined also with failure models of the domain ontology. The system's adding process executes a simple add function, and it is defined in the Adding task. The operation is called performed when the required data context of the former task is provided to the data context of the latter task. To validate operation, scheduler check rules and constraint violation in the error model represented by the Overflow Event State, shown in Table1.

Table 1: Example of incompleteness transition Error

Implementation	Error model
error model implementation OVERFLOW_2.Notionala transitions Error_Free -[overflow]-> Over_Flow; Error_Free -[overflow]-> Ovf_Temp; end OVERFLOW_2.Notionala;	error model OVERFLOW_2 Features Error_Free: initial error state; Ovf_Temp: error state; Over_Flow: error state; overflow: error event; end OVERFLOW_2;

6. Conclusion

We faced in this article the problem of modeling the complexity of reconfigurable systems while taking advantage of its flexibility and dynamic behavior. For this, we proposed a model-driven engineering process that follows a set of steps. The proposed modeling process makes a modular and extensible representation of the system architecture. Model profits from the ontology language capability to represent complex models and address heterogeneous domains information of the dynamically reconfigurable systems. We formalized the known temporal

scheduling problem in reconfigurable systems domain. The ontology knowledge model was implemented using Protégé editor. Moreover, the developed ontology is improved with SWRL inferred rules. Model editing is enhanced, since errors can be detected without the need to perform complex analyses. The model and the rules reasoner and the resulting scheduling decision were verified with a case study.

Future works aim to extend model transformation functionality and involve more dynamic system characteristics such as tasks migrations between hardware and software resources and the total and partial reconfigurability. A filtering process should be added to the reasoner in order to eliminate unnecessary rules and so avoid ambiguity or conflict when making scheduling decision. Another objective is to test more scheduling algorithms with the incorporation of additional types of error models.

Acknowledgment

This work has been approved by the research grant program of the deanship of scientific research at TAIF University.

References

[1] M.A. Cardin, "Enabling flexibility in engineering systems: A taxonomy of procedures and a design framework," Journal of Mechanical Design, Transactions of the ASME, **136**(1), 2014, doi:10.1115/1.4025704.

[2] S. Paul, S. Bhunia, S. Paul, S. Bhunia, "A Survey of Computing Architectures, Springer New York": 11–27, 2014, doi:10.1007/978-1-4614-7798-3_2.

[3] S. Vassiliadis, D. Soudris, "Fine-and coarse-grain reconfigurable computing", Springer Netherlands, 2008, doi:10.1007/978-1-4020-6505-7.

[4] L. Gong, O. Diessel, "Functional verification of dynamically reconfigurable FPGA-based systems", Springer International Publishing, 2015, doi:10.1007/978-3-319-06838-1.

[5] R. Tessier, K. Pocek, A. DeHon, "Reconfigurable computing architectures," Proceedings of the IEEE, **103**(3), 332–354, 2015, doi:10.1109/JPROC.2014.2386883.

[6] G. Gracioli, A.A. Fröhlich, R. Pellizzoni, S. Fischmeister, "Implementation and evaluation of global and partitioned scheduling in a real-time OS," Real-Time Systems, **49**(6), 669–714, 2013, doi:10.1007/s11241-013-9183-3.

[7] M. Alam, A. Khan, A.K. Varshney, "A review of dynamic scheduling algorithms for homogeneous and heterogeneous systems," in Advances in Intelligent Systems and Computing, Springer Verlag: 73–83, 2018, doi:10.1007/978-981-10-8533-8_8.

[8] A. Rodrigues Da Silva, "Model-driven engineering: A survey supported by the unified conceptual model," Computer Languages, Systems and Structures, **43**, 139–155, 2015, doi:10.1016/j.cl.2015.06.001.

[9] L. Yang, K. Cormican, M. Yu, "Ontology-based systems engineering: A state-of-the-art review," Computers in Industry, **111**, 148–171, 2019, doi:10.1016/j.compind.2019.05.003.

[10] B. Selić, S. Gérard, "Model-Based Schedulability Analysis", Elsevier: 201–220, 2014, doi:10.1016/b978-0-12-416619-6.00010-9.

[11] V. Mascardi, A. Locoro, P. Rosso, "Automatic ontology matching via upper ontologies: A systematic evaluation," IEEE Transactions on Knowledge and Data Engineering, **22**(5), 609–623, 2010, doi:10.1109/TKDE.2009.154.

[12] J. Li, J. Tang, Y. Li, Q. Luo, "RiMOM: A dynamic multistrategy ontology alignment framework," IEEE Transactions on Knowledge and Data Engineering, **21**(8), 1218–1232, 2009, doi:10.1109/TKDE.2008.202.

[13] G. Chen, T. Jiang, M. Wang, X. Tang, W. Ji, "Modeling and reasoning of IoT architecture in semantic ontology dimension," Computer Communications, **153**, 580–594, 2020, doi:10.1016/j.comcom.2020.02.006.

[14] M.A. Musen, "The Protégé Project: A Look Back and a Look Forward". AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4), 2015, doi:10.1145/2557001.25757003.

[15] D.L. McGuinness, F. van Harmelen, "OWL Web Ontology Language - Reference," W3C Recommendation [Online], Available at: Http://Www.W3.Org/TR/2004/REC-Owl, 1–53, [Accessed: 02 February 2016], 2004.

[16] P.H. Feiler, B.A. Lewis, S. Vestal, "The SAE architecture analysis & design language (AADL) a standard for engineering performance critical systems," in Proceedings of the 2006 IEEE Conference on Computer Aided Control Systems Design, CACSD, 1206–1211, 2007, doi:10.1109/CACSD.2006.285483.

[17] SAE: Architecture Analysis and Design Language (AADL) AS-5506B. Tech. rep., The Engineering Society For Advancing Mobility Land Sea Air and Space, Aerospace Information Report, Version 2.1 (September 2012).

[18] Prosvirnova, T., Batteux, M., Brameret, P.A., Cherfi, A., Friedlhuber, T., Roussel, J.M., Rauzy, A.: "The altarica 3.0 project for model-based safety assessment". In: Proceedings of 4th IFAC Workshop on Dependable Control of Discrete Systems, DCDS 2013. IFAC, York (Great Britain) (September 2013).

[19] P. Bieber, J. Blanquart, G. Durrieu, D. Lesens, J. Lucotte, F. Tardy, M. Turin, C. Seguin, E. Conquet, "Integration of Formal Fault Analysis in ASSERT: Case Studies and Lessons Learnt," 4th European Congress on Embedded Real-Time Software (ERTS), 1–9, 2008.

[20] P. Hönig, R. Lunde, F. Holzapfel, "Model Based Safety Analysis with smartIflow †," Information, **8**(1), 7, 2017, doi:10.3390/info8010007.

[21] D. Bošnački, A. Wijs, "Model checking: recent improvements and applications," International Journal on Software Tools for Technology Transfer, **20**(5), 493–497, 2018, doi:10.1007/s10009-018-0501-x.

[22] M. Faugère, T. Bourbeau, R. De Simone, S. Gérard, "MARTE: Also an UML profile for modeling AADL applications," in Proceedings of the IEEE International Conference on Engineering of Complex Computer Systems, ICECCS, 359–364, 2007, doi:10.1109/ICECCS.2007.29.]

[23] N. Laaz, K. Wakil, Z. Gotti, S. Gotti, S. Mbarki, "Integrating domain ontologies in an MDA-based development process of e-health management systems at the CIM level," in Advances in Intelligent Systems and Computing, Springer: 213–223, 2020, doi:10.1007/978-3-030-36664-3_25.]

[24] N. Laaz, S. Mbarki, "OntoIFML: Automatic Generation of Annotated Web Pages from IFML and Ontologies using the MDA Approach: A Case Study of an EMR Management Application," in Proceedings of the 7th International Conference on Model-Driven Engineering and Software Development, Scitepress: 353–361, 2019, doi:10.5220/0007402203530361.]

[25] S. Roser, B. Bauer, "Automatic generation and evolution of model transformations using ontology engineering space," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer, Berlin, Heidelberg: 32–64, 2008, doi:10.1007/978-3-540-92148-6_2.

[26] G. Guizzardi, R.A. Falbo, R.S.S. Guizzardi, "The role of Foundational Ontologies for Domain Ontology Engineering: a case study in the Software Process Domain," IEEE Latin America Transactions, **6**(3), 244–251, 2008, doi:10.1109/tla.2008.4653854.

[27] N.F. Noy, M.A. Musen, "Ontology versioning in an ontology management framework," IEEE Intelligent Systems, **19**(4), 6–13, 2004, doi:10.1109/MIS.2004.33.

[28] W. Alakwaa, A. Salah, "Model Transformation from Ontology Model to Content Analysis Model," International Journal of Computer Applications, **7**(3), 5–12, 2010, doi:10.5120/1147-1501.

[29] P. Wang, Z. Jin, L. Liu, G. Cai, "Building toward capability specifications of web services based on an environment ontology," IEEE Transactions on Knowledge and Data Engineering, **20**(4), 547–561, 2008, doi:10.1109/TKDE.2007.190719.

[30] M.R. Khondoker, M.R. Khondoker, P. Mueller, "Comparing Ontology Development Tools Based on an Online Survey," 2010.

[31] T.C. Eskridge, R. Hoffman, "Ontology creation as a sensemaking activity," IEEE Intelligent Systems, **27**(5), 58–65, 2012, doi:10.1109/MIS.2012.101.

[32] S. Baruah, G. Koren, B. Mishra, A. Raghunathan, L. Rosier, D. Shasha, "On-line scheduling in the presence of overload," in Annual Symposium on Foundations of Computer Science (Proceedings), Publ by IEEE: 100–110, 1991, doi:10.1109/sfcs.1991.185354.

[33] G.C. Buttazzo, "Hard Real-Time Computing Systems", Springer US, 2011, doi:10.1007/978-1-4614-0676-1.

[34] Y. Zhao, D. Ma, "Embedded real-time system modeling and analysis using AADL," in ICNIT 2010 - 2010 International Conference on Networking and Information Technology, 247–251, 2010, doi:10.1109/ICNIT.2010.5508520.

[35] T.C. Jepsen, "Just what Is an ontology, anyway?," IT Professional, **11**(5), 22–27, 2009, doi:10.1109/MITP.2009.105.

[36] Y. Ma, L. Liu, K. Lu, B. Jin, X. Liu, "A graph derivation based approach for measuring and comparing structural semantics of ontologies," IEEE Transactions on Knowledge and Data Engineering, **26**(5), 1039–1052, 2014, doi:10.1109/TKDE.2013.120.

[37] K. Jetinai, N. Arch-int, S. Arch-int, "Ontology Mapping and Rule-Based Inference for Learning Resource Integration," Journal of Information and Communication Convergence Engineering, **14**(2), 97–105, 2016, doi:10.6109/jicce.2016.14.2.097.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 811-816 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Algorithm Design for Accurate Steps Counting Based on Smartphone Sensors for Indoor Applications

Hani Muhsen*, Odeh Al-Amaydeh, Rakan Al-Hamlan

Mechatronics Engineering Department, German Jordanian University, Amman, 11180, Jordan

ARTICLE INFO

Article history:
Received: 09 October, 2020
Accepted: 18 November, 2020
Online: 24 November, 2020

Keywords:
Algorithm design
Steps Counting
Pedometer
Accelerometer

ABSTRACT

This paper aims to design an accurate and robust algorithm for counting the steps based on the smartphone's accelerometer for indoor applications. Different daily activities have been considered in the experimental scenarios, including normal walking with various smartphone positions and running activities. The detection process's accuracy is ensured by setting an appropriate segregation approach and defining a fixed threshold for each experiment. The algorithm generates a new envelope signal that mimics the collected steps signal without the vibrations and noise elements to avoid miscounting the exact steps. The followed approach results have been compared with the pedometer applications on the android platform and HUAWEI Watch GT-731. The proposed algorithm showed improved detection accuracy in contrast to the other literature approaches. The obtained accuracy was around 99% for the standard walking scenario and approximately 97% for the running activity scenario.

1. Introduction

Lately, the step count has been extensively used in various applications, such as health monitoring, auxiliary positioning, and gaming. Earlier, the Pedestrian Dead Reckoning (PDR) was used as a localization method based on advanced built-in sensors [1] that exist these days in Smartphones. These sensors found in the smartphones include accelerometers, proximity sensors, gyroscopes, magnetometers, and Global Positioning System. The accuracy of the steps count is one of the challenges of the PDR. Therefore, several studies have been addressed, increasing the accuracy of steps count while walking using the smartphone's accelerometer [2–4].

One of the aspects has been considered in the literature: smartphone placement's impact on measurement accuracy. The smartphone's position and placement might cause jitter during the human motion, which leads to counting false steps [5]. Also, remaining still, watching videos, playing games, and many other activities can lead to a wrong step count [6]. In [7], it was recommended to place the smartphone at the user the fingertips with the view screen point upwards while using the phone. If the smartphone is placed in the jacket or trouser pocket, it is recommended to put the top of the smartphone upright. One of the challenges of this approach is getting rid of the noisy signals accompanying the detected steps. Therefore, a discrete Kalman filter was used in the literature to minimize the smartphone accelerometer's noise during human motion [8]. The Footsteps are usually calculated using an algorithm in the smartphone on the android platform, which can count precisely by dividing the measurement processes into two phases. The first phase includes collecting the linear acceleration and the gravity values from the accelerometer of the smartphone. These values are then used to derive the horizontal components and identify possible start points of regular periodical fluctuations. The second phase includes finding the correlation coefficients to determine if the collected sensing measurements exhibit similar tendencies, and accordingly, it finds the steps count. The results showed higher measurement accuracies if the user holds the phone statically [9]. Although several approaches have been proposed and used for step counting, most of these approaches are not validated with robust methods. Therefore, this work aims to design a robust algorithm for step count purposes with high accuracy. The proposed algorithm will distinguish between false steps and real steps to reduce the measurement error and enhance its credibility.

This paper is structured as follows. In section II, a concise overview of related work. In section III, the proposed approach to count the steps is presented. The obtained results are discussed in section IV, and the paper ends with conclusions.

2. Related Work

Several algorithms have been proposed to enhance the step count based on the smartphone's sensors in-state of the art. In [6, 10], a window peak detection algorithm has been proposed based on the smartphone's accelerometer. The algorithm could afford an average accuracy of 95%, considering the different ways that a

*Corresponding Author: H. Muhsen, Email: Hani.Mohsen@gju.edu.jo

smartphone can be carried. The algorithm's parameters have been optimized and applied to measured data from three different volunteers holding the smartphone at different positions. A new algorithm based on the smartphone sensor has been proposed in [11] to estimate the traveled distance by walking accurately. The algorithm counts the number of peaks in the measured signal by the smartphone's accelerometer. The next step is to detect the step-length, which varies from one step to another. Many adaptive algorithms have been designed to count the footsteps based on LSM (Lifestyle modification). These approaches rely on measuring the movement by a 3-axis accelerometer.

The proposed designs can enhance the step detection accuracy using step amplitude with accuracies, which can reach up to 99%. In [12], two adaptive algorithms have been proposed to count human steps using only the gyroscope. Accordingly, the Fast Fourier Transform (FFT) was applied to the measured data to obtain real-time walking frequency. The proposed algorithms' average accuracy ranged from 95% to 97%, considering different walking speeds. Further, the step counting based on the smartphone's accelerometer has been developed in Indoor Positioning System as proposed in [5]. The approach relies on pattern recognition for step detection and counting. Different scenarios for the smartphone placement and the number of steps higher than one hundred steps have been considered in the validation stage. The result showed high accuracy regardless of the smartphone position. In many step counting approaches, the non-invasive option depends on utilizing the accelerometer to determine the arm movement like a seizure sufferer. In these approaches, the algorithm depends on employing low-power, embedded microcontroller devices in real-time, and they are not intrusive to the patient's life. Therefore, they are less likely to become non-compliant with monitoring [13]. These approaches are applied in wireless biomedical networks that attempt to monitor patients in a completely non-invasive, non-intrusive manner for a specific condition. Nowadays, the Global Positioning System (GPS) is employed to detect and count the steps outdoor. The GPS showed low accuracy in indoor scenarios due to the indoor equipment's electromagnetic interference and the building structures [14]. Therefore, the PDR and smartphone's sensors have been integrated to detect and count the steps accurately in indoor scenarios as in [15].

The power consumption of steps counting algorithms is one of the research topics, especially in smartphones with small battery capacity. In [16], the accelerometer Tap Interrupt algorithm was used to reduce power consumption. The accuracy and measured power consumption at the system level have been compared with a periodic sensor-reading algorithm. The tap interrupt approach presented a lifetime of the battery that is 175 % longer than that of a non-gyroscope 30 ms polling method. The battery life can be extended to 863 % by putting both the processor and the gyroscope in a sleep mode during most operating time. The average power consumption may be reduced by reducing the wake-up frequency from the sleep state and shortening the processing time at each wake-up. Recently, machine learning algorithms have been introduced for step length estimation based on the smartphone's sensors. Utilizing these tools aims to improve the steps detection, counting, and step length estimation accuracy.

In [1], the neural network is implemented in a multi-mode intelligent recognition approach based on the smartphone's sensors. The empirical results presented a maximum positioning error lower than 1% of the total path distance. In [17], an adaptive algorithm has been proposed to enhance the steps counting accuracy in indoor positioning services. The algorithm aims to solve the limitations by introducing four mains parameters. These parameters include the minimal peak distance, minimal peak prominence, thresholding, and vibration elimination. The algorithm presented an improved average accuracy of 97%. Also, the results were better than the results obtained by commercial steps counting applications. This paper aims to introduce an algorithm to obtain accurate and robust step counting for indoor applications. The results of this work will be compared with state of the art to validate the followed approach.

3. Methodology

In this work, the smartphone accelerometer is used to collect the acceleration value during human movement. The collected data is transferred to the MATLAB drive using an application installed on the users' smartphones, called "MATLAB on mobile." MATLAB on mobile offers accessibility on the available sensors in the user smartphone. The obtained data by the accelerometer of the smartphone include four parameters. The first parameter represents the timestamp values, where it will indicate the (date, time in Hours: Minutes: Seconds). The second parameter represents acceleration values on the X-axis. The third parameter defines the acceleration on the Y-axis, and the fourth parameter represents the acceleration on the Z-axis. Most of the available accelerometer modules measure the value of acceleration quantity as a vector, and many modules can detect both linear and angular acceleration [18]. Angular acceleration can be defined as the time rate of change in angular velocity.

In contrast, the linear acceleration is the time rate of change in linear velocity. In this work, only the linear acceleration vector is used to estimate the number of steps elapsed. The angular acceleration caused by the human steps has a small value and can be neglected. Hence, the linear acceleration within a short period can be calculated as follows according to [18]:

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt} \tag{1}$$

where a is the acceleration in [m/s^2], Δv represents the change in velocity in [m/s], Δt is the time interval in [s]. The accelerometer is used to measure the linear acceleration on the x-axis, y-axis, and z-axis independently. It is hard to predict the accelerometer's orientation every time a user operates the count steps functionality. Therefore, the acceleration vector is converted at each time point into an overall scaler value. Hence, Acceleration values will be determined like human walking or running regardless of the accelerometer's orientation. The overall acceleration value is calculated as the square root of three acceleration values, as expressed by equation (2) as in [11].

$$a_{all} = \sqrt{a_x^2 + a_y^2 + a_z^2} \tag{2}$$

Indeed, the gravitational force affects the acceleration value due to the architecture of the accelerometer module. So, it is hard to estimate the axis affected by gravity since it depends on the smartphone's orientation when it is carried by hand or placed in the pocket. Fig.1 depicts the acceleration signals of the three axes. Therefore, the gravitational has been removed from the overall acceleration signal as follows.

$$a_{final} = a_{all} - avg(a_{all}) \tag{3}$$

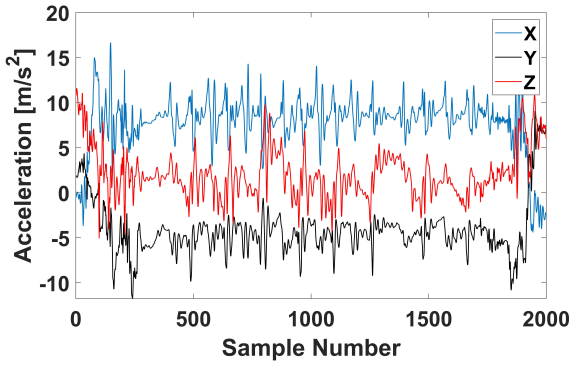


Figure 1: Acceleration values on X, Y, and Z-axis,

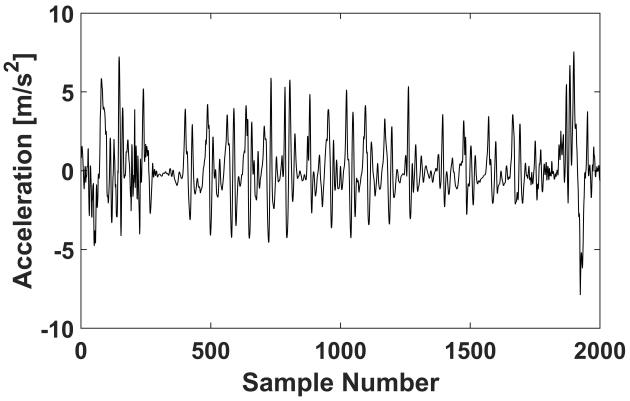


Figure 2: Sum of Acceleration Vectors of X, Y, and Z-axis after subtracting the mean value

The human steps generate little height peaks in the acquired acceleration data. These peaks have limited heights, which might vary depending on the movement's behavior while walking, as shown in figure 2. Also, electromagnetic interference can turn the measured signal into a noisy signal. This noise might result from the surrounding appliances. Therefore, finding maximum local peaks over a specifically tuned threshold can be employed to estimate the number of detected steps.

The previous stage can be fulfilled by generating an envelope function MATLAB, which is determined using spline interpolation over local maxima separated by at least the samples (n_p). The number of samples (n_p) can be calculated using MATLAB. Consequently, the unwanted vibrations and fake steps are removed, making the steps counting much more accessible, as depicted in figure 3.

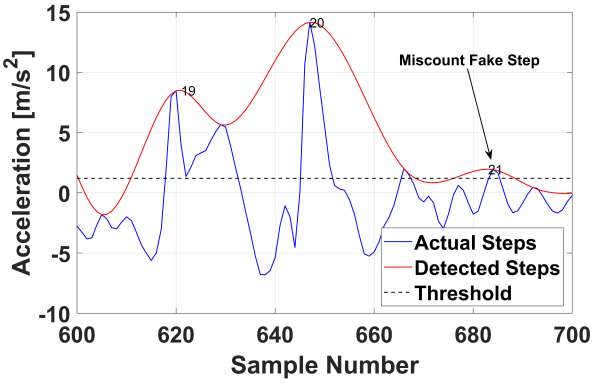


Figure 3: Applying the envelope function on the measured signal to remove the noise and eliminate the vibrations.

In this approach, it is found that the main parameter which impacts the estimation's accuracy is the threshold definition. This threshold depends on the sampling rate, the step rate, and the average of the orthogonal accelerations' envelope function. Therefore, the threshold is defined by equation 4.

$$\varepsilon = \frac{avg\left(a_x^*\right)+avg\left(a_y^*\right)+avg\left(a_z^*\right)}{n_p} - d \ , d \in \{3,5\} \qquad (4)$$

where a_x^*, a_y^*, a_z^* represent the upper peaks envelope, d is a correction factor, n_p is the number of samples, and it is defined as follows:

$$n_p = round(R_s / S_r) \qquad (5)$$

In this work, the step rate S_r is assumed to equal five steps per second. This assumption valid since this value represents the average of the human steps per second. The sampling period τ_s has been calculated using the new signal data's timestamp by finding the difference between every two values in seconds.

The segregation approach in this work is performed by generating a new signal based on the envelope function, which returns the upper limit for the average of the measured accelerations signals to remove the impact of vibrations and noise to ease counting the number of the steps, as shown in Fig.3 in red. This transformation will lead to an accurate detection process. The next step is to find the maximum difference value to represent the sampling period and is given by equation (6).

$$\tau_s = \max\left(t_{n+1} - t_n\right) \qquad (6)$$

Afterward, the sampling rate was calculated as follows:

$$R_s = 1/\tau_s \qquad (7)$$

Comparing the peaks of the final acceleration signal $P_{sk}(i)$ obtained by the envelope function with the threshold ε will estimate the number of steps over the path length. The constraint which must be fulfilled to count the number of peaks higher than the calculated threshold can be expressed mathematically by

$$a_{final}^* > \varepsilon \qquad (8)$$

The stages and the procedures of the proposed algorithm for steps counting purposes are summarized in Table 1.

Table 1: Proposed Algorithm steps.

Proposed Algorithm Steps
1. Collect a_x, a_y, a_z
2. Calculate a_{all} and a_{final}
3. Define step rate $S_r = 5$
4. Calculate the sampling period τ_s, and it is defined as $\max\left(t_{n+1} - t_n\right)$ for $n = 0,1,2,...,k$ where k is the number of samples.
5. Calculate the sampling rate as $R_s = 1/\tau_s$
6. Calculate $n_p = round(R_s / S_r)$

7. Apply envelope function on a_{final}, a_x, a_y, a_z to return the upper peaks envelope $a^*_{final}, a^*_x, a^*_y, a^*_z$.
8. Define the threshold as follows:

$$\varepsilon = \frac{avg\left(a^*_x\right) + avg\left(a^*_y\right) + avg\left(a^*_z\right)}{n_p} - d \ , \ d \in \{3,5\}$$

9. Count the peaks $P_{sk}(i)$ for $a^*_{final} > \varepsilon$
10. Display the step count if finished

4. Results and Discussion

The experiments were conducted at a public institution, which contains many electronic devices. These devices radiate electromagnetic noise, which interferes with measured data. The accelerometer of Huawei P30 has been used to collect the measurement data. Four different scenarios have been considered in this work and summarized in Table 2. The position of the smartphone during the experiments is depicted in figure 4.

Table 2: The symbols of the activities in the empirical scenarios.

Symbol	Activity
A	Walking at varying speeds with the smartphone is carried by the hand
B	Walking at varying speeds with the smartphone is placed in the Jacket pocket
C	Walking at varying speeds with the smartphone is placed in the trousers' front pocket
D	Running at varying speeds with the smartphone is carried by the hand

Figure 4: The position of the smartphone in the experimental scenarios.

4.1. Steps Counting During Different Daily Activities

In this part, the accuracy of the proposed algorithm for steps counting during different daily activities is investigated. The accuracy of steps detection can be expressed as follows:

$$Accuracy = \left(1 - \left(\frac{S_e - S_a}{S_a}\right)\right) \times 100\% \qquad (9)$$

where S_e is the number of the estimated steps, and S_a is the actual step number. Therefore, ten experiments with different speeds and path lengths have been considered for each case study, as summarized in Table 3. The algorithm could detect the number of steps with an average accuracy higher than 99% for normal walking when the smartphone is carried by the participant's hand. The miscount was only for one or a maximum of two steps regardless of the path length. In the next scenario, the smartphone is placed horizontally in the pocket of the participant's jacket. The results presented an average accuracy of about 98%. The third scenario considered putting the smartphone in the front pocket of the participant's trouser.

The result showed a reduction in the average accuracy due to the smartphone's position, and it was around 96.7%. Finally, the running activity has been investigated, considering the smartphone is carried by the participant's hand. The average accuracy of the different tested scenarios is depicted in figure 5. The proposed algorithm showed an average accuracy of 96.9%. The algorithm presented a reduction of 1% accuracy than the standard walking scenario. figure 6 depicts the detected steps and actual steps in the different tested scenarios. In figure 6(a), the number of the detected steps was 19, which is lower than the actual steps by one step in scenario A. This error might result from miscounting one of the real steps by the proposed algorithm due to the set threshold value. figure 6(b) depicts the proposed algorithm's ability to detect the exact number of the actual steps with zero error.

Table 3: Comparison between the actual steps number and the detected steps during different scenarios based on the proposed algorithm.

Test Scenarios					
Exp. No.	**Actual Steps**	**A**	**B**	**C**	**D**
1	10	10	11	9	11
2	20	19	20	21	21
3	30	29	30	29	29
4	40	40	39	39	40
5	50	50	51	51	51
6	60	60	59	60	61
7	70	70	70	73	72
8	80	80	82	80	81
9	90	91	88	92	92
10	100	100	100	96	103

The smartphone placement in the trouser pocket might lead to unwanted vibration, leading to counting more steps than the number of the actual steps, as depicted in figure 6 (c). In the running scenario, the error might result during the transition from the movement to still during the experiments, increasing the number of the detected steps, as shown in figure 6 (d).

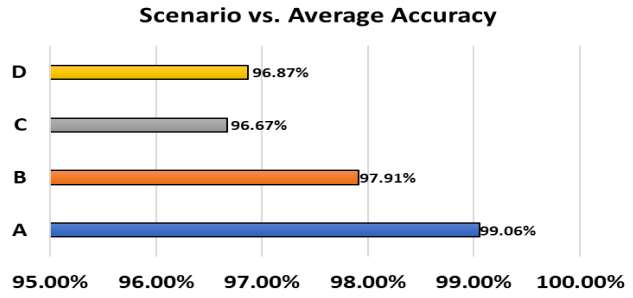


Figure 5: The average accuracy of the ten experiments for each experimental scenario

4.2. Comparison with other Pedometers

In this section, the proposed algorithm has been investigated in terms of its accuracy compared to other pedometers. Therefore, two Android platform applications have been used to detect the number of steps during normal walking and running activity. Also, a HUAWEI Watch GT-731 is used to validate the proposed approach. The first application was Pedometer by King Brain Inc., and the second application was a Pedometer for Walking. Both applications are available on the google play store. The tests have been conducted under the same conditions to validate the

comparison study. The proposed algorithm results showed improved detection compared to other methods with at least 3%.

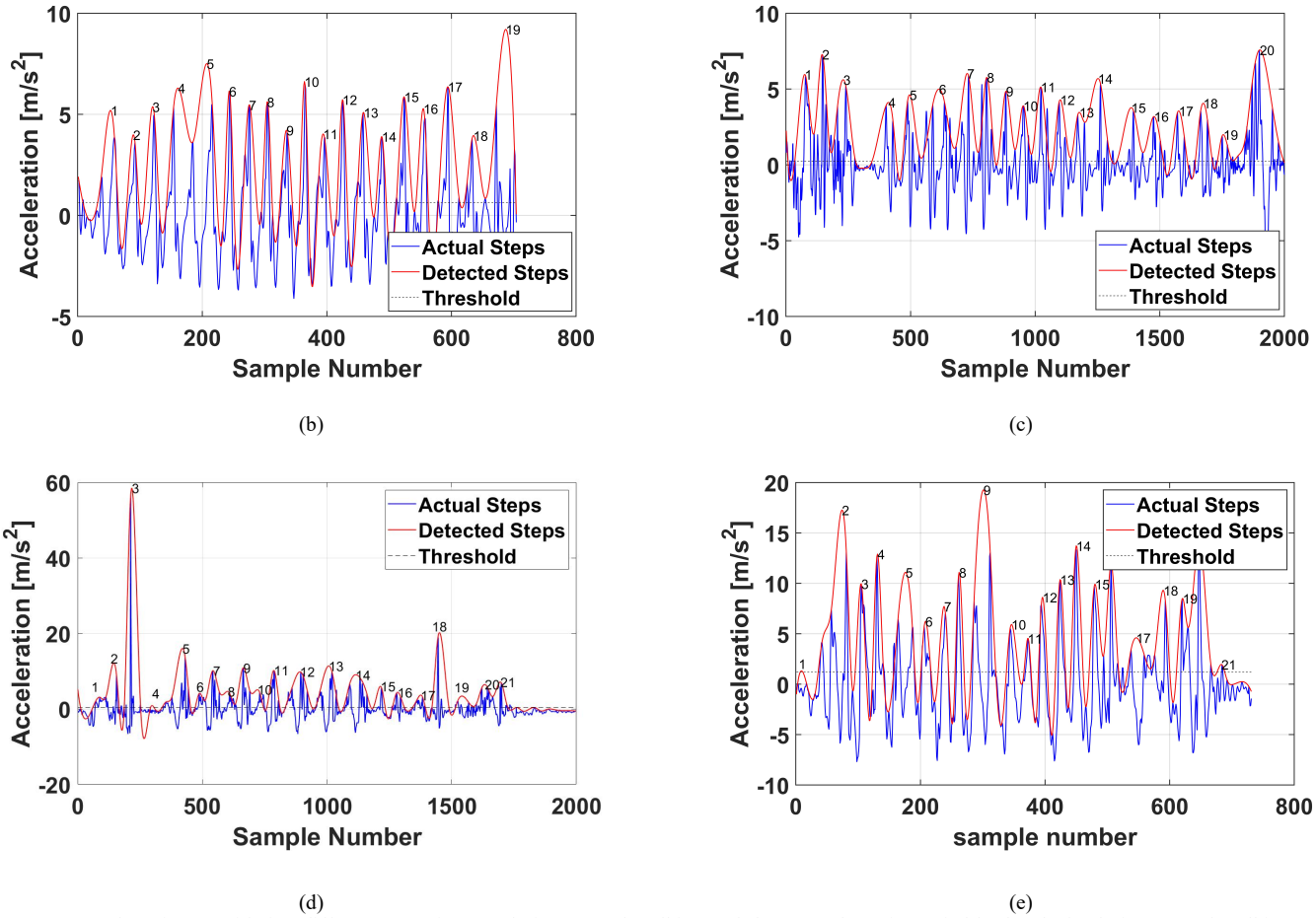


Figure 6: Steps detection considering different scenarios. (a) during normal walking and the smartphone is carried by hand; (b) during normal walking and the smartphone is placed in the right Jacket pocket; (c) during normal walking where the smartphone is placed in the left front pocket of the trouser; (d) during running and the smartphone is carried by hand.

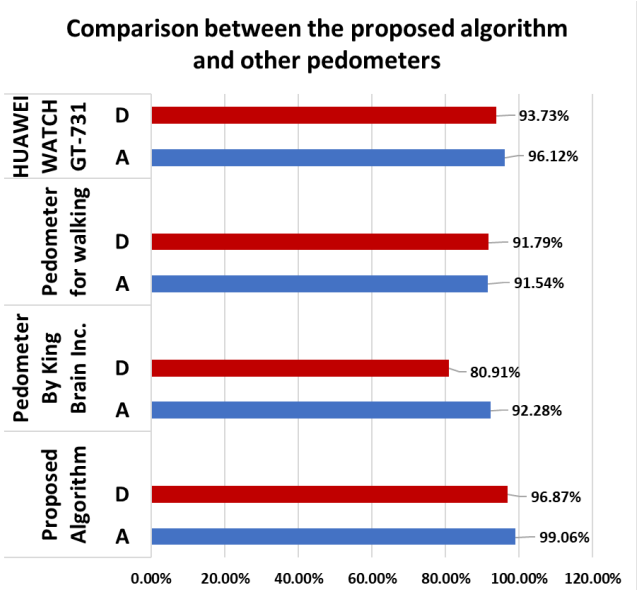


Figure 7: The average accuracy of the proposed algorithm compared to Android applications and HUAWEI Watch GT-731.

The average accuracy of the HUAWEI Watch GT-731 was in second place, with 96.12% for the typical walking scenario and 93.73% for the running scenario. It is worth to mention here the HUAWEI watch does not utilize the smartphone accelerometer, which is the case for the other applications. The result of this study are depicted in figure 7.

5. Conclusion

This work aims to propose a robust algorithm for step detection with high accuracy for indoor systems. Mainly, the proposed algorithm will collect the acceleration data from the smartphone's accelerometer. In the proposed algorithm, a new signal based on the envelope function has been generated to return the upper limit for the average of all measured accelerations signals. The previous phase will remove the impact of vibrations and noise, and it will ease counting the number of steps and obtaining accurate results. It is found that the setting of the threshold will affect the accuracy of the detection process. Therefore, a new formula for the set threshold was created based on the acceleration signals' envelope and the number of samples. The proposed algorithm has been investigated for two case scenarios: the normal walking scenario and the running scenario considering different smartphone's position — the obtained accuracy based on the proposed algorithm

higher than 97% for all considered scenarios, as summarized in Table 4. The results of the proposed approach presented improved accuracy compared to other works. Finally, the proposed algorithm can be used for online and offline step counting with a maximum step rate of 5 steps per second. The sampling rate can be set from 50 Hz to 100 Hz. The threshold's value depends on the measured accelerations signal and the correction factor.

Table 4: Comparison of the average accuracy between the proposed algorithm and state of the art.

	Avg. Accuracy (Normal Walk)	**Avg. Accuracy (Running)**
Proposed Algorithm	99.06%	96.87%
[4]	96.5%	N.A
[6]	93%	N.A
[12]	96.6%	N.A
[5]	94%	78.3%
[3]	93.7%	N.A

Conflict of Interest

The authors declare no conflict of interest.

References

[1] L. Xu, Z. Xiong, J. Liu, Z. Wang, Y. Ding, "A novel pedestrian dead reckoning algorithm for multi-mode recognition based on smartphones," Remote Sensing, **11**(3), 2019, doi:10.3390/rs11030294.

[2] W. Wu, "Research on Algorithm Design and Implementation of Counting Steps Based on Mobile Phone Sensor," **148**(Wartia), 71–74, 2017, doi:10.2991/wartia-17.2017.14.

[3] X. Kang, B. Huang, G. Qi, "A novel walking detection and step counting algorithm using unconstrained smartphones," Sensors (Switzerland), **18**(1), 2018, doi:10.3390/s18010297.

[4] F. Gu, K. Khoshelham, J. Shang, F. Yu, Z. Wei, "Robust and accurate smartphone-based step counting for indoor localization," IEEE Sensors Journal, **17**(11), 3453–3460, 2017, doi:10.1109/JSEN.2017.2685999.

[5] J. Santos, A. Costa, M.J. Nicolau, "Autocorrelation analysis of accelerometer signal to detect and count steps of smartphone users," 2019 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2019, 2019, doi:10.1109/IPIN.2019.8911755.

[6] D. Salvi, C. Velardo, J. Brynes, L. Tarassenko, "An Optimised Algorithm for Accurate Steps Counting from Smart-Phone Accelerometry," Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, **2018-July**, 4423–4427, 2018, doi:10.1109/EMBC.2018.8513319.

[7] K. Tran, T. Le, T. Dinh, "A high-accuracy step counting algorithm for iPhones using Accelerometer," 2012 IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2012, 213–217, 2012, doi:10.1109/ISSPIT.2012.6621289.

[8] G. Rodríguez, F.E. Casado, R. Iglesias, C. V. Regueiro, A. Nieto, "Robust step counting for inertial navigation with mobile phones," Sensors (Switzerland), **18**(9), 1–21, 2018, doi:10.3390/s18093157.

[9] M.S. Pan, H.W. Lin, "A step counting algorithm for smartphone users: Design and implementation," IEEE Sensors Journal, **15**(4), 2296–2305, 2015, doi:10.1109/JSEN.2014.2377193.

[10] W.W. Myo, W. Wettayaprasit, P. Aiyarak, "A more reliable step counter using built-in accelerometer in smartphone," Indonesian Journal of Electrical Engineering and Computer Science, **12**(2), 775–782, 2018, doi:10.11591/ijeecs.v12.i2.pp775-782.

[11] A. Abadleh, E. Al-Hawari, E. Alkafaween, H. Al-Sawalqah, "Step detection algorithm for accurate distance estimation using dynamic step length," Proceedings - 18th IEEE International Conference on Mobile Data Management, MDM 2017, 324–328, 2017, doi:10.1109/MDM.2017.52.

[12] X. Kang, B. Huang, R. Yang, G. Qi, "Accurately counting steps of the pedestrian with varying walking speeds," Proceedings - 2018 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovations, SmartWorld/UIC/ATC/ScalCom/CBDCo, 679–686, 2018, doi:10.1109/SmartWorld.2018.00134.

[13] G.L. Santos, P.T. Endo, K.H. de C. Monteiro, E. da S. Rocha, I. Silva, T. Lynn, "Accelerometer-based human fall detection using convolutional neural networks," Sensors (Switzerland), **19**(7), 1–12, 2019, doi:10.3390/s19071644.

[14] Lai, J. H., Nai, L., Suzumura, T., & Xia, Y. (2018). U.S. Patent No. 10,136,265. Washington, DC: U.S. Patent and Trademark Office.

[15] A. Wang, X. Ou, B. Wang, "Improved Step Detection and Step Length Estimation Based on Pedestrian Dead Reckoning," in 2019 IEEE 6th International Symposium on Electromagnetic Compatibility (ISEMC), 1–4, 2019.

[16] S.Y. Kim, G.I. Kwon, "Interrupt-Based Step-Counting to Extend Battery Life in an Activity Monitor," Journal of Sensors, **2016**, 2016, doi:10.1155/2016/5824523.

[17] V.T. Pham, D.A. Nguyen, N.D. Dang, H.H. Pham, V.A. Tran, K. Sandrasegaran, D.T. Tran, "Highly accurate step counting at variouswalking states using low-cost inertial measurement unit support indoor positioning system," Sensors (Switzerland), **18**(10), 2018, doi:10.3390/s18103186.

[18] H. Crew, The Principles of Mechanics, BiblioBazaar, LLC, 43-45, 2008.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 817-820 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Polarity Switch within Social Networks

Sara Abas*, Malika Addou, Zineb Rachik

Architecture, System and Networks Team (ASYR) - Laboratory of Systems Engineering (LaGeS), Hassania School of Public Works EHTP, Casablanca, 20230, Morocco

ARTICLE INFO

Article history:
Received: 27 June, 2020
Accepted: 21 October, 2020
Online: 27 November, 2020

Keywords:
Online Users
Positive Polarity
Propagation of Information
Social Networks

ABSTRACT

It is the age of information. Social networks are the main reason why, following the increasing activity of online users. With this comes a big impact on the real world, it can be positive and highly negative as well. Therefore, research in this field is highly needed for the betterment of societal behaviors within social networks, to minimize the negative impact. When a negatively polarized message spreads over time, it is eventually exposed to most users of that social network where it is spreading. Social network communities start behaving in a negatively polarized way, due to the negative nature of the propagation of information. In order to optimize technological intervention to the benefit of humanity, in this paper an approach is elaborated in order to help switch the dominant polarity in the case of negativity. The platform of reference is Twitter, since it classifies trends per area, which polarities' dominate within the social network. To achieve that, an extraction as well as a sentiment analysis method are developed to initiate the positive polarity introduction in the network. The idea behind introducing a positive polarity in a network, relies on the studies of the positive effect nature has on users. The positively polarized promoted content is selected from sponsored content of users who consent to spreading their posts.

1. Introduction

In the year 2020, certain events agitated the users of social media platforms, such as: the Covid pandemic and the American social justice riots. As a result, fear spread across the globe not only virtually but also in the real world. This rapid propagation of this type of information has resulted in the spread of misinformation, hatred, cyber-bullying and other unpleasing social behaviors, leading to real life damages. This is where the work of the present paper comes in handy. It becomes unmanageable when the information spreads vastly and one of the key ones who can change this, are the propagators themselves. A study shows that users whose mood shifts are mostly negative generate rants, which are angry forms of expressions [1]. In the case of reaching a big audience, their angry content spreads all over the social platform. As an attempt to reduce the bad effect such users have on those who share the same platform, this paper utilizes research on how nature calms the mood. To promote nature photos or videos on the feed of the potential disruptive users, the selected visual content is extracted form the sponsored data. Sponsored data is data whose owners give consent of publication to online users of the platform. Users who promote their content target people according to many criteria, such as gender, age, geographic area, or center of interests. This paper suggests adding to these criteria, in case the content is nature related, the disruptive users in order to also reach them. To intervene positively; this research suggests a method whose end goal is to switch the user's mood to a positive optimistic one, so that the negative content is less likely to be shared. The article's suggested approach evaluates the current or recently published trends on the Twitter platform, in order to identify the polarity of the dominant news. After this phase, the best contributors in terms of propagation of negatively polarized information are selected. Comes then the procedure of providing them with the polarity switch method as described in the third section of this article. This paper is divided into four sections. The first one is the introduction, the second one is the state of the art in order to mention the previous works. The third one describes the proposed method. The fourth one provides a conclusion.

2. State of the Art

Social media has proven itself to be a fast-growing platform for social exchange [2]. When users express themselves on social media, they insinuate approval, disapproval or a neutral statement. To translate their content polarity wise, the process of sentiment analysis helps. In order to achieve that in the proposed method,

*Corresponding Author: Sara Abas, Email: sara9abas@gmail.com

Natural Language Processing tools are referred to for structural analysis.

2.1. Natural Language Processing

The NLP is an artificial intelligence processing language to create algorithms that can translate human language. Most commonly used NLP platforms are: NLTK (Natural Language Toolkit) [3], spaCy [4], Stanford's NLP tools [5], including the part-of-speech tagger [6], the named entity recognizer [7], the parser, and other provided components.

2.2. Sentiment Analysis

2.2.1. Existing approaches

The rich amount of user-provided data is a gift from social media to content analysis studies. Sentiment analysis, is the study of emotions behind textual content. Review sites are a good example of data made available for research purposes. They are reflective of the clients' receptivity of any provided products. This data source is used for the purpose of testing sentiment analysis solutions. In the available literature, some opinion mining studies investigated different approaches. Turney's Algorithm [8] is a clear example of an unsupervised method for sentiment analysis. The authors use identify the polarity of a phrase, using the Alta Vista [9] as a Web Search Engine resource.

A positive polarity is defined by the identification of the phrase with the word "excellent". Similarly, a polarity, which is classified as negative, is calculated by the co-existence of a phrase with the word "poor".

SO(phrase) = hits(phrase NEAR "excellent") hits("poor") log2 hits(phrase NEAR "poor") hits("excellent") (1)

The hits(phrase NEAR "excellent") "NEAR" operator in the equation represents the co-occurrence of " excellent" and the sentence itself, via the number of search results in Alta Vista. This method is estimated to be 74% accurate. However, there are many words which have different semantic meanings but are spelled the same in the sequence of letters. In this case, this method is likely to lead to the wrong semantic analysis of words. In this paper, a method which identifies a unique meaning to words in tweets, is presented. However, nowadays on social platforms like Twitter, new words are constantly emerging. The social network's users come up with new terms to exchange, which don't have a meaning prior to their first emergence on the platform. These words can't be found on WordNet, which SentiWordNet relies on in the first place. This paper, presents a solution to these newly invented words as a follow up to the main solution. Such words are not available on external database resources. The commonly used sentiment analysis algorithms are Baseline Algorithm [8], Modified Baseline [9], and Turney's Algorithm [4] which were discussed above. However, they are depend on formal phrasing and they don't take into consideration disambiguation. Other works [10, 11] support emoticon analysis, because they are emotionally expressive. An emoticon is a symbol or icon used, to express an emotion without words. The authors' works' relied on a list of pre-classified emoticons or emojis, and emotionally and intensively expressive words. These methods were proven efficient. However, on platforms like Twitter, emoticons are no longer inserted via text, instead they are images provided by the operating system of the phone. A study [12] using available Dutch Tweets, identifies emoticon polarities based on tweets. However, it only studies text-inserted emoticons. Text-inserted emoticons are icons that show up when the user types certain textual character sequences. With the updated phone operating systems of today, users tweet using the available emoticons and do not necessarily type textual symbols.

2.2.2. Nature's effect

The more relieved a user is, the less likely they are to spread negativity. Several researches showed that visualizing images of natural views and landscapes, whether they are viewed as 2D photographs, or scenes, resulted in healthy rewinding effects, such as canceling visual and attentional areas of the brain, removing eye blinking and stress [13, 22]. The study [22] investigated 37 articles showing real evidence of the physiological results of witnessing natural views, through on indoor setting experiments—Exposures to display stimuli, confirmed that viewing natural scenery created a physiological relaxing state.

3. Contribution

The aim of the present approach is to select the social network's propagators which spread negativity within the network. Negativity comes from a content which is negatively polarized, and doesn't align with the positive spectrum of human emotions like: peace, harmony, laughter, love etc. The reason behind detecting the users who spread negativity, is to expose them to nature images, in order to reduce their stress levels. In this work, the social platform Twitter is used for its available "Trends" functionality, which is going to help isolate the most controversial topics, and therefore the most controversial users who discuss these topics. In order to achieve this and to carefully extract trends accordingly to their audience and propagators, the country is first selected on the social platform. The trends functionality is defined by the selection of the geographic area on the Twitter platform. So instead of studying random trends, Twitter selects trends per country. The trends functionality is defined by the selection of the geographic area on the Twitter platform. This helps filter information and circle the social network in which trends are spreading. Once this is achieved, the geographic area where an information is spreading most is considered the social network of base, on which the suggested methods of this paper will run.

3.1. Suggested approach

The macroscopic method is summed up in the figure below.

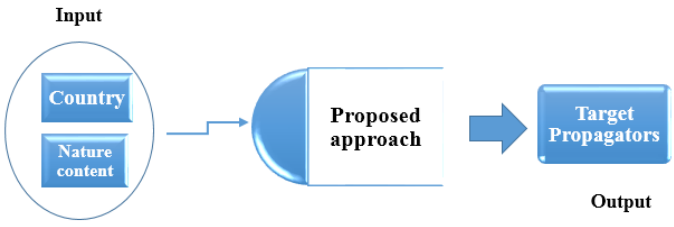


Figure 1: Explained Method Macroscopically

In the entry of the final system, first the country and the nature related sponsored data is provided. This automatically zooms the focus on the content shared within the social network related to that country. As for the nature content, the sponsored database

content is selectively pre-extracted from the sponsored database content of the paltform. On the social network platform, some pages posts are sponsored in order for the page to be marketed and reach a broad audience. From the sponsored content, text mining to find the following keywords is applied: K={ Nature ; Landscape ; Ocean ; Mountain ; Flower ; River ; Beach ; Tree ; Forest ; Sea ; Waterfall }. If the sponsored image's caption contains a word that belongs to the set K, it is a potential advertised content to the target propagators. From this set, the top 20 liked images are chosen, in order to optimize the chance of having a positive influence on the selected users' moods.

After running the method, the final result is a set of target propagators, to whom the selected nature data is displayed, which is sponsored by the marketers of social media seeking promotion of their nature related pages.

3.1.1. Detailed Method

After the country is selected, the process of the suggested method follows four steps.

Figure 2: Detailed Approach

Phase 1: The first phase as illustrated in fig2 is the top trends of the week selection. In this first phase, a simple query on the platform retrieves the top 20 trends of the day, listed in the country: https://twitter.com/i/trends. On the social platform, the top 20 trends are selected and highlighted per country. However, a trend yesterday could be completely neglected today, and no longer have consistent impact. In order to identify the consistently impactful trends, a method is developed that calculates trend relevancy over the past 7 days. The variable "Real_Rank" reflects the real rank of the trend, according to the past 7 days on Twitter.

Let a be one of the top 20 trends of today.

$$Real\_Rank(a)=\sum_0^6 \frac{7-j}{Rank(a\ ;today-j)} \quad (2)$$

Phase 2: Trend polarity Identification. It is necessary to identify which one of the top trends from the first phrase, negatively impacts the social network. In order that, a sentiment analysis method is developed to determine the polarity behind the trend. A trend Tr is a dynamic set of content, which changes every second. In order to identify the polarity of a trend the global polarity of its components is evaluated.

$Tr(j)= \{Tw_1 ; Tw_2 ; .. Tw_n\}$ j is the current day on which the algorithm runs. To identify the polarity of a single tweet Tw1, which is the shared post, the POS Tagger is used in order to retrieve the mentioned adjectives in the sentence. Databases for research purposes are fortunately made available byplatforms, such as Amazon. This includes reviews (ratings, text, helpfulness votes).

Let the following be tweets about the Trend "#destroyThings" extracted from the first phase the platform.

"It is time to destroy things because we hate this horrible system." (Tw_1)

"Destroying is dangerous. Please refrain from doing this bad habit" (Tw_2)

In this sample example there are 4 reviews.

Review 1 : I hate this audiobook.

Review 2 : I don't like it

Review 3 : Boring.

Review 4 : Not very nice.

(Review 1, 0 stars), (Review 2, 2 star), (Review 3, 1 stars), (Review 4, 0.5 stars).

Clearly, the stars given by the reviewers who labelled their review with "horrible" aren't passed 3 stars.

We use the identification method of the polarity of phrase method.

The positivity rate is defined as follows:

$$Pos\ (Trend) = \frac{\sum Polarity(phrase\ k)}{N} \quad (3)$$

N is the number of tweets on the trend "Trend" list.

For adjectives that newly emerge on the social platform, the same method is used on the acknowledged adjectives by the POS Tagger, which occur on their same sentence. In order to explain this, the following examples are cited.

"Amazing. This movie is epic"

"Glad I watched this. Epic."

"This cat is epic. So funny. The owner is odd."

From the first tweet, "amazing" is the identified adjective. This word has a positive polarity. The same applies for the second statement. For the third one, the word "odd" is singled out.

In order to identify the polarity of a newly emerging word, which is a word that does not exist on available knowledge base, the following equation is used. If wk is a positively polarized word, then bk =1. bk(w1) represents an adjective that co-occurs with the word w_1 in the same phrase, m is the number of adjectives in the same tweet j.

$$P\_tweet\_j(w_1)=\frac{\sum bk(w1)}{m} \quad (4)$$

$$Polarity(wk) = \frac{\sum P\_tweet\_j(wk)}{N} \quad (5)$$

where N is the number of tweets containing the word w1.

Phase 3: In this phase, finally the target disruptive users of the social network are selected.

If Pos(Trend) < 50% after it has been selected as a top ranking trend from the first phase, the top 10 most liked tweets belonging to the set of this trend are selected, and the propagators of these tweets are targeted by the selected nature sponsored images.

3.1.2. Algorithm

The global algorithm for the selection of the users to whom nature's sponsored data is exposed is described as follows below. Let T be the social network, and C the country selected in the entry of the system.

Algorithm 1: procedure target Selection

Result: P

S <- Ø, A<- Ø, R <- Ø, P <- Ø

```
S=calculate_trendRank(T)
A=polarity_analyze(S)
R=select(A, 10)
P=select_propagator(R)
```

Calculate_trendRank(T) is a result set which contains 20 vectors. C is the country. In each vector are stored: the trend reference and the real rank of the trend T calculated throughout 7 days, which is calculated by the equation (1). Only trends that are related to the top 20 values of the equation are stored.

Polarity_analyze(S), is a result set which contains a set of vectors. In each vector are stored the trend reference, and the polarity of the trend. The trends are the ones contained in S. The polarity of each trend is calculated via the equation (8).

Select (A, 10) is a method that extracts the top 10 negatively polarized trends from the set A, and results in a set R.

Select_propagators() method uses the set R in order to extract the top 10 most liked posts in the set, and simply identify their users.

4. Conclusion

In the aftermath of a controversial event, social media is usually the most accountable for the chaotic spread of information as well as misinformation. This paper suggests a new method to relieve the online stress, and therefore the real world second effects. In order to do that, available research is used on how nature changes one's emotional state, and calms it. The approach aims to extract the target propagators that need to be stress relieved the most in order to reduce the chaotic information propagation. The goal is to promote nature displays on their timelines. These displays are selected from users whose content is sponsored for promotion. On top of their usual targets, the platform's disruptive users are added in their target potential clients. In order to do that, a sentiment analysis method is developed and information from the social platforms is extracted. In the future works, other aspects are to be covered such as the use of emoticons, the observation of long-term effects of the algorithm on the main propagators and their respective social networks.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] R. Martin, K. Coyier, L. Vansistine, K. Schroeder, "Anger on the Internet: The Perceived Value of Rant-Sites, " Cyberpsychology, behavior and social networking, 16(2), 119–122, 2013, doi: 10.1089/cyber.2012.0130.

[2] E. Muller, R. Peters, "The effect of social networks structure on innovation performance: A review and directions for research, " International Journal of Research in Marketing, **36**(1), 3-19, 2019, doi: 10.1016/j.ijresmar.2018.05.003.

[3] A. Hamandi, A. AbdulHassan, H. Bahjat, "Natural Language Processing Using Natural Language ToolKit ," Iraqi Journal of Information Technology, **7**(2), 2016.

[4] https://spacy.io/

[5] https://nlp.stanford.edu/software/

[6] A. Voutilainen, Part-of-Speech Tagging, The Oxford Handbook of Computational Linguistics, 2012.

[7] https://nlp.stanford.edu/software/CRF-NER.html

[8] P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," in Proceedings of the 40th annual meeting on association for computational linguistics, 417–424, 2002.

[9] B. Pang, L. Lee, S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," in Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, 79–86, 2002.

[10] S.Das, M. Chen, "Yahoo! for amazon: Extracting market sentiment from stock message boards," in 8th Asia Pacific Finance Association Annual Conference, 2001.

[11] D. Cirqueira, M. Fontes Pinheiro, A. Jacob, F. Lobato and A. Santana, "A Literature Review in Preprocessing for Sentiment Analysis for Brazilian Portuguese Social Media, " in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, Santiago, Chile, 746-749, 2018, doi: 10.1109/WI.2018.00008.

[12] A. Hogenboom, D. Bal and F. Frasincar, "Exploiting Emoticons in sent iment analysis," in Proceedings of the 28th Annual ACM symposium on Applied Computing, 703-710, 2013.

[13] K. Gwang-Won, J. Gwang-Won, K. Tae-Hoon, . " Functional neuroanatomy associated with natural and urban scenic views in the human brain:3.0 T functional MR imaging," Korean Journal of Radiology, **11**(5), 507-513, 2013, doi: 10.3348/kjr.2010.11.5.507.

[14] K. Gwang-Won, J. Gwang-Won, K. Tae-Hoon, . "Human brain activation in response to visual stimulation with rural and urban scenery pictures: a functional magnetic resonance imaging study," Sci Total Environ, **408**(12), 2600-7, 2010, doi: 10.1016/j.scitotenv.2010.02.025.

[15] K. Laumann, T. Gärling, K. Stormark, "Selective attention and heart rate responses to natural and urban environments," Environmental Psychology, **23**, 125-134, 2003.

[16] C.Y. Chang, W.E. Hammitt, P.K. Chen, Machnik, "Psychophysiological responses and restorative values of natural environments," in Taiwan. Landscape. Urban Plan, **85**, 79–84, 2008.

[17] V.F. Gladwell, DK. Brown, JL. Barton, "The effects of views of nature on autonomic control, " European Journal of Applied Physiology, **112**(9), 3379-3386, 2012, doi: 10.1007/s00421-012-2318-8.

[18] D.K. Brown et al., "Viewing nature scenes positively affects recovery of autonomic function following acute-mental stress," Environ. Sci. Technol. **47**, 5562–5569, 2013.

[19] M. Igarashi, T. Yamamoto, J. Lee, C. Song, H. Ikei, Y. Miyazaki, "Effects of stimulation by three-dimensional natural images on prefrontal cortex and autonomic nerve activity: A comparison with stimulation using two-dimensional images," Cognitive Processing, **15**, 551–556, 2014.

[20] A.P. Anderson, M.D. Mayer, A.M. Fellow, D.R. Cowan, M.T. Hegel, J.C. Buckey, "Relaxation with immersive natural scenes presented using virtual reality," Aerospace Medicine and Human Performance, **88**, 520–526, 2017.

[21] H. Jo, C. Song, Y. Miyazaki, "Physiological Benefits of Viewing Nature: A Systematic Review of Indoor Experiments," Int J Environ Res Public Health, **16**(23), 4739, 2019, doi: 10.3390/ijerph16234739.

[22] C. Diamantini, A. Mircolli, D. Potena, " A Negation Handling Technique for Sentiment Analysis," in 2016 International Conference on Collaboration Technologies and Systems (CTS), 188-195, 2016, doi: 10.1109/CTS.2016.0048.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 831-837 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Towards a Smart Campus for Qassim University: An Investigation of Indoor Navigation System

Mohammed Hadwan[1,2,3,*], Rehan Uallah Khan[1,2], Khalil Ibrahim Mohammad Abuzanouneh[1,2]

[1]*Department of Information Technology, College of Computer, Qassim University, Buraydah, 51452, Saudi Arabia*

[2]*Intelligent Analytics Group (IAG), College of Computer, Qassim University, Buraydah, 51452, Saudi Arabia*

[3]*Deptartment of Computer Science, Faculty of Applied Sciences, Taiz University, Taiz, 6803, Yemen*

ARTICLE INFO

Article history:
Received: 24 September, 2020
Accepted: 21 November, 2020
Online: 27 November, 2020

Keywords:
Indoor positioning
Smart cities
Bluetooth low energy (BLE)
Internet of Things (IoT)

ABSTRACT

With the advancements in technology, smart entities are becoming increasingly intelligent, therefore, increasing their interaction capabilities with their surrounding environment. Apart from the traditional smart low profile devices, these entities now involve cars, mobiles, televisions, and extend to universities and smart cities. One of the bi-product of the smart cities is the emergence of the concept of smart campus. The smart campus is a teaching environment, where dynamic interaction between students/users and the surrounding devices takes place using the paradigm of the Internet of Things (IoT). The Qassim University (QU) is considerably a large university having thousands of students and hundreds of functional units in the central building and also spread across the different cities of the Al-Qassim province. Therefore, the QU can be considered as a small city where many elements need to be connected and decisions be made. The QU thus represents an optimal and practical scenario for the concept of smart campus. With this spirit, the goal of this research is thus to investigate an Indoor Navigation System (INS) as a suitable platform for the QU. As such, the paper reviews the current technologies and opt for the best available and optimal options. For implementation and simulation, the BLE beacon is selected and user data is analyzed to design a mobile application that includes all the services requested by the users. The system architecture in addition to a 2D map presented in this research will help in identifying the locations of BLE beacons to cover the specific area. The work in this paper is conducted on the main campus of the QU; however, the extension to other setups involves minimum or similar infrastructure.

1. Introduction

Over the last decade, a smart city is one of the emerging concepts that aims to provide a smart livable environment for the resident. Cities as Singapore, Barcelona, San Francisco, London, etc., is an example of cities that apply the concept of smart cities. In [1], the author defined a smart city as "A city well performing built on the 'smart' combination of endowments and activities of self-decisive, independent and aware citizens". The Internet of Things (IoT) paradigm allows cities to manage and control every aspect of the city from buildings, roads, railways, to airports and traffics in a smart manner. IBM smarter planet project adopted the term of the smart city [2]. The smart city relies on six key factors: smart governance, smart people, smart economy, smart environment, smart life, and smart traffic [3]. Nowadays, the typical universities are considered as small cities that include most of the services available in the cities. The advancement of new technologies such as Artificial intelligence, IoT, and the availability of High-speed internet plays a big role in changing the living environment to be smart than ever before [3]. Smart universities also called smart campuses to compose one of the branches of the smart city concept that concern about every aspect inside the university; students, staff, visitors, classrooms, buildings, equipments, etc.[4]. Due to the popularity of the term of smart campus, we will use it in this research to refer to smart university. The smart campus concept can offer many benefits such as save cost and time, automate maintenance, protect the environment, automate students and staff attendance, track occupancy as well and many other benefits [5]. Several

*Corresponding Author: Mohammed Hadwan, Email: m.hadwan@qu.edu.sa

applications have been developed to make our campuses smart such as smart buildings, smart parking, smart learning, smart campus management, and maintenance, etc. [3].

Navigating inside large buildings, such as malls, hospitals, airports, and campuses, using an INS to help users to reach their preferred destination is highly required. INS is investigated in this research due to the huge size of QU main campus that has a similar interior design where finding places is a very difficult task for visitors and even for students and staff as well. The available infrastructures at the QU campus encourage us to explore this idea to make campus life easier and smarter. The main problem of INS is the signal unavailability of a Global Positioning System (GPS) inside buildings because of the walls and other barriers [6]. Smartphones, smart glasses, smart watches are widely spread and used heavily for INS because of the different sensors that is equipped with. Therefore, the aim of this paper is to investigate and identify the obstacles and barriers that could stand against devolving suitable INS for QU main campus.

2. Background

Smart cities with their different technologies and services are one of the active research areas where many survey papers have been published recently [1-5, 7, 8]. The Kingdom of Saudi Arabia is among the countries that apply this concept to change its traditional cities to become smart cities [9]. The smart campus is a concept where many universities working on several projects to adopt this concept such as King Fahd University of Petroleum and Minerals[10], Lancaster University [11], University of Malaga [12], University of New South Wales in Australia [13], etc. The services and applications offered by smart campus are varied. These services aim to make the universities surrounding smart includes but not limited to buildings, laboratories, libraries, classroom, attendance system, grid, learning, parking, navigation, waste and water management and etc. For more information about the smart campus, services refer to [3, 4, 14-15]. Due to the complex mixture of elements and devices to be connected, many expected and unexpected problems might appear and need to be addressed. For instance, internet coverage [16], energy consumption [17], sensory network [18], location-based services [19] etc.

In [20] researchers discussed different types of INS in the literature and highlighted the accuracy problem. In [21], the researchers give recommendations for the user interface design of INS based on the case study conducted by them. They found that it is important not to put heavy details in the map and to simplify the graphical layout. In [22] an android-based application is introduced for indoor navigation inside public buildings using Near Field Communication NFC and Quick Response codes QR. Several features include finding a destination with the shortest path, store the location of car parking, finding the nearest toilet, etc. Smart phones used for INS as in [23] where mobiles used for navigation using wireless access point fingerprints help an individual to get their direction with the indoor space using navigation application. Researchers in [24] Proposed Vision-based approaches for mobile indoor localization based on the combination of Virtual Reality (VR) and Augmented Reality (AR) in order to help in ensuring the accuracy of the localization.

Since the focus of this paper is to develop an Indoor Navigation platform for QU, in the following section we will review the current technologies used for INS. The researches in INS is still active, several survey papers published currently to explore it in detail as in [6, 25-27]. As the signal of GPS cannot reach the devices in indoor areas, several telecommunication technologies were introduced to overcome this problem. For example, Wireless Fidelity (Wi-Fi) [28, 29], Radio Frequency Identification (RFID) [30], Infrared technology (IR) [31], Ultra-Wide Band (UWB) signals [32] and Bluetooth low energy (BLE) beacons[33].

BLE beacons are a relatively new technology with the advantage of low energy consumptions that small batteries can be used for months up to years [34]. Based on the research conducted by [35] to study the localization accuracy between BLE and WIFI. The findings show that BLE was more accurate than WIFI at identical places. According to [36] BLE is a promising technology that has a low deployment cost and appropriate for most mobile devices. iBeacon is the name of Apple's brand of BLE technology based on the micro-localization and the interaction of a mobile device [34]. BLE beacons need to care about the radiation pattern of a given device and possible attenuation element in the chosen places [37].

3. Data Collection and Analyses

User preferences are very important to build effective systems. Therefore, a questionnaire developed and published to collect the user preferences and other necessary data to get insight into the difficulties they faced inside the QU campus. Responses collected from 454 students, lecturers, and general staff. The questions dedicated to familiar and unfamiliar users for the QU campus. The main services and tools requested by the respondent to be available in the QU INS application are presented in table 1.

Table1: The percentage of each place & service required by users

No.	Places & Services	%
1	Places and services around current location (Prayer room, toilet, cafeterias, garden, shops, Photostat, parking, students clubs, gardens etc.)	98%
2	Timetable and office hours information of staff	96%
3	Search by name, Map and around current location	95%
4	Announcements	93%
5	Laboratories	91%
6	Location of new events	91%
7	Classrooms location	90%
8	Examination room location	90%
9	Staff members offices	87%
10	Emergency exits	81%
11	Library	74%
12	Administration offices	31%

The gender, age, user type (student, lecturer, staff, or visitor), preferred services to be available in the system, and other questions been asked and responses obtained. The collected data have been analyzed and important information extracted as follows, refer to Figure 1 and Figure 2:

1- 91% Preferred to have an application that can help to ease the indoor navigation inside the QU campus.
2- 73% of respondents are facing some difficulties in finding locations especially when they visit the QU main building.
3- 100% of respondents never used an INS before.

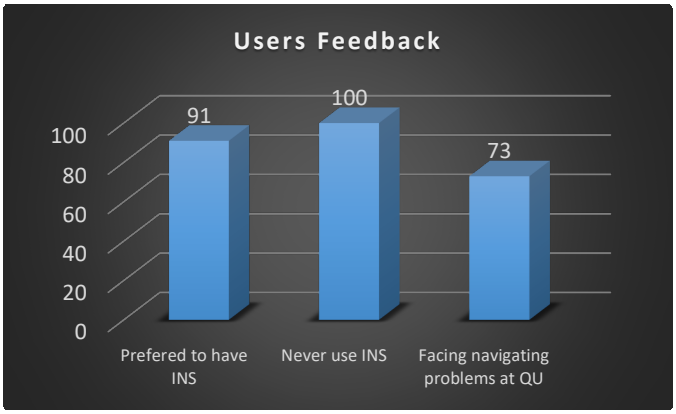


Figure 1: The user's responses to the general questions in the questionnaire.

4. Proposed INS for QU

Based on the extracted information from the questionnaire, a framework for the INS of QU is developed. Figure 3 shows the framework for QU INS, which includes two sections. A section for users where they can search for locations, navigating to desired locations, positioning their current locations, and access the main page where they can modify the settings based on their desire.

The second section is for the developers, where they can develop the INS software, upgrade hardware, add locations, and its information to the database and update the navigation map.

Figure 4 presents the system flowchart for the users where they can search for a location, positioning their selves, and navigate to the desired location. Searching can be through the map our using keyword. For positioning service where it shows the current location of the user in addition to the nearest services around. The navigation service used to guide the users from their location to a specific location inside the building. Based on the literature, BLE beacons are a new, cheap, and popular technology for INS. This encourages us to use BLE beacons to assist in navigation and positioning the users. BLE beacons broadcast their identifier to the nearby devices.

In order to deploy the minimum number of BLE beacon devices, Figure 5 shows the device's location that is enough to cover a single location at the Department of IT, where each beacon covers approximately 62 maters. Each BLE beacon has its own unique ID and for every wall, a boundary segmentation starts from 0 for the first wall, 1 for the second, and so on.

Walls need to determine the sides either left or right. To calculate the position of an object in indoor areas, at least three timestamps from different BLE beacons are needed. A mobile application is designed based on the gathered requirements from the users. Figure 6 shows some examples of screen shots for QU INS. The users can search by names, on a map, or around their current location.

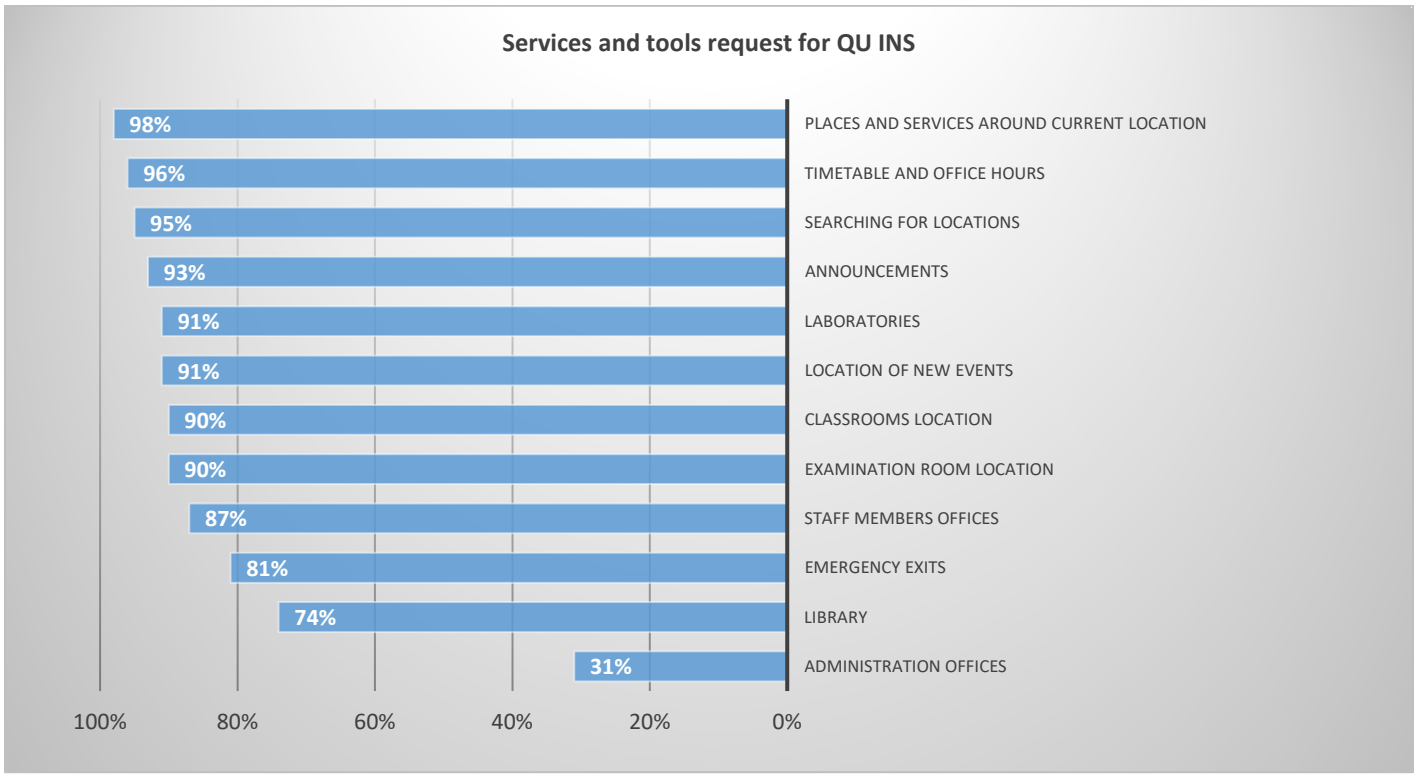


Figure 2: Users' desired services to be included in the INS at QU

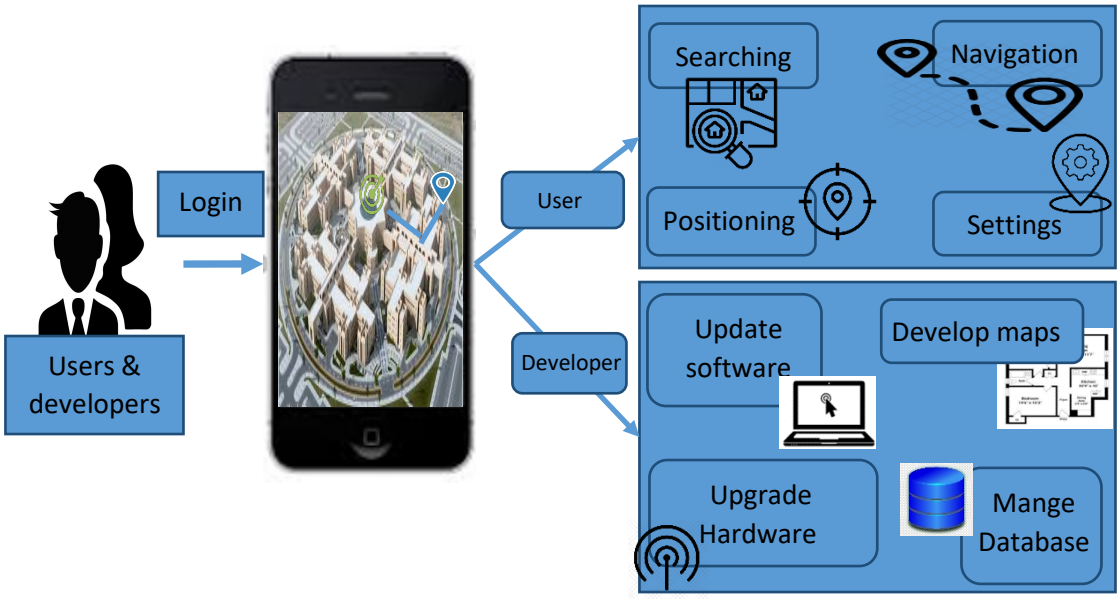


Figure 3: Framework for INS at QU

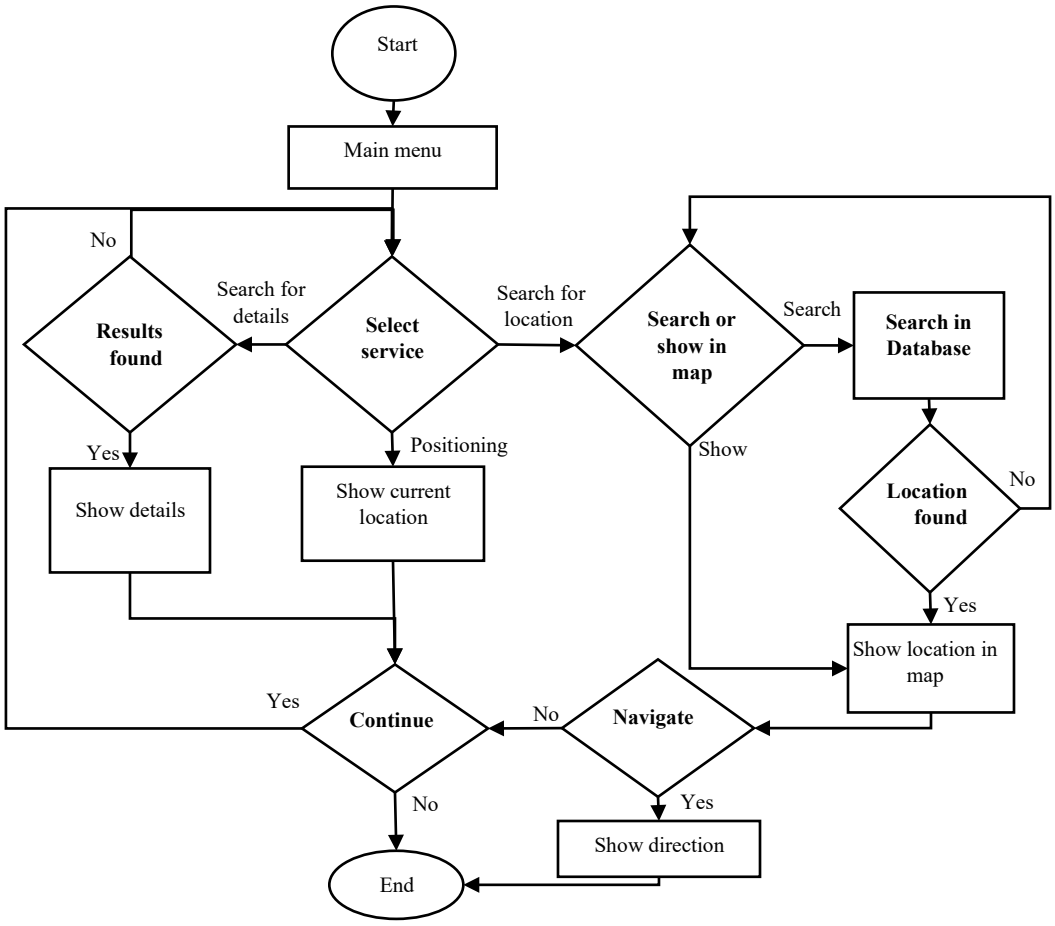


Figure 4: Users' flowchart.

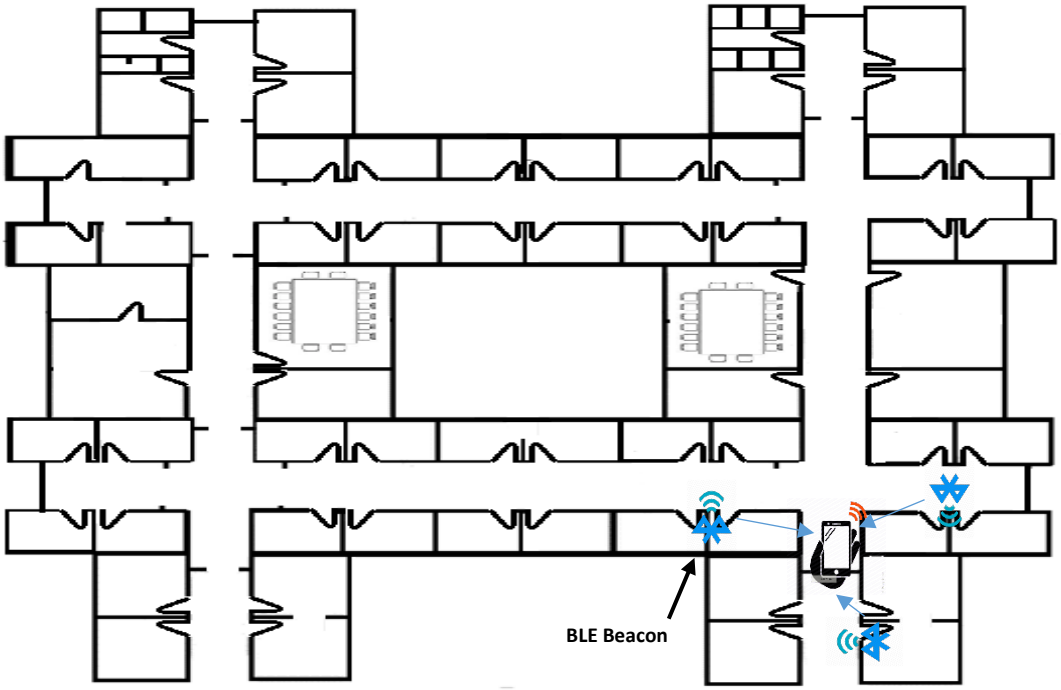


Figure 5: BLE beacons in 2D map at the department of Information Technology at QU

Figure 6: Screenshots of QU Indoor Navigation System

5. Conclusion

The QU is considerably a large university having thousands of students and hundreds of functional units in the central building and spread across the different cities of the Al-Qassim province. Therefore, the QU can be considered as a small city where many elements need to be connected and decisions are made. The QU thus represents an optimal and practical scenario for the concept of a smart campus. As such, we addressed the INS at the QU in this work. We presented a literature review for smart campuses in general and smart universities in particular. In the INS, several technologies are explored for solving navigation and positioning in indoor areas. For an implementation and simulation perspective, the BLE beacon is selected because the BLE beacons are cheap, reliable, having a longer battery life. As such 454 user data is analyzed and is used to design a mobile application that includes all the services required by the users. The system architecture in addition to a 2D map helps in identifying the locations of BLE beacons to cover the whole area of the Information Technology Department at the College of Computer in QU. The system architecture is scalable in order to accommodate any building size. The work in this paper is conducted on the main campus of the QU; however, the extension to other setups involves minimum or similar infrastructure.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The work in this article is funded in its entirety by the Deanship of Scientific Research (SRD), Project number: 5173-coc-2018-1-14-S at the Qassim University, Kingdom of Saudi Arabia.

References:

[1] R. Giffinger, C. Fertner, H. Kramar, R. Kalasek, N. Pichler-Milanovic, and E. Meijers, "Smart cities – Ranking of European medium-sized cities," Science, Centre of Regional, Smart Cities: Vienna, 2007.

[2] C. Yin, Z. Xiong, H. Chen, J. Wang, D. Cooper, and B. David, "A literature survey on smart cities," Science China Information Sciences, **58**(10), 1-18, 2015.

[3] A. Abuarqoub, H. Abusaimeh, M. Hammoudeh, D. Uliyan, M. Abu-Hashem, M. Sharefa, M. Al-Jarrah, and F. Al-Fayez, "A Survey on Internet of Things Enabled Smart Campus Applications," in International Conference on Future Networks and Distributed Systems, Association for Computing Machinery: Cambridge, United Kingdom. 7, 2017.

[4] A. Alghamdi, and S. Shetty, "Survey Toward a Smart Campus Using the Internet of Things," in 2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud), FiCloud: Vienna, 235-239, 2016.

[5] W. Villegas-Ch, X. Palacios-Pacheco, and S. Luján-Mora, "Application of a Smart City Model to a Traditional University Campus with a Big Data Architecture: A Sustainable Smart Campus," Sustainability, **11** , 2019.

[6] F. Zafari, A. Gkelias, and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," IEEE Communications Surveys & Tutorials, **21**(3), 2568-2599, 2019. DOI: 10.1109/COMST.2019.2911558

[7] R. Lea, and M. Blackstock, "Smart Cities: an IoT-centric Approach," in Proceedings of the 2014 International Workshop on Web Intelligence and Smart Sensing, Association for Computing Machinery: New York, NY, USA, 1-2, 2014.

[8] K. Al-Ani, W., Abdalkafor, A. S, and Nassar, A. M, "Smart City Applications: A Survey," in Proceedings of the 9th International Conference on Information Systems and Technologies, Association for Computing Machinery: Cairo, Egypt, 2019.

[9] R. M. Doheim, A. F. Alshimaa, and S. Badawi, "Smart city vision and practices across the Kingdom of Saudi Arabia—a review," Anna Visvizi and M.D. Lytras, Editors, Elsevier, 309-332, 2019. https://doi.org/10.1016/B978-0-12-816639-0.00017-X

[10] S. Alotaibi, and A. Alhussaini, "Smart campus project," in King Fahd University Of Petroleum And Minerals. 2014.

[11] O. Bates, and A .Friday, "Beyond data in the smart city: repurposing existing campus IoT," IEEE Pervasive Comput, **16**(2), 54-60, 2017. DOI: 10.1109/MPRV.2017.30

[12] S. Fortes; J. A. Santoyo-Ramón, D. Palacios, E. Baena, R. Mora-García, M. Medina, P. Mora, R. Barco, R. "The Campus as a Smart City: University of Málaga Environmental, Learning, and Research Approaches," Sensors, 19, 1349, 2019. https://doi.org/10.3390/s19061349

[13] T. Sutjarittham, T., H.H. Gharakheili, and S.S. Kanhere, "Realizing a smart university campus: vision, architecture, and implementation," in Proceeding of International Conference on Advanced Networks and Telecommunication Systems (ANTS), Indore, India, 2015.

[14] M. Cata, "Smart University, a new concept in the Internet of Things," in 14th RoEduNet International Conference, N.i.E.a. Research, Editor, Craiova, Romania, 2015.

[15] R. Faragher, and R. Harle, "Location fingerprinting with bluetooth low energy beacons," IEEE Journal on Selected Areas in Communications, **33**(11), 2418–2428, 2015. DOI: 10.1109/JSAC.2015.2430281

[16] M. Kadadha, H. Al-Ali, M. A. Mufti, A. Al-Aamri, and R. Mizouni, "Opportunistic mobile social networks: Challenges survey and application in smart campus," IEEE 12th International Conference on Wireless and Mobile Computing, in Networking and Communications (WiMob), New York, USA, 1-8, 2016.

[17] A. Kamilaris, A., D. T. H. Ngan, A. Pantazaras, B. Kalluri, S. Kondepudi, and T. K. Wai, "Good practices in the use of ICT equipment for electricity savings at a university campus," in International Green Computing Conference, Dallas, TX, USA, 1-11, 2014.

[18] K. Bandara, P. Jayalath, A.U. Rodrigo, Z. Bandaranayake, Z. Maraikar, and R. G. Ragel, "Smart campus phase one: Smart parking sensor network," in Manufacturing & Industrial Engineering Symposium (MIES), Colombo, 1-6, 2016.

[19] A. Petcovici, and E. Stroulia, "Location-based services on a smart campus: A system and a study," in IEEE 3rd World Forum on Internet of Things (WF-IoT), Reston, VA. 94-99, 2016.

[20] K. A. Nuaimi, and H. Kamel, "A survey of indoor positioning systems and algorithms," in international conference on Innovations in information technology (IIT), IEEE, 185–190, 2011.

[21] A. Puikkonen, A. Sarjanoja, M. Haveri, J. Huhtala and H. Jonna, "Towards designing better maps for indoor navigation: experiences from a case study," in Proceedings of the 8th International Conference on Mobile and Ubiquitous Multimedia (MUM '09), A.f.C. Machinery, Editor, New York, NY, USA, 1-4, 2009.

[22] O. A. Hammadi, A. A. Hebsi, M. J. Zemerly, and J. W. P. Ng, "Indoor Localization and Guidance Using Portable Smartphones," in IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology. IEEE: Macau, 337-341, 2012.

[23] J. Torres-Sospedra, J. Avariento, D. Rambla, R. Montoliu, Sven Casteleyn, Mauri Benedito-Bordonau, Michael Gould, and J. Huerta, "Enhancing integrated indoor/outdoor mobility in a smart campus," Int. J. Geogr. Inf. Sci. 29, 2015.

[24] A. Möller, K., Matthias, H. Robert, D. Stefan and R. Luis. A Mobile Indoor Navigation System Interface Adapted to Vision-Based Localization," In Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia. 2012. Ulm, Germany: Association for Computing Machinery.

[25] L. Zhu, A. Yang, D. Wu, and L. Liu, "Survey of Indoor Positioning Technologies and Systems, in Life System Modeling and Simulation," L.S.M.a. Simulation, Editor. Springer: Berlin, Heidelberg, 2014.

[26] L. Mainetti, L. Patrono, and I. Sergi, "A survey on indoor positioning systems," in 22nd International Conference on Software, Telecommunications and Computer Networks, SoftCOM, Editor. Split, Croatia. 111-120, 2014.

[27] W. Sakpere, M. A. Oshin, and N. B. Mlitwa, "A State-of-the-Art Survey of Indoor Positioning and Navigation Systems and Technologies," South African Computer Journal (SACJ) **29**(3), 145-197, 2017. http://dx.doi.org/10.18489/sacj.v29i3.452

[28] J. Candy, A Mobile Indoor Location-Based GIS Application, in International Symposium on Mobile Mapping Technologie ," Padua, Italy, 2007.
[29] M. Kotaru, K. Joshi, D. Bharadia and S. Katti, Spotfi: "Decimeter level localization using WiFi," ACM SIGCOMM Computer Communication Review, **45,** 269-282 2015.
[30] C. Tsirmpas, A. Rompas, O. Fokou, and D. Koutsouris, "An indoor navigation system for visually impaired and elderly people based on Radio Frequency Identification (RFID) ," Information Sciences, **320**, 288–305, 2015.
[31] E. M. Gorostiza, J. L. L. Galilea, F. J. M. Meca, D. S. Monzú, F. E. Zapata and L. P. Puerto, "Infrared sensor system for mobile-robot positioning in intelligent spaces," Sensors, **11**(5), 5416–5438, 2011. https://doi.org/10.3390/s110505416
[32] S. Gezici, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," IEEE Signal Processing Magazine, **22**, 70-84, 2005
[33] M.Sreeja and M. Sreeram, "Bluetooth 5 beacons: a novel design for indoor positioning," in Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing, Association for Computing Machinery: Cambridge, United Kingdom, 2017.
[34] P. Kriz, F. Maly, and T. Kozel, "Improving Indoor Localization Using Bluetooth Low Energy Beacons," Mobile Information Systems, 2083094, 2016.
[35] X. Zhao, Z. Xiao, A. Markham, N. Trigoni, and Y. Ren, "Does BTLE measure up against WiFi? A comparison of indoor location performance," in European Wireless, 20th European Wireless Conference. 2014.
[36] Z. Zuo, L. Liu, L. Zhang, and Y. Fang, Indoor Positioning Based on Bluetooth Low-Energy Beacons Adopting Graph Optimization. Sensors 2018.
[37] B. Jan, O. Klapka, T. Kozel, and M. Zmitko, "Method of iBeacon Optimal Distribution for Indoor Localization," Cham: Springer International Publishing, 2015.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 838-844 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Effect of Heat Retention Time and Pouring Temperature on Graphite Shape and Mechanical Properties of Gray Cast Iron

Hong-Nga Thi Pham*

Faculty of Mechanical Engineering, HCMC University of Technology and Education, HCM City, 71307, Vietnam

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 11 November, 2020
Online: 27 November, 2020

Keywords:
Graphite
Gray cast iron
Microstructure
Holding time
Pouring temperature

ABSTRACT

In this study, the ASTM A48 Class 30 gray cast iron was cast in a medium frequency induction furnace. The ultimate tensile strength (UTS), ultimate flexural strength (UFS) and microstructure were determined in accordance with ASTM E8/E 8M-80 and ASTM A438-80 (1997). From the results of this study, the graphite size at overheating temperatures 1350°C was smaller than that at 1300°C. Moreover, if the holding time was sufficient, the nested sheet graphite would be completely destroyed. Particularly, in the case of holding time from 5 to 20 minutes, the UTS, UFS are increased and the graphite shape got fine. The smallest graphite size and the highest UTS, UFS values were achieved if 20 minutes holding time is carried out. However, the holding time increased to 30 minutes, the graphite size tended to grow larger, and the UTS, UFS values suffered a decline. In addition, decreasing the average pouring temperature from 1350°C, 1300°C to 1260°C caused the upward trend of the graphite average sizes which were fit with the downward trend in the UTS and UFS values. From the above results, it is possible to determine the casting conditions that can improve the quality of cast iron products in induction furnace are: the optimal average pouring temperature is 1300-1350°C, while the optimal holding time is 10-30 minutes.

1. Introduction

Gray cast iron is a widely-used material in the machinery industry. A typical product of gray cast iron is engine segments. Currently, Vietnamese engine segments are still not high-quality products due to the contamination of cupola furnace technology. In addition, the insoluble coke in the casting process will generate graphite in the form of nest sheet graphite, as shown in Figure 1. As a result, the graphite sheet in the engine segments is also usually in the form of a nest [1]. This is an undesirable form of graphite that negatively impact the mechanical properties of cast iron since the graphite nests break up the continuous of the cast-iron structure, as also the place that the stress concentrated. Therefore, the cast iron has a low flexural and tensile strength [2-4].

Graphite is extremely soft and is considered as a crack that separates the cast iron matrix. The size, shape, and distribution of graphite strongly affect the mechanical properties of cast iron graphite through the microstructure. In [5] the author presented that the main influences on flexural strengths are the shape and the graphite size. Author in [6] obtained that the impact of graphite shape on corrosion vulnerability of gray cast iron in an acidic environment is eclipsed; the vermicular shape of graphite, profitable from the perspective of mechanical and tribological properties might facilitate the destruction of cast iron under the corrosion erosion environment. A similar result was reported by research [7] that the tensile strength of the unidirectionally solidified samples is not as strong as that of spheroidal graphite cast iron because the fracture occurs at the graphite-terminated area in the aligned graphite samples. Study [8] investigates four flaky graphite cast irons of different graphite structures with a pearlitic matrix, results showed that the eutectic graphite structure wore down faster than the flaky graphite structure did, the type of graphite morphology influenced the specimen wear rate as strongly as the graphite volume fraction did in flaky graphite cast irons of this experimental range. The tensile strength of cast iron graphite is only 0.3-0.5 of the tensile strength of steel whose composition corresponds to the base of gray cast iron. Flexural strength and tensile strength of cast iron graphite is only 0.4-0.7 of flexural strength and tensile strength compared to the corresponding values of the steel that has the same composition as the iron matrix.

*Corresponding Author: Hong-Nga Thi Pham, HCMC University of Technology and Education, HCM City, Vietnam, +84-948691160, hongnga@hcmute.edu.vn

In general, the mechanical properties depend on the size, shape, and distribution of graphite which dependent components and pre-cast conditions treatment, additives, slag, furnace, temperature and holding time, the chemical composition of cast iron, cooling rate... [9,10]. Particularly, the graphite with nest sheet shape leads to the low value of the mechanical properties of cast iron [11,12]. However, the studies of the pouring temperature and holding time on the mechanical properties of gray iron are very limited. In [13,14], the authors described the effect of cooling rate on microstructure and mechanical properties of gray cast iron and the effect of pouring temperature on fluidity, porosity, and surface roughness of gray cast iron. According to [5], the pouring temperature plays a key role in constructing the tensile strength and elongation of the cast iron, besides the graphite factor, chemical composition and matrix microstructure also contribute to the tensile strength. Research [15] proves that the faster solidification in the microstructure, the lower the values of the mechanical properties. Results of [16] show that the mechanical properties and microstructure of compacted graphite iron depend on holding time, increasing the holding time from 10 min to 17 min results in lowering the Mg content from 0.031% to 0.021% and as a result lower nodularity was obtained, in addition, lowering the thickness from 20mm to 5mm increases the tendency of the dendritic structure as a result of increasing the cooling rate. In which, the studies on the influence of shape, size, and distribution of graphite on the mechanical properties of gray iron is still very limited. This report focuses on controlling the pouring temperature and holding time to destroy the nested sheet graphite and form a better graphite shape to improve the mechanical properties of cast iron.

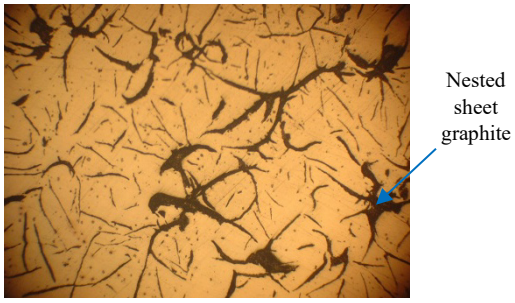


Figure 1: Microstructure of nest sheet graphite cast iron (50X magnification)

2. Experimental

2.1. Materials and Experimental Method

The material composition for the experiment is shown in Table 1. This composition is used to cast gray cast iron with grade 30 of ASTM International standard A48 for small electric motor covers whose capacity is less than 5kW. Cast iron experiments are made at Muoi Hoi Binh Duong Co., Ltd., as shown in Figure 2, without using the nodulizing elements. Initially, the mixed ingredients are fed into a medium frequency induction furnace, with a capacity of 1000kg/batch. The furnace runs for 10 minutes with low amperage to stabilize the frequency. The current then gradually increases to the maximum intensity for 1 hour 30 minutes. The crucible is covered by the acidic powder that allows desulfurization. Due to the rotating magnetic field, cast iron begins to melt and mix relatively well. The furnace is maintained at a constant current for half an hour to allow the cast iron to completely melt and the cast iron temperature is stable. Then, the cast iron temperature is increased to the pouring temperature as required. The pouring pot is pre-dried with a charcoal furnace to a temperature that is greater than 1000°C. The old sand mixture with good ventilation is used, mold is dried naturally. The mold is made by a pneumatic system to ensure uniform tightness of the mold, thus ensuring the heat transfer and exhaust mode.

Table 1: Material composition for experiment

No.	Composition
1	Cast iron scrap, recovery materials in casting production (filling system, shrinkage, broken casting)
2	Ferrosilicon 30-70, with 1-2%
3	Ferromanganese 50-50
4	$CaCO_3$

Figure 2: Medium frequency induction furnace at Muoi Hoi Binh Duong Co., Ltd

Table 2: Overheating temperatures, holding times, and average pouring temperatures

Sample	Average overheating temperatures (°C)	Holding times (minutes)	Average pouring temperatures (°C)
S1	1350	5	1350
S2		10	
S3		20	
S4		30	
S5	1300	5	1300
S6		10	
S7		20	
S8		30	
S9	1300	5	1260

Table 2 shows the cast iron samples are cast with different overheating temperatures, holding times, and average pouring temperatures. The layout of the experimental samples in the sand mold and the samples after casting are shown in Figure 3. The temperature is measured with the AS892 Infrared Thermometer. Some basic specifications are described as following: temperature range 200°C ~2200°C, accuracy ± 2°C, distance spot ratio 80:1, emissivity 0.10~1.00 adjustable, resolution 0.1°C, wavelength 900~1700 µm, response time 500ms, repeatability ±1°C.

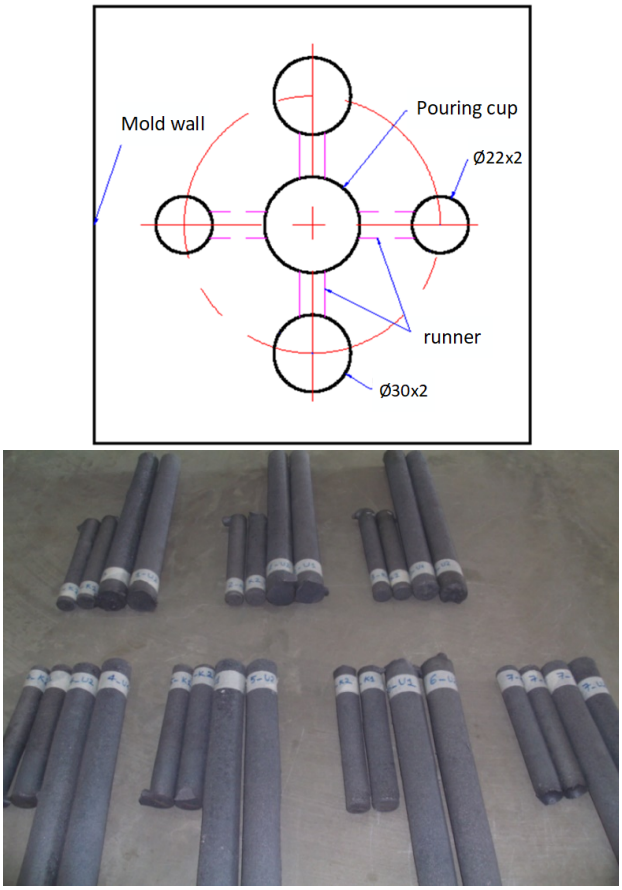


Figure 3: The layout of the experimental samples in the sand mold and the samples after casting

2.2. Methods of Observing Microstructure and Testing Mechanical Properties

Tensile test and Flexural test are conducted at Quatest 3 center, Vietnam, following the ASTM E 8/E 8M-80 (Figure 4) and ASTM A 438-80 (1997) (Figure 5) standards.

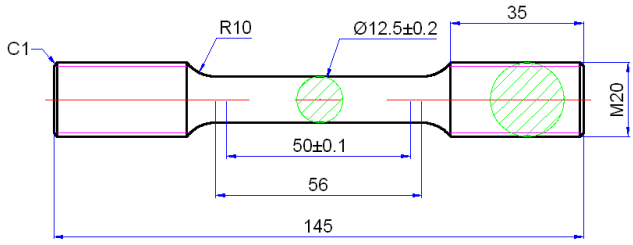


Figure 4: Tensile sample size standard ASTM E 8/E 8M-80

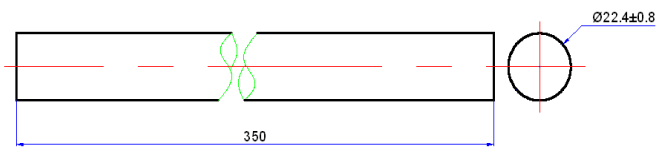


Figure 5: Flexural sample size standard ASTM A 438-80

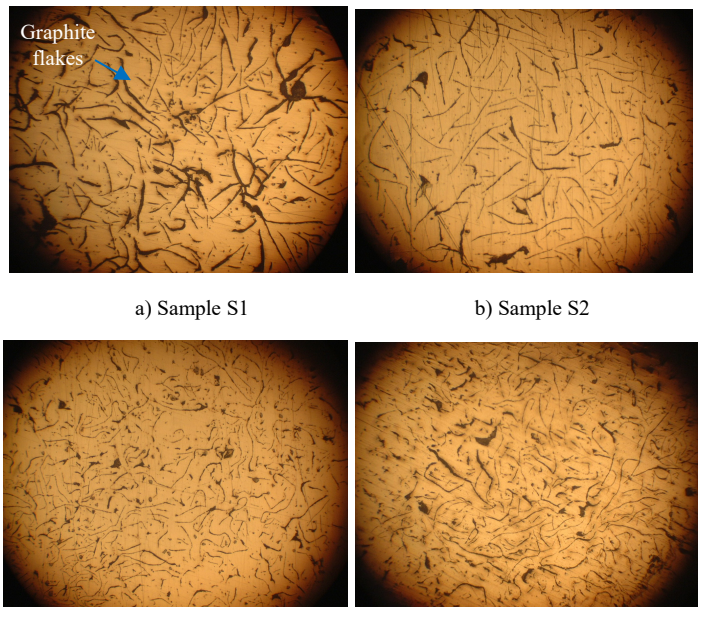
The microstructure of the samples are observed at Materials Laboratory at Mechanical Faculty, University of Technology and Education, HCMC. The graphite size calculation is based on TCVN 3902-84.

Figure 6: Universal testing machine – SANS

3. Results

3.1. Effect of Holding Time on the Graphite Shape

This section considers the effect of holding times and average pouring temperatures to identify the influence of these factors on the amount of graphite and the time needed to destroy the nested sheet graphite. Figure 7 illustrates the microstructure of the samples that overheated at 1350°C and holding at 5-30 minutes. These microstructures indicate that increasing the holding time leads to a decline in the graphite amount. Moreover, the graphite shape is also different with different holding times. Improving the holding time also reduces the number of nested sheet graphite. Interestingly, if the holding time is long enough, the nested sheet graphite will disappear, as shown in Figure 7(d).

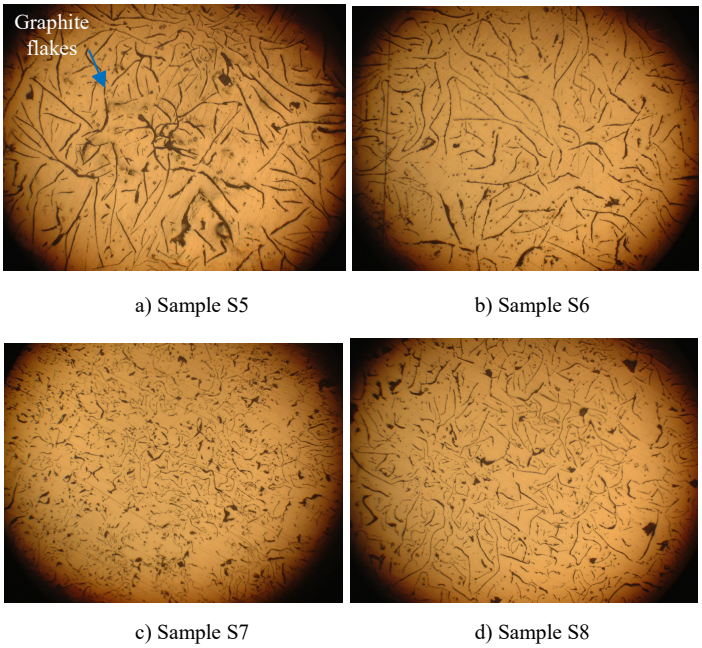


a) Sample S1 b) Sample S2

c) Sample S3 d) Sample S4

Figure 7: Microstructure of samples pouring at 1300oC at different holding time: (a) Sample S1 (holding time 5 minutes), (b) Sample S2 (holding time 10 minutes), (c) Sample S3 (holding time 20 minutes), and (d) Sample S4 (holding time 30 minutes) (50X magnification)

Figure 8 exhibits the microstructure of the samples that overheated at 1300°C and holding at 5-30 minutes. In general, improving the holding time results in a lower amount of graphite. In addition, the graphite shape is also different with different holding times. Similar to 1350°C cases, if the holding time is long enough, the nested sheet graphite will be destroyed, as shown in Figure 8(d).



a) Sample S5 b) Sample S6

c) Sample S7 d) Sample S8

Figure 8: Microstructure of samples pouring at 1350oC at different holding time: (a) S5 (holding time 5 minutes), (b) S6 (holding time 10 minutes), (c) S7(holding time 20 minutes), and (d) S8 (holding time 30 minutes) (50X magnification)

The average size of the graphite sheet at different overheating temperatures and different holding times is showed in Figure 9. Initially, increasing the holding time from 5 to 20 minutes results in finer graphite. However, if the holding time increases to 30 minutes, the graphite size will get bigger. The smallest graphite size is achieved by holding at 20 minutes. Moreover, the graphite size at 1350°C is smaller than at 1300°C.

The reason is that with a higher temperature and a longer holding time, the liquid cast iron is stirred stronger as the viscosity of the cast iron is lower. The stronger stirring results in finer graphite. The higher temperature also causes a more stable state, and a higher trend of overcooling of the casting liquid. However, holding too long could lead to an unstable graphitization process [3,4]. When overheated, the amount and the shape of graphite changes. As the overheating of cast iron increases, the graphite becomes smaller and changes its distribution characteristics. Firstly, the eutectic form of graphite changes, then the number of dendrites will increase. Overheating facilitates the formation of graphite between dendrites which also depends on the composition of the cast iron (carbon and silicon). When the overheating temperature increases, the graphite will strongly smooth out, proving that there is some solubility of these phases and increasing the dispersion of these phases. However, overheating does not facilitate the complete dissolution or agglomeration of these phases. Traditionally, the graphite in gray cast iron could be viewed as a void, and the cracks initiate at the tip of the graphite flakes due to the micro notch stress concentration. The deformed graphite flakes, an optimized microstructure, reduced the stress concentration and prevented the expansion of cracks at the tip of graphite [11]. In [9], the author explains that cast iron slug samples all had average flake lengths of 80-100 µm, and maximum flake lengths of 1000-1400 µm due to the drastically slower cooling rate in this mold configuration compared to the cooling rate in the commercial rotors. This suggests that either there is a limit to how much the diffusivity can be improved by increased flake length or that other factors limited the diffusivity of the gray iron from these large castings. According to [16], holding time displays a reasonable effect on the progression of residual magnesium that identifies with the fading process.

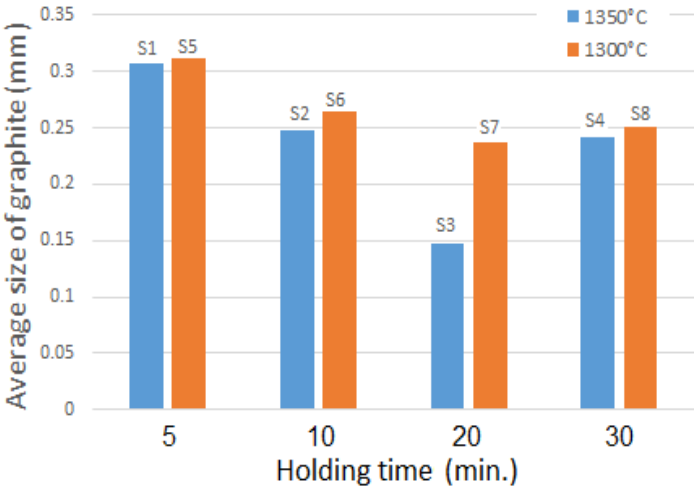


Figure 9: The average size of graphite sheet with different holding times and overheating temperatures

In conclusion, the optimal microstructure is obtained by holding at 10-30 minutes at 1350°C. Casting the gray iron in the medium frequency induction furnace can destroy the nested sheet graphite.

3.2. Effect of the Holding Time on the Mechanical Properties

Table 3 and Figure 10 present the ultimate tensile strength (UTS) of the samples at different holding times and overheating temperatures. These results indicate that the UTS rises as increasing the holding time from 5 minutes to 20 minutes and gaining the best UTS value at 20 minutes. Thereafter, the UTS suffers a decline as the holding time continues to increase from 20 minutes to 30 minutes.

Table 3: Tensile test results

Sample	Fracture force P_b (kN)	Tensile strength $\sigma_b=P_b/F_0$ (MPa)
S1	17.6	145
S2	18.1	150
S3	18.9	154
S4	18.0	148
S5	15.7	129
S6	16.0	132
S7	16.4	135
S8	15.8	130

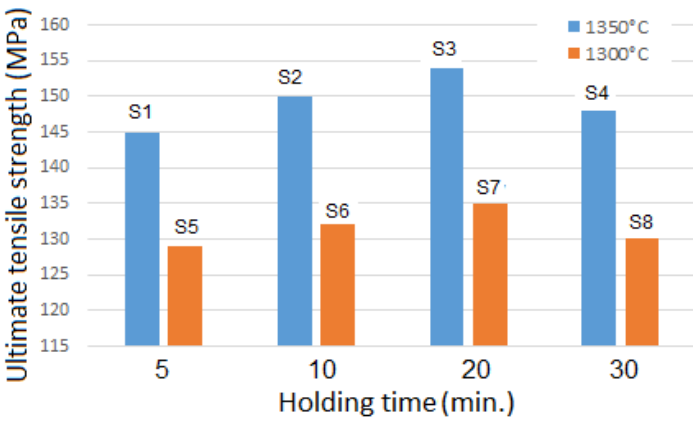


Figure 10: Ultimate tensile strength of the samples with different holding times and overheating temperatures

Table 4 and Figure 11 represent the ultimate flexural strength (UFS) of the samples at different holding times and overheating temperatures. Similar to the UTS value, the UFS increases as improving the holding time from 5 minutes to 20 minutes and reaching the highest value at 20 minutes. Then, this value decreases as the holding time increases from 20 minutes to 30 minutes. Interestingly, the results of both UTS and UFS values are consistent with the change of graphite size in section 3.1. In conclusion, the downward trend of the graphite size is fitted with an upward trend in the UTS and the UFS values. Therefore, decreasing the graphite size could lead to higher mechanical properties.

Table 4: Flexural test results

Sample	Resistance moment $W=\pi d^3/32$ (mm^3)	Ultimate flexural force P_u (kN)	Flexural moment $M_u=P_uL/4$ (kNmm)	Flexural strength $\sigma_{bu}=M_u/W$ (N/mm^2)
S1	1148.36	4.31	328.43	286
S2	1118.27	4.41	336.60	301
S3	1148.36	4.71	359.44	313
S4	1133.25	4.49	342.24	302
S5	1133.25	4.03	307.11	271
S6	1118.27	4.11	313.12	280
S7	1148.36	4.37	333.02	290
S8	1133.25	4.24	322.98	285

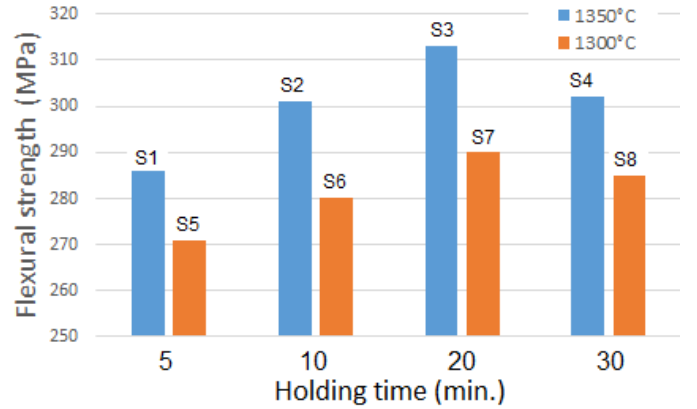


Figure 11: Flexural strength of the samples with different holding times and overheating temperatures

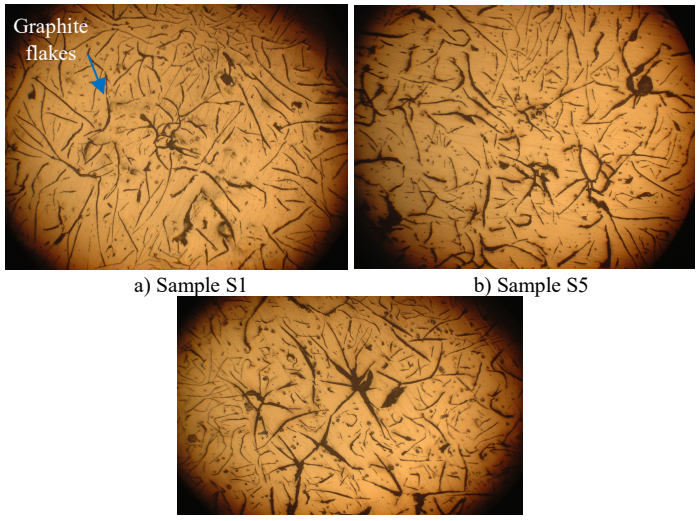
It can be seen that S4 and S8 samples with a holding time of 30 minutes have lower UTS and UFS values than the S3 and S7 samples with a holding time of 20 minutes. The reason is that, as the temperature and holding time increase, the viscosity will increase. The chemical composition and the amount of carbon are evenly distributed, therefore, the quality of the casting increases, the graphitization process happened more fully. However, if the temperature is too high and the holding time is too long, it will prevent the formation of liquid slag, increase the burn rate. In addition, the temperature difference between the liquid iron and the mold causes the graphitization process to occur unstably [4]. Therefore, choosing a holding time of 10-30 minutes is optimal.

The samples overheated at 1350°C and 1300°C gave different mechanical results. The reason is overheating changes the nature of the liquid phase and the differentiation conditions when solidified. On the one hand, overheating increases the stability of the liquid phase and increases its tendency to overcool. On the other hand, under the ideal solidification of a casting in a mold, overheating causes the preheating of the mold and changes the cold condition of the cast iron. Overheating requires a regular increase in the amount of bound carbon (reducing the amount of graphite) due to increased liquid phase stability and its tendency to over-cool. Overheating will limit impurities present in the liquid phase and make them more evenly distributed. In addition, the different overheating temperature will result in different morphology of graphite. The study [11] showed that the distribution and morphology of graphite has a direct influence on the properties of gray cast iron oriented the space orientation of the graphite flakes in gray cast iron and improved both strength and ductility.

3.3. Effect of the Average Pouring Temperature on the Graphite Size and the Mechanical Properties

This section surveys the effect of the average pouring temperature on the graphite size and the mechanical properties of the gray iron. The samples S1, S5, and S9 are chosen and analyzed.

Figure 12 and Table 5 show the microstructure of the cast iron and the average size of their graphite. These results indicate that decreasing the average pouring temperature will cause a rise in the graphite average size.



a) Sample S1 b) Sample S5

(c) Sample S9

Figure 12: Microstructure of samples pouring at different temperatures: (a) Sample S1, (b) Sample S5, and (c) Sample S9 (50X magnification)

Table 5: The effect of the average pouring temperature to the graphite size

Sample	Holding time (minutes)	Average pouring temperature (°C)	Average size of graphite (mm)
S1	5	1350	0.307
S5	5	1300	0.312
S9	5	1260	0.339

Table 6: UTS and UFS values at different average pouring temperatures

Sample	Average pouring temperature (°C)	UTS value (MPa)	UFS value (N/mm^2)
S1	1350	145	286
S5	1300	129	271
S9	1260	126	270

The UTS and UFS of the sample S1, sample S5, and sample S9 values at different average pouring temperatures is demonstrated in Table 6. The result points out that reducing the pouring temperature leads to a decline in the UTS and UFS values. The result is that the lower pouring temperature leads to the larger graphite size. Therefore, the larger graphite size will result in lower mechanical characteristics. Moreover, the higher temperature of the casting liquid will cause a lower viscosity and a better chemical uniformity. At a low pouring temperature of 1260°C (sample S9), the viscosity of the casting liquid is high, the chemical composition is not evenly distributed. To achieve a higher level of uniformity, the casting liquid can behold for a longer time. However, the longer time will cause a loss in the chemical composition as they are burned out or going to the slag. Generally, to improve the mechanical properties of the gray iron, increasing the temperature is better than increasing the holding time. Increasing to too high a temperature will result in a worse slag, a higher level of chemical burning, and the more unstable the graphitization [14].

In conclusion, the optimal pouring temperature to obtain the best mechanical properties is 1300-1350°C.

4. Conclusion

Increasing the holding time from 5 to 20 minutes at 1350°C and 1300°C results in finer graphite, a smaller amount of nested sheet graphite. The finest graphite size is achieved at 20 minutes. Holding up to 30 minutes could destroy the nested sheet graphite, in which at 1300°C the holding time will be longer than overheating and keep the heat at 1350°C to completely lose the nested sheet graphite. However, increasing the holding time to 30 minutes will cause an increase in the graphite size.

- At 1350°C and 1300°C, the UTS and UFS values increase as the holding time improve from 5 to 20 minutes, and reaching the highest values at 20 minutes. Thereafter, these mechanical properties suffer a decline as the holding time increase from 5 to 20 minutes. The finer graphite results in higher UTS and UFS values.

- Declining the average pouring temperature from 1350°C to 1300°C, and then 1260°C (with holding time in 5 minutes and same cooling condition) leads to a finer graphite size and the better UTS and UFS values.

From the above results, the casting conditions that can improve the product quality is the optimal average pouring temperature is 1300-1350°C, while the optimal holding time is 10-30 minutes which will completely lose the nest graphite.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We acknowledge HCMC University of Technology and Education, Dong Nhan Phat Co., Ltd and Material Testing Laboratory (HCMUTE). They gave me an opportunity to join their team, accessed the laboratory and research machines. Without their appreciated support, it would not be possible to conduct this research.

References

[1] T. H. N. Pham, N. T. Tran, P. J. Pritadewi, N. Y. P. Vo. "Research on Factors Influencing the Formation Graphite and Effect of Graphite on Mechanical Properties of Gray Cast Iron". 2019 International Conference on System Science and Engineering (ICSSE), Dong Hoi, Vietnam, 2019, doi:10.1109/ICSSE.2019.8823478.

[2] Y. Liu, J. Xing, Y. Li, S. Wang, D. Tao. "Tomographical Study of the Effect of Graphite on Properties of Cast Iron", Steel research international, 2018, doi:10.1002/srin.201800086.

[3] X. Huang, Y. Ye, X. Shen, X. Chang. "The mechanical properties of gray cast iron and metallographic structure effect on the chip shape", Advanced Materials Research, **339**, 200-203, 2011, doi:10.4028/www.scientific.net/AMR.339.200.

[4] I. Svensson, T. Sjögren. "On Modeling and Simulation of Mechanical Properties of Cast Irons with Different Morphologies of Graphite", International Journal of Metalcasting, **3**(4), 67-73, 2009, doi:10.1007/BF03355460.

[5] S. Kumar, P. Kumar, H. S. Shan. "Optimation of Tensile Properties of Evaporative Casting Process through Taguchi's Method". Journal of Materials Processing Technology, **204**, 59-69, 2008, doi:10.1016/j.jmatprotec.2007.10.075.

[6] E. Lunarska. "Effect of graphite shape on the corrosion of gray cast iron in phosphoric acid", Materials and Corrosion, **47**, 539-544, 1996, doi:10.1002/maco.19960471002.

[7] H. Nakae, H. Shin. "Effect of Graphite Morphology on Tensile Properties of Flake Graphite Cast Iron". Materials Transactions, **42**(7), 1428-1434, 2001, doi:10.2320/matertrans.42.1428.

[8] M. Tsujikawa, K. Nagamine, A. Ikenaga, M. Hino. "Influence of graphite morphology on dry sliding wear of flake graphite cast irons". International Journal of Cast Metals Research, **21**(1-4), 76-80, 2008, doi:10.1179/136404608X361701.

[9] R. L. Hecht, R. B. Dinwiddie, H. Wang. "The effect of graphite flake morphology on the thermal diffusivity of gray cast irons used for automotive brake discs". Journal of Materials Science, **34**(19), 4775-4781, 1999, doi:10.1023/a:1004643322951.

[10] M. C. Rukadikar, G. P. Reddy. "Influence of chemical composition and microstructure on thermal conductivity of alloyed pearlitic flake graphite cast irons". Journal of Materials Science, **21**, 4403-4410, 1986, doi:10.1007/bf01106563.

[11] W. Wang, T. Jing, Y. Gao, G. Qiao, X. Zhao. "Properties of a gray cast iron with oriented graphite flakes". Journal of Materials Processing Technology, **182**(1-3), 593-597, 2007, doi:10.1016/j.jmatprotec.2006.09.028.

[12] R. Hinrichs, M. A. Z. Vasconcellos, W. Österle, C. Prietzel. "Amorphization of Graphite Flakes in Gray Cast Iron Under Tribological Load". Materials Research, 21(4), 2018, doi:10.1590/1980-5373-MR-2017-1000.

[13] M.M. J. Behnam, P. Davami, N. Varahram. "Effect of cooling rate on microstructure and mechanical properties of gray cast iron", Materials Science and Engineering, **A528**, 583-588, 2010, doi:10.1016/j.msea.2010.09.087.

[14] Suyitno, Sutiyoko. "Effect of Pouring Temperature and Casting Thickness on Fluidity, Porosity and Surface Roughness in Lost Foam Casting of Gray Cast Iron", Procedia Engineering, **50**, 88-94, 2012, doi:10.1016/j.proeng.2012.10.011.

[15] T. Sjogren, P. Vomacka, I. L. Svensson. "Comparison of mechanical properties in flake graphite and compacted graphite cast irons for piston

rings". International Journal of Cast Metals Research, **17**(2), 65-71, 2004, doi:10.1179/136404604225017474.

[16] H. Megaheda, E. El - Kashifa, A. Y. Shashab, M. A. Essam. "Effect of holding time, thickness and heat treatment on microstructure and mechanical properties of compacted graphite cast iron", Journal of Materials Research and Technology, **8**(1), 1188-1196, 2019, doi:10.1016/j.jmrt.2018.07.021.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 845-852 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Updated Analysis of Business Continuity Issues Underlying the Certification of Invoicing Software, Considering a Pandemic Scenario

Nelson Russo[*, 1], Leonilde Reis[2]

[1]School of Science and Technology, University of Trás-os-Montes and Alto Douro, Vila Real, 5000-801, Portugal

[2]School of Business Administration, Polytechnic Institute of Setúbal, Setúbal, 2914-503, Portugal

ARTICLE INFO

Article history:
Received: 06 October, 2020
Accepted: 11 November, 2020
Online: 27 November, 2020

Keywords:
Business continuity
Invoicing software
Software certification
Electronic invoicing
Pandemia

ABSTRACT

Portuguese organizations that have invoicing software, certified by the Tax and Customs Authority, need to comply with technical requirements that involve business continuity and disaster recovery. The recent tax legislative changes created conditions for the dematerialization of documents, allowing waiving invoice printing, encouraging the adoption of an electronic invoicing and document archiving system. The pandemic situation boosted the need for organizations to integrate this paradigm in their business processes. However, there are some constraints in the implementation of these requirements, due to technical issues, interpretation of tax legislation or the selection of frameworks or good practices for Information and Communication Technologies. The objective of the work is to present a set of concerns underlying the design of a business continuity plan, supported by current tax legislation, by standards and codes of good practice. In view of the constraints of Portuguese business capacity, it is also presented a minimum solution that meets the legal, regulatory, good practices and conceptual requirements of Information and Communication Technologies for initiating the design of a Business Continuity Plan. The method used in this investigation was based on the analysis of international standards ITIL, ISO, CMMI, COBIT through the assertive interconnection with the subject under study with the dispositions stated in the Portuguese legal framework in the field of invoicing. The main result was the conception of a decision support process for designing a guide, concerning the optimization of the business continuity plan design process. In face of the problematic in study, it is considered that the main expected results were achieved, by fostering the design of Business Continuity Plan in Portuguese organizations, in order to reduce the gap between the practices currently in place and the requirements underlying certification, as a way to prepare organizations to deal with disruptive events in invoicing business processes.

1. Introduction

This work is an extended version of the paper [1] originally presented in 2019 at the 14th Iberian Conference on Information Systems and Technologies (CISTI), where some guidelines for the conception of o Business Continuity Plan (BCP) were presented, having in consideration the organization capabilities and the tax and legal obligations.

Since this is a new study, there are limited implementation results to present. Nevertheless, the companies questioned referred to the need to implement electronic archive to reduce costs, optimize business process and archiving procedures and find information faster. This extended version attends those concerns and goes a step further by integrating those requirements in the analysis, providing a new tool to cope with constraints during the pandemic situation.

The above direction presents further the Business Continuity (BC) components and updates the guidelines for the BCP. Now it is induced a key perspective in the development of the guidelines, including the forced paradigm of remote work and the constraints for business operations, consequence of a pandemic situation. Organizations must endeavor and adapt to resist the changes or

*Corresponding Author: Nelson Russo, al70461@utad.eu

thrive with the opportunities and continue the digital transformation.

Adding to this situation, in recent years, the software development industry has been faced with legislative changes, which include more functional, and tax or fiscal requirements. By legal imposition, there was an expansion on the range of organizations covered by the obligation to have Invoicing Software (ISw) certified by the Autoridade Tributária e Aduaneira (AT), the Portuguese Tax and Customs Authority.

The resulting legislative simplifications created conditions for the dematerialization of documents, providing for the possibility of dismissing invoice printing, encouraging the adoption of an electronic invoicing and electronic document archiving system.

Thus, measures of sustainability and cost reduction were incorporated, namely, the reduction of file space, reduction of paper, consumables and printing hardware, allowing also to expand the optimization and automation of invoicing and archive procedures. These measures may be a stimulus for organizations to invest in the development and use of new technological instruments, incorporating a philosophy of innovation and sustainability.

In this sense, having in mind that ISw are one of the components of Information Systems (IS), they are subject to a specific framework in the tax area and to a set of functional requirements that aim to ensure data integrity, security of information and business continuity. In this work, the diplomas that frame and describe the technical requirements that the ISw must observe to obtain a certificate of conformity will be presented.

Whether due to the investment capacity, the integrated knowledge of the requirements involved, the interpretation of the diversity of existing standards and good practices, the perception of imminent risk or simply the definition of priorities, some organizations have not yet started designing the BCP or have the notion that technology will simply solve problems during, and after, a disruptive event in their activity. A BCP does not necessarily need to have a high degree of complexity, in the case of small organizations. It is commonly discussed that the pertinence of analyzing risks, understanding how to continue the business in the event of a disaster and to recover from that disaster are crucial activities in business continuity and that should be understood as adding value to organizations and not just as a requirement to meet to be compliant. However, the failure to define a BCP makes it difficult to comply with tax legislation in the context of the certification of ISw, increasing the duration in time of the certification process, usually in scenarios where organizations, for commercial reasons, need quickly to obtain the certificate issued by AT. In addition, the present pandemic situation triggered the need for organizations to transform digitally and to integrate their business software with electronic invoicing capabilities, especially for online sales.

Thus, the need arose to create a guide for the design of a BCP, to reduce the gap in organizations, between the perception of business continuity requirements and the requirements underlying the certification of an ISw, translating them into concrete proposals with due legal and fiscal justification, standards and good practices in the area of Information and Communication Technologies (ICT).

2. Literature review

The accomplished literature review presents business continuity as a program to be followed by organizations that intend to be prepared. It included the review of regulations in the field of the thematic, as well as the analysis of Data Centers and legislative needs underlying the certification of ISw.

2.1. Business Continuity

Organizations can be thought of as a set of value chains or business processes transversal to the various organic units that compose them [2]. Economic, technological and human uncertainties have long presented organizations with the possibility of crises arising, impeding their ability to operate and maintain their business processes and, finally, survive [3].

Organizations increasingly realize the importance of adopting proactive approaches, it is essential to safeguard business processes and the relationship between them, in order to achieve business objectives [4]. Thus, organizations must prepare themselves in the event of disruptions to their productivity and competitive capacity, especially in business processes, supported by ICT services, made available to the organization.

With this in mind, organizations must develop a set of policies and procedures to minimize the impact of these disturbances. The importance of these measures is emphasized, whose impact may affect productivity and competitive capacity, especially in the case of organizations whose dependence on ICT services is marked. This is the concept underlying the BC, in which an organization must have the strategic and tactical capacity, to plan and respond to business incidents and interruptions [5], in order to continue business operations, at an acceptable level predetermined.

One of the accepted responses in BC is Disaster Recovery (DR), defined as the process performed by an organization, to resume business services, after a disruptive event [6]. In this sense, the authors advocate a DR planning approach [7], favoring the recovery of business processes, especially those supported by ICT service.

However, organizational resilience can be understood, as the ability to increase the probability of continuity of any system against disruptive incidents, whether are internal or external variations, changes, disturbances, disruptions and surprises [4]. This concept of resilience also covers effective planning for the relaunch of business processes in the short term, with the BCP, in addition to the long-term restoration, with the Disaster Recovery Plan (DRP) of interrupted operations, following disruptive events.

Being prepared for these events requires proactive planning of the organization's resources so that it can deal with disasters effectively and efficiently [8]. Therefore, the organization must plan its actions before a disaster occurs, taking into account the crisis management phases, resulting from a disaster, as proposed in Figure 1.

Thus, a BCP tries to eliminate or reduce the impact of a disaster and is designed to avoid or mitigate risks and reduce the time

necessary to restore conditions to a state of normal operation [10], aiming to preserve the assets of an organization [11].

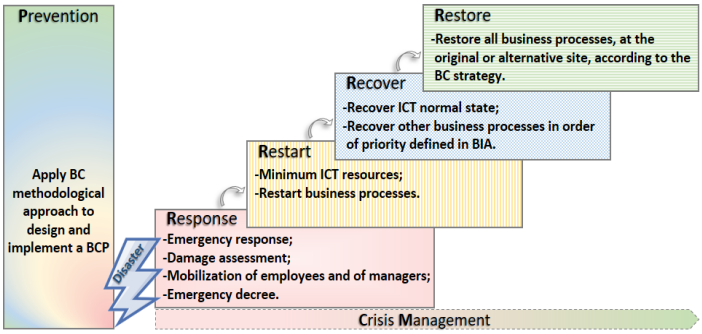


Figure 1: Stages of a business interruption. Adapted from [9].

Some vital activities are identified by Hiles [11], who considers the development of the BCP as a project, with a defined scope, with the analysis of risks and identification of the organization's specific key threats. Through the systematic process of Business Impact Analysis (BIA), vulnerabilities are revealed and it is determined and evaluate the potential effects, or impact on business, of an interruption. The definition of the strategy and requirements for technological resources is a success factor and it is essential to test, exercise and provide training on the BCP [11].

Figure 2 shows the essential components of the Business Continuity Management (BCM), according to the literature review and the standards presented in this work: the contribution of the International Organization for Standardization (ISO) through ISO 22301:2019, ISO 22313:2020 and ISO/IEC 27031:2011, the National Fire Protection Association (NFPA) emergency preparedness standard called NFPA 1600. Includes the vision of the Information Technology Infrastructure Library (ITIL), the perspective of the Control Objectives for Information and related Technology (COBIT) and Capability Maturity Model Integration (CMMI).

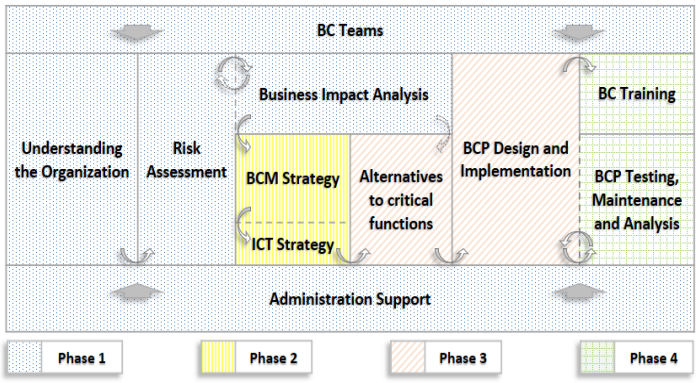


Figure 2: BCM components. Adapted from [12]

Figure 2 presents a sequence of action and interrelationship of the components of the BCM. It also presents the time phasing, in the methodological approach for the design and maintenance of a BCP [12], suggesting the set of components that must be completed, to start another phase.

2.2. Data Center

The Data Center infrastructures must be seen as a guarantee of fail-proof and insurance of business continuity, but if there is a disruptive event, it should be at least as short as possible. The Uptime Institute created the Tier Classification System standard, which is the global language of Data Center performance [13] as a means of effectively evaluating it in terms of business requirements for system availability. It offers a consistent method of comparing unique and customized installations based on the expected performance or uptime of the site's infrastructure.

2.3. Standards

A set of standards was reviewed in order to incorporate good practices on the business continuity topic.

NFPA 1600 2019

Widely used by entities, NFPA 1600 2019 was adopted by the United States Department of Homeland Security as a voluntary consensual standard for emergency preparedness [14].

The standard consists of 10 chapters and 13 annexes in which a common set of criteria for all hazards is established in disaster, crisis and emergency management and business or operations continuity programs. It follows the Plan-Do-Check-Act approach, providing the fundamental criteria of preparation and resilience, including the planning, implementation, execution, evaluation and maintenance of programs for prevention, mitigation, response, continuity and recovery [15]. The standard may be applicable to public, private, non-profit and non-governmental entities.

ISO 22301:2019

The ISO 22301:2019, on the theme “Security and resilience - Business continuity management systems - Requirements”, is a standard for implementing a BC Management System and continuously improving the BC resources based on priorities and feedback management. It specifies the requirements for implementing, maintaining and improving a management system to protect against, reduce the likelihood of, prepare for, respond to and recover from disruptions when they arise [16].

ISO 22313:2020

The ISO 22313:2020 standard, with the theme “Security and resilience - Business continuity management systems - Guidelines on the use of ISO 22301”, provides guidelines and recommendations for the application of the requirements of the Business Continuity Management System (BCMS) indicated in ISO 22301. The guidelines and recommendations are based on international good practices. This standard is applicable to organizations that implement, maintain and improve a BCMS [17], who seek to ensure compliance with the stated business continuity policy, who need to be able to continue delivery of products and services, at acceptable predefined capacities following a disruption and who seek to improve their resilience through the effective application of the BCMS.

ISO/IEC 27031:2011

The International Standard ISO/IEC 27031:2011 entitled “Information technology - Security techniques - Guidelines for information and communication technology readiness for business continuity”. This standard describes the concepts and principles of ICT readiness for business continuity, and provides a framework of methods and processes to identify and specify all aspects (such

as performance criteria, design, and implementation) for improving an organization's ICT readiness to ensure business continuity [18].

2.4. Reference Standards and Good Practices

Within the scope of standardization and improvement of business processes, there is a concern with the application of good practices. Thus, due to its relevance, this chapter presents the set of frameworks and good practices ITIL, COBIT and CMMI.

ITIL

ITIL 4 brings ITIL up to date by re-shaping much of the established ITSM practices in the wider context of customer experience, value streams, and digital transformation, as well as embracing new ways of working, such as Lean, Agile, and DevOps. ITIL 4 provides the guidance organizations need to address new service management challenges and utilize the potential of modern technology. The guidance provided can be adopted and adapted for all types of organization and service. It is designed to ensure a flexible, coordinated and integrated system for the effective governance and management of Information Technology-enabled services [19].

One to the Service management practices, under ITIL management practices, is Service continuity management. Its purpose is to ensure that the availability and performance of a service are maintained at sufficient levels in case of a disaster [19].

COBIT

COBIT 2019 was designed and created by Information Systems Audit and Control Association (ISACA). Is a framework for the governance and management of information and technology, aimed at the whole enterprise. COBIT 2019 is composed of 40 ICT processes in two fundamental areas, Governance and Management, divided into four domains. COBIT describes how an enterprise can design a customized and tailored governance solution for enterprise Information and Technology (IT) and should build its governance system for enterprise ICT [20].

COBIT defines the components to build and sustain a governance system: processes, organizational structures, policies and procedures, information flows, culture and behaviors, skills, and infrastructure. It defines the design factors that should be considered by the enterprise to build a best-fit governance system and addresses governance issues by grouping relevant governance components into governance and management objectives that can be managed to the required capability levels [20].

It was important in this work to characterize the focus on strategic alignment between ICT and business solutions; in the generation of value, but optimizing costs and ICT services; risk management, disaster recovery and continuity of operations.

CMMI

CMMI V2.0 is an integrated set of best practices that enable business to improve performance of their key business processes. It helps a business to understand its current level of capability and performance and its practices can guide improvement [21].

CMMI deals with various Capability Areas like “Managing Business Resilience”, that addresses the ability to anticipate, prepare for, and respond to interruptions in order to continue operations. Incorporates Practice Areas namely “Incident Resolution and Prevention” to resolve and prevent disruptions to sustains service delivery levels. “Risk and Opportunity Management” in order to identify, record, analyze, and manage risks and opportunities and “Continuity” to plan mitigation activities for disruptions to operations so that work can continue or resume [21].

3. Tax legislation

Portuguese tax legislation defines technical requirements for ISw [22] to guarantee data integrity, information security or business continuity and are described, briefly, in the legislation presented below. Some requirements establish the conditions to be observed when scanning and archiving in electronic format, intended to replace the respective paper archive. It also regulates the technical conditions for the issue, upon acceptance by the recipient, of invoices issued electronically and for their conservation and archiving, as well as for the existence and conservation of backup copies of the data supporting the ISw [23].

3.1. Decree Law No. 28/2019

The main objectives of Decree Law no. 28/2019, of February 15, of the Presidency of the Council of Ministers, are to promote legislative simplification and harmonize divergent rules on the conservation of documents for tax purposes. In this context, Decree Law no. 28/2019 created the conditions for the “Paperless invoice” and apposition of Quick Response (QR) code.

It provided for the possibility of waiving the printing of invoices and also strengthened the conditions for the dematerialization of documents, encouraging the adoption of an electronic invoicing system and electronic document archive, being an asset for the development of a more ecological vision, based on reducing the use of paper, inks and printing toners. The conditions created allow organizations to reduce costs with the fulfillment of tax obligations, stimulating the development and use by organizations of new technological instruments, incorporating a philosophy of innovation and debureaucratisation. For this purpose, a substantial reform of the rules applicable to the archiving of books, records, databases and supporting documents of accounting was introduced [23].

3.2. SAF-T (PT)

The Standard Audit File for Tax Purposes - Portuguese Version (SAF-T (PT)) is a standardized tax audit file for data export in Extensible Markup Language (XML) format, mandatory for ISw, created in 2007, with the publication of Ordinance 321-A/2007, of March 26, of the Ministry of Finance and Public Administration.

Its objective is to allow an easy export, at any time, of a predefined set of accounting, invoicing records, and relevant tax documents, in a legible and common format, regardless of the ISw used, without affecting the internal structure of the ISw database or its functionality [24]. The SAF-T (PT) file is intended to facilitate the collection and processing, in electronic format, of the relevant tax data. It allows simplifying procedures and boosting the use of new technologies [24], which may be the file that contains

the list of the relevant tax documents registered on archive. The most current SAF-T (PT) data structure is set out in Annex I to Ordinance No. 302/2016, of December 2, of Finance.

3.3. Decree Law No. 198/2012

The Decree Law no. 198/2012, of 24 August, of the Ministry of Finance, the starting point for the known E-Fatura service, creates a regime that regulates, namely, the electronic transmission of data elements of invoices and other documents with tax relevance, including printing of transport documents, to AT. Establishes measures to control the issuance of invoices and other documents of tax relevance, defines the means for its communication to AT and creates a tax incentive when the acquirer requires such documents [23].

It should be noted that this communication can be made through electronic transmission in real-time integrated with a certified ISw using a web service provided by AT or by sending the summary file, based on the file SAF-T (PT), exported periodically (usually monthly) by a certified ISw.

3.4. Invoicing Software Certification

Compliance with the provisions of Ordinance No. 363/2010, of 23 June, allows obtaining the title of an ISw certified by AT [25]. AT certification is important to foster compliance with the law and legal technical requirements and good ICT practices.

Taxable persons, whose obligation to issue an invoice is subject to the rules established in Value Added Tax Code (VATC), are obliged to use, exclusively, ISw that have been subject to prior certification by AT, whenever have had, in the previous calendar year, a total turnover exceeding 50,000 euros or are using an ISw or have organized accounting.

Thus, if organizations do not choose to use an ISw that have already been certified and intend to use those that they have produced, they should request their certification. To start the certification process, organizations must submit, on the Finance Portal, an official "Modelo 24" form.

The issuance of the certificate may be preceded by compliance tests to observe the requirements resulting from Portuguese tax legislation and a set of defined technical requirements [25]. It should be noted that the certified version of an ISw must comply with the corresponding requirements, which is the responsibility of the ISw producer, even after the compliance tests and throughout the life of the ISw [23].

The compliance tests are conducted by AT, being responsible for analyzing, clarifying and conducting the face-to-face compliance testing meeting an AT Inspector, with functions assigned to certification of ISw.

Since the beginning of the COVID-19 pandemic declaration, faced with the impossibility of free movement and other restrictions, AT has prepared a videoconference solution that would allow to give support to software producers and to safely perform compliance tests. After all the relevant requirements have been verified by videoconference, a meeting would be scheduled to conduct compliance tests, in a face-to-face meeting. Thus, it was possible to keep the certification process active.

AT maintains, on the Finance Portal, an updated list of programs and respective certified versions, as well as the identification of the software producers, proving that during the COVID-19 confinement months, several ISw were certified.

3.5. Ordinance No. 363/2010

The preamble to Ordinance No. 363/2010, of 23 June, states that the increasing use of ISw has undeniable advantages in terms of the speed of information processing. However, it introduces new risks in terms of fiscal control, due to the possibility of subsequent adulteration of recorded data, enhancing situations of tax evasion [25].

In this perspective, the aforementioned ordinance defines rules so that the ISw observe the requirements that guarantee the inviolability of the information initially registered, allowing, consequently, that only the ISw that respect such requirements can be used, after certification by AT. The certification of ISw depends on the cumulative verification of several requirements, among them must be able to export the SAF-T (PT) file, have access control to ISw and respect the technical requirements defined in Order No. 8632/2014.

3.6. Order No. º 8632/2014

Order No. 8632/2014, of July 3, from the Ministry of Finance, was published to comply with the provisions of Ordinance No. 363/2010, which states that one of the requirements for obtaining ISw certification is to observe the technical requirements approved by the AT. Thus, ISw even if already certified, must comply with its approximately 100 technical requirements in different areas, including backup policies, data recovery, access controls or document signing and security [22].

4. Methodology

The research-action methodology in the realistic perspective was the main axis of the investigation protocol. It helped to reduce the complexity of the problem, due to the variety of categories of organizations and their ICT strategies [26], as well as business specificities. It would be a hindrance to integrate the requirements and the hypotheses for solving the problem in a single approach [27]. Thus, a questionnaire was created for ICT managers as a way of capturing relevant information about the invoicing solution implemented in the organization. From the data analysis, it's inferred a BCP proposal that integrates the researcher's knowledge about the problem, the applicable legal and normative requirements, the constraints and the various conceptual technological alternatives of resolution in order to outline a sufficiently comprehensive guide for most of the organizations in Portugal.

5. Results

The guide for the design of a BCP is outlined, starting with a questionnaire with the main questions, structured in specific knowledge areas that the ICT managers must ask in the analysis of the problem, as shown in Figure 3. It culminates in a better understanding of the business processes and procedures involved that will be the target of the BCP, regarding procedures to be designed for disaster recovery and business continuity of the ISw. It enhances the understanding of ISw's data backup policies and

approaches to the implementation of a technological conceptual solution capable of eliminating vulnerabilities and mitigating the effects of a disruptive event in the invoicing business processes.

Structurally, the questionnaire contains a set of multiple-choice questions, each contributing to the perception of the organization invoicing solution and to the compliance with specific requirements defined in the tax legislation, standards and good practices researched.

Figure 3 shows the decision support process with the questions that make up the questionnaire. The word “invoice” refers to all relevant tax documents, including, for example “orders” or, “credit note”. The order in which the questions are answered is not important, since the software that supports the guide adapts to each new question answered. Three questions allow diverting the flow of the program, as its answer invalidates the following interconnected questions:

9. "Is there an Invoicing Software?" - if not, it is not needed to ask about the producer responsible for the ISw and its architecture;

13. "Is the Invoicing Software connected to a central software?" - if not, there is no need to conclude about the location of the central database in the client/server configuration;

15. “Are invoices issued electronically?” - if not, there is no need to question if, or how Electronic Data Interchange (EDI) is used.

All other questions follow a common reasoning flow and are positioned according to the theme of the question, which may be important to request answers to the questionnaire from another employee of the organization. It is important to note that the guide was specially designed for organizations that have an ISw, and it is not an objective to find proposals outside this scope. The questions areas are:

- Organization: Allows the organization to be classified according to the criteria of Small and Medium Enterprises (SME) [28], to know the sector of activity in order to understand the investment capacity [29] and the total turnover to infer, and also about the obligation to have an ISw certified by AT;
- Invoicing volume: Allows to verify the invoicing issuance days; whether the organization has continuous invoicing (for example for scheduling jobs in periods of less activity) or only during normal working hours, as well as the average number of documents issued per day. It aims to reveal the periodicity of backup copies and more appropriate techniques, taking into account the conclusions about investment inferred in the “Organization” area;
- Printing: The objective is to understand whether organizations print the documents issued and keep the accounting archive in paper, in order to suggest scanning and archiving in electronic format, as well as the possibility of recovering documents in case of corruption of the database;
- Invoicing Software: Find out if the organization uses an ISw, even if not required, inferred by the conclusions in "Organization". Know if the ISw was developed in-house or acquired in order to understand the decision-making capability on requirements and technical capability on the ISw. It is important to understand if the ISw is desktop-based, web-based or a terminal application in order to infer the technological infrastructure associated with the physical location of the database;
- Electronic issuance: It aims to infer the processing of the documentation produced and propose the electronic issuance according to the technique used and with the conclusions in “Invoicing volume”;
- Manual issuance: Intends to assess the need to comply with the requirements set out in point 2.4 of Order 8632/2014 [22]. Propose business continuity in case the ISw is inoperable, taking into account the conclusions in “Invoicing volume”.
- Electronic Archive: Intends to assess the need to comply with the requirements set out in Decree Law No. 28/2019. Aims to propose a reduction of archiving costs, the efficiency of archiving procedures, especially those supported on remote work business processes and improve business continuance readiness.

The flow of decisional process and the questions that are part of the questionnaire are presented in Figure 3.

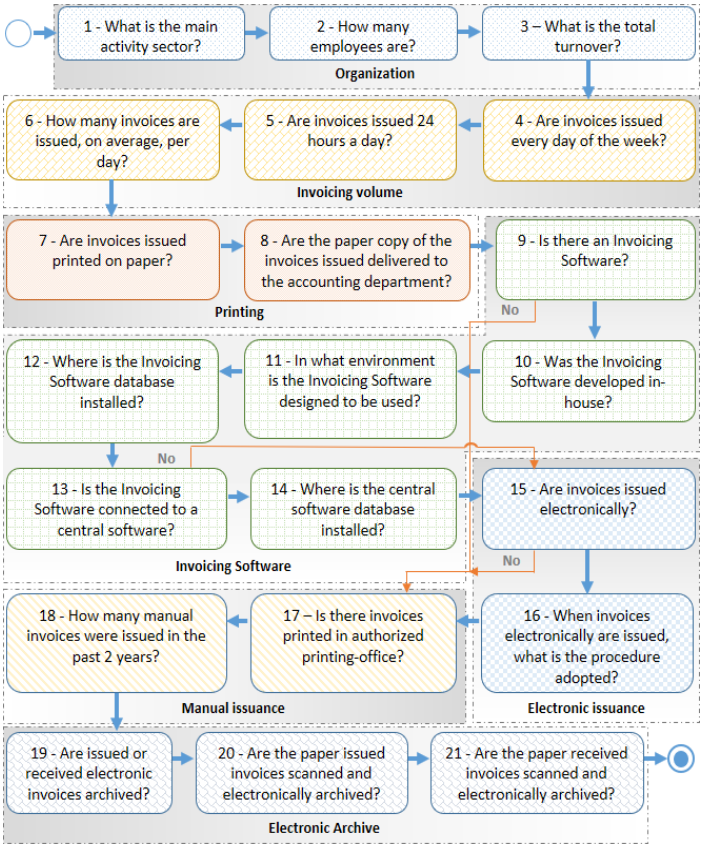


Figure 3: Decision support process for the design of the guide.

The questionnaire presented in Figure 3 has the basic or relevant questions to ask, and the multiple-choice answer will drive the decisional process conclusions. For example, when questioned about the connection to a central software, the answer will have a variety of options to try to determine the architectural solution implemented.

5.1. Discussion

Stablished or entrepreneurial entities, when starting or reformulating their commercial activity and business processes, should take into account the potential of legislative considerations mentioned in this work. The focus on online businesses, amplified by the increased use of telework, resulting from the measures to contain this recent pandemic, allows some management, optimization and control functions to be performed remotely. The effects are reflected in the need to optimize business processes, the time spent on activities, but above all in technological innovation that mitigates the gaps underlying this new vision of the business and digital world. This new vision will depend on the dynamization of ICT in business processes and the quick response to this desire, where there is a need to make quick and effective decisions, but with the necessary efficiency so that the change is adequate.

5.2. Future work

The aim is to design an approach that allows the questionnaire to be published online and to obtain more data for treatment, in order to refine the design of a BCP supported by technological solutions.

Nevertheless, for questionnaire evolution, some other requirements need to be addressed, especially the ones who improve the security of the software system and some other ones that enhance business continuity and disaster recovery.

Another research direction is to design a case study, to evaluate the business processes dynamization or optimization and obtain the return on investment achieved. The study should occur before and after the implementation of electronic invoicing and/or electronic archiving in the organization, in a set of preselected business processes. The target organization is one that intents to transform its business processes for telework, triggered for example, by restrictions due to pandemic situations.

6. Conclusions

The previous identification of the constraints felt by the organizations in the design of a BC solution allowed to conduct the literature review and the identification of the issues, considered relevant, in the context of supporting the design of a guide for the design of BCP. The review of the literature on the theme of BC, including standards and good practices, and the relevant applicable tax legislation was important in this work. The systematization of information in this field adds value to the scientific community, as well as to users who are part of this type of problem.

On the other hand, the change in the paradigm of issuing and communicating invoices and tax relevant documents triggered the need for organizations to prepare for the challenge of innovation or technological adequacy for electronic invoice issuing, as well as for the electronic archive. The current pandemic situation has accelerated this need for digital transformation.

The tax legislation points to the requirements to be met and provides for the possibility of issuing documents by electronic means, the possibility of their electronic transmission to AT, and the possibility of their electronic archive. Understanding what are the correct procedures and what are the requirements involved in the certification of an ISw in Portugal it is considered relevant for organizations that develop the software, as well as for organizations that decide to purchase a solution. This work aimed to help the identification of the main regulatory documents in the thematic, as well as the legal characterization of invoicing in Portugal, in order to support organizations in the design or adaptation of the ISw. The implementation of electronic invoicing and electronic archiving procedures can enhance business continuity and boost the innovation that leads to new software products, digital transformation, marketing approaches, or new ways of making business.

The questionnaire made it possible to systematize the variables involved underlying the tax requirements, the existing technological infrastructures, and the invoicing and archiving procedures implemented. Through a process of analysis of the answers, vulnerabilities can be perceived, as well as the needs for business continuity and disaster recovery. It is considered that the constraints perceived by organizations were reduced by improving the understanding of the variables that they must observe, the areas in the legislation and regulations that they must integrate, as well as the needs of human and technological resources, according to their investment capacity.

This work encourages the adoption of concrete strategies and procedures for business continuity, where invoicing and electronic archiving may be relevant, for organizations with an ISw, certified or not by AT.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] N. Russo, L. Reis, "Análise da Problemática Subjacente à Certificação de Programas Informáticos de Faturação," in CISTI'2019 - 14th Iberian Conference on Information Systems and Technologies. Coimbra, Portugal, 2019. doi:10.23919/CISTI.2019.8760638.

[2] R. Medina-Mora, T. Winograd, R. Flores, "Action Workflow as the Enterprise Integration Technology," Bulletin of the Technical Committee on Data Engineering, IEEEE Computer Society, **16**(2), 1993.

[3] B. Herbane, "The evolution of business continuity management: A historical review of practices and drivers," Business History **52**(6), 978-1002, 2010. doi:10.1080/00076791.2010.511185.

[4] N. Sahebjamnia, S. A. Torabi, S. A. Mansouri, "Building organizational resilience in the face of multiple disruptions," International Journal of Production Economics, **197**, 63-83, 2018. doi:10.1016/j.ijpe.2017.12.009.

[5] R. K.Ramakrishnan, S.Viswanathan, "The importance of Business Strategy in Business Continuity Planingem" in A. Hiles, The definitive handbook of Business Continuity Management Third Edition. West Sussex, England: John Wiley & Sons, Ltd, 2011. doi:10.1002/9781119205883.ch3.

[6] M. M.Al-shammari, A. A. Alwan, "Disaster Recovery and Business Continuity for Database Services in Multi-Cloud" in 1st International Conference on Computer Applications & Information Security (ICCAIS), 1-8, 2018. doi:10.1109/CAIS.2018.8442005.

[7] B. Herbane, D. Elliott, E. M. Swartz, "Business Continuity Management: time for a strategic role?," Long Range Planning, **37**(5), 435-457, 2004. doi:10.1016/j.lrp.2004.07.011.

[8] N. Sahebjamnia, S. A. Torabi, S. Mansouri, "Integrated business continuity and disaster recovery planning: Towards organizational resilience," European Journal of Operational Research, **242**(1), 2015. doi:10.1016/j.ejor.2014.09.055.

[9] A. Guindani, "Gestão da Continuidade dos Negócios," Integração, 2008.

[10] V.Cerullo, M. J. Cerullo, "Business Continuity Planning: A Comprehensive Approach" Journal Information Systems Management, **21**(3), 70-78, 2004. doi:10.1201/1078/44432.21.3.20040601/82480.11.

[11] A. Hiles, "An introduction to business continuity planning" in A. Hiles, The Definitive Handbook of Business Continuity Management, xxiii-xxvii, John Wiley & Sons, Ltd, 2007.
[12] N. Russo, L. Reis, "Methodological Approach to Systematization of Business Continuity in Organizations," Handbook of Research on Multidisciplinary Approaches to Entrepreneurship, Innovation, and ICTs, **24**, 2020. doi:10.4018/978-1-7998-4099-2.ch010.
[13] Uptime Institute, "Tier Classification System", Uptime Institute LLC, September 26, 2020, retrieved 2020-10-12, [Online] Available: https://pt.uptimeinstitute.com/tiers.
[14] National Fire Protection Association, "NFPA 1600® Standard on Continuity, Emergency, and Crisis Management", April 28, 2020, retrieved 2020-04-28, [Online] Available: https://www.nfpa.org/codes-and-standards/all-codes-and-standards/list-of-codes-and-standards/detail?code=1600.
[15] NFPA 1600, "NFPA 1600® Standard on Continuity, Emergency, and Crisis Management 2019 Edition". National Fire Protection Association, 2019.
[16] ISO 22301:2019, "Societal security - Business continuity management systems – Requirements", Switzerland: ISO, 2019.
[17] ISO 22313:2012, "Societal security - Business continuity management systems – Guidance", Switzerland: ISO, 2012.
[18] ISO/IEC 27031:2011, "Information technology - Security techniques - Guidelines for information and communication technology readiness for business continuity", ISO/IEC, 2011.
[19] ITIL, "ITIL Foundation ITIL 4 Edition". Londres: AXELOS, 2019.
[20] ISACA, "COBIT 2019 Framework - Governance and Management Objectives". Schaumburg, IL 60173, USA: ISACA, 2018.
[21] CMMI Institute, "CMMI Model V2.0". CMMI Institute, 2018.
[22] Order No. 8632, "Despacho n.º 8632/2014, de 03 de julho". Lisbon, Portugal: Ministry of Finance, 2014.
[23] Decree Law No. 28, "Decreto-Lei n.º 28/2019, de 15 de fevereiro". Lisbon, Portugal: Presidency of the Council of Ministers, 2019.
[24] AT, SAF-T PT (Standard Audit File for Tax purposes) - Portuguese Version, September 19, 2020. http://info.portaldasfinancas.gov.pt/pt/apoio_contribuinte/SAFT_PT/Paginas/news-saf-t-pt.aspx.
[25] Ordinance No. 363, "Portaria n.º 363/2010, de 23 de junho". Lisbon, Portugal: Autoridade Tributária e Aduaneira, 2010.
[26] KPMG, "Risco e Resiliência - Estudo de Continuidade do Negócio em Portugal", 2018.
[27] R. Yin, Case Study Research - Design and Methods (volume 5, fourth ed). California: Sage Inc, 2009.
[28] Commission, COMMISSION RECOMMENDATION of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises. Official Journal of the European Union. The Commission of the European Communities, 2003.
[29] N. Russo, L. Reis, "Caracterização da faturação em Portugal: sob a perspetiva da certificação de programas informáticos de faturação," Revista de Ciências da Computação, **14**, 67-84, 2019.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 853-864 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

An Enhanced Conceptual Security Model for Autonomous Vehicles

Abdulla Obaid Al Zaabi[*,1,2], Chan Yeob Yeun[1,2],Ernesto Damiani[1,2], Gaemyoung Lee[3]

[1]*Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi 127788, United Arab Emirates*

[2]*Center for Cyber-Physical Systems, Khalifa University, Abu Dhabi 127788, United Arab Emirates*

[3]*Department of Electrical Engineering, Jeju National University, Jeju, 63243, South Korea*

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 01 November, 2020
Online: 27 November, 2020

Keywords:
Autonomous
Vehicles
Security

ABSTRACT

Connected and self-driving cars have emerged over the last decade as a leading example of cyber-physical systems, which seek to considerably enhance traffic safety, reduce emissions, decrease costs, and improve efficiency. Google, TESLA, Uber ATG are becoming pioneers in the autonomous vehicles industry. Autonomous vehicles can have a large codebase and with a large volume of messages exchanged. The concepts of connected driving, cooperative driving, and intelligent transportation systems increase the connectivity of vehicles to the internet or other cloud services. High connectivity, misconfiguration, and insecure coding widen the attack surface of autonomous vehicles. A conceptual model of autonomous vehicles is presented in this paper to better understand autonomous vehicles' attack surface. In addition, specific threat modeling techniques are discussed. Experiments were carried out to demonstrate the security risks to autonomous vehicles due to third party Electronic Control Units.

1 Introduction

The term autonomous vehicles is used to designate technology-oriented vehicles with robotic and self-driving features (self-parking, GPS sensing tools, and higher automation capabilities). Both connected and self-driving cars have emerged over the last decade as a leading example of cyber-physical systems, which seek to enhance traffic safety, reduce emissions, decrease costs and improve efficiency. At the same time, the automobile industry has experienced a major transformation from monolithic manufacturing processes to distributed procurement. The paradigm of distributed procurement allows automobile manufacturers to acquire software and sensor/hardware components from numerous third parties.

Distributed systems usually comprise of a complex connectivity layer with different nodes. The features of connectivity in autonomous vehicles provide real-time connections to the web, sensors, radar, GPS with user privacy, and other measurement units [1, 2]. In other words, autonomous vehicles are empowered to acquire all relevant information needed to perform the self-driving functions [3].

As the market penetration of autonomous vehicles has increased, there is a growing demand among organizations, consumers, and regulatory bodies for the underlying technologies to be tested to guarantee their overall safety, as well as evaluate their cloud and cyber security together with data privacy. The augmented diversity of software, sensors, and communication channels has driven this demand [4]. During the process of critical-information exchange between the vehicle and the external infrastructure, a new attack surface can emerge, as several stakeholders (telematics, hardware, and software components, etc.) and rapid modifications (i.e., software processes) are involved.

Sensors can provide an extensive amount of information on the environment surrounding the vehicle together with sensing the internal components. Electronic Control Units (ECUs) and sensors are the focal point of information exchange inside autonomous vehicles, as they provide critical information to the engine control unit of the vehicle. The data obtained from sensors can be shared and utilized for all types of assessments and calculations by other automotive systems.

Sensors and ECUs contain a large codebase that was enhanced to provide features required by autonomous vehicles. This "functional abundance" has widened the attack surface, adding to the existing vulnerabilities that have not been identified in the old code bases. Sensors and ECUs can talk to each other with different networks equipped in modern vehicles. The design of such networks has focused on efficiently wiring these components rather than on security.

There are 6 different levels of autonomous driving defined by

[*]Corresponding Author: Abdulla Al Zaabi, Abu Dhabi 127788, abdulla.alzaabi@ku.ac.ae

the Society of Automotive Engineers as shown in Figure 1. Modern vehicles nowadays can achieve levels between two and three.

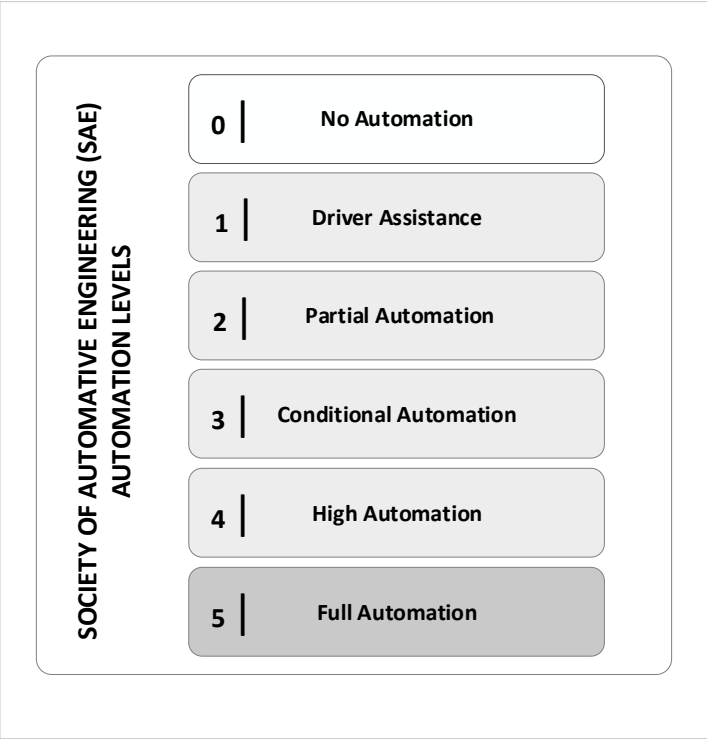


Figure 1: SAE automation levels

The motivation behind this work is to develop a conceptual model for autonomous vehicle to better understand the attack surface of autonomous vehicles and support their security testing. This will allow researchers and manufacturers to better understand the attack surface on autonomous vehicles and bridge the gap in the areas that have not been explored.

This paper develops on our previous work on defining a conceptual model for autonomous vehicles [5], where we modeled the autonomous vehicles' assets, their vulnerabilities and the corresponding threats. Here, we take a closer look to modeling the assets related to the Controller Area Network (CAN) of the vehicles. The rationale behind this choice is that many aftermarket gadgets and smart applications taking advantage of the CAN bus could contain vulnerabilities exploitable by malware. We conduct an experiment on a simulated CAN bus to illustrate different security attack scenarios such as traffic sniffing, replay attack and packet injection. Based on our experiment, we discuss and compare different threat modeling techniques that can be used to identify security threats based on the conceptual model presented.

The contribution of this article is, therefore: (i) enhanced conceptual model of autonomous vehicles to better understand their attack surface, (ii) analysis of Controller Area Network and cyber security attacks on modern vehicles, (iii) analysis of specific threat modeling techniques, (iv) demonstration of the cyber security risks to autonomous vehicles due to third party Electronic Control Units.

The paper is organized as follows. In Section 2 we present an autonomous vehicle conceptual model. Section 3 describes Controller Area Network protocols. Section 4 overviews cyber security attacks on modern vehicles. Section 5 focuses on a remote attack that had an impact on millions of vehicles. Section 6 provides an overview of different threat modeling techniques that could be used for threat modeling of autonomous vehicles. Section 7 illustrates the experimental results of sniffing and injecting traffic on CAN network using a simulated environment. In section 8 we summarize and discuss the future direction of research.

2 Autonomous Vehicle Conceptual Model

This section describes our enhanced conceptual model of autonomous vehicles [5]. The model consists of three main areas; vehicle, communications, and edge-cloud (Figure 2).

2.1 Vehicle

The vehicle component of the conceptual model focuses on the physical boundary of the vehicle. In the conceptual model, we focused on the important items that will support future work in security modeling and analysis of autonomous vehicles. The main categories of the vehicle layer are Electronic Control Units (ECUs), they are the brain of the vehicle, Sensors, and Inter-Vehicle communication. In addition, it is important to highlight the internal connectivity items such as the On-Board Diagnostic port (OBD-II). Different types of networks exist in autonomous vehicles with different speeds. The ECUs in the vehicle are usually connected to at least one or more internal vehicle network. The gateway is used to facilitate communication between different communication buses. The OBD-II connection provides access to the vehicle's internal network and interaction with various low-level sensors.

ECUs are essential components of modern vehicles. Each ECU is responsible for performing actions such as door locks, power steering, fuel injection, and many other critical functions. ECUs can also be categorized into subcategories based on their functionality [6]. The list below provides categories for ECUs found in modern vehicles:

- Powertrain: controls many factors in vehicles and responsible for transmission, emission controls, and onboard control modules.
- Safety: vehicle safety-related operations such as collision detection, active parking, and air-bag system.
- Body: less critical operations of the vehicle such as power windows and mirrors, cooling and heating, etc.

Examples of some of the key ECUs from the categories above are:

- ECM: engine control module
- VVS: vehicle vision system
- IRM: inflatable restrain module
- NCM: navigation control module

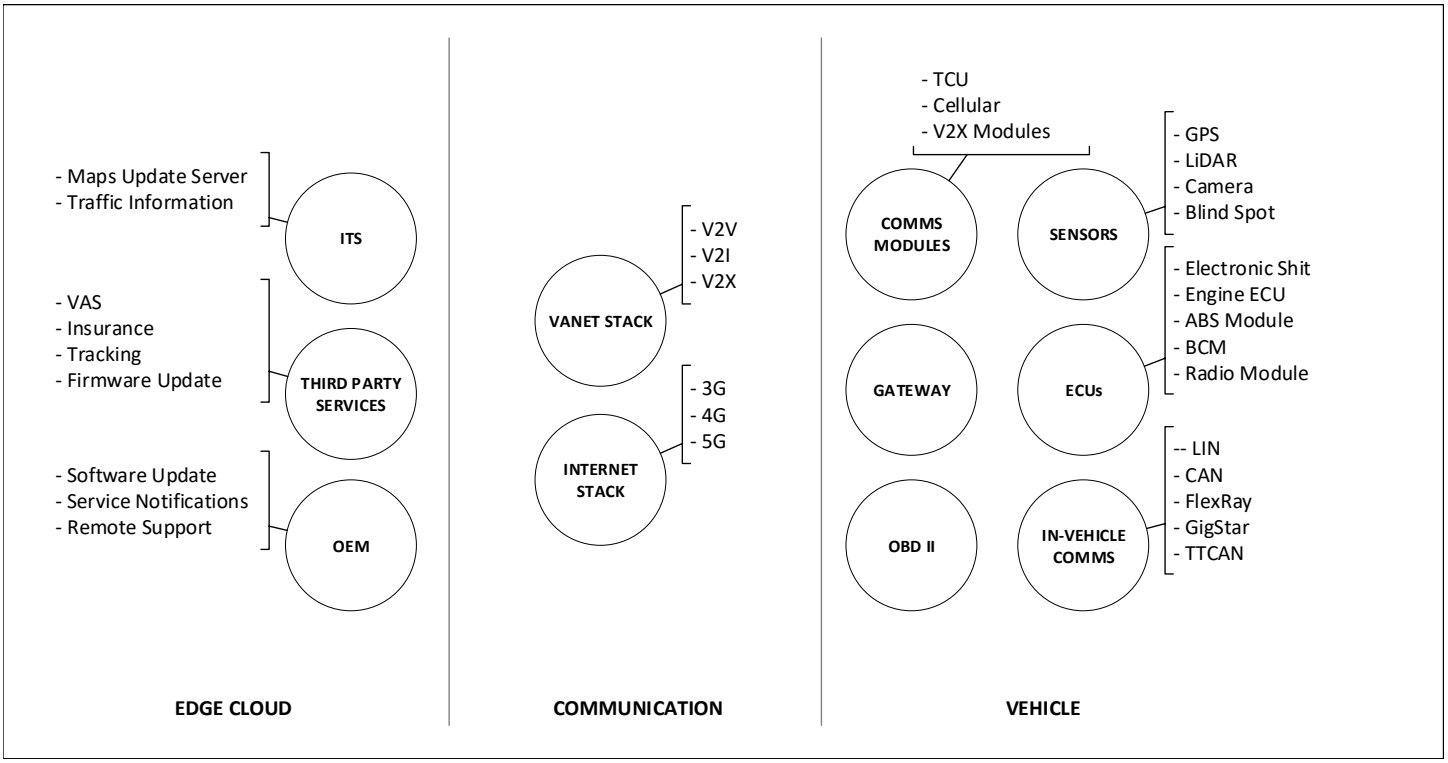


Figure 2: Autonomous Vehicle Conceptual Model

Sensors are used in autonomous vehicles to collect information about the vehicle's surroundings. Vehicles are equipped with various sensors, examples of sensors equipped are GPS, cameras, LiDAR. The vehicle uses the information received from the sensors to perform different actions such as warning messages, navigation capability, distance calculation, and others.

The Inter-vehicle communication component of the conceptual model focuses on the communication networks in the physical boundary of the vehicle. The main purpose of these networks is to allow efficient communication between electronic units in the vehicle. There are different types of networks in vehicles such as CAN, TTCAN, FlexRay, and Local Interconnect Network (LIN) [7, 8]. The CAN is widely used in automotive to support critical applications with a bandwidth of 125 kBits/s. The LIN bus runs at a lower bandwidth (10 kBit/s) to support less critical applications such as power windows and doors lock. FlexRay is a more expensive implementation of CAN, used commonly in BMW. FlexRay has 2 channels at a speed of 5 and 10 Mbit/s [9, 10].

2.2 Communication

The communication layer of the model focuses on the external communication medium for autonomous vehicles. Future transportation system is envisioned to support cooperative driving and information exchanging to enhance the driving experience, increase safety, and reduce collisions. There are two main categories in the communication layer; VANET stack and Internet Stack.

VANET Stack is the vehicular Ad Hoc Networks (VANETs). VANETs can use wireless networking technology to allow vehicles to communicate with each other and share information using Dedicated Short Range Communication (DSRC). There are different types of communication in VANETs. The first one is V2V, which stands for Vehicle to Vehicle communication. V2I: vehicle to station communication. V2R: vehicle to Road Side Unit (RSU) communication. V2X: vehicle to everything.

There are two types of messages in VANETs; beacon messages and special purpose messages [11, 12]. Vehicles broadcast messages that act as heartbeats using beacon messages. Special purpose messages carry information about collisions and warning messages to nearby vehicles [13]. VANETs using DSRC or even cellular network is anticipated as the future of intelligent transportation system [14].

Internet Stack component of the model considers the cellular network as the communication means. Internet stack includes 4G and 5G data link and physical link [15]. Initial research in autonomous driving focused on VANETs as the future of communication for vehicles using DSRC. Modern vehicles are equipped with cellular network connectivity that could enable VANET communication over cellular network.

2.3 Edge Cloud

Edge Cloud layer of the model focuses on the vehicle communication with external entities. Original Equipment Manufacturers (OEM) perform remote updates to vehicles remotely for applying patches, issues fix or enhance features. Many services are currently offered by third party providers such as media services, fleet management, vehicle tracking, and many other applications and services. Below are some examples:

- Maps update service.

- Traffic information.
- OEM firmware update.
- Third-party dongles firmware update.
- Value Added Services.

Some insurance companies nowadays mandate drivers to install OBD-II compatible dongles to monitor driving behavior and adjust insurance tariffs accordingly. Vehicles are expected to be connected to edge cloud services to support smart and green transportation [16]. As part of intelligent transportation system management [17, 18], vehicles are expected to be connected to transportation authorities and share various types of information.

3 Overview of Controller Area Network

The Controller Area Network (CAN) is a standardized network adopted by many car manufacturers for efficiency, safety, and manufacturing purposes. CAN ensures minimum latency, together with other features to handle functional networks of a vehicle. It has been adopted since early 1980s to primarily handle automotive electronics, and more specifically, different types of electronic control units [19]. The CAN protocol has many distinctive features such as being able to discover defects in the communication network and electronic control units (ECUs) [20]. ECUs play a crucial role in operating an autonomous vehicle. By allowing ECUs to exchange information via CAN, the autonomous vehicle can support new intelligent features. In comparison with other networks, however, CAN have some weaknesses such as limited electrical loading, and vulnerability to overflow of undesirable interactions or communications [21].

Known vulnerabilities of CANs have encouraged experts to develop a new security network for ECUs, especially for automated and self-driving vehicles [20]. One of the major motives for developing a new network is that CAN was not designed for communications from outside the vehicle, i.e. vehicle-to-Internet communications, vehicle-to-cloud, vehicle-to-pedestrian, and other types of connectivity. As a result, cyber security challenges, particularly denial-of-service attacks are a major hurdle when using CAN for autonomous vehicles. Moreover, third party dongles have been introduced by different service providers such as insurance or fleet management companies. These devices interface with the vehicle through the On-Board Diagnostics port (OBD-II). Physical OBD-II connection provides access to the vehicle's internal network and interaction with various low-level sensors.

3.1 CAN Protocols

CAN is a simple protocol that was deployed vehicles since 1996. It allows Electronic Control Units (ECUs) to communicate with each other. CAN run on two wires; CAN High (CANH) and CAN Low (CANL). At the physical layer, CANH uses differential signaling, which is used in environments where fault-tolerance to noise is a must. When a signal comes in, CANH raises the voltage on one line and drops the other line of an equal amount.

3.1.1 The OBD-II Connector

Most vehicles are equipped with On-Board Diagnostic Connector named OBD-II. It is usually located under the steering wheel of a vehicle. It is also known as a diagnostic link connector (DLC). CAN Connections can be found in supported vehicles from the OBD port in dual-wire pairs. CANH and CANL can be found on pin 6 and 14 or the OBD-II connector.

- CANH: High-Speed CAN Lines
- CANL: Low Speed or Mid Speed CAN Lines (LS-CAN and MS-CAN).

3.1.2 CAN Packets Layout

There are two types of CAN packets; standard and extended. Extended packets are like the standard ones but with larger space to hold IDs. The Complete frame of standard CAN is shown in Figure 3. Fields of standard packets are described below.

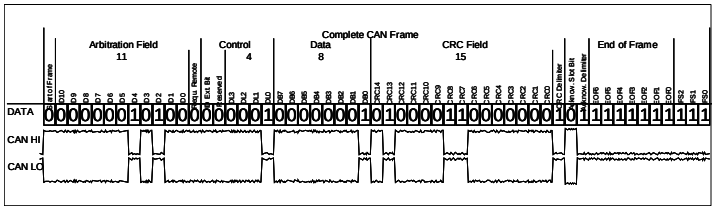


Figure 3: CAN Frame of standard packet [22]

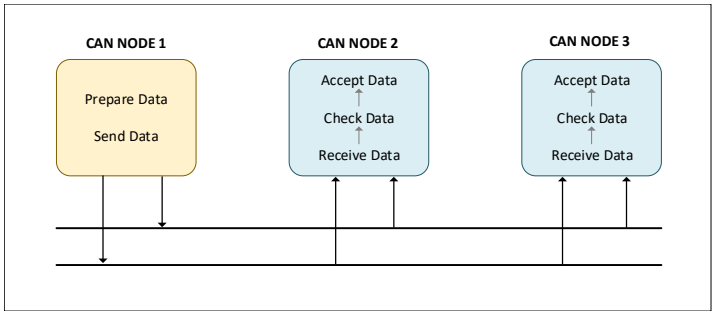


Figure 4: CAN communication based on broadcast messages

- Arbitration ID: a broadcast message that identifies the ID of the target device it tries to communicate with.
- Identifier Extension (IDE): always 0 for standard CAN
- Data length code (DLC): the size of the data, which ranges from 0 to 8 bytes
- Data: data to be transmitted.

Communication on CAN bus is broadcast messages, similar to UDP packets on ethernet networks as shown in Figure 4. All controllers see every packet transmitted on CAN. Extended packets are similar to standard packets, but the can be changed together to create longer IDs. They are designed to fit inside standard CAN packets with backward compatibility.

3.1.3 The ISO-TP Protocol

The ISO 15765-2 is a standard for sending packets over CAN. It extends the limit of sending 8 bytes to 4095 bytes. This protocol is commonly used for diagnostic messages (Unified Diagnostic Services) and KWP messages (alternative protocol to CAN). The first byte is used for extended addressing, leaving only 7 bytes for data per packet. Using ISO-TP can easily flood the CAN bus, thus it should be used carefully [23].

3.1.4 The Local Interconnect Network Protocol

The Local Interconnect Network (LIN) is a master-slave facility that was designed to complement CAN. It can have up to 16 slaves

4 Overview of Cyber Attacks on Modern Vehicles

This section focuses on recent cyber security attacks on modern vehicles. Firstly, the section provides an overview of the interesting vulnerabilities discovered by researchers in different modern vehicles. In the second part of this section discusses in detail one of the most interesting attacks on vehicles, the Jeep Cherokee 2015 attack.

4.1 Vulnerabilities Overview

4.1.1 Jeep Cherokee

Uconnect is an entertainment system that relies on cellular network to deliver entertainment services and applications in Jeep Cherokee entertainment system [24]. The entertainment system had a diagnostic bus port that runs on 6667. Since the port had no authentication requirements, some sensitive information was obtained by connecting to the port, such as GPS information, VIN, and other private information. At the time of the research, around 1.4 million vehicles using the Uconnect system were vulnerable. To identify vulnerable vehicles, a mass scan for the diagnostic bus port could be performed on the IP range of the service provider. The researchers used Sprint Cellular Network IP ranges to identify vulnerable vehicles on 21.0.0.0/8 and 25.0.0.0/8 address ranges. Additional details on the sequence of execution from Uconnect system to cyber-physical actions are explained in next section.

4.1.2 GM Chevy Impala

Another entertainment system related vulnerability was discovered in the GM Chevy Impala by researchers at the University of California at San Diego and the University of Washington [25]. This time, the entertainment system had a buffer overflow vulnerability. The vulnerability was exploited by playing a specially crafted mp3 file in the in-car call system. Successfully exploiting the vulnerability allows an attacker to access the vehicle's system and control most of its critical functions.

4.1.3 BMW

Security researchers identified a vulnerability in BMW ConnectedDrive service to unlock the vehicle remotely. Confidentiality requirement was missing in the communication between the vehicle and the BMW server for ConnectedDrive, the communication was not encrypted. Due to this, the attack performed using a fake cellular network and spoofed the BMW server. The vehicle is equipped with a communication capability using a pre-installed SIM card. An unlock message was sent to the vehicle using the spoofed server and it unlocked the vehicle successfully [26]. At the time, over 2.2 million vehicles were affected by the vulnerability. BMW released a security patch to fix the vulnerability by adding encryption to the communication between the vehicle and the BMW server.

4.1.4 Corvette

Corvette attack was mainly an attack on a third-party device, specifically on an insurance dongle connected to the vehicle via OBD port. The insurance dongle was installed to track driver behavior to determine insurance tariffs. The impact of the vulnerability depends on the commands allowed via OBD port. In the case of the Corvette, the researchers were able to trigger cyber-physical actions on the vehicle via the connected dongle. Through exploiting the vulnerability, the researchers were able to send an SMS to remotely apply breaks on the vehicle. Depending on the commands allowed from the OBD, other vehicles the attack could result in controlling more critical functions [27].

4.1.5 Tesla S Model

Different attackers were performed in Testa S Model, some required physical access to the vehicle and others exploited remotely. For the physical access, researchers were able to connect a laptop to the vehicle's in-car entertainment system and identified multiple vulnerabilities. Researchers were able to successfully implant a trojan in the vehicle's system to control it remotely. All the vulnerabilities identified in the research were patched by Tesla [28].

4.1.6 Nissan LEAF

Nissan application named Companion Application is used to allow users to have access to some information on the vehicle and perform specific actions. Researchers were able to identify vulnerabilities in the Companion Application and gain access to the historical GPS information of the vehicle [29].

5 Analysis of Cyber Security Attack on JEEP Cherokee

In this section, we analyze the attack on Jeep Cherokee that affected over 1.4 million vehicles. The researchers were able to perform the cyber attack remotely, with no physical access to the vehicle. Initial research required reverse engineering some of the components of the vehicle to plan the attack and chain different stages of the attack to reach from remote exploitation to perform physical actions on the vehicle such as applying breaks, vehicle's steering, and others. Although the experiment was conducted on Jeep Cherokee, it was not limited to this model. Vehicles with the same entertainment unit may have suffered similar vulnerabilities and impacts. In research, the attack surface was analyzed and the feasible attack chain was

chosen to perform the full attack chain, from initial access to physical action execution. The first part of this section discusses the CAN connectivity architecture seen in Jeep Cherokee, then discusses the attack phases. This is important to understand how to send messages between two networks in the vehicles (CAN buses).

5.1 CAN Connectivity

In the example of Jeep Cherokee, two CAN buses are used; CAN-HIS and CAN-C connected to different networks. ECUs talk to each other via CAN bus, each CAN bus is connected to a different set of ECUs. The CAN buses of Jeep Cherokee are explained below. The first CAN bus in the vehicle is CAN-IHS, stands for Controller Area Network for Interior High Speed. The CAN-IHS bus is low speed CAN bus that is used for non-critical systems such as climate controls, radios, and others. Examples of the ECUs connected to CAN-IHS bus are as follow:

- AMP Ampilifer Radio Module.
- Body Control Module (BCM).
- AHLM Module – Headlamp Leveling.
- PAM Module – Park Assist.
- DLC Data Link Connector

The second CAN bus is CAN-C bus, which is high-speed bus that is responsible to connect critical ECUs. Examples of the ECUs connected to CAN-C bus is as follow:

- Anti-Lock Brakes Module (ABS)
- Body Control Module (BCM)
- RADIO Module
- Electronic Parking Brake – EPB.
- Electronic Power Steering – EPS.

As we can see in the examples above, some ECUs are connected to both CAN buses, two examples are the radio module and the BCM Module. The direction the researchers took in the research was to target an ECU unit that could talk to both networks to widen the impact of the attack and reach critical ECUs.

5.2 Attack Phases

The attack chain developed by the researchers started from remote device exploitation to execute commands and sending control signals to critical ECUs. In section, the analysis is divided into four phases; target identification, remote code execution, sending CAN messages, and CAN payloads.

5.2.1 Identify Target

Initially, the researchers identified D-Bus service port 6667 by scanning the vehicle internally; connecting to the in vehicle's WiFi network. After examining the networking configurations of Uconnect, the researchers discovered an additional interface with a public IP address, which could result in remote exploitation of Uconnect system. To identify vulnerable Uconnect systems, performing a scan on port 6667 over the provider network. It was initially assumed that the provider blocks connections from other providers. The researchers took an additional step by exploiting a femtocell from the same provider to ensure that traffic is not blocked. It was later discovered that the port is open to all providers which increases the impact of the vulnerability. To avoid false positives, the service error message of services determines if the port belongs to Uconnect system or any other service such as IRC (Internet Relay Chat).

5.2.2 Gain Remote Code Execution

The researchers found more than one way to gain remote execution on the target system via the D-Bus service of Uconnect. Different vulnerabilities identified for remote code execution on the D-Bus service in addition to a feature that was implemented on the D-Bus service to allow code execution for an unauthenticated user. The first vulnerability idenfied was due to unsanitized user input on navTrailService function. The other vulnerabilities are either unauthenticated execution or command injection vulnerabilities.

Since arbitrary command execution achieved using the D-Bus service, researchers developed different payloads to interact with the CAN bus and obtain information using LUA scripts. Examples of the crafted commands sent via the D-Bus service.

- Uconnect system: From the Uconnect system, GPS coordinates can be queried.
- HVAC Module (A/C Heater): controlling heating and A/C of the vehicle.
- Radio Volume: controlling the volume of the radio.
- Knobs: disabling the D-Bus service make knobs unavailable and unresponsive.

5.2.3 Sending CAN Messages

The code execution achieved through the D-Bus service is performed on the OMAP chip. Although the Uconnect system is connected to both CAN buses of the vehicle; CAN-IHS and CAN-C, it cannot send messages to CAN directly. The communication from the OMAP to CAN buses are handled by Renesas V850ES/FJ3 chip. The researchers updated the V850 chip from the OMAP with a modified firmware to be able to send CAN messages. At this point, the execution from OMAP uses V850 chip as a proxy to communicate with CAN buses.

5.2.4 CAN Messages Payloads

The communication between OMAP chip and V850 chip uses a Serial Peripheral Interface (SPI). The V850 chip originally had a

predefined set of command IDs with hardcoded data to communicate with CAN bus. With flushed firmware, the V850 can send arbitrary CAN data from the OMAP chip.

At this point, arbitrary CAN commands can be sent remotely via the D-Bus service. After identifying the messages triggering cyber-physical actions, researchers were able to perform turn signals, locks, RPMS. These messages are considered normal CAN messages, the diagnostic CAN messages could cause greater cyber-physical impact but requires the vehicle to be traveling at a low speed.

6 Analysis of threat modeling frameworks

Threat modeling is in important phase of security engineering process. It involves identifying all security threats to the system regardless of the possibility to exploit them. It requires good understanding of the system. In order to develop meaningful security requirements, it is essential to have threats identified. Proper identification of threats in early stages can greatly reduce the ability of misusing the system [30]. There are different techniques of threat modeling [31]. In general, some of the main threat model techniques are:

- Attacker-centric
- Asset-centric
- Software-centric

These models are general and not specific to autonomous vehicles.

6.1 Attacker-centric approach

This threat modeling technique is built on attacker perspective by profiling an attacker's characteristics, goals and motivation to exploit vulnerabilities. The modeling technique aims to understand potential attackers who are likely to target the system and type of attacks that might be executed. Based on the understanding, mitigation strategies can be developed against studied attackers.

This approach uses tree diagrams for threat modeling. Examples of key elements of this approach include: the goals of the attacker, the techniques used by the attacker, the system components including software and assets and means of detection and mitigation.

A simple example is an attacker trying to obtain data from a database that is hosted in a backend system accessible from a website. Using this approach, the goal of the attacker would be to obtain the information stored in the database. Techniques include SQL injection (using unsensitized input commands) on the website, and use of automated tools by which exploitation could be carried out.

6.1.1 Intel's TARA (Threat Agent Risk Assessment)

This threat modeling technique focuses on prioritizing the most critical areas of the system and most vulnerable to attacks [32]. This ensures that the threat modeling efficiently address the areas that are exposed the most. Moreover, the framework focuses on threat agents that are most dangerous, what techniques and tradecrafts used to achieved their goals. The most exposed areas are then determined by cross-referencing the information obtained with known vulnerabilities. The process is shown in Figure 5. TARA's goal is to build security strategies with focus on areas with highest level of risk. This is achieved by determining the most likely attack vectors and applying efficient security strategy to address them. TARA relies on three main components:

- Threat Agents Library (TAL): unique library containing 22 threat agent and nine threat agent attributes.
- Common Exposure Library (CEL): list of known vulnerabilities and exposures.
- Methods and Objectives Library (MOL): contains threat agent objectives and methods used to achieve goals and objectives.

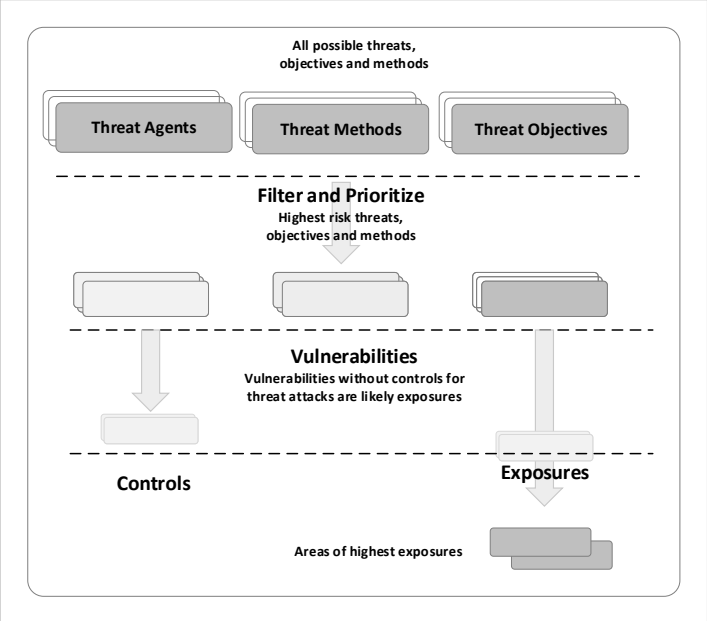


Figure 5: Narrowing down the fields of attacks.

6.1.2 Cyber Kill Chain

The Cyber-Kill threat modeling approach was developed by Lockheed Martin, an American company focusing on global aerospace, defense, security and advanced technologies [33]. The method focuses on the steps taken by an attacker to achieve desired goal. As name implies, the method is chained into multiple phases, each phase has its associated set of goals. Each phase outcomes contribute to the next phase. The cyber kill chain contains the following steps:

- Reconnaissance: research, identification, and selection of targets
- Weaponization: Combining remote access malware with exploit into deliverable payload
- Delivery: Delivering a payload to target (weaponized exploit).
- Exploitation: Exploit vulnerable application or system upon delivery

- Installation: Persistent access via means of installing backdoor in exploited system.
- Command and Control: Outside server communicates with the weapons providing "hands on keyboard access" inside the target's network

6.2 Asset-centric approach

This approach has two main components; assets and threats. The assets taxonomy involves identifying assets, asset groups, asset types, asset details [34]. The threat taxonomy utilizes attack trees, attack graphs and assets' own failure modes in order to prioritize associated risk levels and then mapped to assets. Methodologies based on the asset-centric approach include PASTA, OCTAVE and STRIDE

6.2.1 PASTA

PASTA (Process for Attack Simulation and Threat Analysis) is an asset-centric approach to threat modeling [34]. This methodology was developed by Minded Security and VerSprite. The obvious advantage of this methodology that it providers valuable information about risk reduction and security strategy that could target non-technical staff (example management or stakeholders). The methodology involves seven different stages illustrated below.

- Stage 1: Define Objectives
- Stage 2: Define Technical Scope
- Stage 3: Application Decomposition
- Stage 4: Threat Analysis
- Stage 5: Weaknesses and Vulnerability Analysis
- Stage 6: Attack Modeling and Simulation
- Stage 7: Risk Analysis and Management

The main goals of PASTA methodology are:

- Allows cost and time savings
- Integrates with SDLC
- Better understanding of most likely attack sources
- Better communication between security groups inside the company
- Increases the maturity of the organization in security

6.2.2 OCTAVE

OCTAVE Allegro (Operationally Critical Threat, Asset and Vulnerability Evaluation) a method developed by CERT Survivable Enterprise Management team. This method focuses information assets security in respect to the well-known security requirements, which are Confidentiality, Integrity and Availability (CIA). This method investigates the ways information assets are stored, transferred, used and processed. In addition, it extends to cover threats and vulnerabilities associated with information assets. The OCTAVE Allegro method consists of four different phases, with total of eight steps as shown in Figure 6.

- Phase 1: risk measurement criteria development
- Phase 2: identification and profiling of assets then the identification of containers to the identified assets
- Phase 3: threats to identified containers are identified with respect to assets
- Phase 4: mitigation strategies

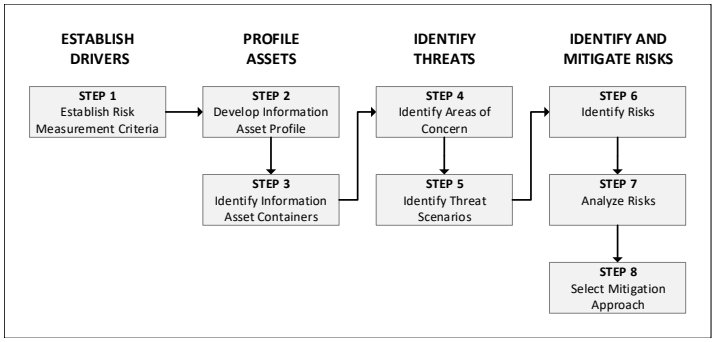


Figure 6: Phases and steps of OCTAVE Allegro.

6.2.3 STRIDE

The acronym STRIDE stands for six categories of security risks: Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privileges.

- Spoofing refers to the act of posing as someone else (i.e. spoofing a user) or claiming a false identity (i.e. spoofing a process). Spoofing threats affect the authenticity property.
- Tampering refers to malicious modification of data in transit, at rest, or within processes. Tampering threats affect the integrity property.
- Repudiation refers to the ability of denying that an action or an event has occurred. Repudiation threats affect the non-repudiation property.
- Information Disclosure refers to data leaks or data breaches. Disclosure threats affect the confidentiality property.
- Denial of Service (DOS) refers to causing a service or a network resource to be unavailable to its intended users. DOS threats affect the availability property
- Elevation of Privileges (EP) refers to gaining access that one should not have. EP threats affect the availability property

By considering these risks for each asset, STRIDE threat model allows to list the applicable risks (also called failure modes) to each asset, and to map them to threats and known attacks. This way measures can be identified to examine and address gaps in the security posture of applications.

7 Experimental Results

This section of the paper discusses and analyzes a proof of concept of two replay attack and packet injection on a simulated CAN bus. In this experiment, we setup a simulation environment to simulate random CAN traffic generation and identifying control signals to perform specific physical actions.

7.1 Simulation Environment Setup

The simulation environment is setup using a virtual CAN bus than communicated over TCP and other tools that interact with the virtual CAN bus. In this section, we described the tools used to set up the simulation environment.

CAN for vehicle communication have different drivers and software utilities with no unified CAN tools and interfaces. However, there is a set of tools with a common interface available in Linux. The SocketCAN is an interface that was created on the Open Source development site BerliOS in 2006. By extending the Berkeley socket API in Linux, SocketCAN can send and receive traffic from the CAN controller in the model of a network device.

SocketCAN is not limited to physical CAN devices and networks, it supports built-in CAN chips and card drivers, external communication via serial or USB, and virtual CAN devices. More importantly, it supports virtual CAN environments.

- Atmel AT91SAM SoCs
- Bosch CC770
- ESD CAN-PCI/331 cards
- Freescale FlexCAN
- Freescale MPC52xx SoCs (MSCAN)
- Intel AN82527
- Microchip MCP251x
- NXP (Philips) SJA1000

Figure 7 compares the implementation of traditional CAN software with the SocketCAN implementation, which is unified. In traditional CAN software, the application has its protocols stack that talks to a character device, whereas SocketCAN is implemented in the Linux kernel with its own CAN protocol family enabling applications to communicate with a CAN bus interface in a generic network interface fashion.

SocketCAN repository contains various utilities for Linux CAN subsystem. These utilities can work with CAN communications within the vehicle from Linux operating system. Examples of the functions supported by the suite:

- Basic tools to display, record, generate can traffic
- Tools to send, generate, sniff CAN frames.
- Serial line discipline (SLC) configuration
- Log files converters
- CAN access via IP sockets
- CAN bus management

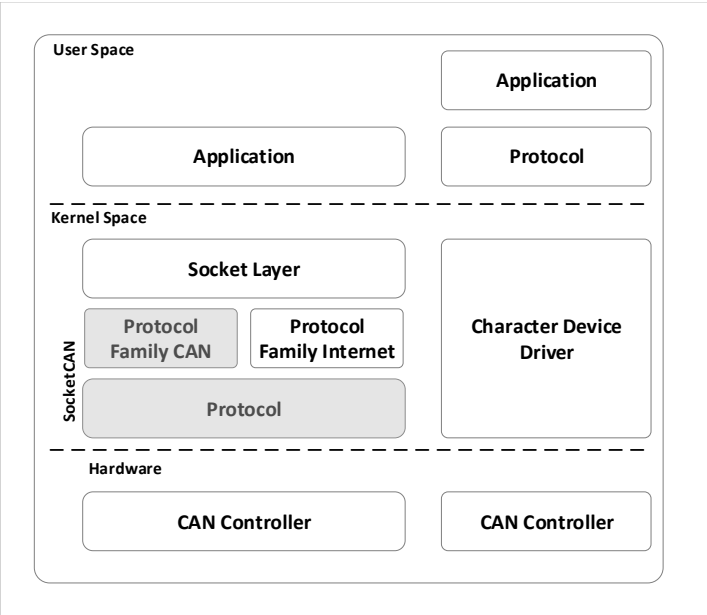


Figure 7: SocketCAN compared to traditional CAN software

Different tools are used in the experiment For CAN bus diagnosis, monitoring, and simulating target vehicle that interacts with the virtual CAN. For CAN diagnostic and monitoring, Kayak application is used. Kayak is a Java-based that can read CAN traffic log files and display information graphically on gadgets.

For target vehicle simulation and CAN traffic fuzzing, the instrument cluster simulator (ICSim) is used, which is an opensource tool developed by Open Garages. The ICSim software is Linux based and relies on virtual CAN devices. ICSim was designed to allow researchers, individuals, developers and others to safely familiarize themselves with CAN reversing. ICSim includes a graphical interface for some of the functions of CAN such as doors, engine RPM, and others.

7.2 Security Attacks on CAN Bus

The following experiment shows how the virtual environment setup can generate, display, and sniff CAN frames. Using the libraries and utilities described earlier in this section, a virtual CAN interface was created in Linux (Debian 4.19-28).

Using the virtual CAN interface is running as “vcan0”, CAN frames can be sent and received using can-utils suite. The current messages are random CAN messages, which are not currently analyzed as per their functions or character device they are trying to communicate with. The first step is generating random CAN messages with default bi trate. The messages generated can be viewed in Figure 8.

Since the messages are communicated over TCP/IP in the virtual environment, messages can be sniffed using different tools such as Wireshark and CANSniffer or Kayak.

```
vcan0  19B   [8]  1F 92 27 58 3A 92 CA 50
vcan0  280   [8]  A6 DD 51 25 FA 32 99 30
vcan0  00B   [8]  11 FD 6A 0B EC B4 3D 5A
vcan0  57A   [8]  40 24 FD 69 F7 30 A7 1A
vcan0  385   [8]  2F E0 CD 65 F5 B4 61 4D
vcan0  244   [8]  37 54 DE 4A 31 A7 9E 5F
vcan0  37D   [8]  5F 38 E1 22 38 AF 26 4B
vcan0  170   [8]  99 8B 43 61 18 FC 0A 5E
```

Figure 8: CAN frames dump using candump of can-utils

7.3 Injecting CAN Frames

This part explains injecting CAN frames into the virtual CAN bus and displays results in the simulator. To identify the CAN message responsible to perform a physical action, the ICSim was used to generate specific actions on the vehicle, and the generated CAN traffic was recorded. Then simply replaying the recorded messages could identify the messages related to the performed actions with other noise that can be discarded for this experiment. The sequence followed in the experiment is explained in Figure 9.

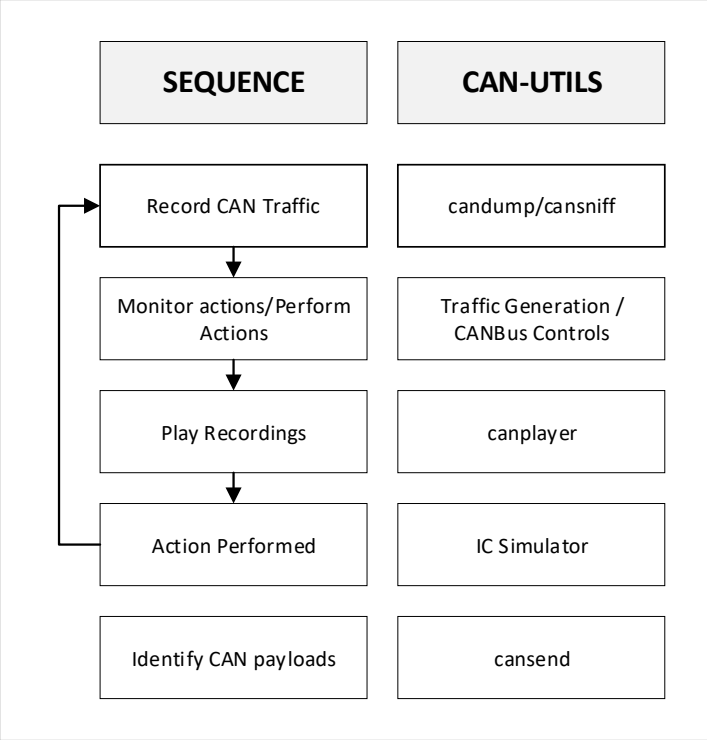


Figure 9: Sequence used to identify CAN messages

Using the generated traffic from the libraries explained earlier, the signals used to perform physical actions were identified. CAN traffic was generated on virtual CAN and recorded in log files. The signals responsible for door locks, RPM and signals identified as shown in Figure 10. At this point, using the CAN ID, we can send specific CAN messages to trigger doors, acceleration or light signals. For example, to accelerate, the CAN ID used is 244 (hex representation).

```
vcan0  517   [8]  09 B4 54 4A F6 4E D9 32
vcan0  188   [8]  B4 59 D4 54 36 DC D3 23
vcan0  0E6   [8]  A3 04 CD 47 E7 19 0C 7B
vcan0  0AF   [8]  8F 04 3A 50 F0 4E D2 18
vcan0  3D6   [8]  FD F6 28 0D 60 F3 28 6D
vcan0  19B   [8]  1F 92 27 58 3A 92 CA 50
vcan0  280   [8]  A6 DD 51 25 FA 32 99 30
vcan0  00B   [8]  11 FD 6A 0B EC B4 3D 5A
vcan0  57A   [8]  40 24 FD 69 F7 30 A7 1A
vcan0  385   [8]  2F E0 CD 65 F5 B4 61 4D
vcan0  244   [8]  37 54 DE 4A 31 A7 9E 5F
vcan0  37D   [8]  5F 38 E1 22 38 AF 26 4B
vcan0  170   [8]  99 8B 43 61 18 FC 0A 5E
```

Figure 10: Specific CAN messages for desired functions

After identifying the CAN messages to perform specific physical actions, two security attacks are performed. The first one is a replay attack. This was achieved by recording CAN traffic in a file. Since CAN communication (in the scenario described in the paper) does not use session IDs or encryption to protect against replay attacks, traffic was recorded were played on a different virtual CAN bus successfully. The second security attack performed is traffic injection. In a virtual CAN bus with random traffic being communicated, specific commands were sent to set RPM to a specific value, open doors, and start turn signals. The results of the signals sent are shown in Figure 11.

The tools used in the experiment do not simulate actual ECUs in terms of behavior. Some ECUs are wired directly and do not accept messages from the CAN bus directly. In such situations, additional effort is required to send signals from allowed ECUs. In diagnostic mode, messages might be accepted from the CAN bus. Additional research is required to fully understand what can be controlled directly from the CAN bus. Aftermarket ECUs expose CAN connectivity and expand the attack surface of modern vehicles. In the experiment, we focused on simple scenarios of packet sniffing and packet injection. The CAN bus is designed to allow real-time communication between ECUs but it is not secured by design. These types of attacks are not limited to aftermarket ECUs or gadgets, any successful exploitation or code execution on ECU connect to CAN bus could achieve similar or higher impact.

In the presence of aftermarket devices, new range of vulnerabilities will be introduced. The inter-vehicle communication was not designed to interact outside the perimeter of the vehicle. With the absence of defined threats, it is extremely difficult to provide assurance on security measures taken.

Starting from a conceptual model for autonomous vehicles is essential to capture different components of autonomous driving ecosystem. Threat modeling could be applied to identify potential threats pertaining autonomous vehicles and support the development of security requirements that cover scenarios discussed in this paper. This should support addressing the knowledge gap in the field of autonomous vehicle security.

Our future plans include the research of a test-based security verification framework for autonomous vehicles. Due to the lack of a centralized tests repository for security testing, we will focus on developing a security test-cases repository based on blockchain technology. In addition, the current research was limited to few

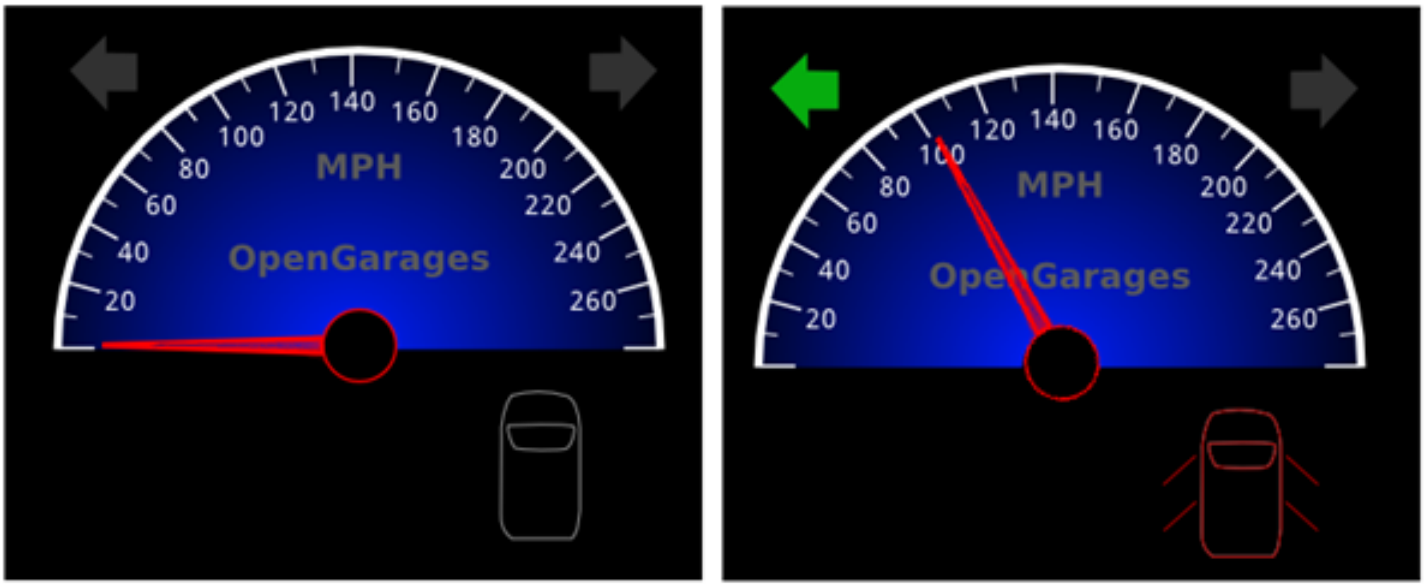


Figure 11: Results of the physical actions displayed on ICS Simulator

scenarios on attacks on CAN bus. We plan to design more attack scenarios against common vulnerabilities in aftermarket ECUs.

8 Conclusion

Several technologies in modern vehicles were designed and built without security in mind. As modern vehicles and future autonomous vehicles are becoming more and more computerized and connected, these technologies present new risks and widen the attack surface.

Threat modeling is an essential step for defining security requirements of autonomous vehicles, or any system in that regard. With the absence of defined threats, it is extremely difficult to provide assurance on security measures taken. The initial work conducted towards the research was focused on establishing a conceptual model for autonomous vehicles. This will support in defining the scope and basing the threat model against it.

The experiment showed briefly how easily CAN traffic can be captured or sniffed. In the presence of aftermarket devices, new range of vulnerabilities will be introduced. The inter-vehicle communication was not designed to interact outside the perimeter of the vehicle. However, these devices open new attack surfaces allowing some components of inter-vehicle communication to be exposed to the outside world.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment This work was supported by the Center for Cyber-Physical Systems, Khalifa University, under Grant 8474000137-RC1-C2PS-T3.

References

[1] A. S. Gajparia, C. J. Mitchell, C. Y. Yeun, "Supporting user privacy in location based services," IEICE transactions on communications, **88**(7), 2837–2847, 2005, doi:10.1093/ietcom/e88-b.7.2837.

[2] D. M. Konidala, C. Y. Yeun, K. Kim, "A secure and privacy enhanced protocol for location-based services in ubiquitous society," in IEEE Global Telecommunications Conference, 2004. GLOBECOM'04., volume 4, 2164–2168, IEEE, 2004, doi:10.1109/GLOCOM.2004.1378393.

[3] A. Talebpour, H. S. Mahmassani, "Influence of connected and autonomous vehicles on traffic flow stability and throughput," Transportation Research Part C: Emerging Technologies, **71**, 143–163, 2016, doi:10.1016/j.trc.2016.07.007.

[4] B. Sheehan, F. Murphy, M. Mullins, C. Ryan, "Connected and autonomous vehicles: A cyber-risk classification framework," Transportation research part A: policy and practice, **124**, 523–536, 2019, doi:10.1016/j.tra.2018.06.033.

[5] A. O. Al Zaabi, C. Y. Yeun, E. Damiani, "Autonomous Vehicle Security: Conceptual Model," in 2019 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific), 1–5, IEEE, 2019, doi:10.1109/ITEC-AP.2019.8903691.

[6] A. M. Wyglinski, X. Huang, T. Padir, L. Lai, T. R. Eisenbarth, K. Venkatasubramanian, "Security of autonomous systems employing embedded computing and sensors," IEEE micro, **33**(1), 80–86, 2013, doi:10.1109/MM.2013.18.

[7] "L. S. P. Revision," 2016.

[8] C. P. Szydlowski, "CAN specification 2.0: Protocol and implementations," Technical report, SAE Technical Paper, 1992, doi:10.4271/921603.

[9] S. C. Talbot, S. Ren, "Comparision of fieldbus systems can, ttcan, flexray and lin in passenger vehicles," in 2009 29th IEEE International Conference on Distributed Computing Systems Workshops, 26–31, IEEE, 2009, doi:10.1109/ICDCSW.2009.15.

[10] M. Wolf, A. Weimerskirch, C. Paar, "Secure in-vehicle communication," in Embedded Security in Cars, 95–109, Springer, 2006, doi:10.1007/3-540-28428-1_6.

[11] L. Buttyán, T. Holczer, I. Vajda, "On the effectiveness of changing pseudonyms to provide location privacy in VANETs," in European Workshop on Security in Ad-hoc and Sensor Networks, 129–141, Springer, 2007, doi:10.1007/978-3-540-73275-4_10.

[12] M. Raya, J.-P. Hubaux, "The security of vehicular ad hoc networks," in Proceedings of the 3rd ACM workshop on Security of ad hoc and sensor networks, 11–21, 2005, doi:10.1145/1102219.1102223.

[13] P. Papadimitratos, L. Buttyan, T. Holczer, E. Schoch, J. Freudiger, M. Raya, Z. Ma, F. Kargl, A. Kung, J.-P. Hubaux, "Secure vehicular communication systems: design and architecture," IEEE Communications Magazine, **46**(11), 100–109, 2008, doi:10.1109/MCOM.2008.4689252.

[14] L. Bariah, D. Shehada, E. Salahat, C. Y. Yeun, "Recent advances in VANET security: a survey," in 2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall), 1–7, IEEE, 2015, doi:10.1109/VTCFall.2015.7391111.

[15] H. Seo, K.-D. Lee, S. Yasukawa, Y. Peng, P. Sartori, "LTE evolution for vehicle-to-everything services," IEEE communications magazine, **54**(6), 22–28, 2016, doi:10.1109/MCOM.2016.7497762.

[16] C. C. Forecast, "Global connected car market to grow threefold within five years," GSMA Connected Living Programme: mAutomotive, 2013.

[17] N. Lu, N. Cheng, N. Zhang, X. Shen, J. W. Mark, "Connected vehicles: Solutions and challenges," IEEE internet of things journal, **1**(4), 289–299, 2014, doi:10.1109/JIOT.2014.2327587.

[18] N. Liu, "Internet of Vehicles: Your next connection," Huawei WinWin, **11**, 23–28, 2011.

[19] A. K. Singh, A. Negi, S. Azad, S. Mudali, "Fuzzy based controller for lidar sensor of an autonomous vehicle," Energy Procedia, **117**, 1160–1164, 2017, doi:10.1016/j.egypro.2017.05.241.

[20] D. Fényes, B. Németh, P. Gáspár, "A predictive control for autonomous vehicles using big data analysis," IFAC-PapersOnLine, **52**(5), 191–196, 2019, doi:10.1016/j.ifacol.2019.09.031.

[21] D. J. Fagnant, K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," Transportation Research Part A: Policy and Practice, **77**, 167–181, 2015, doi:10.1016/j.tra.2015.04.003.

[22] C. Smith, The car hacker's handbook: a guide for the penetration tester, no starch press, 2016, doi:10.4271/1593277032.

[23] I. Foster, K. Koscher, "Exploring controller area networks," login. USENIX Association, **40**(6), 2015.

[24] C. Miller, C. Valasek, "Remote exploitation of an unaltered passenger vehicle," Black Hat USA, **2015**, 91, 2015.

[25] A. Greenberg, "GM took 5 Years to fix a full-takeover hack in millions of Onstar cars," Wired, September, 1059–1028, 2015.

[26] "J. Banter, "Open Thyself! – Security vulnerabilities in BMW's Connected-Drive," C'T Magazine for Computer Technique," 2020.

[27] "A. Charlton, "Hackers disable Corvette brakes by texting dongle meant to lower insurance risk," International Business Times," 2020.

[28] "D. Gilbert, "Tesla Model S hacked: Researchers discover six security flaws in popular electric car," International Business Times," 2020.

[29] "J. Billington, "Hacker takes control of Nissan electric vehicle from other side of the world through Leaf app," International Business Times," 2016.

[30] S. Myagmar, A. J. Lee, W. Yurcik, "Threat modeling as a basis for security requirements," in Symposium on requirements engineering for information security (SREIS), 2005, 1–8, Citeseer, 2005.

[31] S. F. Burns, "Threat modeling: A process to ensure application security," GIAC security essentials certification (GSEC) practical assignment, 2005.

[32] T. Casey, "Threat agent library helps identify information security risks," Intel White Paper, **2**, 2007, doi:10.13140/RG.2.2.30094.46406.

[33] D. Kiwia, A. Dehghantanha, K.-K. R. Choo, J. Slaughter, "A cyber kill chain based taxonomy of banking Trojans for evolutionary computational intelligence," Journal of computational science, **27**, 394–409, 2018, doi:10.1016/j.jocs.2017.10.020.

[34] T. UcedaVelez, M. M. Morana, Risk centric threat modeling, Wiley Online Library, 2015, doi:10.1002/9781118988374.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 865-870 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Low Power Fast Settling Switched Capacitor PTAT Current Reference Circuit for Low Frequency Applications

Muhammed Mansoor C. B.*, Hanumantha Rao G., Rekha S.

Department of Electronics and Communication Engineering, National Institute of Technology Karnataka, Surathkal, 575025, India

ARTICLE INFO

Article history:
Received: 29 August, 2020
Accepted: 06 November, 2020
Online: 27 November, 2020

Keywords:
PTAT
Switched Capacitor
Log-Domain Filter

ABSTRACT

This paper presents a low voltage, low power, fast settling switched capacitor based Proportional to Absolute Temperature (PTAT) current reference circuit. Unlike in a conventional resistor based PTAT current source, the proposed circuit saves a significant amount of silicon area on chip and hence the circuit becomes less susceptible to process variations. It creates a reference current of 1 nA from a 0.5 V power supply at room temperature (27°C). It has PTAT characteristics in the temperatures from −10°C to 80°C. The circuit draws a very low power of 1.5 nW and exhibits a good supply voltage sensitivity of 3.2 %/V. A startup circuit connected to the PTAT source improves the transient response by reducing the settling time. To test the PTAT current reference circuit, a low power log-domain filter which can be used for biomedical applications is realised and biased with the proposed PTAT current source. Results show that the filter cutoff frequency is constant over temperature variations. The CMOS technology used for designing the circuits is UMC 65 nm and tool used for simulations is Cadence Virtuoso.

1 Introduction

Portable and wearable/implantable devices are gaining worldwide acceptance among consumer/biomedical electronics. Circuit designs in such devices have a bottleneck on power dissipation and chip area. Log-domain circuits are one of the best possibilities in exploring low power, high dynamic range analog designs. Log-domain filters and transconductance-capacitor ($G_m - C$) filters can be used in biological data acquisition system which involves low frequency signals like ECG (electrocardiogram) and EEG (electroencephalogram). Since the transistors in these log-domain filters are working in sub-threshold mode of operation (weak-inversion), a low cutoff frequency can be achieved with low values of bias currents (order of *nA*). This paper is an extension of work originally presented in [1].

The cutoff frequency (f_o) of log-domain and $G_m - C$ filters is given by (1) [2, 3].

$$f_o = \frac{G_m}{2\pi C} = \frac{I_o}{2\pi\eta V_T C} \tag{1}$$

where η, I_o and V_T are sub-threshold slope factor, bias current and thermal voltage respectively. Equation (2) gives the expression for thermal voltage.

$$V_T = \frac{kT}{q} \tag{2}$$

where k, T and q have their usual meanings.

Equation (1) shows the direct relation between cutoff frequency and bias current. I_o is usually generated from a current reference circuit. The current reference should be stable against variations in process and supply voltage. Supply voltage sensitivity is the parameter which shows the stability of the current reference against supply voltage variations. Lower supply voltage sensitivity indicates a better current reference. Depending on the mode of operation of transistors in the circuit or the topology of the filter architecture, the current reference can be a temperature independent or Proportional to Absolute Temperature (PTAT) current reference. From (1) and (2), in log-domain and $G_m - C$ filters, it is evident that f_o has an inverse relationship with temperature. To fix this problem, I_o is generated in such a way that it has also a linear relationship with temperature. Thus a PTAT current reference is used to bias log-domain/$G_m - C$ filters. As a cumulative effect, the filter cutoff frequency remains constant over temperature variations.

Many PTAT current reference designs are explored in the literature. Beta-multiplier based circuits discussed in [4]–[5] have adequate PTAT current characteristics. However, to generate low reference currents under low supply voltage for low frequency filters used in biological data acquisition system, larger values of resistors are needed which occupies large area on chip. Moreover the control over temperature coefficient of resistance also becomes

*Corresponding Author: Muhammed Mansoor C. B., Research scholar, Dept. of Electronics and Communication Engineering, National Institute of Technology Karnataka, Surathkal, India, +91 9497368483, mansoorcb@gmail.com

difficult. A comprehensive research has been carried out on the conventional PTAT designs to tackle such shortcomings [6]–[7]. In [7], self-cascode transistors are used instead of resistor. This improves supply variation rejection, but linearity in current versus temperature characteristic has been compromised. In [8], composite transistors are used in the PTAT circuit to enhance the supply voltage sensitivity, but large resistor values are needed to generate smaller currents. In [2, 9], switched capacitors are used in place of resistor. But PTAT current is generated from PTAT voltage using opamp circuits like voltage to current converter which in turn increases the overall power dissipation of the circuit. [1] shows a good switched capacitor based PTAT reference. But the settling time of the reference current is compromised.

Settling time of current reference becomes a significant aspect especially in high speed applications. When the settling time of the reference current increases, the circuit which uses this reference current to generate its bias current also takes a significant amount of time to settle its bias current to steady state value. In case of a filter used in a wireless communication system, it takes a significant time to fix its cutoff frequency which may lead to attenuation of relevant signal information. For example, the circuit presented in [1] requires 8 ms to stabilise its current which means that it takes around 8 ms for the subsequent filter to fix its cutoff frequency which is biased by this reference current source. This work proposes a feasible solution to this problem.

In this paper, a fast settling PTAT current source is proposed. A startup circuit added to the PTAT source improves the settling time of reference current. The proposed circuit exhibits good PTAT characteristics and better stability against power supply fluctuations. This can be used to bias low frequency filters in biomedical applications. A low power log-domain filter is also designed and biased using the proposed PTAT current source. Results show that the cutoff frequency of the filter remains constant irrespective of temperature variations. This paper is organized as follows. Details of the proposed PTAT circuit along with the results are given in Section 2. Log-domain filter realisation and filter response are detailed in Section 3. Finally conclusions are drawn in Section 4.

2 Proposed Improved PTAT Reference

A conventional beta-multiplier based current source circuit [5] is shown in Figure 1. Depending on the mode of operation of transistors, it can be a constant-G_m biasing circuit [5] or a PTAT current source. It acts as former if all the transistors work in strong-inversion saturation region ($V_{GS} > V_{th}$ and $V_{DS} \geq V_{GS} - V_{th}$) and as latter if all the transistors work in weak-inversion saturation ($V_{GS} < V_{th}$ and $V_{DS} \geq 3V_T$).

When transistor operates in weak-inversion region, its drain current [10] can be expressed as following.

$$I_D = \frac{W}{L} I_{D0} e^{\frac{V_{GS}-V_{th}}{\eta V_T}} (1 - e^{\frac{-V_{DS}}{V_T}}) \tag{3}$$

where $I_{D0} = \mu_n C_{ox} (\eta - 1) {V_T}^2$; μ_n and C_{ox} are electron mobility and gate oxide capacitance per unit area respectively. Other terms have their usual meanings. For $V_{DS} \geq 3V_T$ (weak-inversion saturation region), (3) can be approximated to (4).

$$I_D = \frac{W}{L} I_{D0} e^{\frac{V_{GS}-V_{th}}{\eta V_T}} \tag{4}$$

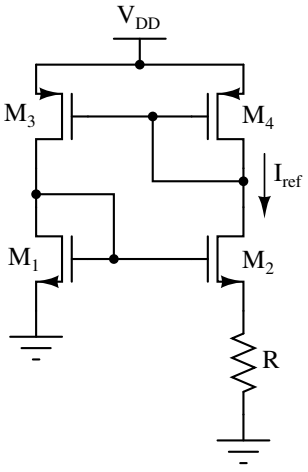


Figure 1: Beta-multiplier circuit

From Figure 1, the reference current I_{ref} which is also the drain current of M_2 (I_{D2}) can be written as (5).

$$I_{ref} = \frac{V_R}{R} \tag{5}$$

Voltage across resistor R (V_R) can be given as (6).

$$V_R = V_{GS1} - V_{GS2} \tag{6}$$

Using (4), gate-source voltages of M_1 and M_2 can be expressed as (7) and (8) respectively.

$$V_{GS1} = \eta V_T ln(\frac{I_{D1}}{I_{D0}(\frac{W}{L})_1}) + V_{th1} \tag{7}$$

$$V_{GS2} = \eta V_T ln(\frac{I_{D2}}{I_{D0}(\frac{W}{L})_2}) + {V_{th2}}' \tag{8}$$

${V_{th2}}'$ can be written as (9) [11] to include the body effect of M_2.

$${V_{th2}}' = V_{th2} + (\eta - 1) V_{SB} = V_{th2} + (\eta - 1)(V_R) \tag{9}$$

By substituting (7) and (8) in (6) and by assuming $V_{th1} = V_{th2}$, V_R can be derived as following.

$$V_R = I_{ref}.R = V_T ln(p) \tag{10}$$

where

$$p = \frac{(W/L)_2}{(W/L)_1}$$

Thus, I_{ref} can be expressed as (11).

$$I_{ref} = \frac{V_T ln(p)}{R} \tag{11}$$

Assuming R is thermally stable, I_{ref} is directly proportional to temperature thereby giving PTAT characteristics. For lower values of currents (order of nAs), large values of resistors are needed. For example, to generate a reference current of 1 nA using circuit shown in Figure 1, a 10 MΩ resistor is required if p is chosen as 1.5. To accommodate such large resistors on chip, huge area is needed

which makes the circuit more process dependent, hence reducing the accuracy of I_{ref}. In a typical SoC (System-on-Chip), where digital circuits and analog circuits co-exist together, a stable and precise clock will be available. In such systems, the resistor R in Figure 1 can be replaced by a switched capacitor driven by this clock which significantly enhances accuracy of I_{ref}.

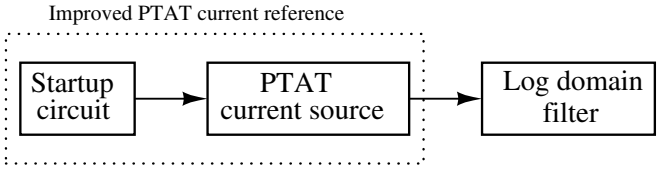


Figure 2: Block diagram

The entire block diagram of the circuit setup explained in this paper is shown in Figure 2. The startup circuit is added to the PTAT current generator to reduce the settling time of I_{ref} thereby improving the transient response.

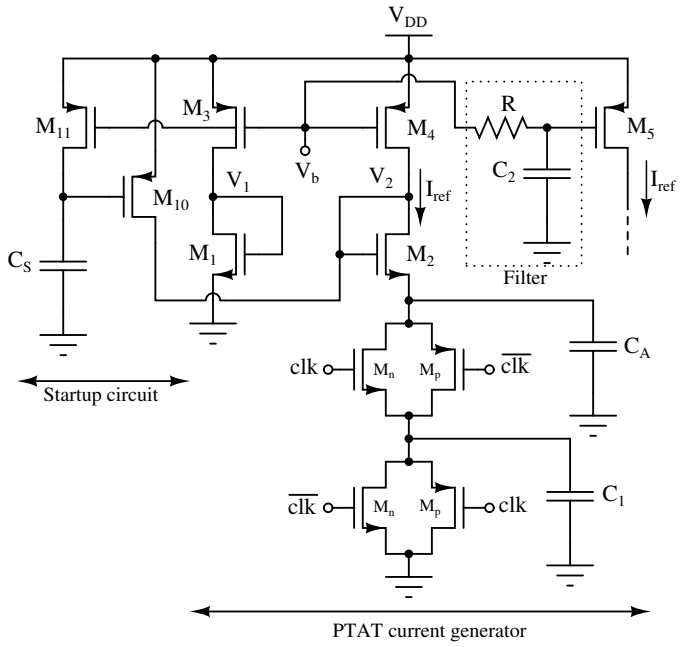


Figure 3: Proposed improved PTAT current reference

The transistor level implementation of PTAT current reference along with startup arrangement is shown in Figure 3. In the proposed circuit, every transistor is working in weak-inversion region. The value of p is chosen as 1.5 which means that the size of M_2 is 1.5 times the size of M_1. A switched capacitor comprising of capacitor C_1 and transmission gate switches (M_n and M_p) is used in place of resistor. M_p is selected 4 times larger than M_n in the transmission gate so that its on resistance does not depend on the voltage across it. Two clocks ('clk' and '$\overline{clk}$') which are complementary to each other and having frequency $\frac{f_{clk}}{}$ are used for the switching operation. When 'clk' becomes high ('$\overline{clk}$' becomes low), capacitor C_1 charges and when 'clk' becomes low ('$\overline{clk}$' becomes high), it discharges. The average resistance (R_{eq}) of the switched capacitor is given as in (12) [5].

$$R_{eq} = \frac{1}{f_{clk} C_1} \tag{12}$$

By substituting R_{eq} (12) in place of R in (11), we get (13).

$$I_{ref} = V_T f_{clk} C_1 ln(p) \tag{13}$$

From (13) it can be seen that, I_{ref} is directly proportional to C_1. If f_{clk} and p are chosen properly, smaller capacitor values can be obtained for the currents in the order of nA.

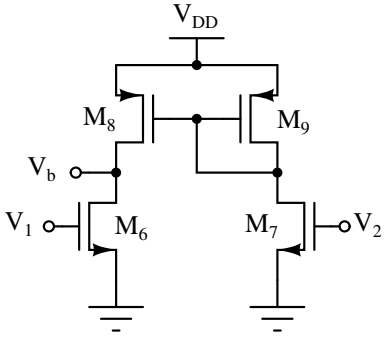


Figure 4: Error amplifier

The purpose of error amplifier which is shown in Figure 4 is to make the drain voltages of M_1 and M_2 equal thereby making currents through them equal. This kind of biasing minimises the dependency of I_{ref} on V_{DD} which improves supply voltage sensitivity.

It takes around *8 ms* for the reference current to settle to the steady state value of *1 nA* in the circuit presented in [1]. To reduce this delay, a startup arrangement comprising of M_{10}, M_{11} and C_S is added to the circuit. A similar kind of startup circuit is discussed in [12]. In Figure 3, when the circuit is switched on and before it reaches the steady state, initially the gates of $M_{1,2}$ and $M_{3,4}$ are at zero volt and V_{DD} respectively. Assuming zero initial voltage across the capacitor C_S, M_{10} turns on and pulls up the gates of $M_{1,2}$ to V_{DD}. $M_{1,2}$ start conducting and I_{ref} rises. Then, potential at the gates of $M_{3,4}$ drops and M_{11} turns on. C_S starts to charge towards V_{DD} and potential at the gate of M_{10} increases and once the circuit reaches steady state, M_{10} turns off. M_{10} and M_{11} do not have a static current path from V_{DD} to ground. Hence startup circuit dissipates zero static power.

Value of capacitor C_S and sizes of M_{10} and M_{11} determine how fast startup circuit can initialise the main circuitry and how fast it can be disconnected once main circuitry reaches the steady state. Transistors M_{10} and M_{11} operate in linear region and the linear on resistance of these transistors are given by (14).

$$R_{on} = \frac{1}{\mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{th})} \tag{14}$$

How fast startup circuit is switched on depends on how fast transistor M_{10} turns on. The on time of M_{10} is limited by the input capacitance of M_{10} which comprises of gate to source and gate to drain capacitances. By properly choosing the size of M_{10}, the on time can be optimised.

When it comes to disconnection of startup circuitry, assuming M_{11} as a linear resistance given by (14), the RC time constant (τ) of the branch consisting of M_{11} and C_S is given by (15).

$$\tau = \frac{C_S}{\mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{th})} \tag{15}$$

When capacitor C_S charges above the voltage $(V_{DD} - V_{th})$, M_{10} turns off. $(V_{DD} - V_{th})$ is approximately 150 mV in this design and time for C_S to charge to this voltage (30 % of final value) from 0.5 V supply through M_{11} is nearly 0.4τ.

A low pass filter which comprises of R and C_2 is added to maintain the ripple content in the output reference current within the accepted level. Capacitor C_2 is chosen as 0.1 pF and resistor R is tuned to obtain the ripple content in I_{ref} less than *1 %*. To remove high frequency noise during switching operation, capacitor C_A is used at the source of M_2. A 100 kHz clock is selected for this work. The capacitors, C_1, C_A, C_S are taken as 0.8 pF, 10 pF, 1 pF respectively. By substituting (13) in (1), f_o can be expressed as (16).

$$f_o = \frac{f_{clk} C_1}{2\pi\eta C} ln(p) \tag{16}$$

From (16), it can be seen that f_o is independent of temperature.

2.1 Simulation results

The proposed circuit is designed using UMC 65 nm CMOS process with a supply voltage of 0.5 V. Periodic Steady State (PSS) analysis is carried out to plot the reference current. Figure 5 shows the output current generated (1 nA) for one clock period (10 μs) at room temperature ($27°C$).

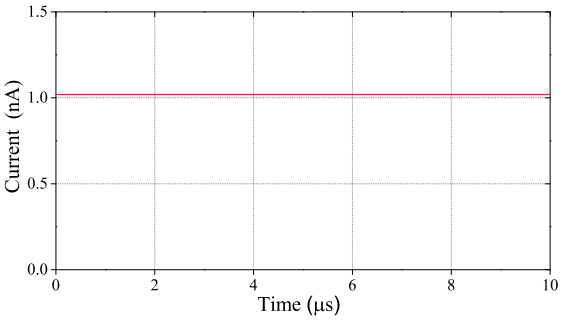


Figure 5: I_{ref} at V_{DD}= 0.5 V

Figure 6 shows I_{ref} variations when temperature is varied from $-10°C$ to $80°C$ for V_{DD} = 0.5 V. It can be observed that I_{ref} changes linearly with temperature and thus has a good PTAT characteristics over the range of temperature.

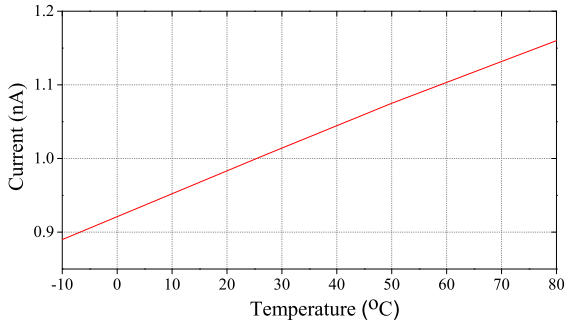


Figure 6: I_{ref} vs. Temperature plot

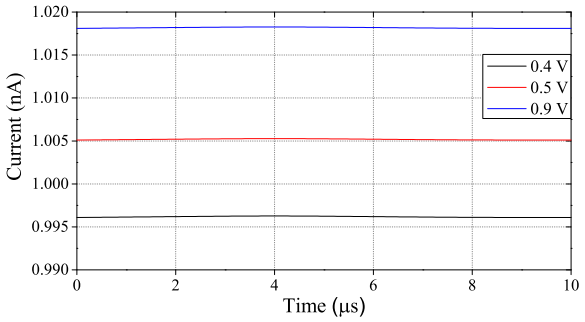


Figure 7: I_{ref} for different supply voltages

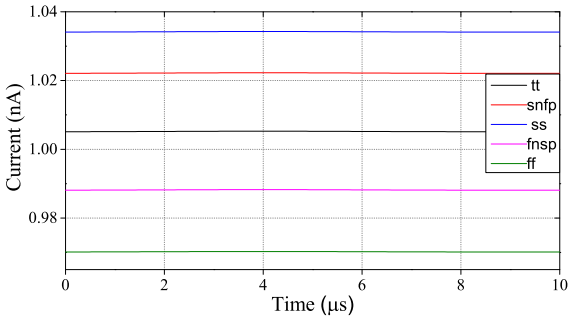


Figure 8: I_{ref} across the process corners

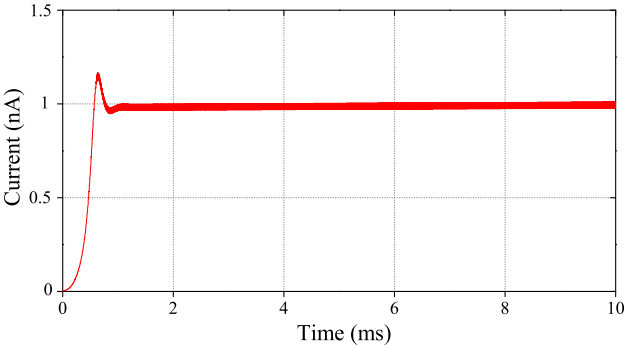


Figure 9: Transient response of the current flowing through M_2

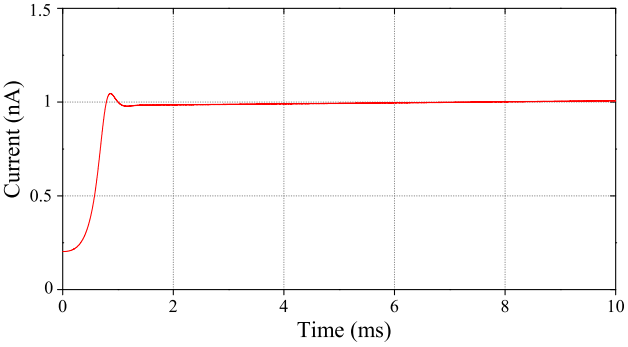


Figure 10: Transient response of the current flowing through M_5

I_{ref} is plotted for different values of V_{DD} at room temperature as shown in Figure 7. I_{ref} is *0.996 nA* and *1.018 nA* for V_{DD} of 0.4 V and 0.9 V respectively. The supply voltage sensitivity of I_{ref} is found to be *3.23 %/V* by considering these extreme values.

Figure 8 plots I_{ref} for various process corners at room temperature and V_{DD} = 0.5 V. I_{ref} deviates maximum in *ss* and *ff* corners (*1.034 nA* and *0.968 nA*). The maximum deviation is less than ±*3.6 %*.

The current flowing through the source of M_2 is shown in Figure 9. It has ripples having peak to peak magnitude of approximately *0.1 nA*. The ripple free current flowing through M_5 is shown in Figure 10. The settling time is found to be approximately *1 ms*.

Table 1: Comparison of the proposed work with other similar works

	[7]	[13]	[14]	[1]	This work
Technology (nm)	1500	180	40	65	65
Supply voltage (V)	1.2	1.2	1	0.5	0.5
Reference current (nA)	0.4	6	100	1	1
Settling time (ms)	-	-	-	8	1
Power (nW)	2	-	350	1.5	1.5
Supply voltage sensitivity (%/V)	6	6.47	2.9	2.7	3.2

The improved PTAT reference consumes a power of *1.5 nW* at room temperature from a supply voltage of 0.5 V. The proposed circuit along with existing PTAT current references in the literature are listed in Table 1 and a fair comparison is done. The proposed PTAT circuit functions under the lowest supply voltage and has a very low supply voltage sensitivity and settles faster than [1] (nearly 87 % improvement in settling time).

3 Log-Domain Filter Realization

The proposed PTAT current source shows excellent PTAT characteristics. To verify this, a low frequency log-domain filter is designed and biased with the proposed switched capacitor based PTAT current reference. This log-domain filter can be used in low power biomedical applications like processing of ECG, EEG signals. The expectation is that the cutoff frequency of the filter remains the same irrespective of the variations in temperature.
The transfer function of a first order low-pass filter is

$$H(s) = \frac{Y(s)}{X(s)} = \frac{\omega_o}{s + \omega_o} \tag{17}$$

where ω_o is the cutoff frequency of the filter. The corresponding time domain differential equation is given by

$$\frac{dy(t)}{dt} + \omega_o y(t) = \omega_o x(t) \tag{18}$$

where x(t) and y(t) represent the input and output currents of the filter respectively. By solving the differential equations and relating the current-voltage in transistor operating in subthreshold operation, we get

$$I_K I_{out} = I_{dc} I_{in} \tag{19}$$

where I_{in} contains both dc current I_{dc} and input current I_{signal}. I_K is the current flowing through M_3 which contains both capacitor current I_C and I_{dc}. I_{out} is the output current. Equation (19) can be realized using a translinear loop having four transistors. These equations can be implemented in the form of a translinear circuit as shown in Figure 11 [15].

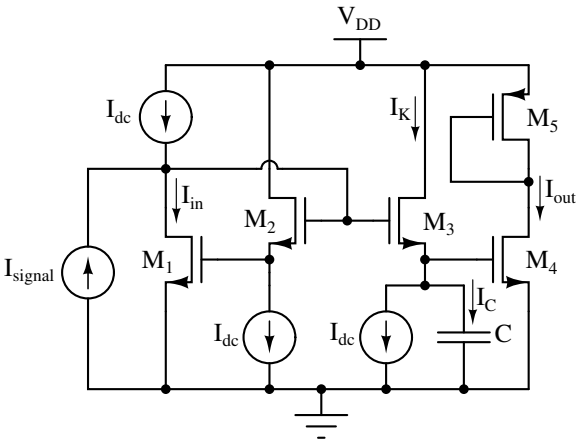


Figure 11: First order log-domain filter

In this circuit, the MOSFETs M_1–M_4 form the translinear loop. These transistors are working in sub threshold saturation region. The cutoff frequency of the filter can be tuned by varying the DC current and is given by following [16].

$$f_o = \frac{I_{dc}}{2\pi\eta V_T C} \tag{20}$$

The ideal DC current sources shown in Figure 11 can be replaced by current sources implemented using MOSFETs. In this work I_{dc} is taken as I_{ref} (1 nA). The cutoff frequency of 100 Hz is obtained by choosing C=37 pF.

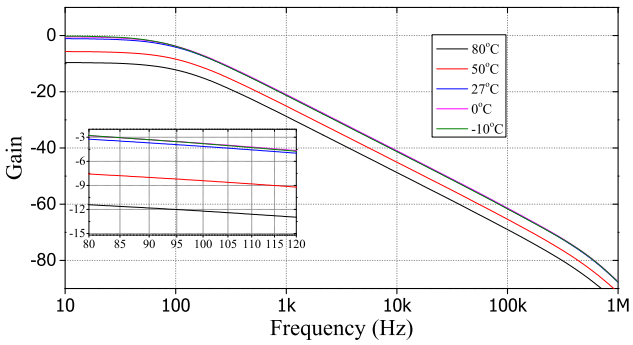


Figure 12: AC response of the filter biased with temperature independent current reference

3.1 Simulation results

The filter is designed using UMC 65 nm CMOS technology with a supply voltage of 0.5 V. For a comparison, I_{dc} is generated from a temperature independent current reference and ac response of the filter is plotted. Figure 12 shows the variation of cutoff frequency

in the temperatures ranging from $-10°C$ to $80°C$. The maximum deviation is found to be *16 %*. To verify the PTAT characteristics, I_{dc} is generated from the proposed PTAT source and ac response is plotted. Figure 13 shows that deviation of cutoff frequency is negligible and maximum deviation is found to be less than *2 %*. The band transitions are enlarged in corresponding inset graphs. Table 2 shows the deviation of filter cutoff frequency for both temperature independent and the proposed PTAT references.

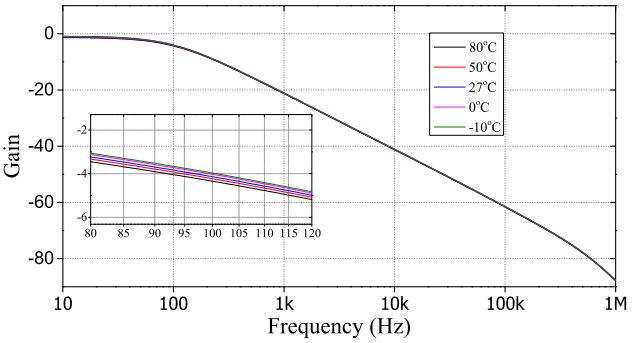


Figure 13: AC response of the filter biased with PTAT current reference

Table 2: Cutoff frequency deviation

Temperature (°C)	Cutoff frequency (Hz)	
	Temp. independent	PTAT
-10	116	101.9
0	107	101.2
27	100	100
50	93	99
80	88	98.4

Table 3 lists the parameters of the filter designed in this work.

Table 3: Parameters of the filter

Technology	65 nm CMOS
Supply voltage	0.5 V
Power	2.5 nW
Cutoff frequency	100 Hz
DC gain (at $27°C$)	0.94
I_{in_pp} at 25 Hz for 1% THD	0.82 nA
Input referred noise from 1 Hz to 100 Hz	3.26 pA_{rms}
Dynamic range	48 dB

4 Conclusion

This paper proposes a low voltage, low power, fast settling switched capacitor PTAT current reference circuit. Settling time is reduced with the aid of a startup circuit. The PTAT circuit exhibits good PTAT characteristics in the temperatures ranging from $-10°C$ to $80°C$ while operating under a very low power. The circuit is less susceptible to power supply fluctuations and can work at supply voltages as low as 0.4 V. To test the PTAT source, a low power log-domain low frequency filter is also designed and biased with PTAT reference. Simulation results imply that the filter cutoff frequency remains unchanged irrespective of temperature variations and this validates the theory explained in the paper.

References

[1] H. G. Rao, S. Rekha, "A 0.5 V, 1 nA Switched Capacitor PTAT Current Reference Circuit," in 2019 IEEE Conference on Modeling of Systems Circuits and Devices (MOS-AK India), 51–54, 2019, doi:http://idr.nitk.ac.in/jspui/handle/123456789/7046.

[2] A. Gerosa, A. Maniero, A. Neviani, "A fully integrated dual-channel log-domain programmable preamplifier and filter for an implantable cardiac pacemaker," IEEE Transactions on Circuits and Systems I: Regular Papers, **51**(10), 1916–1925, 2004, doi:https://doi.org/10.1109/TCSI.2004.835027.

[3] R. Schaumann, H. Xiao, V. V. Mac, Design of Analog Filters 2nd Edition, Oxford University Press, Inc., 2009.

[4] E. Vittoz, J. Fellrath, "CMOS analog integrated circuits based on weak inversion operations," IEEE journal of solid-state circuits, **12**(3), 224–231, 1977, doi:https://doi.org/10.1109/JSSC.1977.1050882.

[5] B. Razavi, Design of analog CMOS integrated circuits, Tata McGraw-Hill, 2002.

[6] L. Magnelli, F. Crupi, P. Corsonello, C. Pace, G. Iannaccone, "A 2.6 nW, 0.45 V temperature-compensated subthreshold CMOS voltage reference," IEEE Journal of Solid-State Circuits, **46**(2), 465–474, 2011, doi: https://doi.org/10.1109/JSSC.2010.2092997.

[7] E. M. Camacho-Galeano, C. Galup-Montoro, M. C. Schneider, "A 2-nW 1.1-V self-biased current reference in CMOS technology," IEEE Transactions on Circuits and Systems II: Express Briefs, **52**(2), 61–65, 2005, doi: 10.1016/j.trc.2016.07.007.

[8] H. G. Rao, C. B. M. Mansoor, S. Rekha, "A 0.8 V, 5 nA PTAT current reference circuit with improved supply voltage sensitivity," in 2019 Global Conference for Advancement in Technology (GCAT), 1–4, 2019.

[9] S. Malik, M. Schlarmann, R. Geiger, "A low temperature sensitivity switched-capacitor current reference," in Proceedings of the European Conference on Circuit Theory and Design (ECCTD), 269–272, Citeseer, 2001.

[10] Y. Tsividis, C. McAndrew, Operation and Modeling of the MOS Transistor, Oxford Univ. Press, 2011.

[11] L. H. Ferreira, T. C. Pimenta, R. L. Moreno, "An ultra-low-voltage ultra-low-power weak inversion composite MOS transistor: Concept and applications," IEICE transactions on electronics, **91**(4), 662–665, 2008, doi: https://doi.org/10.1093/ietele/e91-c.4.662.

[12] S. Kumar, H. G. Rao, S. Rekha, "A 1-V, 8.6-nA Resistor-less PTAT Current Reference with Startup Circuit," in 2018 15th IEEE India Council International Conference (INDICON), 1–4, 2018.

[13] Y. Osaki, T. Hirose, N. Kuroki, M. Numa, "1.2-V supply, 100-nW, 1.09-V bandgap and 0.7-V supply, 52.5-nW, 0.55-V subbandgap reference circuits for nanowatt CMOS LSIs," IEEE Journal of Solid-State Circuits, **48**(6), 1530–1538, 2013, doi:https://doi.org/10.1109/JSSC.2013.2252523.

[14] V. Jorge, A. L. Aita, "A 1-V PTAT current reference circuit with 0.05%/V current sensitivity to VDD," in Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS), 502–505, 2016.

[15] A. Patil, C. B. M. Mansoor, H. G. Rao, S. Rekha, "Low power log-domain filter," in 2019 IEEE 16th India Council International Conference (INDICON), 1–4, 2019.

[16] J. Mulder, W. A. Serdijn, A. C. van der Woerd, A. H. van Roermund, Dynamic translinear and log-domain circuits: analysis and synthesis, volume 481, Springer Science & Business Media, 2012.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 871-882 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Blockchain Application in Higher Education Diploma Management and Results Analysis

Fernando Richter Vidal*, Feliz Gouveia, Christophe Soares

Faculty of Science and Technology, University Fernando Pessoa, Porto, Portugal

ARTICLE INFO

Article history:
Received: 28 August, 2020
Accepted: 14 November, 2020
Online: 08 December, 2020

Keywords:
Blockchain
Blockchain Revocation
Higher Education
Digital Diplomas
Academic records
Blockcerts
Electronic document

ABSTRACT

Academic certifications are achievements desired by people, because they have a direct impact, positively, on their social lives. Such an important document, still widely issued in paper format, may be subject to forgery or impossibility of verification due to the unavailability of the issuing entity. This work consists of identifying, analyzing and testing some of the blockchain-based tools that are emerging, to offer more efficiency, reliability and independent degrees. A concept proof is presented, through the implementation of a prototype capable of issuing, verifying and sharing certificates. The results of this experiment are presented, analyzing the use of blockchain technology for this purpose. Finally, the work presents an overview of the current state of development and maturity in which these tools are found, reporting the advances and limitations, and exposing issues that still need to be resolved.

1 Introduction

This work is an extension of the work originally presented in 2019 at the international conference "Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)" [1], which aims to present an analysis of the use of blockchain technology in the education area, for issuing and verifying academic degrees in higher education.

Academic certificates attest to the certificate holder's abilities and skills and are accepted internationally [2]. These qualifications have a major impact on income and social position, both in emerging countries and in developed countries.

For example, in Brazil, data from *Pesquisa Nacional por Amostra de Domicilios Continuas* (PNADC) showed that the level of education is decisive for the Brazilian's income [3]. According to the results presented, those who complete higher education achieve almost triple the remuneration compared to those who only have a high school education.

In Europe, according to Eurostat [1], every year higher education graduates in Europe exceed 4.5 million, in which France and the United Kingdom are the leading countries with over 740,000 graduates per year. The fraction of people with university degrees between 30 and 34 years old almost doubled in fourteen years, going from 23 % in 2002, to 39 % in 2016 [4].

As professionals become qualified, knowledge is directly transformed into income, and consequently in a better quality of life. Furthermore, these numbers confirm a continuous increase and constitute a solid basis that justifies the creation of solutions to verify the authenticity of university degrees.

The weakness of the paper model, which is still widely used, was even more evident during the crisis caused by COVID-19, in which the use of digital resources became indispensable.

Although digital initiatives have emerged to address these weaknesses [5], these solutions still depend on the issuing entity, concerning authenticity verification. Fortunately, blockchain promises to offer, besides this innovation, other improvements not yet achieved by digital solutions. Using the resources that the technology offers, it is possible to make the check process disconnected from the issuer and still guarantee authenticity.

The second innovation is related to privacy. Certificates are documents that contain personal information and are sensitive to data leakage. If, on the one hand, student privacy needs to be preserved, on the other hand, its distribution benefits those interested [6].

*Fernando Richter Vidal, Sao Joao da Boa Vista-SP Brazil,37705@ufp.edu.pt
[1] `http://ec.europa.eu/eurostat`

Finally, blockchain solutions offer the third innovation compared to digital certificates, about timestamp. Due to the information being immutable and stored chronologically within the network, they accurately express the dates on which the events occurred.

This paper discusses the use of blockchain technology, aiming to benefit academic certificates with these possible innovations. The document is organized as follows: the next section provides a brief overview of blockchain technology; next, we describe the main differences between the blockchain solution and digital certificates; right after, we present the CertEdu prototype developed by *University Fernando Pessoa* (UFP), as well as a brief discussion of the results obtained and we conclude with some final remarks.

2 Background

2.1 Blockchain

Blockchain is a distributed system, formed by records that are organized in blocks and are linked to each other through cryptographic mechanisms. The technology became well known through the cryptocurrency Bitcoin [7], and soon expanded to several other applications.

There are three types of blockchain: public, private or consortium [8]. Public networks, also called permissionless blockchains, provides a free access environment to any participant who wants to join the network, however, the transaction validation rules are predefined and cannot be changed by any member. Typically, this type of blockchain implements the *Proof of Work* (PoW) consensus mechanism. In private networks, also called permissioned blockchain, the rules are defined according to the business interest. The organization that controls the network can define, for example, which users will decide on consensus or how to manage the network to accept new members. Finally, the consortium network is a category that merges properties from the public and private networks in the same environment. For example, in some scenarios, it would be interesting to keep the access public to the network, but also allow to make some data encrypted to preserve privacy and the anonymity of a participant.

Due to the consensus mechanisms and their *Peer-to-Peer* (P2P) topology, the network has gained strong protection against tampering. The success of an attempt of violation is conditioned to changes in all other previous blocks. Besides, attacks that would be able to take control of the network, such as a 51 % attack, would be extremely costly and could cost USD 500K per hour of processing [9]. In this aspect, the bigger the network, the more secure it becomes, which explains why Bitcoin is more secure when compared to other smaller networks [10]. This blockchain features results in an important property, which is immutability.

A block is always connected to a previous block, except for the genesis block, forming, then, a sequential chronological chain. The maximum size that a block can reach, depends on how each blockchain platform implements it. On Bitcoin, this value cannot exceed 1 Mb [11], while, on Ethereum there is no fixed limit, and the size change according to the number of gas units that can be spent per block (block gas limit).

As shown in figure 1, the block is organized in two areas: header and content. The header stores eighty bytes of control information. We find information like the hash header and the previous block; a software version, the target of the difficulty, a nonce, the root of Merkle [12], and a timestamp. Regarding the content area, they are stored as records. On average, a content area can store over five hundred transactions [5].

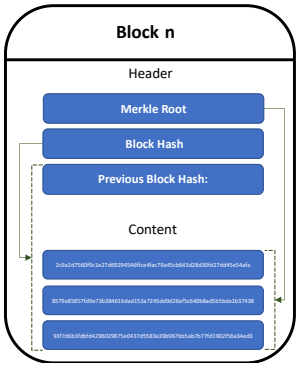


Figure 1: Overview of a Block Structure

Transactions are used to transfer an asset from one account to another account. The owner of an asset can move it inside the blockchain, through an account identified by a public address, which is controlled by its corresponding private key [13].

There is a conceptual difference between the public key and public address. The public key is used to verify the signatures generated by the private key's owner, while the address serves to identify the account. Using the example of Bitcoin, the process of generating the address is as follows: the public key is obtained by applying the function *Elliptic Curve Digital Signature Algorithm* (ECDSA) over the private key; the public key is subjected to two functions, *Secure Hash Algorithm* (SHA)-256, and *Race Integrity Primitives Evaluation Message Digest* (RIPEMD)-160; the resulting string size is reduced by passing the function *Base58Chech*, resulting in an address like *mgAzKQZZi*47*g*4*UMvmGJCsicbJ*4*P*3*B*8*SHRr* [14].

In the next sections, will be discussed how academic certificates can be used as assets in the blockchain. Through the CertEdu application implemented by UFP, questions such as authenticity will be analyzed, as well as universities and students can have their identities verified. Finally, the results obtained by systematic tests are analyzed.

2.2 Certification

Certification involves three processes: issuing, verification, and sharing [15]. This paper evaluates the application of blockchain in each one of the processes, mainly in the verification and sharing. The digital certificates innovate compared to the paper model but do not guarantee verification and sharing because they depend on central points. As Schär notes, academic certificates are useful, only if they can be verified [6].

The disruptive technology of the blockchain allows creating a structure capable of making the verification process independent [16]. The figure 2 presents a typical scenario, in which the university, the student, and the employer are involved. Note that Jane's certificates can be verified directly on the blockchain, without contacting the university.

Table 1: Comparative between digital certificates vs. blockchain solutions

Properties	Digital Certificate	Blockchain Certificate
Reliability	Relies only in the digital signature	Several cryptography mechanisms are used
Privacy	All information are available in the certificate	Only the hash is public
Autonomy	Depend on central regulatory bodies	Technology eliminates the need for third parties
Data Loss	Depend on backup mechanisms	Standard normal distribution
Proof of existence	Dates relies on the suitability of the subscriber	Timestamp represent the date of the facts

Table 2: Comparison between the tools analyzed

Criteria	BTCert	Hyperledger	EduCTX	Blockcerts
Blockchain agnostic	no	no	no	yes
Self-sovereign identity	yes	yes [2]	yes	yes
Community with active user	no	yes	no	yes
Public network	yes	no	no	yes
Privacy concerns	yes	yes	yes	yes

The public information available on the blockchain says nothing about the student. This occurs thanks to the use of the hash, which records the certificate in a unidirectional way on the network, not allowing to retrieve any personal information from Jane. As a result, Jane has the autonomy to share the digital certificate file, only with whomever she wants, and everyone who receives the file will also be able to check it independently on the blockchain.

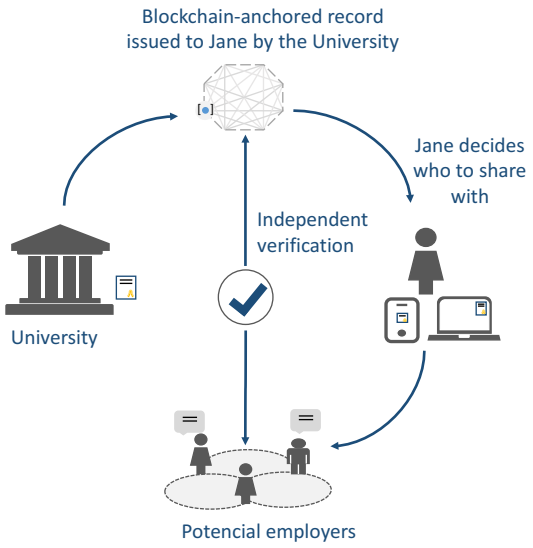


Figure 2: Diploma verification [1]

The table 1 presents a comparison between the properties of the digital certificate solution and the blockchain certificate solution. The blockchain innovation is realized in several aspects. The technology is more secure because unlike the digital certificate, in which all security depends relies only on the digital signature, the use of different cryptographic mechanisms, combined with the adoption of a distributed ledger, offers certificates a higher level of security.

Regarding privacy [17], digital certificates are much more sensitive to data leakage, considering that all personal information is contained in the certificate. This fact can restrict its replication. Analyzing this property in the blockchain solution, the distributed information says nothing about the student, and the certificate can be published without worries.

Digital signatures depend on central regulatory bodies. In some countries, there is not even an authority capable of certifying a signature [15]. In this aspect, blockchain offers complete autonomy, eliminating the need for third parties.

The blockchain, mainly in public architectures and consortium [8], offers a genuine backup mechanism since all information is replicated in pairs. Digital certificates, on the other hand, are easy to destroy electronically and depend on sophisticated backup mechanisms to avoid disappearing [15].

Finally, about the proof of existence, the reliability of the dates generated by digital certificates relies on the reliability of the subscriber. This can be a problem when a university has its private key stolen, and an attacker uses it to sign valid certificates with dates retroactive to the reported date of the theft. In this regard, the blockchain allows identifying the timestamp. As Ronning says, ”every credential issued with a stolen key must fail” [18].

3 CertEdu

3.1 Frameworks analyzed

The CertEdu project begins with the choice of the tool which will be used in the development. In 2018, when the project started, there were not many tools available for this purpose. The use of blockchain in the issue of academic certificate management is re-

[2] Shout combine Hyperledger Fabric + Hyperledger Indy

cent and the first students to receive their diplomas anchored in the blockchain, were from *Massachusetts Institute of Technology* (MIT), in 2017 [19].

We identified four potential platforms. The table 2 shows the comparison between them, and the used evaluated criteria.

Between the evaluated solutions, the Blockcerts tool was chosen to implement the prototype of this paper. The reason was that it better serves the requirements placed, especially about working on any blockchain. In reality, Blockcerts was the only certificate issuing solution that was born with the requirement to work for any blockchain. This requirement greatly increases the complexity of the solution, but it is important because it keeps the application life cycle long. Another point that drew attention is its active community of developers, which makes the project receive constant updates.

Interesting points were also noted on other platforms.

BTCerts, a project inspired by the MIT solution, address to the problem of centralizing the revocation process of Blockcerts. The model proposed by BTCerts solves the issue and can be easily adapted to any type of blockchain, but the costs are concernedly, mainly because it does not explain how the complementary revocation information would be registered, since the OP_RETURN_DATA field has a limited size of 83 bytes [20].

EduCTX, although having a more focused approach to digitizing credits *European Credit Transfer and Accumulation System* (ECTS), the solution brings an approach about how to use the multiple signatures, involving the student and institution to validate the transaction. The solution also raises an interesting question regarding the consensus mechanism used. Depending on the type of application, one consensus mechanism is more suitable than another. Considering education, it makes no sense to mine blocks to record transactions that contain academic certificates or ECTS credits. Universities are reliable nodes and responsible for the data they provide. Operating in a permissive way to create transactions, universities maintain autonomy in the issuing process, but they are unrelated about verification and sharing.

On Hyperledger's evaluation, its flexibility was identified as a strong point, and as a weak point, the lack of models capable of speeding up development. It would be interesting if pre-assembled open-source libraries existed for that purpose. Another issue involving Hyperledger is that to identify the university or the student, it would be necessary to combine the use of two greenhouse technologies (Fabric + Indy), which can make the development of the application even more complex.

Recently, the use of smart contracts has been evaluated for academic record systems[13, 21, 22, 6, 23, 24]. Prototypes are implementing this type of mechanism in the Blockcerts [3], but they are new and still being discussed. This type of approach is interesting but needs to be observed, because not every blockchain implements this type of contracts (example Bitcoin). Besides, the languages implemented by different blockchains can be different, which can make it difficult to create a standard. In the examples shown, the vast majority are based on Ethereum. We believe that because blockchain technology is recent, with many networks likely to disappear and others appearing in the future, the proposed solution should be compatible with any blockchain technology [5] [1]. Also, the blockchain scalability issue [25], may encourage applications to migrate to smaller networks.

3.2 Prototype

UFP has built a prototype and has been testing the application of academic certificates with blockchain technology. The application called CertEdu was built based on the Blockcerts platform and has its architecture designed according to the figure 5. As you can see, CertEdu issues electronic documents on Bitcoin and Ethereum networks. The objective of implementing two networks is precisely to assess the prototype's ability to achieve the desired blockchain compatibility property.

The implementation described in this work shows that even existing different ways to operate the blockchains (permissioned, permissionless, consortium), the implemented solution is easily adapted to operate on any type of network because of these reasons:

- The solution uses its own Merkle Proofs mechanism
- The technique of anchoring the diploma hash on the blockchain, allows verification and overcomes the space limitation
- The solution doesn't need the use of smart contracts

The technical standard of Blockcerts was designed to work with any blockchain, thus preventing the success of the project from being conditioned to the evolution of another product. In 2017 when the project was started, integration was only possible with Bitcoin, but it soon extended to Ethereum. In 2018, Universidad del Rosario, in Colombia, [26] built the integration with Hyperledger.

Blockcerts uses different layers that work together to create the hashes for each batch of certificates, issuing them on the blockchain and subsequently allowing web platforms to print the certificates by using *JavaScript Object Notation* (JSON) objects and verifying them on the blockchain [4].

The figure 3 shows a process of creating a digital certificate on the blockchain by using the components of Blockcerts. As we can see, the asset stored is the hash of the JSON generated file, which in practice means linking the digital file with the blockchain.

The next section describes CertEdu implementation.

3.2.1 Certificate templates

Cert-tools is responsible for creating the certificate templates that will later be signed and anchored on the blockchain. For each model, it is possible to customize information such as the title, logo, description, history. There are also customizable fields, so that peculiar information can be treated. These fields can be created globally (they will appear for all the certificates generated from the model in question) or by the recipient (they will only appear for a specific group of recipients). Next, part of the code of interaction with the component is presented.

[3] https://community.blockcerts.org/t/introducing-smart-contracts-to-blockcerts/2362

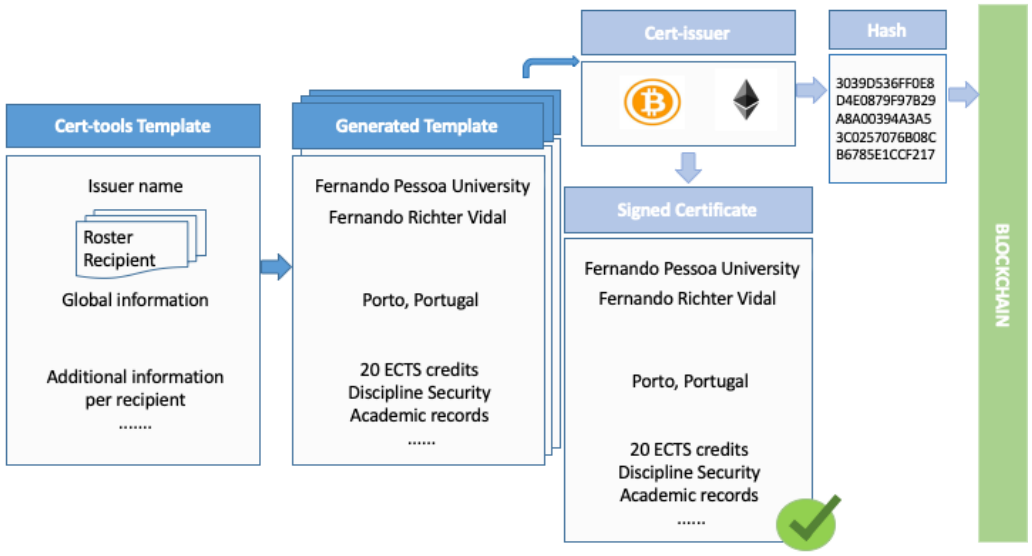


Figure 3: Diploma configuration [1]

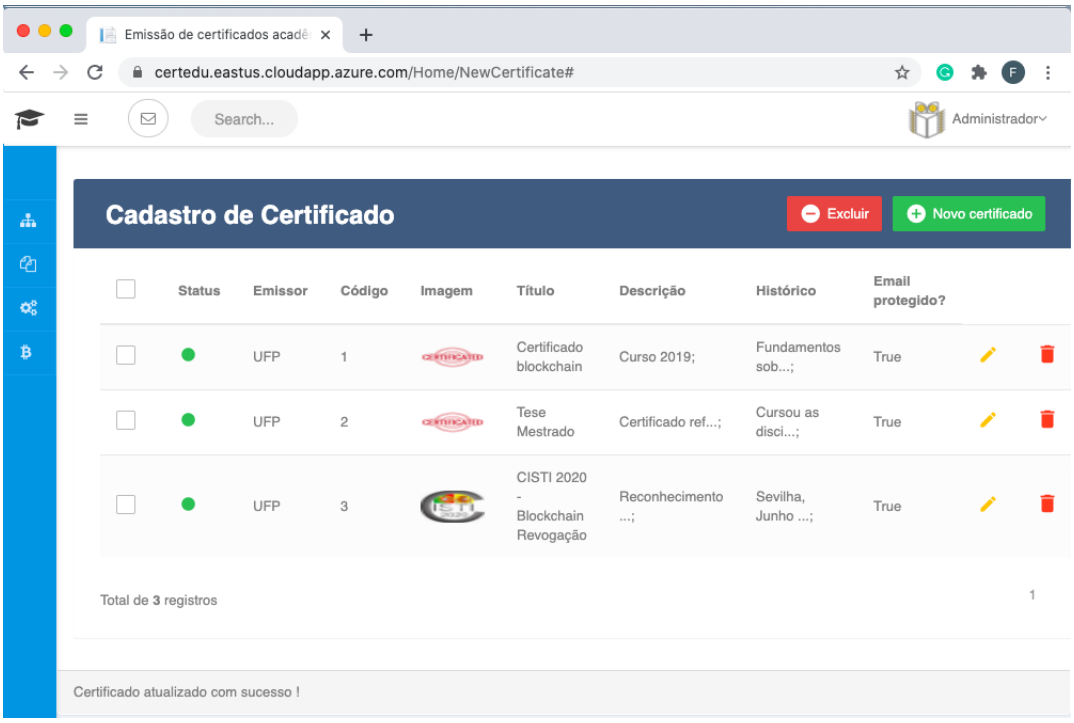


Figure 4: CertEdu's Certificate Template

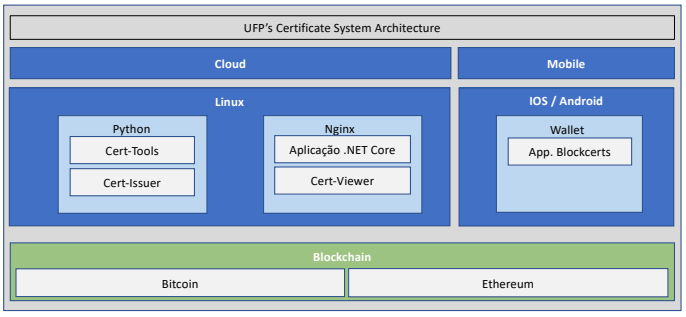


Figure 5: UFP Architecture [1]

```
create-certificate-template -c /ufp/model/ConfigJson_UFP.ini
--data_dir /ufp/ --template_dir /ufp/model
--template_file_name /ufp/model/UFP_3.json
--issuer_logo_file /ufp/image/UFP.png
--cert_image_file /ufp/image/UFP_3.png
--issuer_url https://certedu.eastus.cloudapp.azure.com/
--issuer_email ufp@ufp.edu.pt
--issuer_name 'Universidade Fernando Pessoa'
--issuer_id https://certedu.eastus.cloudapp.azure.com/UFP_1.json
--revocation_list
https://certedu.eastus.cloudapp.azure.com/rev_UFP_1.json
--issuer_certs_url https://certedu.eastus.cloudapp.azure.com
--certificate_description 'Reconhecimento de apresentação no congresso.'
--certificate_title 'CISTI 2020  - Blockchain Revogação'
--criteria_narrative 'Sevilha, Junho de 2020'
--issuer_public_key ecdsa-koblitz-pubkey:
```

```
mgAzKQZZi47g4UMvmGJCsicbJ4P3B8SHRr
--badge_id 370820c5-59ff-4a0b-8c9c-55faa0b9f431
--unsigned_certificates_dir /ufp/unsigned
```

Cert-tools receives as an input the information that will form the certificate and as output, returns a file JSON, ready to be signed by the next component. The file in question is generated by a Blockcerts module called cert-schema [4], which is based on the *Verifiable Credentials* (VC) [5]. The data types mapped by the standard follow the norm *Internationalized Resource Identifiers* (IRIs), the same used by XMLSchema [6]. All this concern in following these standards is to offer entities and interested parties a standardized format for certifications.

The figure 4 shows the implementation of cert-tools in CertEdu. The model stored in the database contains the necessary parameters for the issuance of a certificate. In another system interface, all it takes is to link the model code with a list of students in order to publish the certificates. Note that there is a status indicator, which checks when saving the record if all settings have been applied successfully. For example, there is a parameter that allows you to indicate the public address of the certificate. When it is activated, the application tries to publish on the indicated website with the parameterized access credentials. If it fails, the model remains pending and cannot be used until the error is fixed.

3.2.2 *Publishing area*

Cert-issuer is the component responsible for generating the transaction on the blockchain. The input receives the diploma file generated by cert-tools and returns as output the diploma hash published on the blockchain. Its role, in addition to the signing, is to allow blockchain compatibility, by providing a structure that allows connectors from other networks to be implemented. In the component area called blockchain_handlers, all it takes it to create three functions to connect a new network: connection (connectors.py), transaction (transaction_handlers.py), and subscriber (signer.py). On top of that, it is necessary to change the block below the main function of the component.

```
issue_certificate.py

def main(app_config):
    chain = app_config.chain
    if chain == Chain.ethereum:
        from cert_issuer.blockchain_handlers import ethereum
        certificate_batch_handler, transaction_handler,
        connector = ethereum.instantiate_blockchain_handlers(app_config)
    elif chain == Chain.bitcoin:
        from cert_issuer.blockchain_handlers import bitcoin
        certificate_batch_handler, transaction_handler,
        connector = bitcoin.instantiate_blockchain_handlers(app_config)
    else:
      new blockchain
      ...

    return issue(app_config, certificate_batch_handler,
    transaction_handler)
```

The standard is maintained by the open-source community, to support networks the Bitcoin and Ethererum. The integration to other networks is emerging as initiatives do, such as the University of Rosario, in Colombia [26], which built the connector for Hyperledger, and is testing it experimentally.

Another notable point of the figure 3 is the possibility of signing a group of models at once. Technically, cert-tools generates several certificate files and calculates the group's Merkle root, recording this value on the blockchain.

The figure 6 shows the structure of a certificate file that makes up a batch. All files in the batch have the same value as the merkleRoot field, and additionally store, in addition to the hash itself, the hashes of the nodes needed to verify the root of Merkle (proof 0, 1 and 2 of 6). In practice, when the verifier receives a file, it calculates the hash and checks on the blockchain whether this file belongs to the generated batch. With that, you only need to spend a single time, to be possible to check *n* certificates. Also, this check is very useful for revocation, because canceling a single blockchain registration automatically cancels the entire batch.

Figure 6: Certificate batch structure

Blockcerts implements an independent version of the Merkle root, the 2017 Merkle Proof Signature Suite [7]. This means that the field calculated in the JSON file, always follows the same pattern, contributing at this point for the tool to work on several networks. Each blockchain can implement the field differently, as is in the case Ethereum, which unlike Bitcoin, uses Merkle Patricia [27].

3.2.3 *Embedded authenticity checker*

The verifier has two roles: to inform the authenticity of a certificate and to represent it graphically to the user. The first versions of this component in Blockcerts were called cert-viewer, but later changed to blockcerts-verifier [8].

The technology is based on JavaScript, which makes it easier for applications, such as CertEdu (Figure 8), to embark on a universal certificate verifier within its structure.

[4] https://github.com/blockchain-certificates/cert-schema
[5] https://www.w3.org/TR/vc-data-model/
[6] http://www.w3.org/TR/xmlschema-2/
[7] https://w3c-dvcg.github.io/lds-merkleproof2017/
[8] https://github.com/blockchain-certificates/blockcerts-verifier

How the issuance process is implemented, directly impacts on the complexity of implementing this component. Notably, Blockcerts seeks to use standardized components of the blockchain, such as hashes, transaction recording, Merkle root. With this, the inclusion of new blockchain in the operation of the blockcerts-verifier becomes a simpler process. Considering a different scenario, in which the mechanisms use smart contracts, depending on the way it is implemented, this process of incorporating new networks can become complex.

There is still no defined standard on the display format of the digital certificate (figure 7 and 8). Although this is not a technical security problem, it can cause some distrust for an appraiser who receives the same diploma issued by the same broadcaster but in different formats. Fortunately, the verification function can guarantee the data veracity to the interested party, and can even validate the file in different universal verifiers, such as the one offered by Blockcerts [9]

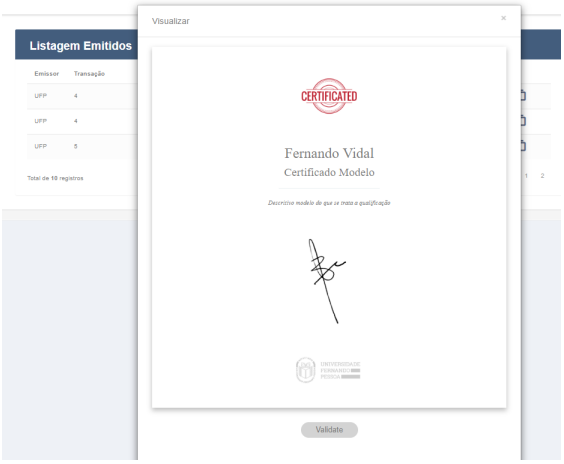


Figure 7: Model 1 for viewing a digital certificate issued by Blockcerts

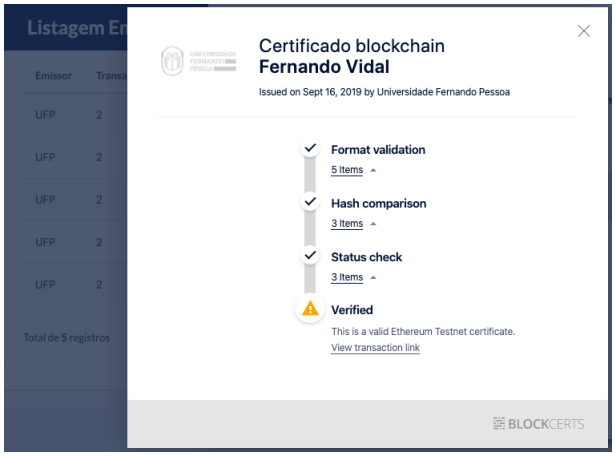


Figure 8: Model 2 for viewing a digital certificate issued by Blockcerts

[9]https://www.Blockcerts.org/

3.2.4 *Mobile application*

Cert-wallet is a Blockcerts application, also available in an open-source format, which aims to offer users autonomy over their records. Once the user downloads the application on their device, their identity is linked to a phrase created automatically by Blockcerts, which becomes a kind of private key.

The student is sovereign to register an issuer in his account and redeem the certificates that were issued by him. Autonomy is also evidenced by the fact that the student has the freedom to send his diploma to whomever he wishes, directly, without the intervention of any third parties. The difference between the digital process and the paper is that the registration received by the third party can be verified directly on a public blockchain network (in the case of this paper, Bitcoin or Ethereum), without any consultation with the issuing university.

The application has three main functions:

1. Automatically generate and send the student's public address to the sending systems
2. Store the digital certificates
3. Share the digital certificates

The figure 9 shows the operation of the first function of the cert-wallet. Note that the user informs an address of a sender profile file, and also a unique and disposable code, in English, called the nonce. This code serves as a kind of student credential to access the available service. Right after this procedure, cert-wallet accesses another address that is embedded in the profile file, and sends it to that address, the public key generated by the application, and also the code entered.

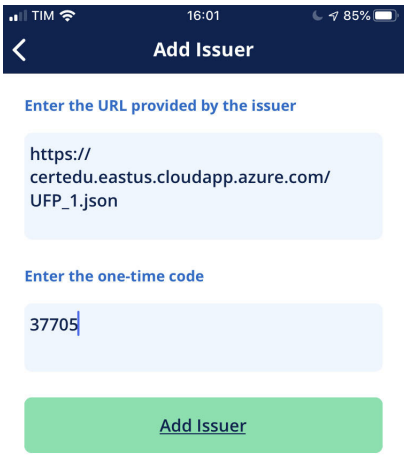


Figure 9: Cert-wallet application interface responsible for generating the student's public address

On the other side of the operation, in the second address mentioned software that is prepared to receive the public key and compare the received code is operating, validating, or not the message. If so, the cert-wallet receives a signal to register the issuer in its

database, and the student's public address is stored in the issuing application so that it can be used in future issues.

The second function of the application refers to the storage of certificates, which, as noted in the figure 10, can be done in two ways: informing an address *Uniform Resource Locator* (URL) in which the certificate is published, for example, `https://certedu.eastus.cloudapp.azure.com/certdata/2137705.json`, or by directly importing a JSON file.

Immediately after importing a credential, the application starts a process of validating and verifying the authenticity of the file, informing at the end of the process whether the certificate is valid or not.

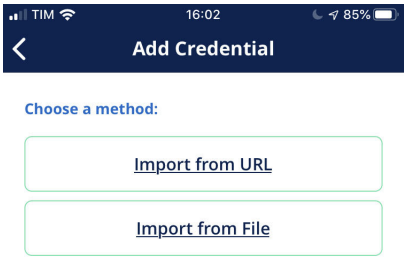


Figure 10: Ways to store a certificate in the cert-wallet

The third and final function of the cert-wallet, refers to the possibility given to the users to share their certificates. When selecting a certificate, it is possible to physically send the file to another user or simply provide the address that references it.

Table 3: Processes and challenges related to the blockchain solution

Process(s)	Challenge	Satisfies
Issuing	A feasible financial solution to large issuing	Yes
Issuing	Identify of universities	Yes
Issuing	Allow revocation	Partial
Verifying	Verifying em local database (decentralizing)	Partial
Verifying	Verifying in universal verifier	Yes
Verifying	High availability	Yes
Sharing	Identify of students	Yes
Sharing	Sovereign to share the certificates	Yes
Sharing	Graphical display of recorded information	Yes
All	Operating compatibility on any type of blockchain	Yes

[10] `https://faucet.ropsten.be/`
[11] `https://faucets.blockxlabs.com/ethereum`
[12] `https://coinfaucet.eu/en/btc-testnet/`
[13] `https://bitcoinfaucet.uo1.net/send.php`

3.2.5 Goals

The purpose of the application is to simulate an environment close to a real situation, for this reason, it was decided to configure the TestNet (Bitcoin) and Rospten (Ethereum) networks. Faucets [10], [11], [12], [13] were used to transfer credits to the application's accounts, and thus emulate a scenario close to reality, in which there are limited financial resources to generate certificates.

The table 3, presents an analysis by the process, about the main challenges to be addressed. In this way, the work assesses such issues through CertEdu, and highlights which issues are addressed and which still need to be addressed. Concerning the points not yet solved, some possible solutions to these problems are presented.

Below, the topics partially covered are discussed.

3.2.6 Disconnected checking and revocation

One of the innovations proposed by blockchain certificate solutions is disconnected verification. For example, it would be possible to verify the authenticity of the diplomas, using an off-line local copy. All certificates that are already there are perfectly verifiable.

However, the prototype found that this feature was not met, due to the way Blockcerts implements the verification process. There is a dependency on two external files (hosted on the issuer's server), one for university identification, and another for checking certificate revocation.

Regarding the first dependency, Learning Machine and NextID in late 2019, published an article [18] with a proposal to replace the issuer profile. This functionality must be present in versions 3.0 of blockcerts.

Regarding revocation, there is still no definitive solution to the problem. The paper [5], presents different approaches, such as smart contracts [28], control data [29] or even the combined use of *Interplanetary File System* (IPFS) and blockchain [30] , [31], [32].

3.3 Systemic tests

The following tests simulate some of the situations that can occur when issuing a blockchain certificate. Firstly, attempts are made to tamper the certificate, either through data editing or through the usurpation of property. Right after this test, it is evaluated how the validation system behaves in a disconnected way.

3.3.1 Tamper

In this simulation, a diploma issued by CertEdu is changed. Using a simple text editor, the original information of the diploma is modified, including a letter R in the student's name. It is notable that the digital diploma file is not encrypted, and can be easily edited by users who have a minimum of knowledge in text editing tools.

The result of the modified document validation operation can be seen on the figure 11, and the rules processed by the verifier can be observed through the table 4.

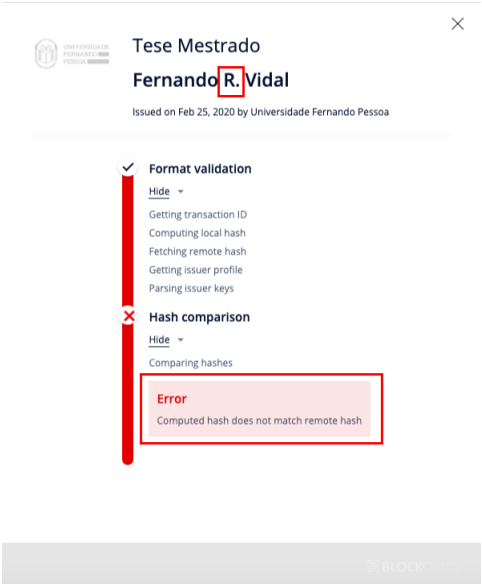


Figure 11: Tampered certificate validation

Table 4: Result of validations performed by the universal verifier

Nº	Rule	Satisfies ?
1	The transaction number contained in the certificate (sourceId) exists in the referenced blockchain ?	Yes
2	The computed hash over the information, match with the targetHash field?	Yes
3	The *targetHash* match with the blockchain stored content ?	No
4	The issuer's public address (server) is available ?	Yes
5	The embedded issuer's public key in the certificate, match with the key hosted in the issuer's server	Yes

We realized in this experience that the verifier can identify any adulterations of the file's content. Analyzing the verifier's rules, even if an attacker tampered a certificate and recalculated the hash data, the tamper would be detected, because the content of the *targetHash* field would not match the value stored in the blockchain.

3.3.2 *Unauthorized appropriation*

This other fraud simulation discusses the possibility for an attacker to change the issuers and /or recipients of a genuine certificate. For example, it would be interesting for criminals to replicate real university certificates with lesser visibility, changing them as if they were from prestigious universities. Likewise, a profitable fraud would be to offer genuine degrees to malicious recipients, who only intend to accumulate "achievements".

Blockcerts embed the student's public key in the certificate data, in the *recipientProfile* section, in the *publicKey* field. Likewise, in the *verification* section, the sender's information is embedded. This means that such fields are part of the hash calculation, which guarantees protection against tampering.

However, an attacker could alter the certificate data, informing other public keys of issuers and recipients, and recalculate the certificate hash. Additionally, to prevent rule 3 (table 4) from identifying the fraud, the attacker could change the *sourceId* field for a transaction that he created, which contains exactly the value of the hash of the defrauded certificate.

In this case, rule 3 will pass as true, because both the local calculation of the hash and the comparison with the information stored in the blockchain will be considered valid. However, Blockcerts can identify this fraud, because in addition to validating the content of the transaction, it checks whether the address of the transaction is owned by the issuing university. This is possible because the system compares the public address of the transaction with the public address hosted on the sender's server.

The figure 12 presents excerpts from the university's profile file, highlighting the public key information. The adulterated certificate, and the forged transaction, point to the *mgAzKQZZi47g4UMvmGJCsicbJ4P3B8SHRr* hash. However, the university's real public key points to the value *JKmg4UDWqBcnhklMEsDgdDHsRrF*.

As shown in figure 13, the validator detects the inconsistency of this information, preventing the fraud from being carried out.

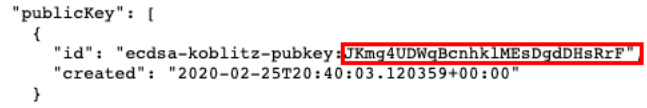


Figure 12: Issuer identification file

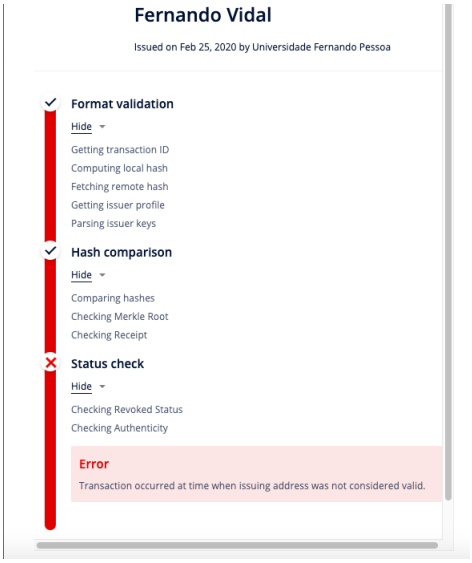


Figure 13: Certificate check with the issuer tampered with

3.3.3 *Stolen key and revocation*

This simulation evaluates Blockcerts' ability to identify frauds that may occur in situations where the university has had its private key stolen.

The digital diplomas issued by non-blockchain solutions, such as those that sign using a *Rivest-Shamir-Adleman* (RSA) key, do not guarantee in their signature proof that the creation date associated with their signature is consistent with reality. In most cases, the

issuer who signs with the keys to their property, must sign with the correct date and time [18].

In situations where the private key is stolen, the attacker may want to issue retroactive diplomas, thus managing to guarantee illegal conquests for any recipient.

When the university realizes that its key was stolen, it revokes the validity from the date on which the theft occurred. However, in non-blockchain solutions, since certificate dates cannot be trusted, it is very difficult to determine which diplomas are still valid.

Therefore, by using a blockchain's reliable timestamp, you can calculate the actual date of issuance of a diploma, and thus, determine the revocation/expiration. With that being said, it becomes more reliable to separate valid and invalid credentials. For example, all credentials that are referenced on a blockchain, before the reported date of the theft, remain securely valid and are not affected by any revocation process.

The figures 14 and 15, demonstrate how the date of the issuance of a certificate can be easily compared with its registration on the blockchain. By consulting the date of the corresponding transaction (indicated in the *sourceId* field), you can easily perform the comparison.

Figure 14: Certificate issue date

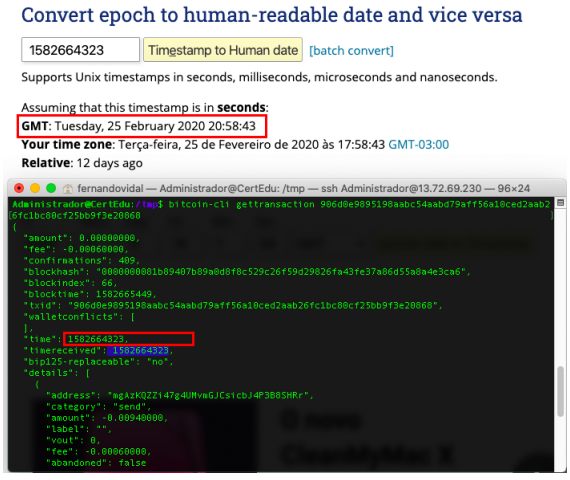


Figure 15: Registration date of certificate issuance on the blockchain

In this matter, it is concluded, that the blockchain solution offers a differential about the solutions without blockchain, since it guarantees students, that their old diplomas will not be affected, even if unexpected situations like a private key theft happens.

3.3.4 Validation without server access

This simulation is divided into two parts. First, the possibility of the issuer's server failing is evaluated and then the functionality of continuing to validate in a non-synchronized way, records on a local blockchain are tested.

In this first test, the sender's server is inaccessible. This simulation intends to evaluate the system's behavior when faced with a failure of the university server. For this, the server on which the profile and revocation files are hosted was turned off and was performed a check for a genuine certificate.

It can be seen through figure 16, that the verifier was unable to tell whether the certificate was valid or not.

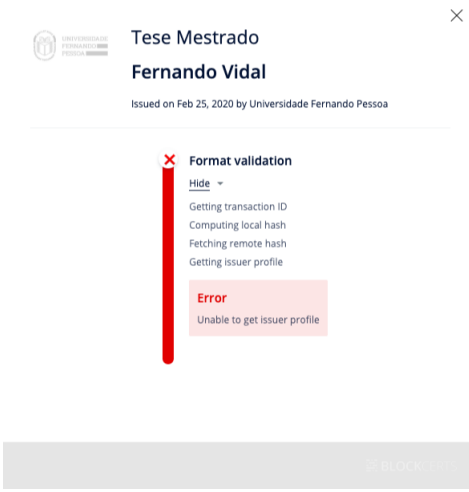


Figure 16: Local check, error when connecting transmitter

The test concludes that the solution is not yet capable of offering independence from the issuer. Besides, if an attacker manages to invade the server, even if it is temporarily, it would be possible to carry out the attacks described in the 3.3.1 section, due to the possibility of changing the information stored in the sender's profile.

In the second test considers a scenario in which the local blockchain is not synchronized. One of the benefits of using the blockchain would be the possibility to continue to validate certificates, on a local blockchain, even if it no longer exists.

Analyzing the libraries of the implemented prototype, it is noticed that the verifier uses interfaces *Application Programming Interface* (API), made available by block query tools, to verify them. This means that for a local validation it would be necessary to modify the verification address for the local base. As this is a relatively easy operation, it can be said that the tool meets this requirement.

Another situation that can be simulated is the verifier validating the issuance of a certificate that has not yet been authorized by the consensus mechanisms. As mentioned in the section, the validation of the blocks is one of the pillars that brings all the security of the technology. Validating a diploma from a registration that has not yet been confirmed could be considered a very serious failure.

Figure 17 shows the result of this operation, and as the result is shown, it can be said that the system also passed this security test.

Figure 17: Block not yet confirmed

4 Conclusion

Diploma fraud is far from being over. The combat mechanisms proved to be extremely inefficient, mainly in the paper models. Digital degrees have evolved with digitalization, but they do not offer privacy to the students, so they can have confidence in freely distributing their certificates.

Blockchain's disruptive technology offers a breakthrough in distribution and ensures protection from tampering. It was also shown that the blockchain offers the best resources to act in case of loss or theft of the university's private key, protecting the entity against undue retroactive emissions. Besides, the level of privacy offered by the technology, by recording only the hash certificate on the blockchain, makes the solution less prone to data leakage than the digital certificate solutions without blockchain.

The materialization of the objective of using the blockchain for the management of diplomas was achieved during the construction of the CertEdu prototype, in which it was possible to operate issuances, revocations, shares, and verifications of academic certificates.

However, tests also pointed out that disconnected operation is still an issue that needs to be worked on. Although some blockchains already offer resources, such as the smart contract, which would allow to easily resolve the centralization points placed, the premise of the solution operating on any type of blockchain has not been met, so those issues that still prevent decentralization are still raised by this work is pending.

There is also concern about the unpredictability of the issuance costs. Thinking of public networks like Bitcoin and Ethereum, you cannot predict the rate of transactions in the long run. This issue can inhibit the adhesion of universities.

Finally, it is concluded that the application of blockchain in the management of academic certificates is perfectly possible and that the technology already offers benefits in terms of privacy, distribution and revocation, compared to digital solutions. However, to be able to take advantage of the full potential of the technology, the centralization points addressed by this work, such as the validation of the issuer's profile and revocation, need to be migrated to features that operate within the blockchain itself.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors acknowledge the support of DuoTi Solucoes.

References

[1] F. Vidal, F. Gouveia, C. Soares, "Analysis of Blockchain Technology for Higher Education," in 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 28–33, IEEE, 2019, doi:10.1109/CyberC.2019.00015.

[2] N. Smolenski, D. Hughes, Academic Credentials in an Era of Digital Decentralization, Learning Machine Research, Anaheim, CA, 2016.

[3] IBGE, "Rendimento médio real mensal do trabalho principal, por níveis de instrução," 2018.

[4] M. Oliver, J. Moreno, G. Prieto, D. Ben\'\itez, "Using blockchain as a tool for tracking and verification of official degrees: business model," in Using blockchain as a tool for tracking and verification of official degrees: business model, 29th European Regional Conference of the International Telecommunications Society (ITS): "Towards a digital future: Turning technology into markets?", Trento, Italy, 1st - 4th August 2018, International Telecommunications Society (ITS), Trento, 2018.

[5] F. R. Vidal, F. Gouveia, C. Soares, "Revocation Mechanisms for Academic Certificates Stored on a Blockchain," in 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), 1–6, IEEE, Sevilha, 2020, doi:10.23919/CISTI49556.2020.9141088.

[6] F. Schär, M. Fabian, "Blockchain diplomas: Using smart contracts to secure academic credentials," Beiträge zur Hochschulforschung, **48**, 2019.

[7] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," Www.Bitcoin.Org, 2008, doi:10.1007/s10838-008-9062-0.

[8] M. Turkanović, M. Hölbl, K. Košič, M. Heričko, A. Kamišalić, "EduCTX: A blockchain-based higher education credit platform," IEEE Access, 2018, doi:10.1109/ACCESS.2018.2789929.

[9] M. Saad, J. Spaulding, L. Njilla, C. A. Kamhoua, S. Shetty, D. Nyang, A. Mohaisen, "Exploring the Attack Surface of Blockchain: {A} Systematic Overview," CoRR, **abs/1904.0**, 2019.

[10] J. a. Kroll, I. C. Davey, E. W. Felten, "The Economics of Bitcoin Mining, or Bitcoin in the Presence of Adversaries," The Twelfth Workshop on the Economics of Information Security (WEIS 2013), 2013, doi:June11-12,2013.

[11] A. M. Antonopoulos, Mastering Bitcoin: Unlocking Digital Crypto-Currencies, O'Reilly Media, Inc., 1st edition, 2014.

[12] R. C. Merkle, "A digital signature based on a conventional encryption function," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 1988, doi:10.1007/3-540-48184-2_32.

[13] L. M. Palma, M. A. G. Vigil, F. L. Pereira, J. E. Martina, "Blockchain and smart contracts for higher education registry in Brazil," International Journal of Network Management, **29**(3), e2061, 2019, doi:10.1002/nem.2061.

[14] D. Yaga, P. Mell, N. Roby, K. Scarfone, "Blockchain technology overview," Technical report, National Institute of Standards and Technology, Gaithersburg, MD, 2018, doi:10.6028/NIST.IR.8202.

[15] A. Grech, A. F. Camilleri, Blockchain in Education, Publications Office of the European Union, 2017, doi:10.2760/60649.

[16] H. Haugsbakken, I. Langseth, "The Blockchain Challenge for Higher Education Institutions," European Journal of Education, **2**(3), 2019.

[17] N. Bore, S. Karumba, J. Mutahi, S. S. Darnell, C. Wayua, K. Weldemariam, "Towards Blockchain-enabled school information hub," in ACM International Conference Proceeding Series, 2017, doi:10.1145/3136560.3136584.

[18] A. Ronning, W. W. Chung, "Blockcerts V3 Proposal," 2019.

[19] M. L. MIT Media Lab, "Blockcerts-An Open Infrastructure for Academic Credentials on the Blockchain," 2016.

[20] M. Bartoletti, L. Pompianu, "An analysis of Bitcoin OP_RETURN metadata," CoRR, **abs/1702.0**, 2017.

[21] J. C. Cheng, N. Y. Lee, C. Chi, Y. H. Chen, "Blockchain and smart contract for digital certificate," in Proceedings of 4th IEEE International Conference on Applied System Innovation 2018, ICASI 2018, 2018, doi: 10.1109/ICASI.2018.8394455.

[22] T. Kanan, A. T. Obaidat, M. Al-Lahham, "SmartCert BlockChain Imperative for Educational Certificates," in 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), 629–633, IEEE, 2019, doi:10.1109/JEEIT.2019.8717505.

[23] K. Mori, H. Miwa, "Digital university admission application system with study documents using smart contracts on blockchain," in Advances in Intelligent Systems and Computing, 2020, doi:10.1007/978-3-030-29035-1_17.

[24] K. Patel, M. L. Das, "Transcript Management Using Blockchain Enabled Smart Contracts," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2020, doi:10.1007/978-3-030-36987-3_26.

[25] M. Jirgensons, J. Kapenieks, "Blockchain and the Future of Digital Learning Credential Assessment and Management," Journal of Teacher Education for Sustainability, 2018, doi:10.2478/jtes-2018-0009.

[26] C. Iragorri, "Academic certificates on Hyperledger," 2018.

[27] G. Wood, "Ethereum: A Secure Decentralised Generalised Transaction Ledger. EIP-150 REVISION," 2017, 2017.

[28] J. Santos, K. H. Duffy, "A Decentralized Approach to Blockcerts Credential Revocation," 2019.

[29] L. Rujia, D. Galind, "BTCert," 2017.

[30] R. Kumar, R. Tripathi, "Implementation of Distributed File Storage and Access Framework using IPFS and Blockchain," in 2019 Fifth International Conference on Image Information Processing (ICIIP), 246–251, IEEE, 2019, doi:10.1109/ICIIP47207.2019.8985677.

[31] S. Wang, Y. Zhang, Y. Zhang, "A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems," IEEE Access, 2018, doi:10.1109/ACCESS.2018.2851611.

[32] A. Rajalakshmi, K. Lakshmy, M. Sindhu, P. Amritha, "A Blockchain and IPFS based framework for secure Research record keeping," International Journal of Pure and Applied Mathematics, **119**(15), 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 883-893 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Evaluating the Effectiveness of Query-Document Clustering Using the QDSM Measure

Claudio Gutiérrez-Soto[*,1],Marco Palomino[2], Arturo Curiel[3], Héctor Riquelme Cerda[1], Fernando Bejar Rain[1]

[1]*Universidad del Bío-Bío, Department of Information Systems, Casilla, 5-C, Chile*

[2]*University of Plymouth, School of Engineering, Computing and Mathematics, PL4 8AA, England*

[3]*CONACYT-Universidad Veracruzana, Facultad de Estadística e Informática, 91020, México*

ARTICLE INFO

Article history:
Received: 07 July, 2020
Accepted: 19 November, 2020
Online: 08 December, 2020

Keywords:
Average link
Query logs
Query-sensitive similarity measure (QSSM)
Ward's method
Web search engines

ABSTRACT

It is well documented that the average length of the queries submitted to Web search engines is rather short, which negatively impacts the engines' performance, as measured by the precision metric. It is also well known that ambiguous keywords in a query make it hard to identify what exactly search engine users are looking for. One way to tackle this challenge is to consider the context in which the query is submitted, making use of query-sensitive similarity measures (QSSM). In this paper, a particular QSSM known as the query-document similarity measure (QDSM) is evaluated, QDSM is designed to determine the similarity between two queries based on their terms and their ranked lists of relevant documents. To this extent, F-measure and the nearest neighbor (NN) have been employed to assess this approach over a collection of AOL query logs. Final results reveal that both the Average Link Algorithm and Ward's method present better results using QDSM than cosine similarity.

1 Introduction

This paper is an extension of work originally presented in the International Conference of the Chilean Computer Science Society (SCCC) [1]. Nowadays, exist a large amount of information available on the Web; additionally, web search engines (WSE) daily index thousands of pages, by which finding relevant and timely information over this growth without constraints becomes quite a challenge. According to [2], this unrestricted growth has not been accompanied by corresponding technical advances in approaches to extract relevant information. Unsuccessful searches are common in WSEs and can be given by several reasons among which we can mention the following. First, the lengths of submitted queries by users are mostly short (e.g., the average size of a web search is 2.4 words [3]). Owing to queries are conformed by a few keywords, it is complicated to determine the specific topic in which the query is inserted, therefore, a result with a few keywords in the query can be formed by unrelated topics. Furthermore, users ignore how to formulate a correct query [4]. The problem is made more complex when the user does not have a specific idea about which should be the result that he/she is looking for (i.e., which turns in relevant information for him/her). In light of the foregoing, it is not easy for WSEs to interpret the meaning of what users are looking for. One way to tackle this issue, is to consider the context in which the queries are submitted at WSEs [5], [6]. To capture the context, WSEs should consider what queries are related among them (i.e., determining if the queries are similar) and how their results have been beneficial for users. A way to determine the relationship between similar queries and relevant documents for these queries is given by the cluster hypothesis, which establishes that all documents considered as relevant for a query are similar to each other (i.e., similar documents can be relevant for the same query)[7]. Accordingly, it can be assumed that relevant documents for a query q, are relevant for a query q', such as q' is similar to q. Thus, having methods that provide the similarity among documents and queries can bring a better characterization about the meaning of a new query, and as a consequence, it entails more effective results.

Whether the WSE is able to establish how similar is a new query regarding queries recently submitted, then the search engine

[*]Corresponding Author: Claudio Gutiérrez-Soto, Casilla 5-C Email: cogutier@ubiobio.cl

should provide documents, which were relevant in previous searches. Hence, the recent past queries along with the relevant documents provide a context, in which is feasible to improve the answers to new searches [8]. Nevertheless, measuring how similar are two queries using their documents is not easy, suitable metrics that allow representing the context are needed.

Aiming to capture the context in which queries are submitted, Tombros and Van Rijsbergen [9] present a pioneering approach introducing the measuring called query-sensitive similarity measure (QSSM). This measure establishes the following; two documents are more similar than others, whether both are more similar regarding a given query. Following this argumentation line, QSSMs can be used as a metric to measure the similarity between two queries considering the context. As such, a WSE using additional information can improve its effectiveness to answer a new query. To achieve this goal, queries alongside their documents should be stored in clusters. Currently, few approaches store the queries along with their documents [10]–[14] (these approaches recommend to the user a list of similar queries, which are related to the submitted query by the user). However, the approaches mentioned previously are not directly related to the approach presented in this paper.

1.1 Contribution

The main contribution of this paper is the effectiveness evaluation of QDSM. Roughly speaking, effectiveness is related to the quality of recovered documents. Better effectiveness occurs when more relevant documents are retrieved (from a total of N documents retrieved, there are more relevant documents than non-relevant documents). By contrast, worse effectiveness occurs when more non-relevant documents are retrieved. With this in mind, grouping similar queries alongside their relevant documents (in clusters) should directly affect the effectiveness. Therefore, improving clusters' effectiveness implies having more relevant documents by clusters, which aligns with the cluster hypothesis.

To evaluate QDSM effectiveness, the F-measure alongside the nearest-neighbor (NN) cluster hypothesis tests were used. Both tests were applied over the following algorithms; Single Link, Complete Link, Average Link, Bisection K-means, and Ward's method. Three relevance models were simulated with the aim to determine which documents are relevant for a specific query. Besides, a wide range of experiments was carried out over five sets of queries. Final results are presented as QDSM improvement percents regarding cosine measure (S_c) (or cosine similarity), which were contrasted with the values obtained applying the Student Paired T-test (two samples).

The remainder of this paper is organized as follows: In Section 2, a detailed review of articles related to the problem of capturing the context using clustering is presented. In Section 3, the methodological description, is exposed. Section 4 presents the experimental environment. Section 5 displays the empirical results, which are then discussed in Section 6. Finally, Section 7 gives some perspectives of future work along with the conclusions.

2 Related Work

There have been many works that deal with the use of clustering in Information Retrieval (IR). Clustering in IR has been employed to improve the effectiveness (i.e., quality of clusters). Overall, clustering-based approaches that intend to capture the context in which the queries are submitted can be classified into two categories, considering the underlying repositories (these are also known as collections or datasets). The first category involves using traditional IR datasets. Some of them use QSSM similarities such as [15]–[17]; meantime, the second category uses log-file data from search engines.

To improve effectiveness in the retrieval process, an approach relies on hierarchic query-specific clustering is presented in [15]. In pursuing this goal, a wide range of experiments was performed. According to the authors, given a specific query, the hierarchy should be adapted to increase the likelihood to situate relevant documents to the query in nearby clusters. Two characteristics stand out in this research. First, an analysis of optimal clusters variation considering the number of top-ranked documents allowing better effectiveness is exposed. Finally, a comparison between their results and inverted file search (IFS) is provided. Five traditional IR collections alongside four hierarchic agglomerative methods were employed in all experiments. Final results indicate that query-specific clustering outweighs static clustering in each of the experiments. On the other side, a framework based on probabilistic co-relevance, which gives a query-sensitive similarity, is presented in [17]. The similarity between two documents corresponds to the ratio between the co-relevance probability and a specific query. Two cases were considered to identify the co-relevance. First, the document's relevance is independent of the rest of the documents. Second, the document's relevance is dependent on the rest. Several experimental scenarios were studied using the nearest neighbor test on TREC collections. The final results reveal that the framework outperforms term-based similarity.

The approaches mentioned above expand the users' judgments grounded on the following assumption. All terms included in a relevant document for a specific query are relevant too. Consequently, it is assumed that all documents that include some of these terms are also relevant. Besides, these approaches do not deal with the similarity among queries, with the exception of [16],[18]. In [18], a method called Scatter/Gather, which explores clusters based on documents, is proposed. The method returns a ranked title's list for the organization and viewing of retrieval results. Scatter/Gather is used as a tool for retrieval of browsing results, which presents summaries to users. Towards that goal, documents are joined in similar topics. A fractional algorithm provides k clusters on TREC/Tipster dataset. As a result of experimentation, the authors assert that their method gives tailored clusters according to the query's characteristics. In such a way, their results corroborate the cluster hypothesis since relevant documents are more similar to each other than non-relevant documents. In [16], the authors introduce the Weighted Borda (WBorda) model, which determines the co-relevance of a document using different similarities' types. To this end, a Support Vector Machine (SVM) was trained to get the estimated co-relevance, fusing the induced rankings using several functions. Each function considers the similarity between documents and the query. Several similarity measures were considered in experiments such as cosine BM25, M1, and M3. The final results in tasks such as nearest-neighbor clustering, cluster-based, and graph-based document retrieval indicate that WBorda provides better results than several proposed

co-relevance models.

On the other hand, approaches such as [10],[19], and [20] belong to the second category. Users' log-based, an approach of query clustering is proposed in [19]. Towards that end, the documents previously read by users are employed to construct cross-references among documents and queries. According to the researchers exist a strong relationship between the selected documents and queries. This approach underlies two fundamental aspects: First, two queries are similar if users clicked on the same documents; Second, if a set of documents was selected for the same queries, then the documents' terms are related to the queries' terms. The empirical results were obtained using the DBSCAN algorithm and the Encarta encyclopedia dataset. The final results show that many similar queries are gathered in the same clusters utilizing this approach. A query-clustering classification, which compares various query similarity measures, is presented in [10]. Three groups: content-based approaches, feedback-based approaches, and results-based approaches are suggested in this classification. In content-based approaches, the common terms of queries are used to describe query clusters. Similarity functions such as Jaccard, Cosine, and Dice were employed to build the clusters. In that regard, the authors claim that this method is not convenient for search engines due to many queries have few terms. On the other side, in feedback-based approaches, the similarity measure is grounded users' selections over search results; therefore, two queries are similar whether they encourage the selection of similar documents. In turn, results-based approaches evaluate the similarity between queries through the overlap of returned documents. In this case, the researchers point out that this approach's principal drawback corresponds to high processing times. Notable results are obtained using the three approaches in parallel. In [20], a WSE provides a user with a list of similar queries regarding the user's submitted query. Semantically similar queries give support to the clustering process. Clusters are formed, taking into account the historical preferences of registered users in the WSE. To build the clusters, term-weight vector representation of queries considering the clicked URLs was employed. The method exhibits two benefits, (1) it discovers the related queries, and (2) sorts the queries rely on a relevance criterion. It is important to mention that the queries are sorted using the following criteria: (a) the similarity between the clusters' queries and the new query and (b) the support, which is related to how much the query answers capture the user interest. The experiments were conducted using the combination of (a) and (b). The results display improvements on average precision.

In summary, the first category is based on traditional IR datasets. A traditional IR dataset is formed by three sets, a set of documents (D), a group of queries (Q), and a set of users' judgments (JU). The user's judgments contain the relevant documents for a query in Q. Note that all works mentioned in this category include new relevant documents (if some document has some relevant term, then it is considered relevant), which are not part of the original JU. On the other hand, it should be noted that there are no JU in the second category (log-files from search engines). Consequently, subject matter experts evaluate the pertinence of a document given a query. Note that all approaches mentioned in this related work modified some documents' relevance, which directly impacts effectiveness. Contrary to these approaches, in this paper, three types of users' judgments are simulated without altering the documents' original-relevance.

The overall procedure and a discussion about the results are presented in the following sections.

3 Methodology

The methodology overview is as follows. Initially, a user submits a query to the WSE. Thereafter, the WSE returns the documents as a result of the query. These documents are ranked from the most similar to the least similar regarding the query. Once this is done, the documents are stored along with the query in clusters inside the WSE. Aiming to form the clusters considering documents and queries, QDSM is used. In this way, when a user submits a new query, it is contrasted with the past queries (these are the queries previously stored) in the query-document clusters. Accordingly, an effectiveness improvement should occur due to the clusters closest to the new query containing relevant documents for the new query.

Document relevances become a crucial factor in enhancing effectiveness. In a traditional IR dataset, document relevances are given by subject matter experts, who determine what documents are relevant given a query. These documents are reflected in the users' judgments. On the other hand, the relevance of documents in a WSE is given by ranking functions. Overall, ranking functions attempt to capture the relevance through users' clicks on documents, which are expressed in the ItemRanks. In this manner, a retrieved document (i.e., URL or web page) with an ItemRank high could be considered as relevant.

It is essential to keep in mind that most approach clustering-based extend or use subject matter experts to give relevance to the documents, due to none of these datasets have been designed to work with similar queries (past queries are part of clusters). According to [21], a good way to tackle this problem is by using simulation. In this paper, document relevances have been simulated. Two notable advantages are obtained with the simulation use. First, it is neither necessary to use subject matter experts nor extend the user's judgments. Second, several models of relevance (a model can be seen as a ranking function) can be used; for instance, given a query, a document can be relevant or non-relevant depending on the model. In this paper, this is given by a relevance function, which determines the relevance of a document considering both its corresponding ItemRank and relevance probability.

Aiming to shed light on how QDSM is evaluated using a relevance function, suppose the following example. Assume that three documents (d_5, d_{10}, and d_{12}) have been recovered for a query q, such as d_5 is the most similar document concerning the query. The respective ItemRank for each document is 20, 25, and 30. Likewise, the probabilities of relevance according to their respective ItemRanks are 90%, 40% 70%. Additionally, suppose a relevance function that only considers the last recovered document (in this case, d_{12}). As d_{12} has an ItemRank of 30, the probability of being relevant is 70%. To simulate the relevance probability of d_{12}, a binary array of 100 elements is used. Initially, this array is instantiated with 0 values; subsequently, 70 random positions with value 1 are assigned in the array using Uniform Distribution. In order to give the relevance to d_{12}, an array position is selected using Uniform Distribution; thus, if this value is 1, then d_{12} is relevant; in another case, d_{12} is

non-relevant (Note that d_5 and d_{10} are non-relevant).

Following the same example, suppose a relevance function that assigns the relevance individually, then the same procedure to give relevance is performed for each document (d_5, d_{10}, and d_{12}). Thus, a possible result could be that d_5 and d_{12} being relevant, meantime d_{10} could be non-relevant. Finally, suppose a relevance function that provides the average relevance, then the average of ItemRanks is obtained, and its relevance probability is used to give relevance to the three documents.

Note that different relevance functions could provide different results on QDSM, due to QDSM considers the relevant documents as part of its metric.

3.1 The Query-Document Similarity Measure

The Query-Document Similarity Measure (QDSM) is a Query-Sensitive Similarity Measure (QSSM), which has as a fundamental purpose to capture the semantic similarity between queries, taking into consideration terms that belong to the queries as well as the position in which appear the relevant documents in both lists. Indirectly, the terms associated with the relevant documents should contribute to providing context. Specifically, each list of documents is presented in descending order according to the similarity of documents regarding the query. From the semantic point of view, two queries are closer if they share more relevant documents in their lists. This can be appreciated by observing the number of relevant documents that form the intersection between the two lists. Therefore, while more relevant documents make up the intersection, the more similar the queries will be. Thus, this paper's primary assumption is that using similar queries alongside their relevant documents should provide clusters with better effectiveness than S_c, since additional information can be captured from documents, including the queries (i.e., information is not complete in each query individually). Specifically, this additional information is given by the union of queries and documents' terms but does not belong to the intersection among them. Using this rationale, QDSM is in line with the cluster hypothesis, which claims that relevant documents for a particular query tend to be close, whereby these relevant documents should tend to be in the same cluster for a specific query.

QDSM takes advantage from the place in which relevant documents appear on the list. As reported by [22], the most similar documents concerning the query tend to appear at the beginning of the list. On this basis, the order in which relevant documents appear in both lists gives information about the context (particularly the terms of relevant documents). QDSM deals with the order of relevant documents through the use of the Longest Common Subsequence (LCS) algorithm. LCS allows acquiring the relative similarity keeping the order in which simultaneously appears a relevant document in both queries. By doing so, the context capturing in which the queries are submitted is possible.

This measure is convenient in two situations:

- When terms of a query are few, as is currently happening in the WSEs.
- In a dynamic environment, where the documents' relevance could change (i.e., the position of a document in the list could change as well as its ItemRank), non-relevant documents could become relevant documents.

Accordingly, the queries are either short length (i.e., few keywords in the query) or ambiguous. Nevertheless, these can be enriched with more information associated with their relevant-documents retrieved.

Aiming to give formality, some definitions are detailed below.

Definition 1

Let D be a set of documents, such as every document in D, is formed by a set of terms (i.e., words contained in d). D is stored in a WSE W. Besides, let q be a single query, such as $q \in Q$, where Q is a set of queries interpretable by W.

Definition 2

The cosine measure between q and d_i, is defined as:

$$S_c(q, d_i) = \frac{\sum_{j=1}^{m} t_{qj} \cdot t_{ij}}{\sqrt{\sum_{j=1}^{m} t_{qj}^2 \cdot \sum_{j=1}^{m} t_{ij}^2}}$$

shuch as $d_i \in D \wedge q \in Q$, t_q are the query'terms and t_i are the document's terms.

Definition 3

Let $|d_i|$ be the number of terms in a document d_i, such as $d_i \in D$. Note that this definition can be applied to obtain the number of terms for a query q.

Lemma 1

$S_c(q, d_i) \neq 0 \iff (q \cap d_i) \neq \emptyset$

Proof by contradiction: Suppose $(q \cap d_i) = \emptyset$. Then, there are no terms in common between q and d_i. Hence, applying the *Definition 2* implies $S_c(q, d_i) = 0$. Likewise, if $S_c(q, d_i) = 0$, then $(q \cap d_i) = \emptyset$. □

Definition 4

Let $\mathcal{L}(q) = \{d_i | S_c(q, d_i) \neq 0 \ \forall d_i \in D\}$ be the set of documents whose similarity with $q \in Q$ is different to 0.

Definition 5

Let $\mathcal{L}_N(q)$ be a list of N retrieved documents from W, such as $\mathcal{L}_N(q)$ is ranked by decreasing order (i.e., they are ordered from highest to lowest according to S_c).

Definition 6

Let $F(ItemRank(d_i), M)$ be a binary function, which provides relevance to a document d_i, such as 1 is relevant and 0 is non-relevant. The function is defined as follows:

$$F(ItemRank(d_i), M) = \begin{cases} 1: & Pr(1) = \mathcal{M}(M, ItemRank(d_i)) \\ 0: & Pr(0) = 1 - Pr(1) \end{cases}$$

where M corresponds to the relevance model (i.e., PartialRel, AvRel or LastRel). *ItemRank* is a function that provides the rank for the document d_i, and $\mathcal{M}$ is the function that gives the probability considering M and *ItemRank*. A binary array formed by 100 elements is used to represent the probability in $\mathcal{M}$. The Uniform Distribution is employed to instantiate the values (the percentage is represented with values 1 in the array) and determine the relevance (if the array's selected position contains a 1, then the document is relevant).

Definition 7

Let $\mathcal{L}_{N,R}(q)$ be a list of retrieved documents along with their relevances, then:

$\mathcal{L}_{N,R}(q) = \{(d_i, F(ItemRank(d_i), M))|d_i \in \mathcal{L}_N(q)\}$.
Definition 8

Given two queries q and q' such as both queries are in Q, and their corresponding lists of documents $\mathcal{L}_{N,R}(q)$ and $\mathcal{L}_{N',R}(q')$. Then QDSM is defined as follow:

$QDSM(q,q') = \frac{S_c(q,q') + LCS(\mathcal{L}_{N,R}(q), \mathcal{L}_{N',R}(q'))}{1 + max(\|\mathcal{L}_{N,R}(q)\|, \|\mathcal{L}_{N',R}(q')\|)}$

where:

- $S_c(q,q')$ corresponds to the cosine measure between the queries q and q'.
- $LCS(\mathcal{L}_{N,R}(q), \mathcal{L}_{N',R}(q'))$ is the LCS algorithm applied over $\mathcal{L}_{N,R}(q)$ and $\mathcal{L}_{N',R}(q')$ [23].
- *max* gives the greatest number of relevant documents between the lists $\mathcal{L}_{N,R}(q)$ and $\mathcal{L}_{N',R}(q')$.

Theorem 1. $QDSM(q,q') = 0 \iff S_c(q,q') = 0 \wedge LCS(\mathcal{L}_{N,R}(q), \mathcal{L}_{N',R}(q')) = 0$.
Proof by contradiction.

- Suppose $S_c(q,q') \neq 0 \vee LCS(\mathcal{L}_{N,R}(q), \mathcal{L}_{N',R}(q')) \neq 0$. If $S_c(q,q') \neq 0$ is enough for $QDSM(q,q') \neq 0$ (by *Definition 8*). Likewise, as $LCS(\mathcal{L}_{N,R}(q), \mathcal{L}_{N',R}(q')) \neq 0$ then $QDSM(q,q') \neq 0$ (by *Definition 8*).
- Finally, as $QDSM(q,q') \neq 0$ then either $S_c(q,q') \neq 0 \vee LCS(\mathcal{L}_{N,R}(q), \mathcal{L}_{N',R}(q')) \neq 0$ (by *Definition 8*). □

Lemma 2
$LCS(\mathcal{L}_{N,R}(q), \mathcal{L}_{N',R}(q')) \neq 0 \iff \exists\, t \in (d \cap q \cap q')$
Proof.

- $LCS(\mathcal{L}_{N,R}(q), \mathcal{L}_{N',R}(q')) \neq 0$, then $\exists\, d \in (\mathcal{L}_{N,R}(q) \cap \mathcal{L}_{N',R}(q'))$ (by the LCS invariant [23]). Thus, $d \in \mathcal{L}_{N,R}(q) \wedge d \in \mathcal{L}_{N',R}(q')$. Since $d \in \mathcal{L}(q) \wedge d \in \mathcal{L}(q')$, then $S_c(d,q) \neq 0 \wedge S_c(d,q') \neq 0$ (by *Definition 4*). Thus, $(d \cap q) \neq \emptyset \wedge (d \cap q') \neq \emptyset$ (by *Lemma 1.*). Hence, $(d \cap q \cap q') \neq \emptyset$. Therefore, $\exists\, t \in (d \cap q \cap q')$.
- Finally, as $\exists\, t \in (d \cap q \cap q')$ then $S_c(q,q') \neq \emptyset \wedge S_c(d,q) \neq \emptyset \wedge S_c(d,q') \neq \emptyset$ (by *Lemma 1.*). Thus $d \in \mathcal{L}_{N,R}(q) \wedge d \in \mathcal{L}_{N',R}(q')$ (by *Definition 7.*). Therefore, $LCS(\mathcal{L}_{N,R}(q), \mathcal{L}_{N',R}(q')) \neq 0$ (by *Theorem 1*). □

Lemma 2 asserts that exists at least a common document in both lists, and therefore at least there is one term in common among queries and the document.

Lemma 3
$|d| \neq |q| \implies \exists t \in (d \triangle q)$
Proof. Suppose $|d| > |q|$, which implies that d has at least one term more than q. On the other hand, can ocurr that $|q| > |d|$, then q has at least one term more than d. Thus, $\exists t \in ((d-q) \cup (q-d)$. Therefore, $\exists t \in (d \triangle q)$. □

Lemma 3 points out that if different numbers of terms form the document and the query, then at least there is one term that does not belong to the intersection between them. Note that d and q are in $\mathcal{L}_{N,R}(q)$.
Theorem 2
$QDSM(q,q') \neq 0 \wedge |d| \neq |q| \neq |q'| \iff \exists\, t \in ((d \triangle q) \triangle q')$; such as d is in both $\mathcal{L}_{N,R}(q) \wedge \mathcal{L}_{N',R}(q')$.
Proof:

- $QDSM(q,q') \neq 0 \wedge |d| \neq |q| \neq |q'|$, then $LCS(\mathcal{L}_{N,R}(q), \mathcal{L}_{N',R}(q')) \neq 0 \wedge S_c(q,q') \neq \emptyset \wedge |d| \neq |q| \neq |q'|$ (by *Definition 8*). As the first part, consider $LCS(\mathcal{L}_{N,R}(q), \mathcal{L}_{N',R}(q')) \neq 0 \wedge |d| \neq |q| \neq |q'|$ then, $\exists\, t' \in (d \cap q \cap q') \wedge |d| \neq |q| \neq |q'|$ (by *Lemma 2*). Thus $\exists\, t' \in (d \cap q \cap q') \wedge ((|d| \neq |q|) \neq |q'|)$. Subsequenlty, $\exists\, t' \in (d \cap q \cap q') \wedge \exists t \in ((d \triangle q) \neq q')$. (by *Lemma 3.*). Hence, $\exists\, t \in ((d \triangle q) \triangle q')$. The another case is $S_c(q,q') \neq \emptyset$, then $\exists\, t \in (q \triangle q')$, considering d in the hypothesis, $(q \triangle q') \subset ((d \triangle q) \triangle q')$. Therefore $\exists\, t \in ((d \triangle q) \triangle q')$.
- Finally and without loss of generality, if $\exists\, t \in ((d \triangle q) \triangle q')$ then $\exists\, t \in ((d-q) \cup (q-d)) \vee ((d-q') \cup (q'-d)) \vee ((q-q') \cup (q'-q)) \wedge ((S_c(q,q') \neq \emptyset \wedge \exists\, t \in (d \cap q \cap q'))$ (by *Lemma 1. and Lemma 2.)* then $QDSM(q,q') \neq 0 \wedge |d| \neq |q| \neq |q'|$. □

Definition 9

Let RQ be a set of queries, along with their retrieved documents and their corresponding relevances, then RQ is defined as follow:

$$RQ = \bigcup_{i=1}^{|Q|} \mathcal{L}_{N,R}(q_i)$$

Definition 10

Let $\widehat{D}$ be a benchmark query set, which is formed by Z subsets of queries, such as:

$$\widehat{D} = \bigcup_{i=1}^{Z} RQ$$

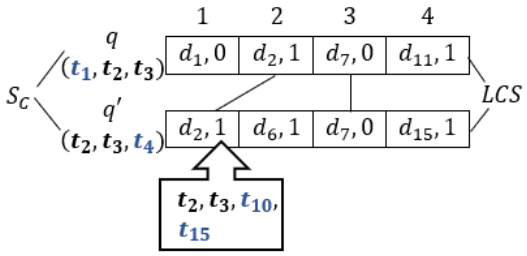


Figure 1: QDSM measure

Theorem 2 ensures that the context in which the queries are submitted in the WSE can be captured by the complementary terms to both queries and their relevant documents (i.e., the symmetric difference of sets d, q and q' ($(d \triangle q) \triangle q'$)). An example of the Theorem 2 essence and how QDSM is computed, is displayed in Figure 1. In Figure 1, the common terms for both queries are presented in bold (i.e., these are t_2 and t_3), note that both terms are common in d_2, which is a relevant document (i.e., all document d_i, 1 are relevant meanwhile d_i, 0 are non-relevant). LCS is applied over both lists of retrieved documents considering only the relevant documents in the lists (i.e., even though d_7 is in both lists, only d_2 is considered). Finally, all terms that give context are in blue color (i.e., t_1, t_4, t_{10} and t_{15}).

It is worth noting that QDSM takes the value 1 (see *Definition 8*) when q and q' are the same queries, and all retrieved documents are relevant. Specifically, this latter can be itemized in two parts. In the first part, $S_c(q,q')$ provides 1 because q and q' are the same. In the second part, both results lists are equal, and all retrieved documents

are relevant; therefore, both lists hold the relevant documents in the same positions.

In summary, QDSM provides a metric that captures the semantic relationship between two queries (context), considering the relevant documents' order in both lists. A wide range of experiments was conducted in order to compare the effectiveness between S_c and QDSM. The experimental setup is displayed in the following section.

4 Experimental Environment

A benchmark query set extracted from the well-known dataset of query log "AOL Query Logs Dataset (AOL) [24]", was used to carry out the experiments. This collection has more than 20 million web query logs stored, submitted by around 650 thousand users in more than 36 thousand lines of data. These queries were stored at an interval of three months in the year 2006. Broadly speaking, queries in AOL are depicted as rows in the database files, which contains five columns with the following fields:
{*AnonID*, *Query*, *QueryTime*, *ItemRank*, *ClickURL*}, where:

- AnonID: an anonymous user ID number.
- Query: the query submitted by the user in the WSE.
- QueryTime: the exact time at which the user submitted the query.
- ItemRank: if the user clicked on a result, it keeps the rank of the selected document; holds empty otherwise.
- ClickURL: The domain portion of the URL is showed as a result if the user clicked on a search result.

4.1 The Benchmark set of Related Queries

Aiming carrying out the clustering experiments, a benchmark set of related queries (*RQ*) (see *Definition 8.*) was processed randomly from AOL. To verify that the queries were partially related, each time a query was chosen, it was checked that at least existed another query, in such a way S_c was neither one nor zero. To achieve this goal, the queries with ClickURLs empty were removed due to these do not have answers associated with the queries. Furthermore, stopwords processing was previously performed before to apply S_c. The core insight is that ClickURLs allow depicting a list of retrieved documents for q (see Definition 5.). It should be noted that register with the same query q (i.e., the same terms), logged by the same user around the same time, corresponds to a single query, which was split into several registers. Providing the maximum amount of information implies to use the longest session, which at least contains one register (i.e., at least one result or document).

On the other side, it is important to highlight that AOL does not possess users' judgments. Note that the users' judgments play a fundamental role in order to know what documents are relevant for a specific query [25]. Furthermore, these relevant documents are necessary to evaluate precision, recall, and, therefore, effectiveness. To tackle this issue, users' judgments were simulated following the approach presented by [26]. Simulations of relevance judgments are presented in the following section.

4.2 Simulation of Relevance Judgments

Simulating document relevance regarding a query is not a trivial task. This task embraces a great variety of aspects, such as users' literacy [27], needed information at any one point of time, and the user's profile [28] among others. To address this problem, the approach proposed by [26], which provides the relevance probabilities for documents depending on their ItemRanks on AOL, is applied in this paper. Towards that end, $F(ItemRank(d_i), M)$ (see *Definition 6.*) is simulated using $\mathcal{M}(M, ItemRank(d_i))$ in *Definition 7.* In simple words, the relevance is assigned using a value 0 (non-relevant) or 1 (relevant), which is obtained considering the values presented in Table I (i.e, M in $F(ItemRank(d_i), M)$). The relevance probabilities were calculated using the ItemRanks, assuming the user clicks provide information about how the user interprets the query [30]. In Table I, two relevance models are presented by "AllRel" and "LastRel" columns. "AllRel" implies all clicked documents are considered as relevant; meantime "LastRel" reports that only the last clicked document is relevant. Regarding "AllRel", two variants were used for it. The first variant is named "PartialRel", which considers the individual ItemRank of each document obtained from Table 1. The second variant ("AvRel") corresponds to the average of ItemRanks of the query's recovered documents. For example, suppose three documents (d_5, d_{10}, and d_{12}) that have been recovered for a query q. The respective ItemRanks are 20, 30, and 40; therefore, the average is 30. Subsequently, the relevance probability is determined by the "ItemRank" (average) row and the "AllRel" column, so for this example, the relevance probability for each document is 0.5106. Although the three documents have the same relevance probability, the relevance for each document is individually obtained.

4.3 Clustering Experiments

Five algorithms were evaluated considering the three relevance models over the same $\widehat{D}$. Five sets of RQ were used; the smallest set of RQ contains 123 queries alongside their documents (i.e., for each document, the relevance has been assigned), meantime the biggest set comprises 2,141 queries. Aiming to compare the clusters' quality between S_C and QDSM, two well-known measures have been used, F-measure and the nearest neighbor (NN) cluster hypothesis test. F-measure was proposed by [29]; the idea behind this measure is to evaluate effectiveness in the post-processing step, in which each cluster is assigned to a class. The F-measure can be seen as a way of combining the precision and recall for a retrieval specific model, and it is defined as the harmonic mean of the model's precision and recall. In simple words, F-measure has as purpose to provide a binary classification as positive or negative according to the belonging of objects to determined classes in the clusters. F-measure allows giving more importance to precision, recall, or both. On the other hand, the nearest neighbor (NN) cluster test (which is also well-known as the (NN) test) was proposed by Voorhees ([30], [31]). In simple terms, the (NN) test reviews each of the retrieved documents for a specific query, identifying how many of its n close neighbors are relevant. The (NN) test is also used as a non-parametric classification and regression technique.

Table 1: Probability of a document being relevant in the AOL dataset if it has rank k, for two different click-based relevance interpretations, as calculated by [26]

ItemRank	AllRel	LastRel
≤ 20	0.4365	0.5702
≤ 120	0.5106	0.6278
≤ 300	0.5395	0.6493
> 300	0.4705	0.3507

Turning towards the cluster hypothesis, QDSM should provide better effectiveness than S_c if it is possible to find more relevant documents per cluster. Each experiment was executed ten times, and results are displayed as percentages of increasing or decreasing of QDSM regarding S_c. Specifically, F-measure was used giving the same weight for precision and recall; meantime, The (NN) test was instantiated with value three in all experiments. In addition, the Student's Paired t-Test (Two Samples test) was used to support the results. Five algorithms Single Link, Complete Link, Average Link, Bisection K-means, and Ward's Method, were used in each experiment.

All experiments were carried out on a server with: Intel Xeon Processor E3-1220 3.00 GHz; 16 GB Ram memory of 2133 MHz; 1 TB 7200 RPM Hard Drive; and Linux Operating System Debian Jessi 8.4.

5 Experimental Results

In this section, the quality of clusters (effectiveness) produced by QDSM and S_c is compared. To achieve this goal, the F-measure and the (NN) test were used considering the Single Link, Complete Link, Average Link, Bisection K-means, and Ward algorithms. Effectiveness was obtained using the relevance models PartialRel, AvRel, and LastRel. Note that the number of documents considered in the (NN) test corresponds to 3, it means that the relevances of the three closest documents with respect to the query were evaluated. To do that, the similarities between documents are checked alongside their relevance regarding the query. Overall, all results presented in each Table corresponds to QDSM, which are expressed in terms of percentages regarding the S_c. In Table 2, the QDSM effectiveness over the three relevance models was evaluated using the Single Link Algorithm. Note that the three relevance models were tested considering five sets of queries (# of q). From this Table, it is possible to appreciate that there is not an improvement of QDSM concerning S_c. Furthermore, Single Link presents better effectiveness for S_c than QDSM; this is consistent with the p-value obtained using The Student Paired T-test (two samples), which was 0.0029 for this Table. Generally speaking, Single Link exhibits the best results considering the "AvRel" relevance model, following by "LastRel" and finally "PartialRel" model. Continuing the same trend, the results for F-measure are exposed in Table 3. Here S_c shows again better results than QDSM, which is in line with the p-value: 0.00036, and notably the best results are presented by "AvRel" relevance model.

Regarding the Complete Link algorithm, similar results to the Single Link algorithm are presented in Tables 4 and 5, where S_c has better results than QDSM. In Table 4, the best results are provided by "LastRel" relevance model using the (NN) test, meantime that the best results using the F - measure are given by "PartialRel" model. The p-value for the (NN) test was 0.720, whilst the p-value for the F - measure was 0.00003.

On the other hand, the Average Link algorithm displays different results to Single Link and Complete Link algorithms, where QDSM is better than S_c. In Table 6, the best results are given by "PartialRel", followed by "LastRel" and "AvRel" respectively. Results showed in Table 6 are in line with the p-value: 0.00041. Following the same trend, in Table 7, QDSM presents better results than S_c for F-measure, which is in accordance with the p-value: 0.000002. It should be noted that there is no substantial difference between "AvRel" and "LastRel".

In turn, the results for the Bisection K-means algorithm are exposed in Tables 8 and 9. Both Tables provide conflicting results since Table 8 two relevance models ("PartialRel and AvRel") give good results for QDSM, meantime these relevance models provide opposing results in Table 9. These latter results are coherent with their respective p-values. The p-value for Table 8 corresponds to 0.124, whilst p-value for Table 9 is 0.904. These values are not significative due to they are not greater than 0.05.

Finally, in Tables 10 and 11, the results for the Ward's method are displayed. Both Tables provide excellent results for QDSM in contrast to S_c (except when the number of queries is 129, using the relevance model "PartialRel"). Furthermore, the p-values associated with both Tables present the more significative value among all algorithms. The p-value associated with Table 10 is 0.00007; meanwhile, the p-value for Table 11 corresponds to 0.00001.

To sum up, considering both measures, the three relevance models, and the p-values obtained, the best results are provided by Average Link algorithm and the Ward's method.

Table 2: The (NN) test over Single Link Algorithm with *PartialRel*, *AvRel*, and *LastRel*.

# of q	*Single Link*		
	PartialRel	*AvRel*	*LastRel*
123	-30.92%	-33.73%	-42.36%
666	-11.76%	-106.36%	-69.62%
1,145	4.29%	-13.82%	4.36%
1,843	-14.70%	-22.68%	-32.44%
2,141	-40.20%	-4.74%	-11.20%

Table 3: F-measure test over Single Link Algorithm with *PartialRel*, *AvRel*, and *LastRel*.

# of q	*Single Link*		
	PartialRel	*AvRel*	*LastRel*
123	-25.15%	-12.23%	-11.89%
666	-23.96%	-44.88%	-0.35%
1,145	-22.54%	-25.10%	-5.63%
1,843	-28.60%	-43.90%	-8.04%
2,141	-20.16%	-17.20%	-6.67%

Table 4: The (NN) test over Complete Link Algorithm with *PartialRel*, *AvRel*, and *LastRel*.

# of q	*Complete Link*		
	PartialRel	*AvRel*	*LastRel*
123	-20.97%	26.76%	-41.72%
666	-14.90%	7.47%	-8.79%
1,145	-33.38%	28.44%	-29.72%
1,843	-30.80%	45.38%	-45.34%
2,141	-27.41%	41.98%	-55.13%

Table 5: F-measure test over Complete Link Algorithm with *PartialRel*, *AvRel*, and *LastRel*.

# of q	*Complete Link*		
	PartialRel	*AvRel*	*LastRel*
123	-39.50%	-11.94%	-7.34%
666	-19.09%	7.48%	-33.90%
1,145	-32.67%	-47.29%	-22.01%
1,843	-15.06%	-14.54%	-14.28%
2,141	-8.71%	-15.88%	-17.39%

Table 6: The (NN) test over Average Link Algorithm with *PartialRel*, *AvRel*, and *LastRel*.

# of q	*Average Link*		
	PartialRel	*AvRel*	*LastRel*
123	48.36%	4.50%	5.17%
666	38.47%	7.70%	18.89%
1,145	47.37%	23.69%	28.59%
1,843	47.22%	26.17%	37.29%
2,141	43.38%	32.89%	30%

Table 7: F-measure test over Average Link Algorithm with *PartialRel*, *AvRel*, and *LastRel*.

# of q	*Average Link*		
	PartialRel	*AvRel*	*LastRel*
123	8.35%	6.76%	17.37%
666	7.83%	25.27%	12.14%
1,145	2.06%	20.44%	13.39%
1,843	3.17%	11.61%	13.19%
2,141	7.38%	12.85%	22.52%

Table 8: The (NN) test over Bisection K-means Algorithm with *PartialRel*, *AvRel*, and *LastRel*.

# of q	*Bisection K − means*		
	PartialRel	*AvRel*	*LastRel*
123	15.81%	41.97%	-15.46%
666	31.73%	12.88%	-1.31%
1,145	24.22%	15.26%	-1.82%
1,843	29.48%	34.40%	-4.75%
2,141	37.98%	24.00%	-41.03%

Table 9: F-measure test over Bisection K-means Algorithm with *PartialRel*, *AvRel*, and *LastRel*.

# of q	*Bisection K − means*		
	PartialRel	*AvRel*	*LastRel*
123	-5.86%	-6.34%	5.25%
666	4.79%	-16.31%	-0.52%
1,145	-17.68%	-13.87%	2.83%
1,843	-2.50%	8.45%	7.77%
2,141	-7.82%	13.51%	12.96%

Table 10: The (NN) test over Ward's Method with *PartialRel*, *AvRel*, and *LastRel*.

# of q	*Ward′s Method*		
	PartialRel	*AvRel*	*LastRel*
123	36.47%	16.32%	39.44%
666	21.75%	8.09%	17.94%
1,145	30.61%	40.12%	28.12%
1,843	31.22%	29.21%	23.39%
2,141	21.62%	27.39%	18.80%

Table 11: F-measure test over Ward's Method with *PartialRel*, *AvRel*, and *LastRel*.

# of q	*Ward′s Method*		
	PartialRel	*AvRel*	*LastRel*
123	-4.38%	15.04%	15.37%
666	7.78%	9.34%	4.81%
1,145	8.22%	5.68%	6.85%
1,843	10.86%	7.07%	8.66%
2,141	13.08%	17.97%	8.55%

6 Discussion

The main reason traditional IR datasets were not used; it was because autonomous queries build them. As mentioned early, a traditional IR dataset is made up of a set of documents D, a set of queries Q, and a set of users' judgments JU. In a strict sense, evaluating the effectiveness considering similar queries (queries form part of

the clusters) means having a set of similar queries (Q') for Q. Note that Q' must consider a set of users' judgments (JU'). Indeed, the effectiveness of two similar queries should be different. Accordingly, these datasets are not suitable for evaluating approaches based on similar queries because they have neither Q' nor JU'. On the other hand, approaches based on log files employ subject matter experts to extend and evaluate whether a document is relevant or non-relevant given a query. In short, the extension of relevant documents is common in both types of collections. In this manner, aiming to avoid using subject matter experts or extending document relevances using relevant terms, three models of relevance have been simulated in all experiments.

Concerning the effectiveness evaluation for both measures (S_c and QDSM), it is noteworthy that F-measure has been widely used in several approaches, which deal with post-retrieval clustering. Nevertheless, the use of this measure provides two drawbacks. The first one is that the result associated with this measure comprises the number of relevant and non-relevant documents related to recall and precision in its mathematical formula. Thus, the initial effectiveness changes its value once new documents are considered relevant in the post-retrieval process. The second one refers to how the clusters are conformed taking into account the different classes of objects that these contain. Consider that objects can belong to predetermine classes, and the ideal situation is given when the clusters are formed only by objects of the same class. Two terms well-known in the cluster evaluation reflect this situation, homogeneity and completeness. The idea behind homogeneity is that each cluster has few classes; meantime, completeness intends each class to be contained in a few clusters. Thus, two-cluster forming using the same objects and the same classes can have the same F-measure, while their homogeneity and completeness are different. In turn, like F-measure, the NN-test has been extensively used in several approaches to assess effectiveness. Nevertheless, this measure is not sensitive to homogeneity and completeness, since it contemplates the direct search of the n-nearest neighbors. Hence, this test is more appropriate to corroborate the cluster hypothesis, which considers the relevant documents that form the clusters.

Regarding the results presented in section "Empirical Results", it is essential to point out that there is no significant difference between values provided by F-measure and the (NN) test, excepts for the Complete algorithm (Table 4 and 5), in particular for "AvRel" where for the (NN) test, Table 4 presents favorable results for QDSM in contrast to Table 5. Besides, the p-values for both Tables differ. On the other hand, exists a substantial difference with some results presented by [1]. In particular, regarding the relevance models used in that research. There, the relevance "AllRel" is used considering the proposed by [26], meantime in this research "AllRel" has been modified by "PartialRel" and "AvRel", it means that no all documents have been considered relevant such as occurs in [1]. It is important to point out that it is unlikely that all recovered documents are relevant, such as happens in the real world. Nevertheless, the Average Link algorithm presents interesting results in both works. Concerning the results provided by the algorithms in this research, the best results are provided by Average Link and Ward algorithms using both tests (F-measure and the (NN) test). The main Ward characteristic is that it minimizes the variance of the objects belonging to a particular cluster using the "error sum of squares". In this way, each cluster should tend to have objects of a few classes (relevant and non-relevant). Carried to the hypothesis cluster context should have a clear separation between clusters of relevant documents and clusters of non-relevant documents. Therefore, the nearest closest neighbor of a relevant document should be relevant too. On the other hand, the distance (S_c or QDSM) between two clusters for the Average Link Algorithm is determined as the average distance between each object in one cluster to every object in another cluster., by which it is feasible to avoid extreme measures obtaining more homogenous clusters. The latter is in contrast with the way to built clusters in Single and Complete algorithms. Finally, Bisection K-means is a hybrid approach between agglomerative and hierarchic clustering. This algorithm exhibits favorable results in the (NN) test except when the relevance is "LastRel", recall that in this case, only the last recovered document could be relevant.

Although the running times escape from the scope of this paper, it is worth noting that most time complexities are not high. To obtain the time complexities is necessary to consider visiting a distance matrix (i.e., one matrix for S_c and QDSM respectively) with the aim to find the n nearest-neighbor. Furthermore, calculating LCS implies to visit another matrix with M files and N rows. Note that M corresponds to a list of retrieved documents for a query q, whilst N is another list of retrieved documents for a query q'. Therefore, LCS takes $O(MN)$. Recall that LCS is used to evaluate QDSM. According to [32], the optimal implementation of Ward based on the algorithms, nearest neighbor chain, and reciprocal nearest neighbor, takes $O(N^2)$. In turn, the time complexity of Bisection K-means algorithm is $O(N^2 log_2 N)$. As for the implementation of Single Link algorithm takes $O(N^2)$ in time complexity [33]. On the other hand, Complete Link implies $O(N^2 log_2 N)$ [34]. The time complexity for the Average Link algorithm takes $O(N^2 log_2 N)$ [35]. It is important to mention that both matrixes of distances are previously built before using each clustering algorithm.

7 Conclusion and Future Work

This paper is intended to check the quality of clusters (effectiveness) built using Query-Document Similarity Measure (QDSM). To achieve this goal, the F-measure and the nearest neighbor (NN) test were used to evaluate clusters' quality. The clusters of documents were built using the AOL Query Logs Dataset. In order to provide relevance to the documents, three variants related to the ItemRanks over recovered documents were simulated. Extensive experimentation was carried out using the algorithms Single Link, Complete Link, Average Link, Bisection K-means, and Ward. According to results obtained, applying the nearest neighbor (NN) test, QDSM presents significant results using the Average Link, Ward, and Bisection K-means. On the other hand, in accordance with the results obtained by the F-measure; Ward and Average Link algorithms provide better results using QDSM than Cosine Similarity (S_C). The best results are provided by the Average Link algorithm, followed by Ward's method using QDSM, considering the three variants of relevance. Ideas for future research comprises the comparison between QDSM and other state-of-art measures.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The Group of Smart Industries and Complex Systems (gISCOM) under grant DIUBB 195212 GI/EF supported the work presented in this article. Marco Palomino acknowledges the funding provided by the Interreg 2 Seas Mers Zeeën AGE'IN project (2S05-014).

References

[1] C. Gutiérrez-Soto, A. C. Díaz, G. Hubert, "Comparing the Effectiveness of Query-Document Clusterings Using the QDSM and Cosine Similarity," in 2019 38th International Conference of the Chilean Computer Science Society (SCCC), 1–8, 2019, doi:10.1109/SCCC49216.2019.8966432.

[2] A. Mikroyannidis, "Toward a Social Semantic Web," Computer, **40**(11), 113–115, 2007, doi:10.1109/MC.2007.405.

[3] "Satista,Average number of search terms for online search queries in the united states as of august 2017," 2017.

[4] A. Kanavos, P. Kotoula, C. Makris, L. Iliadis, "Employing query disambiguation using clustering techniques," Evolving Systems, **11**(2), 305–315, 2020, doi:10.1007/s12530-019-09292-7.

[5] M. Alshomary, N. Düsterhus, H. Wachsmuth, "Extractive Snippet Generation for Arguments," in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '20, 1969–1972, Association for Computing Machinery, New York, NY, USA, 2020, doi:10.1145/3397271.3401186.

[6] Z. A. Merrouni, B. Frikh, B. Ouhbi, "Toward Contextual Information Retrieval: A Review And Trends," Procedia Computer Science, **148**, 191 – 200, 2019, doi:10.1016/j.procs.2019.01.036, the Second International Conference on Intelligent Computing in Data Science, ICDS2018.

[7] N. Jardine, C. van Rijsbergen, "The use of hierarchic clustering in information retrieval," Information Storage and Retrieval, **7**(5), 217 – 240, 1971, doi:10.1016/0020-0271(71)90051-9.

[8] C. Gutiérrez-Soto, A. C. Díaz, "Improving Precision in IR Considering Dynamic Environments," in Proceedings of the 21st International Conference on Information Integration and Web-Based Applications and Services, iiWAS2019, 458–462, Association for Computing Machinery, New York, NY, USA, 2019, doi:10.1145/3366030.3366101.

[9] A. Tombros, C. J. van Rijsbergen, "Query-sensitive Similarity Measures for Information Retrieval," Knowl. Inf. Syst., **6**(5), 617–642, 2004, doi:10.1007/s10115-003-0115-8.

[10] L. Fu, D. H.-L. Goh, S. S.-B. Foo, "The effect of similarity measures on the quality of query clusters," Journal of Information Science, **30**(5), 396–407, 2004, doi:10.1177/0165551504046722.

[11] K. W. Leung, W. Ng, D. L. Lee, "Personalized Concept-Based Clustering of Search Engine Queries," IEEE Transactions on Knowledge and Data Engineering, **20**(11), 1505–1518, 2008, doi:10.1109/TKDE.2008.84.

[12] D. Beeferman, A. Berger, "Agglomerative Clustering of a Search Engine Query Log," in Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '00, 407–416, Association for Computing Machinery, New York, NY, USA, 2000, doi:10.1145/347090.347176.

[13] S. Chawla, "A novel approach of cluster based optimal ranking of clicked URLs using genetic algorithm for effective personalized web search," Applied Soft Computing, **46**, 90 – 103, 2016, doi:10.1016/j.asoc.2016.04.042.

[14] J. Sankhavara, "Feature Weighting in Finding Feedback Documents for Query Expansion in Biomedical Document Retrieval," SN Computer Science, **1**(5), 75–82, 2020, doi:10.1007/s42979-020-0069-x.

[15] A. Tombros, R. Villa, C. Van Rijsbergen, "The effectiveness of query-specific hierarchic clustering in information retrieval," Information Processing and Management, **38**(4), 559 – 582, 2002, doi:10.1016/S0306-4573(01)00048-6.

[16] F. Raiber, O. Kurland, F. Radlinski, M. Shokouhi, "Learning Asymmetric Co-Relevance," in Proceedings of the ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR), ACM - Association for Computing Machinery, 2015.

[17] S.-H. Na, "Probabilistic co-relevance for query-sensitive similarity measurement in information retrieval," Information Processing and Management, **49**(2), 558 – 575, 2013, doi:10.1016/j.ipm.2012.10.002.

[18] M. A. Hearst, J. O. Pedersen, "Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results," in Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '96, 76–84, Association for Computing Machinery, New York, NY, USA, 1996, doi:10.1145/243199.243216.

[19] "Query Clustering Using User Logs," ACM Trans. Inf. Syst., **20**(1), 59–81, 2002, doi:10.1145/503104.503108.

[20] R. Baeza-Yates, C. Hurtado, M. Mendoza, "Query Recommendation Using Query Logs in Search Engines," in W. Lindner, M. Mesiti, C. Türker, Y. Tzitzikas, A. I. Vakali, editors, Current Trends in Database Technology - EDBT 2004 Workshops, 588–596, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[21] C. Gutiérrez-Soto, G. Hubert, "Evaluating the Interest of Revamping Past Search Results," in H. Decker, L. Lhotská, S. Link, J. Basl, A. M. Tjoa, editors, Database and Expert Systems Applications, 73–80, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[22] C. Gutiérrez-Soto, G. Hubert, "On The Reuse of Past Searchesin Information Retrieval:Study of Two Probabilistic Algorithms," International Journal of Information System Modeling and Design, **6**(2), 71–90, 2015.

[23] L. Bergroth, H. Hakonen, T. Raita, "A survey of longest common subsequence algorithms," in Proceedings Seventh International Symposium on String Processing and Information Retrieval. SPIRE 2000, 39–48, 2000, doi:10.1109/SPIRE.2000.878178.

[24] G. Pass, A. Chowdhury, C. Torgeson, "A Picture of Search," in Proceedings of the 1st International Conference on Scalable Information Systems, InfoScale '06, 1–es, Association for Computing Machinery, New York, NY, USA, 2006, doi:10.1145/1146847.1146848.

[25] J. Guo, Y. Fan, L. Pang, L. Yang, Q. Ai, H. Zamani, C. Wu, W. B. Croft, X. Cheng, "A Deep Look into Neural Ranking Models for Information Retrieval," CoRR, **abs/1903.06902**, 2019.

[26] S. Jiang, S. Zilles, R. Holte, "Empirical Analysis of the Rank Distribution of Relevant Documents in Web Search," in 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, volume 1, 208–213, 2008, doi:10.1109/WIIAT.2008.200.

[27] A. Crystal, J. Greenberg, "Relevance criteria identified by health information users during Web searches," Journal of the American Society for Information Science and Technology, **57**(10), 1368–1382, 2006, doi:10.1002/asi.20436.

[28] A. Spink, J. Bateman, J. B. J., "Searching the Web: a survey of EXCITE users," Internet Research, **9**(2), 117–128, 1999, doi:10.1108/10662249910264882.

[29] D. C. Blair, "Information Retrieval, 2nd ed. C.J. Van Rijsbergen. London: Butterworths; 1979: 208 pp. Price: $32.50," Journal of the American Society for Information Science, **30**(6), 374–375, 1979, doi:10.1002/asi.4630300621.

[30] E. M. Voorhees, The effectiveness and efficiency of agglomerative hierarchical clustering in document retrieval, Ph.D. thesis, Cornell University, 1985.

[31] E. M. Voorhees, "The Cluster Hypothesis Revisited," in Proceedings of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '85, 188–196, Association for Computing Machinery, New York, NY, USA, 1985, doi:10.1145/253495.253524.

[32] R. Amorim, V. Makarenkov, B. Mirkin, "A-Ward pβ : Effective hierarchical clustering using the Minkowski metric and a fast k-means initialisation," Information Sciences, **370**, 342–354, 2016, doi:10.1016/j.ins.2016.07.076.

[33] C. J. van Rijsbergen, "An algorithm for information structuring and retrieval," The Computer Journal, **14**(4), 407–412, 1971, doi:10.1093/comjnl/14.4.407.

[34] O. Aichholzer, F. Aurenhammer, "Classifying Hyperplanes in Hypercubes," SIAM Journal on Discrete Mathematics, **9**(2), 225–232, 1996, doi:10.1137/S089548019426348X.

[35] H. Schütze, C. Silverstein, "Projections for Efficient Document Clustering," in Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '97, 74–81, Association for Computing Machinery, New York, NY, USA, 1997, doi:10.1145/258525.258539.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 894-903 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Method of Analysis and Classification of Acoustic Emission Signals to Identify Pre-Seismic Anomalies

Marapulets Yury, Senkevich Yury, Lukovenkova Olga, Solodchuk Alexandra*

Laboratory of Acoustic Research, Institute of Cosmophysical Research and Radio Wave Propagation, Far Eastern Branch of the Russian Academy of Sciences (IKIR FEB RAS), 684034, Russia

ARTICLE INFO

Article history:
Received: 19 August, 2020
Accepted: 25 November, 2020
Online: 08 December, 2020

Keywords:
Acoustic Emission
Symbol Description
Signal Alphabet
Pre-Seismic Anomaly

ABSTRACT

A new method of analysis and classification of rock acoustic emission signals is proposed. It is based on symbol description of signals and involves the following processing. First, signal segments containing pulses are detected. Second, noise of the detected pulses is reduced by the wavelet filtration method. Fourth-order symlets and adaptive threshold scheme based on empirical Bayes method are used. Application of wavelet filtration allows us to increase the signal-to-noise ratio by 9 dB, on the condition that initial SNR is 8.9 dB. Each pulse is described by a descriptive matrix which represents a square binary matrix reflecting pulse amplitude-phase structure and characterizing the positions of pulse extrema relatively each other. Based on the similarity degree of descriptive matrices, classes are formed. The proposed algorithm classifies correctly 78% of pulses. Each class is considered as a symbol and a set of obtained symbols forms an alphabet. Then, changes of alphabet parameters on subsequent fragments of the signal are studied. The developed method was realized in the form of an application program with the help of which acoustic emission signals were analyzed. The signals were recorded at "Karymshina" site (IKIR FEB RAS), located in a seismically active region, Kamchatka peninsula, in 2016–2019. It was discovered during the analysis that in case of an acoustic emission anomaly, alphabet content redistributed in favor of the symbols described by larger-size descriptive matrices (up to 350 extrema). At the same time, the number of such symbols increased. Analysis of alphabet averaged dynamics discovered a well exhibiting tendency of alphabet cardinality decrease relatively an average 10–18 days before earthquakes and recovery of the value over the period from 2 to 8 days after them. The obtained estimates of the pre-seismic anomaly occurrence periods are consistent with the results of early studies. The proposed method updates the existing methodology of short-term earthquake precursor formation. Thus, applying the developed method for analysis, it is possible to identify pre-seismic anomalies of acoustic emission that is topical in the creation of strong earthquake preparation warning system operating in automatic mode.

1. Introduction

Acoustic emission (AE) in solid bodies is elastic oscillations occurring in the result of dislocation changes in a medium. In such a case, the characteristics of generated pulse emission are directly associated with plastic process features. That determines the interest in the emission investigation with the aim to develop methods for medium acoustic diagnostics.

AE is applied for different problems of diagnosis in seismology, geophysics as well as in non-destructive testing of objects and materials in industry. At the beginning of the XXI century it was shown that acoustic emission anomalies are observed before strong earthquakes and can be consider as their operative precursors [1-3]. In further investigations [4, 5] the nature of generation of pre-seismic acoustic anomalies and the features of their manifestations were considered. In the course of the research carried out at IKIR FEB RAS, it was shown that these anomalies are generated by the change of geomechanics stress field in earthquake preparation zone and appear as the result of the transformation, stress – near-surface rock deformation [6, 7]. The peculiarity of anomaly manifestation is geoacoustic emission intensity increase in the frequency range from hundreds of hertz to

*Corresponding Author: Solodchuk Alexandra, Email: aleksandra@ikir.ru

the first tens of kilohertz. It usually lasts for to several hours and occurs 1–3 days before an earthquake.

Statistics of pre-seismic acoustic anomaly observations in a seismically active region of Kamchatka peninsula, which is a part of the Pacific Ring of Fire, shows high efficiency and prospectivity of AE method to forecast catastrophic earthquakes [6]. Detailed analysis of AE signals shows that they a represented by a sequence of relaxation oscillation pulses with steep edge and smooth drop of different frequency filling [6]. In these circumstances, both separate pulses and their groups with significant overlapping are observed. In case of intensive emission, pulses merge into a continuous signal.

During a background (seismically calm) period, acoustic pulses with the repetition frequency within 0.1–0.5 pulses per a second are usually observed. During the increase of stress and of rock deformation velocity before an earthquake, both pulse amplitude and their number per a time unit grow. Pulse repetition frequency reaches the value of tens and even hundreds per a second [6].

At the present time one of the main and still unsolved problems is automation of search and operative detection of AE pre-seismic anomalies. Application of popular (widely spread in the world) methods for data analysis becomes complicated by several reasons, they are:

- wide dynamic range of recorded data is associated with great volumes of recorded information requiring special methods for storage, transfer, processing and analysis;
- strong irregularity of natural mediums and bad propagation of elastic oscillation causes signal significant distortion and attenuation which limits the capabilities of investigation methods;
- strong noisiness both by natural and industrial noises makes the analysis of inner (morphological) structure of AE pulses more complicated;
- wide variety of signal waveforms complicates the problem of classification and requires analysis methods adapting for a given signal.

The paper proposes a new approach to analyze and classify rock acoustic emission signals in order to identify AE pre-seismic anomalies. The approach is based on the description of acoustic pulses from which a signal alphabet is formed. Then the changes of alphabet parameters of signal subsequent fragments are studied.

2. AE signal model

In a general case, acoustic emission signals can be represented by a classical adaptive model, i.e. as a sum of noise $\varepsilon(t)$ and some function $s(t)$, the analytical expression of which is unknown,

$$x(t) = s(t) + \varepsilon(t), \tag{1}$$

where $\varepsilon(t)$ describes nonstationary background noise containing noises of natural and artificial origin and $s(t)$ is a "useful" AE signal. It is a flux of pulses of different waveform, amplitude and frequency. Moreover, we accept the following constraint $||s(t)|| > ||\varepsilon(t)||$. Figure 1 illustrates an example of AE signal.

Within the framework of the proposed approach, pulses of a "useful" signal $s(t)$ can be described by the means of their singular points. As it is known from the functional analysis theory, when there are no fractures of the first and of the second kind, these points coincide with local extrema and local inflections. It is assumed that signal function intermediate values lying between the singular points are less informative and may be neglected in the course of the analysis.

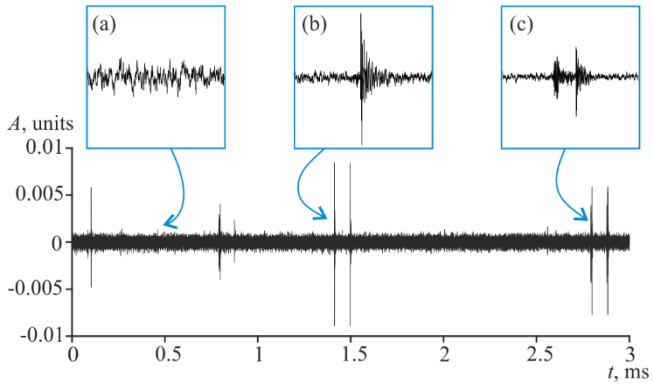


Figure 1: Acoustic emission signal: (a) – noise, (b), (c) – pulses

In order to understand the essence of further transformations, we consider an acoustic emission pulse signal (its graph is illustrated in Figure 2). We fix the values pulse local extreme amplitudes $\{a_i\} = a_0, a_1, \ldots, a_N$ and the time intervals between neighboring extrema $\{\tau_i\} = \tau_0, \tau_1, \ldots, \tau_{N-1}$. On the basis of these values we create a matrix of mutual relation of the fixed values (2). Matrices of **A** and **T** relation reflect pulse amplitude-phase structure. They characterize the positions of local extrema relatively each other and can be considered as a pulse model.

$$\mathbf{A} = \begin{pmatrix} a_0/a_0 & a_0/a_1 & \cdots & a_0/a_N \\ a_1/a_0 & a_1/a_1 & \cdots & a_1/a_N \\ \vdots & \vdots & \ddots & \vdots \\ a_N/a_0 & a_N/a_1 & \cdots & a_N/a_N \end{pmatrix}$$

$$\mathbf{T} = \begin{pmatrix} \tau_0/\tau_0 & \tau_0/\tau_1 & \cdots & \tau_0/\tau_{N-1} \\ \tau_1/\tau_0 & \tau_1/\tau_1 & \cdots & \tau_1/\tau_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{N-1}/\tau_0 & \tau_{N-1}/\tau_1 & \cdots & \tau_{N-1}/\tau_{N-1} \end{pmatrix} \tag{2}$$

We should note that if one knows any one value a_i, any one value τ_i and the relational matrix (2), it is possible to recover the initial waveform of a pulse more or less without loss.

To compose signal symbol description, the obtained pulse diversity is classified by comparing the relational matrices (2), i.e. based on the pulse waveform similarity criterion. However, such selection should take into account two key points: firstly, amplitude attenuation and phase distortions of acoustic pulses occurring during the propagation from a source to a receiver point; secondly, the distortions caused by the interference of additive noise on pulse waveforms at a receiver point. These problems inevitably result in significant increase of the number of pulse classes (redundant classification). The following transformations and computing operations allow us to solve, to a large extent, the problems arising during the classification.

When analyzing a pulse waveform, signal function numerical values at singular points often do not play the major role. In a

greater degree, to detect a pulse waveform, mutual position of these points is important [8]. In the framework of the applied approach, search for the differences in the observed pulse waveforms is of the main interest. Thus, relational matrices can be reduced to matrices of rank relation. In order to do that,it would be enough to transform matrices (2) into binary ones according to the following rule:

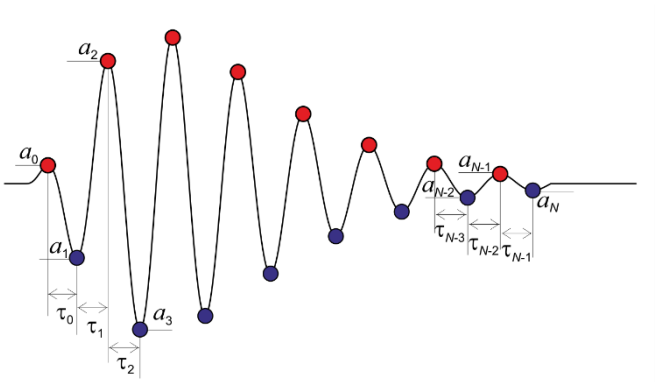


Figure 2: Local extrema of AE pulse

$$r_{i,j} = \begin{cases} 1, & a_i > a_j \\ 0, & a_i \le a_j \end{cases}, \qquad \omega_{i,j} = \begin{cases} 1, & \tau_i > \tau_j \\ 0, & \tau_i \le \tau_j \end{cases}, \tag{3}$$

where $r_{i,j}$ is the result of comparison of the i-th and j-th vales of extreme amplitudes; $\omega_{i,j}$ is the result of comparison of the i-th and j-th values of the intervals between the extrema. We create binary matrices of the comparison (3) results for amplitudes **R** and time intervals **W** separately.

$$\mathbf{R} = \begin{pmatrix} 0 & r_{0,1} & \cdots & r_{0,N} \\ r_{1,0} & 0 & \cdots & r_{0,N} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N,0} & r_{N,1} & \cdots & 0 \end{pmatrix}$$
$$\mathbf{W} = \begin{pmatrix} 0 & \omega_{0,1} & \cdots & \omega_{0,N-1} \\ \omega_{1,0} & 0 & \cdots & \omega_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{N-1,0} & \omega_{N-1,1} & \cdots & 0 \end{pmatrix} \tag{4}$$

The obtained binary square matrices **R** and **W** (4) of the sizes N+1 and N, respectively (intervals between local extrema are less by 1) have redundant character since, owing to inequality asymmetry (if $a > b$, then $b \le a$), we have $r_{i,j} = \neg r_{j,i}$ and $\omega_{i,j} = \neg \omega_{j,i}$, and all the elements lying on the main diagonal are equal to 0. In order to represent the pulse waveform more compactly, we join significant elements of both matrices. Transform the matrices **R** and **W** into upper triangular ones zeroing the redundant elements according to (5)

$$r_{i,j} = \begin{cases} r_{i,j}, & i < j \\ 0, & i \ge j \end{cases}, \qquad \omega_{i,j} = \begin{cases} \omega_{i,j}, & i < j \\ 0, & i \ge j \end{cases}. \tag{5}$$

As the result we obtain triangular binary matrices $\mathbf{R}_\Delta$ and $\mathbf{W}_\Delta$ (6).

$$\mathbf{R}_\Delta = \begin{pmatrix} 0 & r_{0,1} & r_{0,2} & \cdots & r_{0,N} \\ 0 & 0 & r_{1,2} & \cdots & r_{1,N} \\ 0 & 0 & 0 & \cdots & r_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & r_{N-1,N} \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$
$$\mathbf{W}_\Delta = \begin{pmatrix} 0 & \omega_{0,1} & \omega_{0,2} & \cdots & \omega_{0,N-1} \\ 0 & 0 & \omega_{1,2} & \cdots & \omega_{1,N-1} \\ 0 & 0 & 0 & \cdots & \omega_{2,N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \omega_{N-2,N-1} \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{6}$$

We reduce the matrix $\mathbf{R}_\Delta$ size to the value N by removing zero elements of the lower line and the left column. We make transposition operation of the matrix $\mathbf{W}_\Delta$ and as the result obtain the matrix $\mathbf{W}_\Delta^{\mathrm{T}}$. After summing up the matrices $\mathbf{R}_\Delta$ and $\mathbf{W}_\Delta^{\mathrm{T}}$, we obtain a square binary matrix **D** of size N (7). It can be used as a model of pulse amplitude-phase structure.

$$\mathbf{D} = \begin{pmatrix} r_{0,1} & r_{0,2} & \cdots & r_{0,N-1} & r_{0,N} \\ \omega_{0,1} & r_{1,2} & \cdots & r_{1,N-1} & r_{1,N} \\ \omega_{0,2} & \omega_{1,2} & \cdots & r_{2,N-1} & r_{2,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \omega_{0,N-2} & \omega_{1,N-2} & \cdots & r_{N-2,N-1} & r_{N-2,N} \\ \omega_{0,N-1} & \omega_{1,N-1} & \cdots & \omega_{N-2,N-1} & r_{N-1,N} \end{pmatrix} \tag{7}$$

We call (7) a pulse descriptive matrix. An example of a pulse and its descriptive matrix are illustrated in Figure 3.

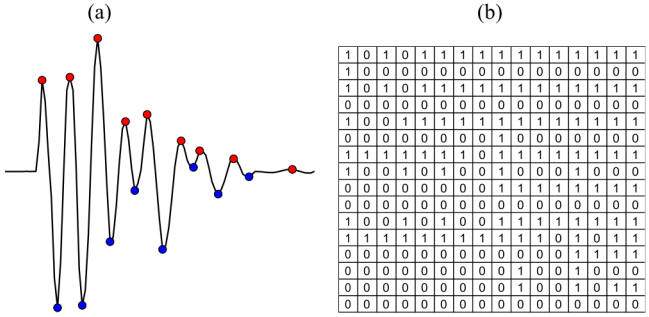


Figure 3: Representation of a pulse by descriptive matrix: (a) – pulse with detected extrema; (b) – pulse descriptive matrix

Descriptive matrices have invariant character under time and amplitude transformations that is proved by the following properties:

- invariance under time compression or expansion (if $k > 0$ and $\tau_i > \tau_j$, then $\tau_i \cdot k > \tau_j \cdot k$);
- invariance under amplitude compression or expansion (if $k > 0$ and $a_i > a_j$, then $a_i \cdot k > a_j \cdot k$).

The invariance is retained, if the coefficient k is substituted by any monotone positive function $f(x) > 0$. Attenuation processes

and nonlinear amplitude-phase distortions, changing AE pulse signal waveforms during its propagation in heterogeneous anisotropic mediums from a source to a receiver, have the character of monotone functions in the majority of practical cases. Thus, descriptive matrix invariance is of great importance when creating algorithms for identification and classification of acoustic emission pulses.

3. Model adequacy

To verify the proposed model adequacy, we generated two artificial AE signals.

It was shown in the paper [9] that AE typical pulses are described by modulated Berlage functions with a reasonable degree of accuracy

$$u(t) = A \cdot t^{n(p_{\max})\cdot\Delta} \cdot \exp\left(-\frac{n(p_{\max})\cdot\Delta}{p_{\max}\cdot t_{end}}\cdot t\right)\cdot\sin(2\pi ft+\varphi_0), \quad (8)$$

where A is the pulse amplitude; t is the time, $0 \le t \le t_{end}$; p_{max} is the pulse envelope maximum position defined relating to t_{end}; $n(p_{max})$ is the minimum value of the parameter affecting pulse envelope slope which is chosen so that $u(p_{max} \cdot t_{end}) \le 0.05 \cdot u(t_{end})$; Δ is the coefficient of pulse envelope slope variation, the more Δ value is, the steeper the envelope is, $\Delta > 1$; f is the pulse modulation frequency; φ_0 is the initial phase. To obtain different pulse waveforms, the parameter values p_{max}, Δ, t_{end}, f, φ_0 vary.

The first artificial signal (Figure 4a) contains five pulsesof similar waveform and the second one contains five pulsesof different waveform (Figure 4b). As long as the model should be invariant under amplitude and time transformations of the signal, the length and the amplitude of each pulse were drawn at random. Instead of the frequency f, the number of periods k on the time interval from 0 to t_{end} was defined. Parameters of the generated pulses are listed in Table 1.

Table 1: Parameters of test pulses

Wave-form	Number of periods, k	Maximum position, p_{max}	Steepness, Δ	Initial phase, φ_0
I	10	0.2	1.5	0
II	5	0.1	1.2	0
III	15	0.25	2	0
IV	11	0.01	1	$\pi/3$
V	20	0.4	1.5	0

Signal graphs and modeling results are illustrated in Figure 4a–b. Pulses having similar descriptive matrices are shown by the same color. Thus, application of the proposed model allowed us to determine correctly the similarity/difference of pulse waveforms. We should note that the pulses of the first signal have different frequency-time structures (Figure 4c), and would be attributed to different classes when applying frequency-time methods. Pulses of the second signal have the same frequency (Figure 4d), but different waveforms and would be attributed to the same class when applying frequency-time methods.

Figure 5 shows the real AE signal containing five pulses. Three pulse waveforms are distinguished visually. The suggested model describes real signals adequately. Pulses of the same waveform have similar descriptive matrices.

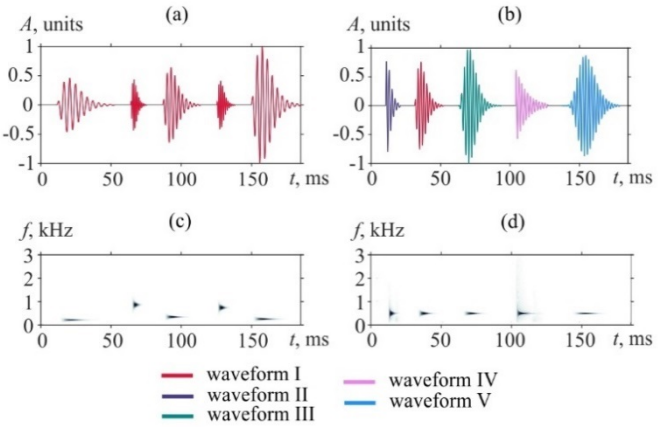


Figure 4: Modeling results for two signals. Pulses described by the same descriptive matrix are of the same color: (a) – signal containing five pulses of the same waveform; (b) – signal containing five pulses of different waveform; (c) – Wigner-Ville transformations of the first signal pulses; (d) – Wigner-Ville transformations of the second signal pulses

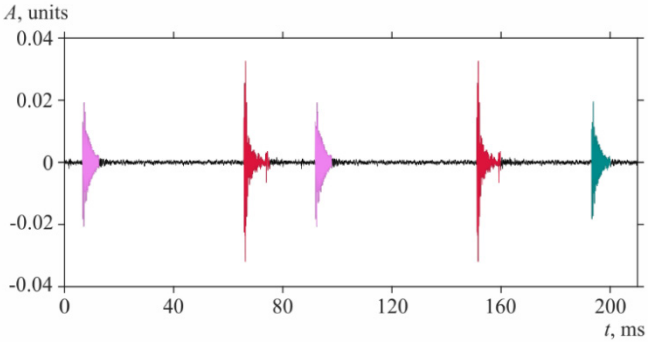


Figure 5: Modeling results for real AE signal. Pulses described by similar descriptive matrices are of the same color

In the considered examples we analyze the signals in which noises are either absent or insignificant. Real AE signals often contain noises of artificial and natural origin, thus perturbing a signal. Distortion may result in a gap, false detection or amplitude change of some extrema. Then the size and some elements of pulse descriptive matrices differ (Figure 6).

4. Classification of pulse diversity

We define initial conditions for classification. We put bounds on the size of descriptive matrices detected during the analysis by the values N_{min} and N_{max} from the bottom and from the top, respectively. Analysis epoch is the length of the time window in which an AE signal is observed. In practice, this index should provide formation of stable classes. N_{min} is determined by the minimum number of oscillations during pulse duration and cannot be set less than 3. N_{max} is defined according to the capabilities of a computing device and the maximum run-time.

Pulses are classified accorded to the results of calculation of descriptive matrix similarity coefficient. First, we restrict ourselves to descriptive matrices of the same size. In order to do that we divide descriptive matrices into groups $\mathbf{D}^{(Z)}$ of matrices of the same size Z, $N_{min} < Z < N_{max}$.

For the two descriptive matrices of the same size $\mathbf{D}_1^{(Z)}$ and $\mathbf{D}_2^{(Z)}$ the similarity coefficient $g(\mathbf{D}_1^{(Z)}, \mathbf{D}_2^{(Z)})$ is determined as a ratio of elements, coinciding in value and taking the same positions, to the

total number of elements in the matrix. In case of exact identity, the similarity coefficient is 1.

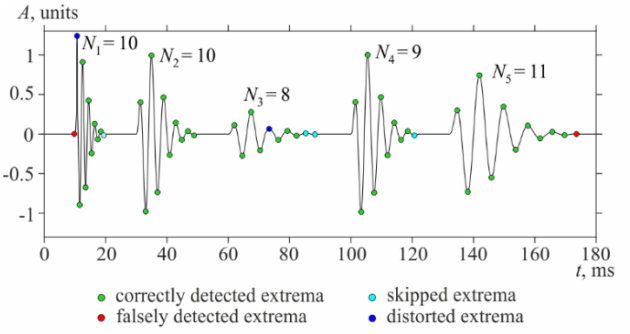


Figure 6: Examples of errors occurring during extreme detection

$$g(\mathbf{D}_1^{(Z)}, \mathbf{D}_2^{(Z)}) = \frac{\#(d_{1ij} = d_{2ij})}{Z^2}, \quad (9)$$

where #(*condition*) is the operator for calculation of the number of elements satisfying the condition, d_{1ij} are the elements of the matrix $\mathbf{D}_1^{(Z)}$, d_{2ij} are the elements of the matrix $\mathbf{D}_2^{(Z)}$.

The calculated similarity coefficient $g(\mathbf{D}_1^{(Z)}, \mathbf{D}_2^{(Z)})$ is compared with the selected threshold value G_0 (it is chosen empirically). If

$$g(\mathbf{D}_1^{(Z)}, \mathbf{D}_2^{(Z)}) > G_0,$$

then the pulses described by $\mathbf{D}_1^{(Z)}$ and $\mathbf{D}_2^{(Z)}$ belong to the same class.

Classification is the comparison of descriptive matrices of the same size and selection of the empirical threshold G_0 ($0 < G_0 \leq 1$) with which the similarity coefficient (9) is compared.

The possibility that two pulses of the same source will have exactly the same number of local extrema is very low. So, the obtained classification is too restrictive. Thus, we developed a procedure to compare descriptive matrices of different sizes. It takes into account false extrema or gaps of some extrema and allows us to reduce the number of detected classes. However, when descriptive matrix sizes are widely scattered, they are compared in erratic manner that may cause random merging of classes described by descriptive matrices of smaller sizes. We put bounds on the possibility for merging of descriptive matrices of smaller sizes by the threshold coefficient S_0:

$$N_L / N_M \geq S_0, \quad 0 < S_0 \leq 1,$$

where N_L and N_M are the sizes of smaller and larger descriptive matrices, respectively, S_0 is the empirically defined threshold.

Assume that there are two descriptive matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ of the sizes N_1 and N_2, respectively ($N_1 \geq N_2$). If $N_2/N_1 \geq S_0$ ($N = S_0 \cdot N_1$), then the matrices are comparable. For the meaningful comparison, the matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ are matched so that the latest N elements of the main diagonal $\mathbf{D}_2$ are overlapped on the first N elements of the main diagonal $\mathbf{D}_1$. Then we calculate the similarity coefficient (9) of matrices of size N in the domain bounded by the intersection. Elements are further compared after $\mathbf{D}_2$ shift along the main diagonal $\mathbf{D}_1$. The similarity coefficient (9) is calculated for each subsequent shift so far as the matrices in the domain bounded by the intersection have the size of not less than N (Figure 7). The comparison result is the maximum value of the similarity coefficient.

If the maximum similarity coefficient (9) exceeds the empirically defined threshold G_0 or is equal to it, the decision on sufficient closeness of descriptive matrices of compared pulses is made. In this case, the pulse with descriptive matrix of the smaller size is included into the class described by larger-size descriptive matrix (larger object absorbs a smaller one).

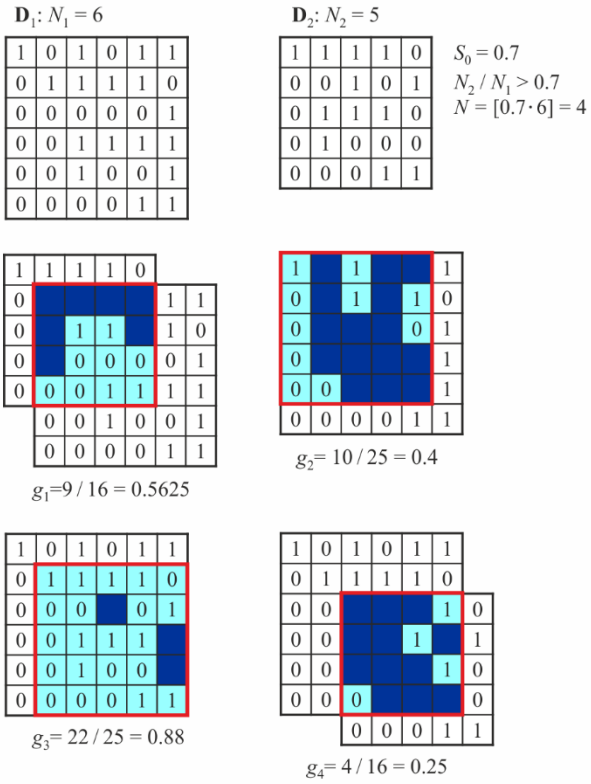


Figure 7: Comparison of descriptive matrices of sizes 6 and 5

In the result of classification many classes are formed. Each of them unites the pulses close in amplitude-phase structure with the accuracy up to the values determined by the coefficients G_0 and S_0. Each class can be considered as some symbol and the whole set of obtained symbols composes a signal alphabet.

In order to estimate the classification algorithm noise immunity for different values of the thresholds G_0 and S_0, we realized a simulation experiment. We processed the data containing 100 pulses of similar waveform (the alphabet cardinality is 1 symbol) with signal/noise ratio (SNR) from ∞ to 2.7 dB.

At the first stage of the experiment we compared the descriptive matrices with sizes differing by less than 30% ($S_0 \geq 0.7$). Similarity coefficient threshold was chosen to be $G_0 = 1$ (case when a matrix of smaller size is the sub-matrix of a larger-size matrix).

Figure 8 shows the dependences of alphabet cardinality on SNR for different values of the coefficient S_0 (0.7–1). Insignificant difference of the graphs proves that descriptive matrix element values significantly depend on local influence of noise on pulse waveform and are weakly sensitive to changes in the threshold S_0 at a high value G_0. When SNR values are less than 30 dB, the

classification does not work since matrices do not match with 100% agreement ($G_0 = 1$).

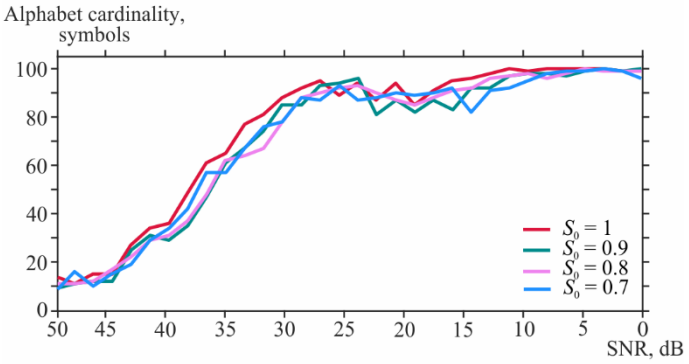


Figure 8: Dependence of alphabet cardinality on SNR for different values of the threshold S_0 (0.7–1) and for a fixed valued of the threshold G_0 =1

At the next stage of the computational experiment, we investigated class formation during threshold G_0 changes (from 0.6 to 0.9). In this case, matrices of the same size ($S_0 = 1$) were compared. The results of the second stage of the computational experiment are illustrated in Figure 9. The graphs show that decrease of the threshold G_0 causes significant reduction of alphabet cardinality. Moreover, the number of detected classes reduces with G_0 decrease.

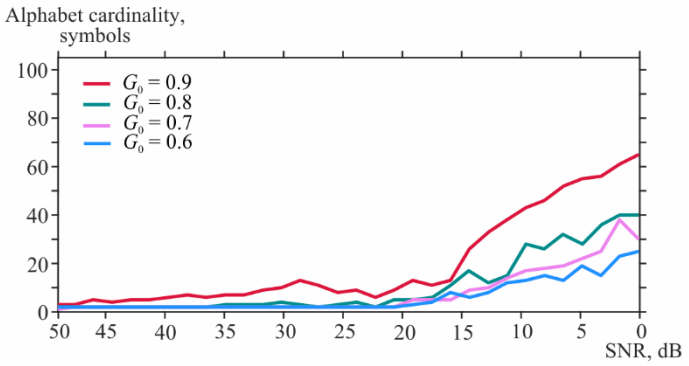


Figure 9: Dependences of alphabet cardinality on SNR for different values of the threshold G_0 (0.6–0.9) and for a fixed value of the threshold $S_0 = 1$

At the third stage of the computational experiment we investigated the dependence of alphabet cardinality on SNR for $S_0 = 0.7$ and different values of G_0. The results are shown in Figure 10.

The experiment shows that the classification method is nonlinearly sensitive to the choice of threshold functions of G_0 and S_0: decrease of S_0 causes pulse redistribution into classes with larger-size descriptive matrices and decrease of G_0 makes it possible to compact classes. Based on the experiment results, the recommended values of the thresholds S_0 (from 0.6 to 0.9) and G_0 (from 0.7 to 0.9) were selected.

As long as the developed method of classification is expectedly sensitive to the performance quality of the algorithms for pulse detection and noise reduction, one of the important tasks in AE signal investigation is the improvement of methods for signal pre-processing.

Classical digital filtration, empirical mode decomposition (EMD), and wavelet filtration were considered as methods of noise reduction and detection of signal useful component. Digital filtration was not used because the AE pulses have short duration (100–400 samples per pulse). Problem of mode functions selection complicated the usage of EMD for signal denoising (for example, in [10]), since rejected modes are different for different pulses. So, the authors chose wavelet filtration technology with adaptive threshold scheme. Wavelets are widely used in various signal processing tasks [11], in particular in noise reduction tasks [12, 13]. Denoising has three stages: direct wavelet transforms, application of a selected threshold scheme to every detailing component, inverse wavelet transforms. In order to choose a wavelet family, a threshold scheme and an algorithm for calculating threshold value, a computational experiment was carried out on model and real AE signals. The best results were obtained applying fourth-order symlets (maximum possible level of decomposition), median threshold scheme and empirical Bayes rule to calculate the threshold value [14, 15]. The selected wavelet filtration method allows us to increase the signal-to-noise ratio by about 9 dB, on the condition that initial SNR is 8.9 dB.

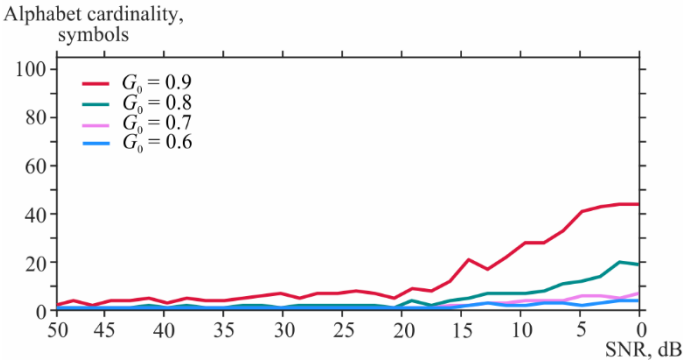


Figure 10: Dependences of alphabet cardinality on SNR for different values of the threshold G_0 (0.6–0.9) and for a fixed value of $S_0 = 0.7$

Figure 11a–b illustrates the process of detection of local extrema in a noised pulse and in a pulse with additive noise. Figure 11c shows how the wavelet filtering algorithm recovers the waveform of a perturbed signal.

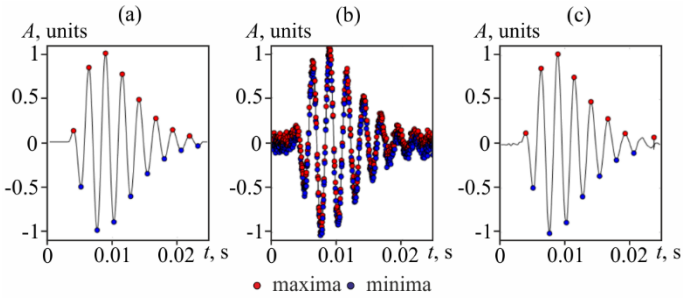


Figure 11: Noise effect on detection of local extrema: (a) – initial pulse; (b) – pulse with overlapped white noise, SNR = 12.4 dB; (c) – pulse denoised by wavelet filtration, SNR = 22.7 dB

Application of wavelet filtering with adaptive threshold schemes allows us to apply the proposed methods of analysis to perturbed and noisy signals.

Figure 12 illustrates the results of classification of model signal pulses with the signal/noise ratio of 8.9 dB. The signal contains 50 Berlage pulses described by three waveforms the parameters of which are listed in Table 3. Before being classified, each detected pulse was denoised by the selected method of wavelet filtering (after wavelet filtration, the SNR of a single pulse is about 20 dB).

When constructing descriptive matrices, we did not take into account the extrema the amplitudes of which were less than 15% from the maximum one. Classification algorithm thresholds were selected as $S_0 = 0.6$ and $G_0 = 0.7$. For a noisy signal (Figure 12a) we succeeded to define correctly the number of classes and to structure the pulses into classes (Figure 12b). In the course of multiple repetition of the experiment it was discovered that the obtained structuring into the classes is stable to the changes of initial conditions of white noise generator.

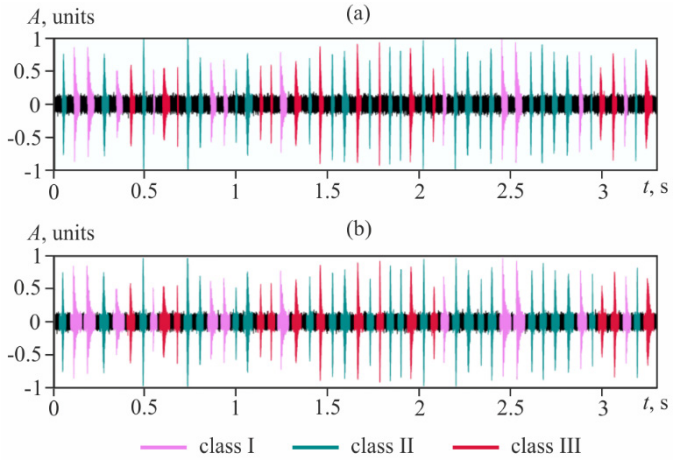


Figure 12: Classification results: (a) – signal with overlapped white noise and initial structuring into classes; (b) – classification of $S_0 = 0.6$, $G_0 = 0.7$; 3 classes were defined

Table 2: Waveform parameters

Class	Number of periods, k	Maximum position, p_{max}	Steepness, Δ	Initial phase, φ_0
I	7	0.05	1.1	0
II	20	0.35	3	$\pi/3$
III	40	0.2	2	$\pi/2$

The proposed algorithm classifies correctly 78% of pulses, on the conditions that thresholds are selected as $S_0 = 0.6$, $G_0 = 0.7$ and SNR is 8.9 dB before wavelet denoising. This estimate is based on the results of 100 experiments.

Efficiency of the proposed classification method can be estimated by a contraction coefficient

$$b = \frac{|A_{input}|}{|A_{output}|}$$

where $|A_{input}|$ is the alphabet cardinality (number of pulses) before classification, $|A_{output}|$ is the cardinality of the alphabet obtained as the result of classification.

Table 3 illustrates the classification results for 12 files containing 15-minute records of AE signals from IKIR FEB RAS archive data for January 1, 2018. The classification was carried out with the coefficients $G_0 = 0.8$, $S_0 = 0.8$.

As the result of further processing of AE data for the period 2016-2019, it was determined that contraction coefficient (10) varies from 50 to 200.

The developed methods of pulse description and classification were realized in the form of software. Preprocessing procedures such as centering, normalization, and noise reduction were performed by MATLAB system.

Table 3: Classification results

File number	Alphabet cardinality, $\|A_{input}\|$	Alphabet cardinality, $\|A_{output}\|$	Contraction coefficient, b
1	2864	37	77.4
2	3233	36	89.8
3	3581	46	77.8
4	3986	43	92.6
5	4593	45	102.0
6	3132	46	68.0
7	3778	43	87.9
8	4424	38	116.4
9	4836	36	134.3
10	3968	39	107.7
11	4734	36	131.5
12	3475	44	78.9

5. Examples of identification of pre-seismic AE anomalies

The developed method of classification was realized in the form of an application program. It was used to process and to analyze acoustic emission signals recorded in 2016–2019 at "Karymshina" site of IKIR FEB RAS (52.83º N, 158.13º E). AE signals are stored in the database in the form of 15-minute sound files (WAVE format). In the course of processing of each file, signal alphabet was constructed. Then alphabet symbols were ranked according to the size of their descriptive matrices and frequency (number of pulses described by the symbols).

Based on the conception of the relation of AE signal parameters with the state of a system generating a signal, we should analyze quantitative and qualitative parameters of symbol description of a signal for different time periods (from a day to a year) and to compare the defined anomalies with occurred earthquakes according to the data of regional seismic catalogue (Kamchatka Branch of the FRC GS RAS, Earthquake Catalogue for Kamchatka and Kuril Islands (1962 – present time), http://sdis.emsd.ru/info/earthquakes/catalogue.php).

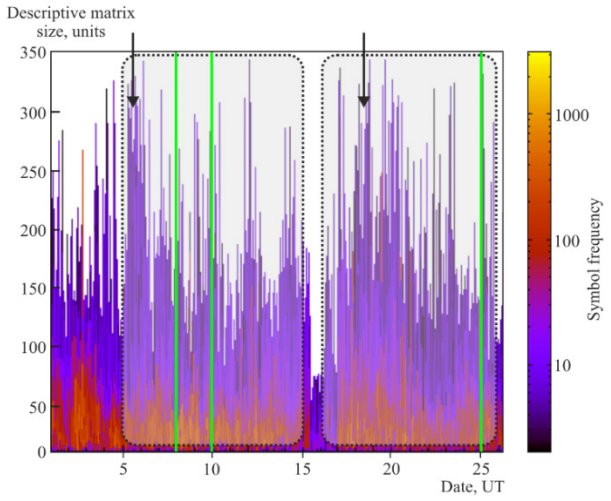


Figure 13: Anomalous dynamics of AE signal alphabet in October 2017

In order make the analysis, alphabets of each 15-minute AE signal were formed. Then three-dimensional images of a special type were drawn. Descriptive matrix size was measured on the vertical axis. Time (with 15-minute step) was measured on the horizontal axis. Color denotes symbol frequency.

Examples of anomalies observed 1–12 days before earthquakes and 1–6 days after earthquakes are illustrated in Figures 13–15.

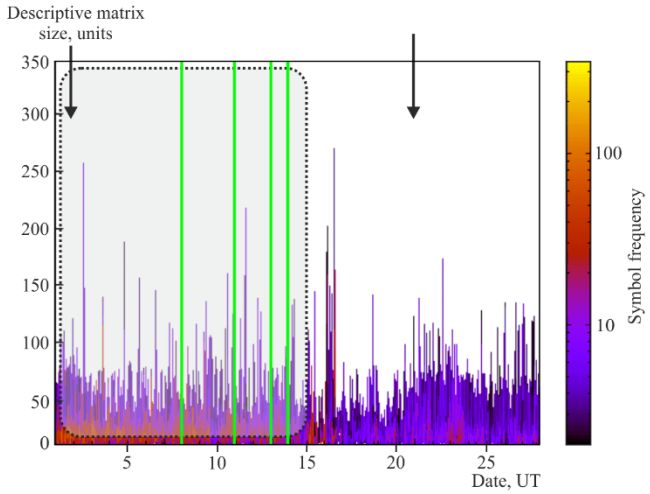


Figure 14: Anomalous dynamics of AE signal alphabet in February 2019

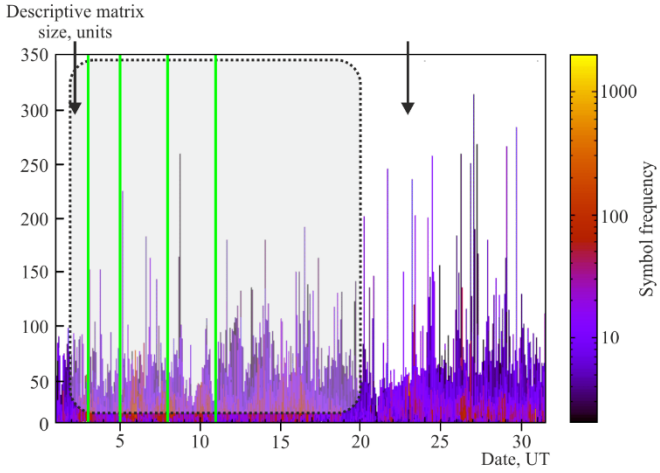


Figure 15: Anomalous dynamics of AE signal alphabet in October 2019

Anomaly peaks in the Figures are marked by black arrows. Green lines indicate earthquake times. These anomalies are alphabet content redistribution in favor of the symbols described by larger-size descriptive matrices (up to 350 extrema). Simultaneously with that the number of such symbols increases. Alphabet cardinality also grows from 20 to 150. Assumed areas of anomalies associated with earthquakes are marked by rounded rectangles. Short-term changes of alphabet content (less than 305 peaks) were not taken into consideration.

We should note that not all the analyzed AE anomalies in the form of redistribution of alphabet symbol frequency are associated with earthquake preparation processes. In the experiment the acoustic emission is recorded in sedimentary rocks and is a sensitive indicator of their stress-deformed state. Both global tectonic processes (for example, earthquake preparation) and local deformations, associated with the impact of different natural factors, may influence that state. Thus, not all the recorded AE anomalies are pre-seismic.

Figures 13–15 illustrate examples of AE anomalies observed during the periods when earthquakes occurred quite rarely, from one to four strong events per a month. Such a choice allows us to demonstrate the relation of detected anomalies with earthquakes. When earthquakes repeat quite frequently, the anomalous behavior is not so clear that may be explained by overlapping of fragments of evoked anomalous effects. Example of such behavior of alphabet content dynamics is shown in Figure 16.

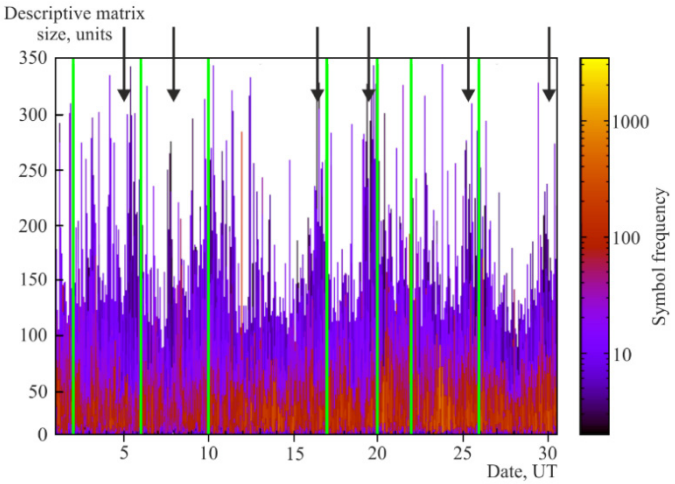


Figure 16: Anomalous dynamics of AE signal alphabet in April 2017

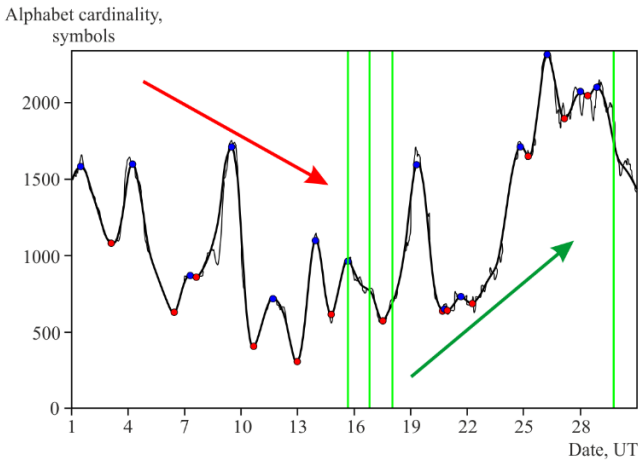


Figure 17: Graphs of AE signal alphabet cardinality dynamics in September 2017

Further analysis of the dynamics of AE signal alphabet parameters allowed us to find some regularities associated with the anomalies of the envelope of day-averaged alphabet cardinality. For the averaging we applied a sliding window of the width of 96 values of alphabet current cardinality. As an example, Figures 17–19 illustrate alphabet cardinality dynamics for one month, September 2017, December 2018 and September 2019. Thin gray line shows the result of alphabet averaging and the thick black line shows the result of approximation of the least square method. The graphs clearly show an oscillating process with small uncertain deviations of phase and amplitude the average period of which is from 2 to 3 days. The nature of this oscillating process was not determined and is under the study now. Blue dots on the graph are local maxima of detected oscillations. Red dots are their

minimums. Vertical green lines show the earthquake times. Red arrows indicate alphabet transformation tendencies before a series of earthquakes. Green arrows indicate transformation tendencies after a series of earthquakes. We can clearly distinguish the tendency of decrease of alphabet cardinality relatively the average 10–18 days before earthquakes and recovery of the value over the period from 2 to 8 days after them. This decrease is evidently determined by the fact that acoustic process before an earthquake becomes significantly more nonstationary. More pulses of similar waveforms appear and, consequently, alphabet set gets narrow.

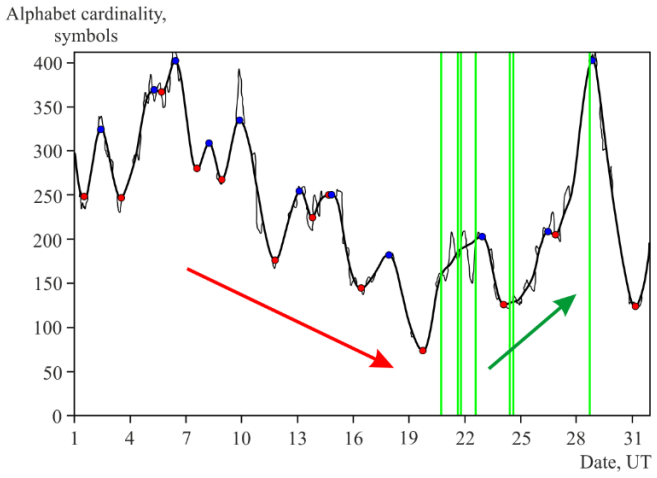


Figure 18: Graphs of AE signal alphabet cardinality dynamics in December 2018

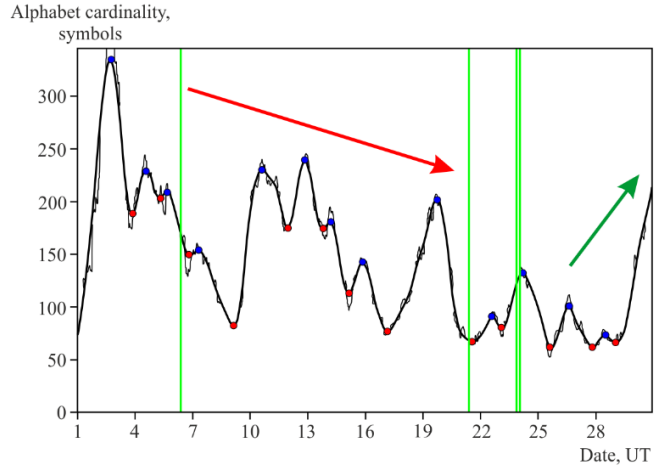


Figure 19: Graphs of AE signal alphabet cardinality dynamics in September 2019

The obtained estimates of the pre-seismic anomaly occurrence periods are consistent with the results of early studies. It was shown in [1, 2] that AE intensity increases at the frequency of 25 kHz 2–3 days before strong earthquakes. According to [16, 17], anomalies in the form of signal level increase appear in the frequency range of more than 6.5 kHz in the AE signals accumulated by 1 s. The results of time-frequency analysis of single AE pulses registered 1–7 days before earthquakes are presented in [18, 19]. According to these results, the number of pulses of a certain frequency increases before strong earthquakes. The above methods analyze signal frequency spectrum and, unlike the one described in the article, do not take into account waveform features. The proposed method of AE signals analysis updates the existing methodology and allows us to form more accurate short-term precursors of strong earthquakes.

Note that at all stages of the investigation, we had to face known nonlinear manifestations of the natural system under the study. These manifestations reflected in the chaotic character of AE signal dynamic characteristics. As it had been expected, the attempt of application of traditional methods for signal statistical processing, such as calculation of probability density function, ensemble averaging, estimation of statistic parameters of different orders, to nonstationary AE signals did not provide stable results. Thus, the attention was focused on the search for invariant transformations which would be applied to AE chaotic character properly. On the whole, we think that we are at the beginning of a perspective survey which, owing to the chosen approach and the developed computer tools for analysis, will allow us to investigate deeper and to extend general and personal knowledge on seismic process behavior and, consequently, to predict catastrophic effects of earthquakes at different stages and with more confidence.

6. Conclusions

A new method for analysis and classification of rock acoustic emission signals has been proposed. It is based on their symbol description. First of all, signal sections containing pulses are determined. Each of the sections is described by a descriptive matrix. Based on the similarity degree of the descriptive matrices, pulses are structured into classes. Each class is considered as a symbol and a set of obtained symbols composes an alphabet. Then we study the changes of alphabet parameters on signal subsequent fragments.

The developed method allows us to:

- distinguish informative features in AE signals based on their qualitative peculiarities;
- solve the problem of processing of a large volume of data by unifying pulse description and systematization;
- identify anomalies in AE signals distinguishing pre-seismic ones.

The method was implemented as an application program by the means of which we analyzed acoustic emission signals which were recorded at “Karymshina” site of IKIR FEB RAS in the seismically active region, Kamchatka peninsula, in 2016-2019. It was discovered during the analysis that in case of acoustic emission anomaly occurrence, alphabet content is redistributed in favor of the symbols described by larger-size descriptive matrices (up to 350 extrema). Simultaneously with that the number of symbols increases. Analysis of alphabet averaged dynamics detected a clear tendency of alphabet cardinality decrease relatively the average 10–18 days before earthquakes and recovery of the value over the period from 2 to 8 days after them.

Thus, a new approach to the analysis of acoustic emission signals has been developed. The approach updates the existing methodology of short-term earthquake precursor formation. On the basis of it, a system for automatic identification of pre-seismic AE anomalies can be created. It is topical for prediction of strong earthquakes which are catastrophic natural phenomena.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The research was supported by the Russian Science Foundation, Project No. 18-11-00087.

References

[1] G. Paparo, G. P. Gregori, U. Coppa, R. De Ritis, A. Taloni, "Acoustic Emission (AE) as a diagnostic tool in geophysics" Ann. Geophys-Italy, **45**(2), 401–416, 2002. https://doi.org/10.4401/ag-3511

[2] G. P. Gregori, G. Paparo, M. Poscolieri, A. Zanini, "Acoustic emission and released seismic energy" Nat. Hazard Earth Sys., **5**, 777–782, 2005. https://doi.org/10.5194/nhess-5-777-2005

[3] A. V. Kuptsov, "Variations in the geoacoustic emission pattern related to earthquakes on Kamchatka" Izv-Phys. Solid Eart+, **41**(10), 825–831, 2005.

[4] G. I. Dolgikh, V. A. Shvets, V. A. Chupin, S. V. Yakovenko, A. V. Kuptsov, I. A. Larionov, Y. V. Marapulets, B. M. Shevtsov, O. P. Shirokov, "Deformation and acoustic precursors of earthquakes" Dokl. Earth Sci., **413**(1), 281–285, 2007. https://doi.org/10.1134/S1028334X07020341

[5] G. P. Gregori, M. Poscolieri, G. Paparo, S. De Simone, C. Rafanelli, G. Ventrice, "Storms of crustal stress" and AE earthquake precursors" Nat. Hazard Earth Sys., 10, 319–337, 2010. https://doi.org/10.5194/nhess-10-319-2010

[6] Yu. V. Marapulets, B. M. Shevtsov, I. A. Larionov, M. A. Mishchenko, A. O. Shcherbina, A. A. Solodchuk, "Geoacoustic emission response to deformation processes activation during earthquake preparation" Russ. J. Pac. Geol., **6**(6), 457–464, 2012. https://doi.org/10.1134/S1819714012060048

[7] I. A. Larionov, Y. V. Marapulets, B. M. Shevtsov "Features of the Earth surface deformations in the Kamchatka peninsula and their relation to geoacoustic emission" Solid Earth, 5, 1293–1300, 2014. https://doi.org/10.5194/se-5-1293-2014

[8] J. Serra, Image Analysis and Mathematical Morphology, Academic Press, 1982.

[9] A. B. Tristanov, Yu. V. Marapulets, O. O. Lukovenkova, A. A. Kim, "A new approach to study of geoacoustic emission signals" Pattern Recognit. Image Anal., 26, 34–44, 2016. https://doi.org/10.1134/S1054661816010259

[10] Yu. I. Senkevich, "The use of the empirical mode decomposition method to clean and restavration acoustic emission signal" E3S Web Conf., **62**, 03008, 2018. https://doi.org/10.1051/e3sconf/20186203008

[11] L. Alperovich, L. Eppelbaum, V. Zheludev, J. Dumoulin, F. Soldovieri, M. Proto, M. Bavusi, A. Loperte, "A new combined wavelet methodology applied to GPR and ERT data in the Montagnole experiment" J. Geophys. Eng., **10**(2), 1–17, 2013. https://doi.org/10.1088/1742-2132/10/2/025017

[12] O. Mandrikova, Yu. Polozov, N. Fetisova, T. Zalyaev, "Analysis of the dynamics of ionospheric parameters during periods of increased solar activity and magnetic storms" J. Atmos. Sol. Terr. Phy., **181**, 116–126, 2018. https://doi.org/10.1016/j.jastp.2018.10.019.

[13] O. Mandrikova, N. Fetisova, "Modeling and analysis of ionospheric parameters based on multicomponent mode" J. Atmos. Sol. Terr. Phy., **208**, 105399, 2020. https://doi.org/10.1016/j.jastp.2020.105399

[14] I. M. Johnstone, B. W. Silverman, "Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences" Ann. Statist., **32**(4), 1594–1649, 2004. https://doi.org/10.1214/009053604000000030

[15] I. M. Johnstone, B. W. Silverman, "Empirical Bayes selection of wavelet thresholds" Ann. Statist., **33**(4), 1700–1752, 2005. https://doi.org/10.1214/009053605000000345

[16] Y. Marapulets, O. Rulenko, "Joint Anomalies of High-Frequency Geoacoustic Emission and Atmospheric Electric Field by the Ground–Atmosphere Boundary in a Seismically Active Region (Kamchatka)" Atmosphere, **10**(5), 267, 2019. https://doi.org/10.3390/atmos10050267

[17] O. P. Rulenko, Yu. V. Marapulets, Yu. D. Kuz'min, A. A. Solodchuk, "Joint Perturbation in Geoacoustic Emission, Radon, Thoron, and Atmospheric Electric Field Based on Observations in Kamchatka" Izv-Phys. Solid Eart+, **55**(5), 766–776, 2019. https://doi.org/10.1134/S1069351319050094

[18] O. Lukovenkova, A. Solodchuk, "Analysis of geoacoustic emission and electromagnetic radiation signals accompanying earthquake with magnitude $M_w = 7.5$" E3S Web Conf., **196**, 03001, 2020. https://doi.org/10.1051/e3sconf/202019603001

[19] O. Lukovenkova, A. Solodchuk, A. Tristanov, E. Malkin, "Complex analysis of pre-seismic geoacoustic and electromagnetic emission signals" E3S Web Conf., 127, 03001, 2019. https://doi.org/10.1051/e3sconf/201912703001

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 904-911 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

A Fast Adaptive Time-delay-estimation Sliding Mode Controller for Robot Manipulators

Dang Xuan Ba*

Department of Automatic Control, Ho Chi Minh City University of Technology and Education (HCMUTE). Ho Chi Minh City, 71307, Vietnam

ARTICLE INFO

Article history:
Received: 14 August, 2020
Accepted: 25 November, 2020
Online: 08 December, 2020

Keywords:
Adaptive time-delay estimation
Gain adaptation
Robotic manipulators
Sliding mode control

ABSTRACT

A nonlinear adaptive robust controller is proposed in this paper for trajectory-tracking control problems of robot manipulators. On one hand, to effectively approximate the systematic dynamics, a simple time-delay estimator is first adopted. On the other hand, to minimize the control error, the controller is designed based on a sliding mode structure using the obtained estimation results. A fast learning mechanism is then proposed for automatically tuning control gains. Another proper adaptation law is furthermore developed to support the nominal inertia-matrix selection of the time-delay estimation. Effectiveness of the closed-loop system is intensively discussed using Lyapunov-based constraints and extended simulation results.

1. Introduction

Nowadays, robots play a crucial role in modern industry thanks to excellent performances in conducting complicated, repetitive, and dangerous tasks [1–4]. Day by day, high-precision intelligent robots are required to fulfill the given demands [2], [5]. To design such controllers, two important steps of the controller design are taken into account: nonlinearities and uncertainties in system models as well as unexpected effects of complex working environments have to be tackled to push the closed-loop system in a stationary state; in the second step, a driving robust control signal is generated to force the control objective stabilizing around origin [6–8].

Robot dynamics could be derived using Newton-Euler analyses, Lagrange formulations or virtual decomposition methods [6,9]. Although general systematic models could be obtained, but they contain lots of uncertainties due to particular structures of robots and disturbances [10,11]. To cope with the mathematical problems, soft-computing approaches are reasonable solutions [7,8]. Nonlinear uncertain dynamics could be well estimated by fuzzy-logic approximators and neural networks but heavy-computation burdens along with lots of tuning parameters are new obstacles in real-time applications. The time-delay estimation (TDE) is increasingly adopted as one of the most effective tools to reckon unknown dynamics in modern control, owing to the simplicity in design and easiness in real-time implementation [12,13]. This technique uses measured information of the acceleration and control signals to result in offset terms of the dynamics based on predefined nominal values of inertia matrices [5,14]. Effectiveness of the TDE methods has been gradually confirmed in both theoretical proofs and experimental validations [15,16].

To effectively complete control objectives in industrial applications, a proper control strategy with impressive adaptation and robustness ability is next required [17–19]. The sliding mode control (SMC) technique has a special attraction to designers thanks to the simplicity, widespread applications, and robustness to unknown system dynamics [20–23]. However, excellent control performances of the SMC controllers may be hard to maintain in different working cases. As a solution, gain-adaptive SMC schemes have been recently noted for servo systems [18,24–26]. Driving gains or robust gains were actuated by various learning mechanisms to effectively force the control objective to zero as fast as possible [25,26]. In these learning schemes, only one degree-of-freedom learning law is proposed, and the control error is stabilized to a bound around zero [17,24]. In another direction, intelligent approaches based on Type-2 fuzzy logic theory could be found as possible solutions of the gain adaptation [27,28]. Solving optimal problems of the closed-loop system under robustness constraints in a horizontal time yields the expected gains along with excellent control performances [29]. However, intensive behaviors of the gain variation that could judge the real-time control quality were ignored.

This paper is an extension of work originally presented in

*Corresponding Author: Dang Xuan Ba, HCMC University of Technology and Education (HCMUTE), Ho Chi Minh city, Vietnam, badx@hcmute.edu.vn

2019 International Conference on System Science and Engineering (ICSSE) [1]. In the previous work, both driving and robust gains of a conventional TDE sliding mode controller were actively self-tuned subjecting to a minimal control error. Its effectiveness was confirmed by intensive simulation results. The controller was operated with a fixed nominal value of the inertia matrix of the system dynamics. Improperly selecting this term could lead to instability of the closed-loop system [15]. An excellent control performance is obviously associated with a perfect value of the matrix [16]. Nevertheless, the best nominal matrix is not same for different running states. As a sequence, a dynamical learning mechanism for such the term is the main motivation of this research for maintaining high control precision in various working conditions.

In this study, an adaptive TDE sliding mode controller is developed for trajectory-tracking control of serial robot manipulators with the following contributions:

- The system behaviors are first approximated by a basic time-delay estimator (TDE). To support selection of a proper inertia matrix in the estimation process, an adaptation rule is designed to search the optimal nominal one.
- The control objective is then realized by a robust sliding mode control method. To provide the flexibility in tuning the control gains in divergent operation cases, a fast gain-learning law is proposed. The adaptation mechanism is worked as a damping low-pass filter that could yield smooth responses of the closed-loop system. Different from the previous work [1], nonlinear leakage functions are employed to speed up the learning process.
- Control performances of the overall system are analyzed by a Lyapunov-based approach and comparative simulation results.

The rest of this paper is structured as follows. Section 2 presents system modeling and problem statement. Section 3 describes design of the proposed controller incorporated with the adaptation laws. Section 4 discusses validation results, and the paper is then concluded in Section 5

2. System Dynamics and Problem Statement

Dynamics of a *n*-joint robotic manipulator are generally expressed as follows [6,9,17,24]:

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{g}(\mathbf{q}) + \mathbf{f}(\dot{\mathbf{q}}) + \boldsymbol{\tau}_\mathbf{d} = \boldsymbol{\tau} \quad (1)$$

where $\mathbf{q},\dot{\mathbf{q}},\ddot{\mathbf{q}} \in \Re^n$ are the joint angles, velocities, and accelerations, respectively, $\boldsymbol{\tau} \in \Re^n$ is the joint torques or control inputs, $\mathbf{M}(\mathbf{q}) \in \Re^{n\times n}$ is the inertia matrix, $\mathbf{C}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}} \in \Re^n$ is the Centripetal/Coriolis vector, $\mathbf{g}(\mathbf{q}) \in \Re^n$ is the gravitational term, $\mathbf{f}(\dot{\mathbf{q}}), \boldsymbol{\tau}_\mathbf{d} \in \Re^n$ is the frictional torque, and $\boldsymbol{\tau}_\mathbf{d} \in \Re^n$ denotes disturbances.

The dynamics (1) could be simply rewritten in a following form:

$$\bar{\mathbf{M}}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{v}(\mathbf{q}) = \boldsymbol{\tau} \quad (2)$$

where $\bar{\mathbf{M}} \in \Re^{n\times n}$ is an proper nominal diagonal inertia matrix, $\mathbf{v} = (\mathbf{C}\dot{\mathbf{q}} + \mathbf{g} + \mathbf{f} + \boldsymbol{\tau}_\mathbf{d} + (\mathbf{M} - \bar{\mathbf{M}})\ddot{\mathbf{q}}) \in \Re^n$ is a lumped term that is defined as a combination of other dynamic terms such as the Centripetal/Coriolis, gravity, inertia variation, frictions, disturbances, modeling errors and unmodeled terms.

Remark 1: Assume that the system outputs $(\mathbf{q},\dot{\mathbf{q}})$ are measurable and a desired trajectory $(\mathbf{q_d})$ is bounded and differentiable up to the second order. The main objective of this paper is to find out an appropriate control signal to drive a tracking error of the system output $(\mathbf{q})$ and the given profile $(\mathbf{q_d})$ to zero or as small as possible. However, maintaining high control quality of the closed-loop system is a not-easy yet interesting work due to existence of the uncertain nonlinear dynamics and complicated working environments. Additional characteristics of the controller are required to be model-free, adaptive, and robust with high control performances.

3. An Adaptive Time-delay Nonlinear Controller

To accomplish the aforementioned features, the motion controller is built up based on a simple sliding mode framework incorporated with an adaptive time-delay estimator. Fast learning laws are then integrated to search proper control parameters. The proposed controller is sketched in **Figure 1.**

3.1. Simple Sliding Mode Control

The main control error is mathematically defined as

$$\mathbf{e} = \mathbf{q} - \mathbf{q_d} \quad (3)$$

A sliding manifold is employed as an indirect control objective of the closed-loop system:

$$\mathbf{s} = \dot{\mathbf{e}} + \boldsymbol{\Lambda}\mathbf{e} \quad (4)$$

where $\boldsymbol{\Lambda}$ is a selected positive-definite diagonal gain matrix.

For obtaining high-precision control results, the controller is designed with a model-free compensator - that could cope with the systematic uncertain nonlinearities and disturbances - and a robust control signal, as follows:

$$\boldsymbol{\tau} = \boldsymbol{\tau}_\mathbf{MOD} + \boldsymbol{\tau}_\mathbf{SMC} \quad (5)$$

Here, $\boldsymbol{\tau}_\mathbf{MOD}$ is a model-dynamical eliminator that is used to compensate for unexpected effect of the offset vector ($\mathbf{v}$) on the control behaviors and other terms of the surface dynamics (4). The signal is thus selected as follows:

$$\boldsymbol{\tau}_\mathbf{MOD} = \hat{\mathbf{v}} + \hat{\mathbf{M}}(\ddot{\mathbf{q}}_\mathbf{d} - \boldsymbol{\Lambda}\dot{\mathbf{e}}) \quad (6)$$

Note that the estimate $(\hat{\mathbf{v}})$ of the offset dynamics ($\mathbf{v}$) is computed using the basic time-delay estimator [15,22]:

$$\hat{\mathbf{v}} = \boldsymbol{\tau}_{t-T_s} - \hat{\mathbf{M}}\ddot{\mathbf{q}}_{t-T_s} \quad (7)$$

where the subscript $(t-T_s)$ is a time-delay value, T_s is the sampling time, and $\hat{\mathbf{M}} = \text{diag}(\hat{\mathbf{m}}) \in \Re^{n\times n}$ is estimate of the nominal matrix $\bar{\mathbf{M}}$.

The remaining term $\boldsymbol{\tau}_{\mathbf{SMC}}$ is designed to drive the sliding manifold from the stationary state back to around zero and stabilize thereafter by dealing with a model perturbance $\boldsymbol{\delta}$:

$$\boldsymbol{\tau}_{\mathbf{SMC}} = -\mathbf{K}_1\mathbf{s} - \mathbf{K}_2\,\text{sgn}(\mathbf{s}) \tag{8}$$

where $\mathbf{K}_{i|i=1,2} = \text{diag}(\mathbf{k}_i) = \text{diag}([k_{i1};...;k_{in}])$ is positive-definite diagonal gain matrices.

Remark 2: With the robust sliding mode design, excellent control outcome can be resulted in by a proper selection of the control gains $(\mathbf{K}_1, \mathbf{K}_2)$ and nominal inertia matrix $(\hat{\mathbf{M}})$ regardless of the existence of the disturbances. It is however sometimes a challenge to perfectly tune the gains to achieve good transient performance and especially ensure high control accuracies for unpredictable execution conditions in the real-time control.

3.2. Gain-adaptation Laws

To treat difficulty of the gain selection, the strategy for gain tuning is proposed in that the control gains $(\mathbf{K}_i|_{i=1,2})$ are separated into two terms: nominal gains $(\bar{\mathbf{K}}_i|_{i=1,2})$ and variation gains $(\breve{\mathbf{K}}_i|_{i=1,2})$. The nominal ones are chosen for stability of the closed-loop system. The variation ones are automatically tuned for the desired transient performance.

A fast adaptation rule for the automatic control-gain tuning is designed as follows:

$$\begin{cases} \dot{\breve{\mathbf{k}}}_1 = \boldsymbol{\Xi}_k\,\text{diag}(\mathbf{s})\mathbf{s} - \text{diag}(\boldsymbol{\eta}_1)\left(\mathrm{I}+\text{diag}\left(\mathbf{s}^2\right)\right)^{-1}\text{sat}\left(\breve{\mathbf{k}}_1\right) \\ \dot{\breve{\mathbf{k}}}_2 = \boldsymbol{\Xi}_k\,\text{diag}\left(\text{sgn}\left(\mathbf{s}\right)\right)\mathbf{s} - \text{diag}\left(\boldsymbol{\eta}_2\right)\left(\mathrm{I}+\text{diag}\left(\mathbf{s}^2\right)\right)^{-1}\text{sat}\left(\breve{\mathbf{k}}_2\right) \end{cases} \tag{9}$$

where $\boldsymbol{\Xi}, \text{diag}\left(\boldsymbol{\eta}_1\right)$ and $\text{diag}\left(\boldsymbol{\eta}_2\right)$ are positive-definite diagonal constant matrices.

On the other hand, properly choosing the nominal inertia matrix could reduce bound of the system perturbance ($\boldsymbol{\delta}$). To efficiently support this task, a new learning law is proposed as follows:

$$\dot{\hat{\mathbf{m}}} = \boldsymbol{\Xi}_m\,\text{diag}(\mathbf{s})\left(\ddot{\mathbf{q}}_{t-T_s} - \ddot{\mathbf{q}}_{\mathbf{d}} + \boldsymbol{\Lambda}\dot{\mathbf{e}}\right) - \text{diag}\left(\boldsymbol{\eta}_m\right)\text{diag}\left(\mathbf{s}^2\right)\hat{\mathbf{m}} \tag{10}$$

where $\boldsymbol{\Xi}_m$ and $\text{diag}\left(\boldsymbol{\eta}_m\right)$ are positive-definite diagonal constant matrices.

3.3. Stability Analysis

The theoretical effectiveness of the closed-loop system is investigated using the following theorem:

Theorem 1: Employing the model-free sliding mode control signals (3)-(8) to the robot dynamics (2), in which the driving-robust gains and nominal inertia matrix are automatically updated using the laws (9)-(10), asymptotic convergence of the closed-loop system is resulted in if the following condition holds:

$$\begin{cases} \min\left(\text{eig}\left(\bar{\mathbf{K}}_1\right)\right) > \max\left(\text{eig}\left(0.25\boldsymbol{\Xi}_m^{-1}\text{diag}(\boldsymbol{\eta}_m)\bar{\mathbf{M}}\bar{\mathbf{M}}\right)\right) \\ \bar{k}_{2i|i=1..n} \geq \max\left(|\delta_i|\right) \end{cases} \tag{11}$$

Proof:

The time derivative of the sliding manifold (4) along the system (2) is expressed

$$\dot{\mathbf{s}} = \bar{\mathbf{M}}^{-1}\left(\boldsymbol{\tau} - \mathbf{v}\right) - \ddot{\mathbf{q}}_{\mathbf{d}} + \boldsymbol{\Lambda}\dot{\mathbf{e}} \tag{12}$$

By substituting the control signal (5)-(8) into the dynamics (12), we have

$$\dot{\mathbf{s}} = \bar{\mathbf{M}}^{-1}\left(\boldsymbol{\delta} + \tilde{\mathbf{M}}\left(\ddot{\mathbf{q}}_{\mathbf{d}} - \ddot{\mathbf{q}}_{t-T_s} - \boldsymbol{\Lambda}\dot{\mathbf{e}}\right) - \mathbf{K}_1\mathbf{s} - \mathbf{K}_2\,\text{sgn}(\mathbf{s})\right) \tag{13}$$

where $\boldsymbol{\delta} = [\delta_1;\delta_2;...;\delta_n] = \bar{\mathbf{v}} - \mathbf{v} = \left(\boldsymbol{\tau}_{t-T_s} - \bar{\mathbf{M}}(\mathbf{q})\ddot{\mathbf{q}}_{t-T_s} - \mathbf{v}\right) \in \Re^n$ is the system perturbance mentioned above, and $\tilde{\mathbf{M}} = \text{diag}(\tilde{\mathbf{m}}) = \hat{\mathbf{M}} - \bar{\mathbf{M}} \in \Re^{n\times n}$ is the estimation error of the nominal matrix $\bar{\mathbf{M}}$.

Now, we consider a Lyapunov function:

$$V = 0.5\mathbf{s}^T\bar{\mathbf{M}}\boldsymbol{\Xi}_k\mathbf{s} + 0.5\sum_{i=1}^{2}\left(\breve{\mathbf{k}}_i^T\breve{\mathbf{k}}_i\right) + 0.5\tilde{\mathbf{m}}^T\boldsymbol{\Xi}_k\boldsymbol{\Xi}_m^{-1}\tilde{\mathbf{m}} \tag{14}$$

Substituting (9), (10), and (13) into the time derivative of the Lyapunov function leads to

$$\begin{aligned} \dot{V} &= \mathbf{s}^T\boldsymbol{\Xi}_K\left(\boldsymbol{\delta} - \bar{\mathbf{K}}_1\mathbf{s} - \bar{\mathbf{K}}_2\,\text{sgn}(\mathbf{s})\right) \\ &\quad - \sum_{i=1}^{2}\left(\breve{\mathbf{k}}_i^T\,\text{diag}(\boldsymbol{\eta}_i)\left(\mathrm{I}+\text{diag}\left(\mathbf{s}^2\right)\right)^{-1}\breve{\mathbf{k}}_i\right) \\ &\quad - \tilde{\mathbf{m}}^T\boldsymbol{\Xi}_k\boldsymbol{\Xi}_m^{-1}\text{diag}(\boldsymbol{\eta}_m)\text{diag}\left(\mathbf{s}^2\right)\tilde{\mathbf{m}} \\ &\quad + \tilde{\mathbf{m}}^T\boldsymbol{\Xi}_k\boldsymbol{\Xi}_m^{-1}\text{diag}(\boldsymbol{\eta}_m)\text{diag}\left(\mathbf{s}^2\right)\bar{\mathbf{m}} \end{aligned} \tag{15}$$

From the condition (11), there always exist two positive constants α_k and α_m such that

$$\begin{aligned} \dot{V} &= -\alpha_k\mathbf{s}^T\boldsymbol{\Xi}\bar{\mathbf{K}}_1\mathbf{s} - \sum_{i=1}^{2}\left(\breve{\mathbf{k}}_i^T\,\text{diag}(\boldsymbol{\eta}_i)\left(\mathrm{I}+\text{diag}\left(\mathbf{s}^2\right)\right)^{-1}\breve{\mathbf{k}}_i\right) \\ &\quad - \alpha_m\tilde{\mathbf{m}}^T\boldsymbol{\Xi}_k\boldsymbol{\Xi}_m^{-1}\text{diag}(\boldsymbol{\eta}_m)\text{diag}\left(\mathbf{s}^2\right)\tilde{\mathbf{m}} \le 0 \end{aligned} \tag{16}$$

Theorem 1 has been proven:

Remark 3: As seen in Figure 1 as well as stated in Theorem 1**,** the controller is just designed under a simple TDE sliding mode control framework to ensure the stability of the overall system**.** However, control performance of the system is significantly improved by injecting excellent adaptation laws for the control gains and nominal inertia matrix. The proposed controller hence possesses strong robustness, model-free, and self-learnable ability.

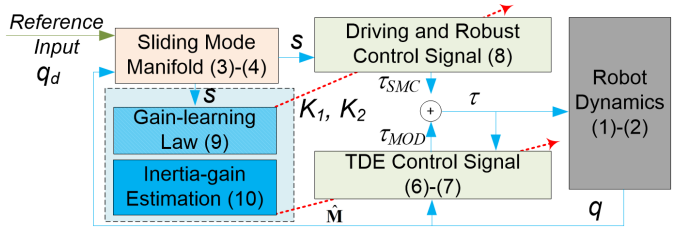


Figure 1: An overview of the designed controller.

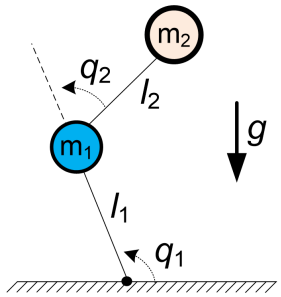


Figure 2: Configuration of a 2DOF manipulator.

Table 1: Selected Parameters for the simulation model

Description	Parameters	Values	Unit
Link length	l_1, l_2	0.2, 0.1	m
Gravitational Accel	g	9.81	m/s2
Friction coefficient	a_1, a_2	1	
Mass of links	m_1, m_2	10, 5	kg

4. Simulation Validation

4.1. Setup

To assess tracking performance of the proposed controller, extended simulations were carried out on a 2DOF robot. Configuration of the robot is depicted in Figure 2. The robot was selected to work in the vertical plane in which the gravity force was critically affected on the system performance. Dynamics of the robot are presented as follows [6,24]:

$$\mathbf{M} = \begin{bmatrix} m_2(l_2^2 + l_1^2) + 2l_1 l_2 m_2 \cos(q_2) + m_1 l_1^2 & m_2 l_2^2 + l_1 l_2 m_2 \cos(q_2) \\ m_2 l_2^2 + l_1 l_2 m_2 \cos(q_2) & m_2 l_2^2 \end{bmatrix}$$

$$\mathbf{C}\dot{\mathbf{q}} = \begin{bmatrix} -m_2 l_1 l_2 \sin(q_2)\dot{q}_2 (\dot{q}_2 + 2\dot{q}_1) \\ m_2 l_1 l_2 \sin(q_2)\dot{q}_1^2 \end{bmatrix}$$

$$\mathbf{g} = \begin{bmatrix} m_2 l_2 g \cos(q_1 + q_2) + l_1 g \cos(q_1)(m_1 + m_2) \\ m_2 l_2 g \cos(q_1 + q_2) \end{bmatrix}$$

$$\mathbf{f} = \begin{bmatrix} a_1 \dot{q}_1 \\ a_2 \dot{q}_2 \end{bmatrix}$$

where $q_{1,2}$ are the robot joint angles. The detailed parameters of the robot model are given in **Table 1**.

Another robust-gain-learning TDE controller (RLTDE) was also applied to the same system as a benchmark for evaluating the effectiveness of the proposed controller.

The design of the RLTDE controller was referred as in the previous work [24], as follows:

$$\begin{cases} \mathbf{s} = \dot{\mathbf{e}} + \mathbf{\Lambda}_c \mathbf{e} \\ \boldsymbol{\tau} = \bar{\mathbf{M}}\left(-\ddot{\mathbf{q}}_{t\text{-}T_s} + \bar{\mathbf{M}}^{-1}\boldsymbol{\tau}_{t\text{-}T_s} + \ddot{\mathbf{q}}_\mathbf{d} - \mathbf{\Lambda}_c \dot{\mathbf{e}} - \mathbf{K}_{c1}\mathbf{s} - \mathbf{K}_{c2}\,\mathrm{sgn}(\mathbf{s})\right) \\ \dot{\hat{k}}_{c2i} = \begin{cases} \varphi_i \left(\alpha_i^{-1}|s_i(t)|\right)^{\theta(t)} \theta(t), \text{if } \hat{k}_{c2i} > 0 \\ \varphi_i \alpha_i^{-1}|s_i(t)| \qquad \text{otherwise} \end{cases} \\ \theta(t) = \mathrm{sgn}\left(\|\mathbf{s}\|_\infty - \varepsilon\right) \end{cases}$$

The basic control gains of the two controllers were inherently selected as [24]:

$$\begin{cases} T_s = 0.001, \mathbf{\Lambda}_c = \mathrm{diag}([10;10]); \\ \boldsymbol{\alpha} = [0.6;1.57], \boldsymbol{\varphi} = [4000;3000]; \varepsilon = 0.015 \end{cases}$$

For a pair comparison, the RLTDE controller was operated with four distinct parameter sets as given in Table 2, denoted as RLTDE1, RLTDE2, RLTDE3 and RLTDE4, respectively.

Table 2: Different Control Parameters Selected for the RLTDE Controller Implementation

Name	Nominal Inertia Matrix ($\bar{\mathbf{M}}$)	Control Gain ($\mathbf{K}_{c1}$)
RLTDE1	diag([0.01;0.01])	diag([1;1])
RLTDE2	diag([0.04;0.04])	diag([1;1])
RLTDE3	diag([0.01;0.01])	diag([60;60])
RLTDE4	diag([0.04;0.04])	diag([60;60])

Besides, the additional parameters of the proposed controller were manually tuned as follows:

$$\begin{cases} T_s = 0.001; \mathbf{\Lambda} = \mathrm{diag}([10;10]), \\ \bar{\mathbf{K}}_2 = \mathrm{diag}([0.5;0.5]), \bar{\mathbf{K}}_1 = \mathrm{diag}([1;1]); \\ \mathbf{\Xi}_k = \mathrm{diag}([300;300]); \boldsymbol{\eta}_1 = \boldsymbol{\eta}_2 = [1;1] \\ \mathbf{\Xi}_m = \mathrm{diag}([0.003;0.003]); \boldsymbol{\eta}_m = [3\times10^{-4};3\times10^{-4}]. \end{cases}$$

4.2. Simulation Results

In the first scenario, reference inputs of the robot joints were selected to be sinusoidal signals plotted in Figure 3. By applying the controllers to the system, the simulation results obtained are presented in Figures 4 and 5.

As seen in Figure 4, the RLTDE1 provided very good control performances, 0.53 (RMS deg) and 0.44 (RMS deg) of the control errors at joints 1 and 2, thanks to employment of the strong robust gain ($\mathbf{K}_2$) generated by a nonlinear learning algorithm [24]. Behaviors of the robust gain adaptation are demonstrated in Figure 5. The key purpose of using such the gain in the RLTDE control method was to attenuate effect of the incomplete TDE estimation. Figure 4 also reveals that the control performances were further improved if the more proper control gains were found out. When increasing only the nominal inertia matrix ($\bar{\mathbf{M}}$), or only the driving gain ($\mathbf{K}_{c1}$), or both of them ($\bar{\mathbf{M}}$ and $\mathbf{K}_{c1}$), higher

control accuracies were exhibited such as: the RLTDE2 controller obtained 0.17 (RMS deg) and 0.12 (RMS deg) of the control errors at joints 1 and 2, respectively; the RLTDE3 controller got 0.2 (RMS deg) and 0.11 (RMS deg) ones; and the RLTDE4 controller provided 0.09 (RMS deg) and 0.06 (RMS deg) ones. It seems larger the control gains would create better the control performance. In fact, structures of the RLTDE and proposed controllers are almost same.

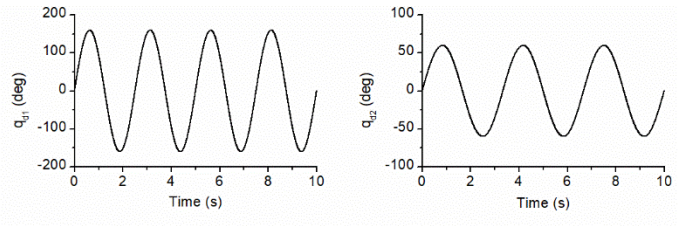


Figure 3: Reference inputs of the robot joints in the first simulation.

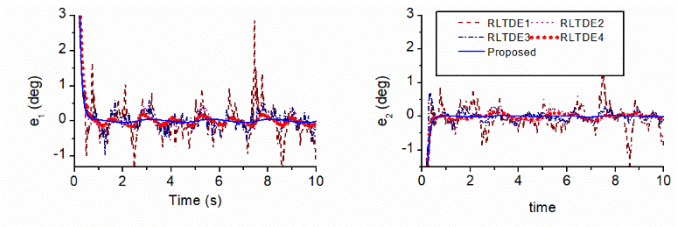


Figure 4: Control errors obtained in the first simulation

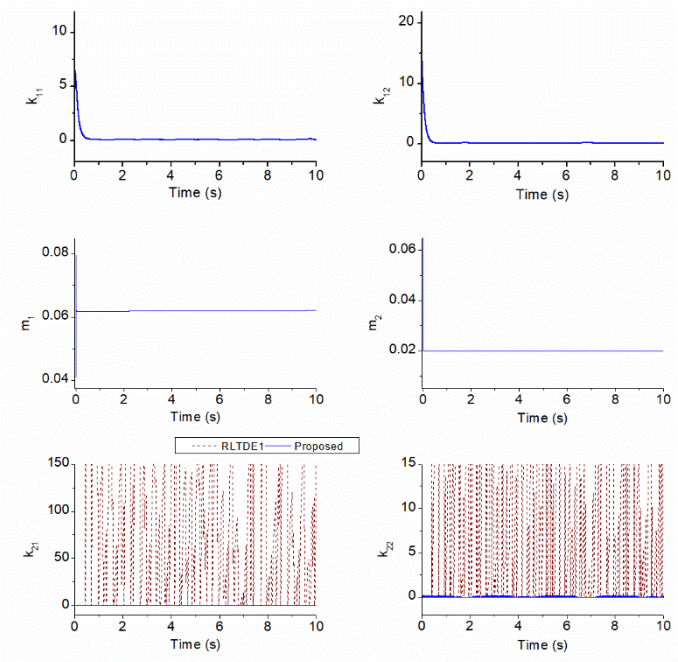


Figure 5: Comparative gains learning in the first simulation.

The core difference is that the proposed controller adopts the adaptation rules for the driving-robust gains and optimal nominal inertia matrix. Properly searching an optimal set of the driving gain and nominal inertia matrix could effectively reduce bound of the gain ($\mathbf{K}_2$) that could be clearly observed in Figure 5. Hence, as shown in Figure 4, the highest control precision, 0.03 (RMS deg) and 0.014 (RMS deg) of the control errors at joints 1 and 2, was resulted in by the proposed controller. Also noted in Figure 5, the optimal values did not need to be so large for the best control performance: the variation of the driving gain converged back to zero and the stationary nominal inertia matrix was found to be *diag(0.062; 0.021)*. This implies the flexibility of the proposed controller over the previous one.

In the second scenario, the robot joints were controlled to track smooth step signals. Applying the same controllers to the system, the simulation results are presented in Figure 6, 7 and 8.

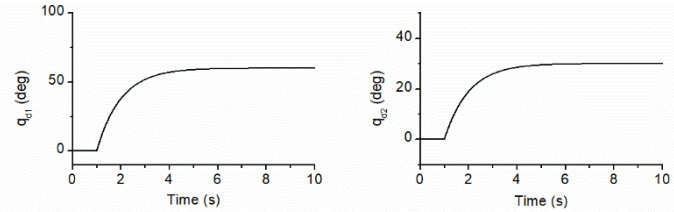


Figure 6: Reference signals of the robot joints in the second simulation.

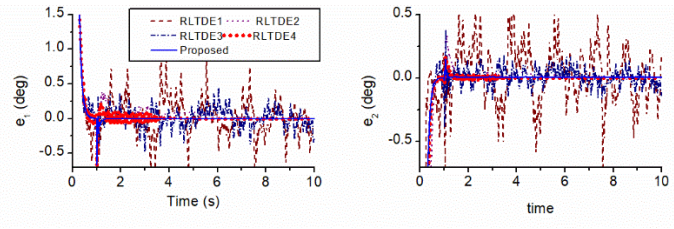


Figure 7: Control errors obtained in the second simulation.

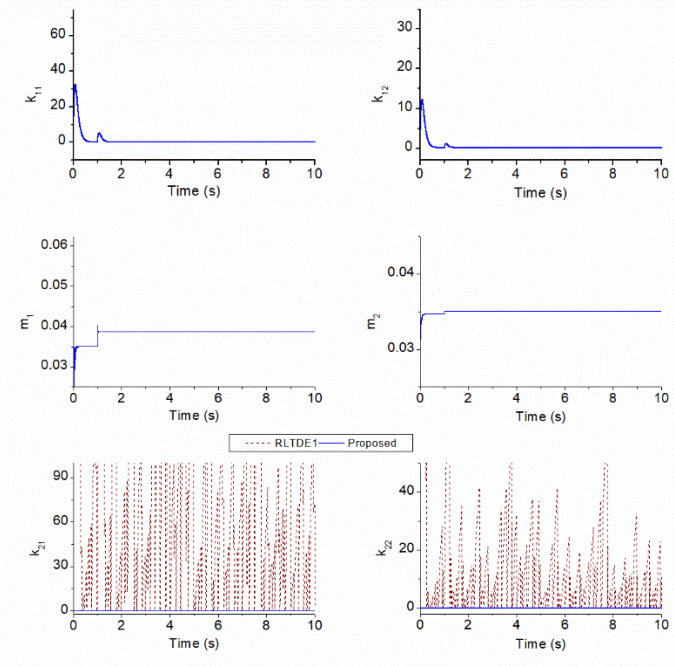


Figure 8: Comparative gains learning in the second simulation.

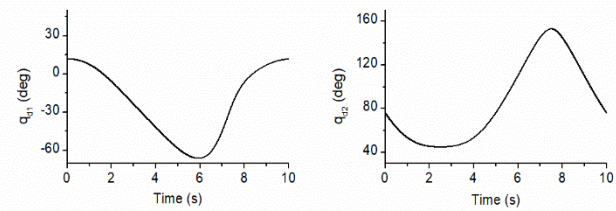


Figure 9: Desired joint angles computed from the eclipse trajectory of the end-effect in the third simulation.

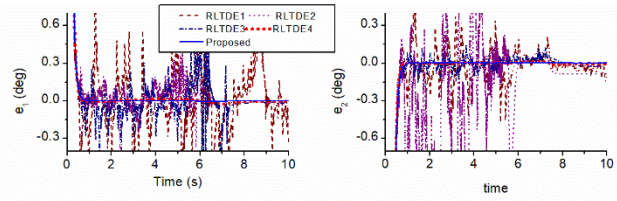


Figure 10: Control errors obtained in the third simulation.

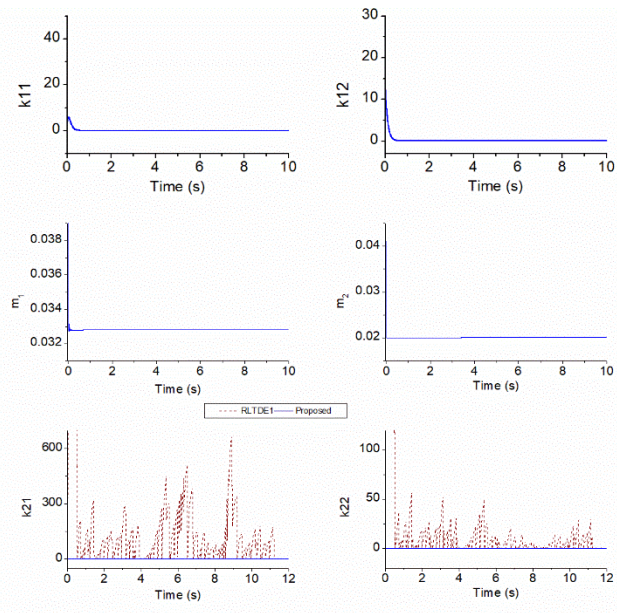


Figure 11: Comparative gains learning in the third simulation.

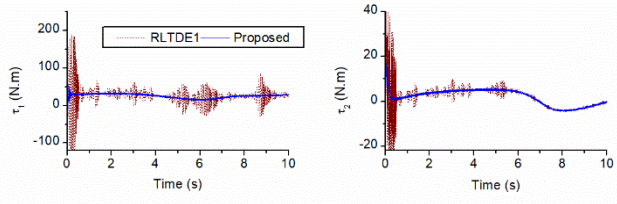


Figure 12: Control signal generated in the third simulation.

As shown in Figure 7, the control performances of the RLTDE1, 1.1 (MA deg) and 0.73 (MA deg) of the control errors at joints 1 and 2, were degraded as comparing to those of the other controllers. An improper selection of the nominal inertia matrix would make the systematic perturbance ($\boldsymbol{\delta}$) become large. Meanwhile, the driving gain ($\mathbf{K}_{c1}$) was fixed with a small value, it means that the control performance was covered only by the robust gain ($\mathbf{K}_{c2}$). Another reason came from the special learning mechanism of the RLTDE: the robust gain ($\mathbf{K}_{c2}$) was respectively increased and decreased as the sliding surfaces were outside and inside of specific regions. It implies that the RLTDE control outcome in this case depended on the regions predefined. New features integrated in the proposed controller could make the system avoid the aforementioned problems: as observed in Figure 8, the adaptation laws of the inertia matrix and control gains worked well to ensure the closed-loop system stabilize at origin or force the control error to be as small as possible. Hence, outperformance of the proposed controller, 0.005 (MA deg) and 0.01 (MA deg) of the control errors at joints 1 and 2, over the RLTDE1, 2, and 3 has been exhibited in this simulation scenario. Interestingly, once the perfect values of the driving gain ($\mathbf{K}_{c1}$) and nominal inertia matrix ($\bar{\mathbf{M}}$) were used, the RLTDE4 obtained an excellent control accuracy: 0.002 (MA deg) and 0.0004 (MA deg) of the control errors at joints 1 and 2.

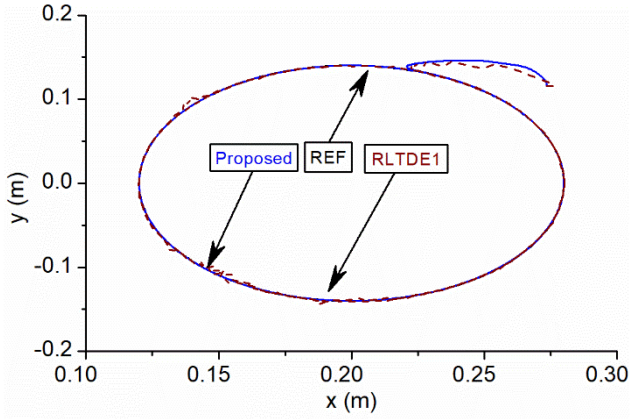


Figure 13: Comparison of control results in Cartesian space

In the third verification case, the end-effector of the robot was controlled to track an eclipse trajectory. By applying inverse kinematic computation for the testing robot, the desired joint angles are displayed in Figure 9. Tracking control errors of the controllers for the new trajectory are plotted in Figure 10. The control results indicate that the RLTDE controllers tried to automatically adjust their gains to minimize the control error. However, using only one degree-of-freedom adaptation of the control gains was not adequate to result in the minimal sliding manifold even though they were supported with various control gains and nominal inertia matrices fine-tuned. The best control quality, 0.024 (MA deg) and 0.015 (MA deg) of the control errors at joints 1 and 2, were accomplished by the RLTDE4 which had just shown the outstanding performance in the second simulation. The control accuracy was significantly improved, 0.004 (MA deg) and 0.005 (MA deg) of the control errors at joints 1 and 2, by using three degree-of-freedom in the gain learning proposed by this paper.

Indeed, control input data generated by the controllers shown in Figure 12 reveal that sharing the robustness burden of the control mission into three learning spreads provided a faster response and a smaller control error. Its effectiveness and feasibility are demonstrated more clearly by plots of the comparative end-effector positions obtained by the proposed controller and a certain RLTDE one, as shown in Figure 13.

4.3. Discussion

The key properties of the proposed adaptation laws could be evidently expressed by data shown in Figure 5, 8 and 11: in transient time, the closed-loop system needed strong power to drive the control objective back to desired position regardless of uncertain nonlinearities and disturbances. To serve this mission, the driving and robust gains were promptly activated under the designed constraint (9), and the optimal nominal gain matrix was also searched according to the strategy (10). After that, the whole system would relax in steady-state time with the small-activated gains. By further considering these figures, much high-frequency oscillation could be observed in the learning gains of the RLTDE1

controller. In real-time operation with smooth control signals, the proposed controller could maintain the lifetime of the actuators.

Table 3: Statistical Evaluation of the Control Performances in the Simulation Scenarios

Control Error		Joint 1		Joint 2	
		MA	RMS	MA	RMS
Test 1	RLTDE1	2.87	0.53	1.98	0.44
	RLTDE2	0.6	0.17	0.26	0.12
	RLTDE3	0.9	0.2	0.38	0.11
	RLTDE4	0.22	0.09	0.11	0.06
	Proposed	0.05	0.03	0.04	0.014
Test 2	RLTDE1	1.1	0.25	0.73	0.2
	RLTDE2	0.09	0.03	0.024	0.01
	RLTDE3	0.48	0.13	0.18	0.06
	RLTDE4	0.002	0.0004	0.0004	0.0001
	Proposed	0.005	0.001	0.01	0.007
Test 3	RLTDE1	2.02	0.4	0.47	0.09
	RLTDE2	0.39	0.08	1.17	0.2
	RLTDE3	2.2	0.27	0.13	0.03
	RLTDE4	0.024	0.01	0.015	0.009
	Proposed	0.004	0.0018	0.005	0.002

Table 3 summarizes static values of the control errors using maximum absolute (MA) and root-mean-square (RMS) computation for a specific time (5s to 12s). As seen in the table and by recalling the second simulation results, if the optimal set of the control gains and nominal inertia matrix were known, an exceptional control performance would be achieved. However, the optimal one was not same in different working conditions. Instead of searching the optimal values with time consuming, the designed controller integrated an automatic gain-tuning algorithm for suboptimal ones. The statistical data in Table 3 show that the proposed controller yielded the excellent MA and RMS results as comparing to the RLTDE ones in most of the simulation cases. Hence, the influence of such the suboptimal gains and inertia matrices on the control performance was acceptable. Here, advantages of the proposed controller are confirmed throughout both the analytical and validation results.

5. Conclusion

In this paper, a new adaptive robust control method is developed for position tracking-control problems of robot manipulators. The controller design is mainly based on the sliding mode control framework. Effect of the systematic uncertain nonlinearities and disturbances are well treated by an adaptive time-delay estimator. The driving-robust control signal is then employed for realizing the control objective. Fast gain-learning laws are properly designed to increase flexibility of the controller in various working conditions. The control gains are automatically tuned to provide both the desired transient performance and the steady-state behavior. Effectiveness of the closed-loop system is intensively investigated by theoretical proofs and comparative simulations.

Conflict of Interest

The author declares no conflict of interest.

References

[1] D.X. Ba, M.-H. Le, "Gain-learning sliding mode control of robot manipulators with time-delay estimation," in 2019 International Conference on System Science and Engineering (ICSSE), 478–483, 2019.

[2] Y. Park, I. Jo, J. Lee, J. Bae, "A Dual-cable Hand Exoskeleton System for Virtual Reality," Mechatronics, **49**(November 2017), 177–186, 2018, doi:10.1016/j.mechatronics.2017.12.008.

[3] T.L. Nguyen, S.J. Allen, S.J. Phee, "Direct torque control for cable conduit mechanisms for the robotic foot for footwear testing," Mechatronics, **51**(August 2016), 137–149, 2018, doi:10.1016/j.mechatronics.2018.03.004.

[4] G. Chen, B. Jin, Y. Chen, "Nonsingular fast terminal sliding mode posture control for six-legged walking robots with redundant actuation," Mechatronics, **50**(August 2016), 1–15, 2018, doi:10.1016/j.mechatronics.2018.01.011.

[5] M. Tavakoli, P. Lopes, L. Sgrigna, C. Viegas, "Motion control of an omnidirectional climbing robot based on dead reckoning method," Mechatronics, **30**, 94–106, 2015, doi:10.1016/j.mechatronics.2015.06.003.

[6] J.J. Craig, Introduction to Robotics: Mechanics and Control, 3rd ed., Pearson Prentice Hall, USA, 2005.

[7] L. Wang, T. Chai, L. Zhai, "Neural-network-based terminal sliding-mode control of robotic manipulators including actuator dynamics," IEEE Transactions on Industrial Electronics, **56**(9), 3296–3304, 2009, doi:10.1109/TIE.2008.2011350.

[8] Z. Liu, G. Lai, Y. Zhang, C.L.P. Chen, "Adaptive Fuzzy Tracking Control of Nonlinear Time-Delay Systems with Dead-Zone Output Mechanism Based on a Novel Smooth Model," IEEE Transactions on Fuzzy Systems, **23**(6), 1998–2011, 2015, doi:10.1109/TFUZZ.2015.2396075.

[9] W.-H. Zhu, Virtual Decomposition Control: Toward hyper degrees of freedom robots, Springer-Verlag Berlin Heidelberg, 2010.

[10] S. Islam, X.P. Liu, "Robust Sliding Mode Control for Robot Manipulators," IEEE Transactions on Industrial Electronics, **58**(6), 2444–2453, 2011.

[11] W. Kim, C.C. Chung, "Robust output feedback control for unknown non-linear systems with external disturbance," IET Control Theory and Applications, **10**(2), 173–182, 2015, doi:10.1049/iet-cta.2014.1299.

[12] T.C. Hia, S. Jung, "A Simple alternative to Neural Network Control Scheme for Robot Manipulators," IEEE Transactions on Industrial Electronics, **42**(4), 414–416, 1995.

[13] S. Kim, J. Bae, "Force-mode Control of Rotary Series Elastic Actuators in a Lower Extremity Exoskeleton using Model-inverse Time Delay Control (MiTDC)," IEEE/ASME Transactions on Mechatronics, **22**(3), 1392–1400, 2017, doi:10.1109/TMECH.2017.2687979.

[14] T.C. Hia, "A New Technique for Robust Control of Servo Systems," IEEE Transactions on Industrial Electronics, **36**(1), 1–7, 1989.

[15] K. Youcef-toumi, O. Ito, "A Time Delay Controller for Systems With Unknown Dynamics," Journal of Dynamic Systems, Measurement, and Control, **112**(1), 133–142, 1990.

[16] Y. Wang, D. Yu, Y. Kim, "Robust Time-Delay Control for the DC – DC Boost Converter," IEEE Transactions on Industrial Electronics, **61**(9), 4829–4837, 2014.

[17] M. Jin, J. Lee, N.G. Tsagarakis, "Model-free Robust Adaptive Control of Humanoid Robots with Flexible Joints," IEEE Transactions on Industrial Electronics, **64**(2), 1706–1715, 2016, doi:10.1109/TIE.2016.2588461.

[18] F. Mazenc, P.-A. Bliman, "Backstepping Design for Time-Delay Nonlinear Systems," IEEE Transactions on Automatic Control, **51**(1), 149–154, 2006.

[19] C. Hu, B. Yao, Q. Wang, "Performance-Oriented Adaptive Robust Control of a Class of Nonlinear Systems Preceded by Unknown Dead Zone With Comparative Experimental Results," IEEE Transactions on Industrial Electronics, **18**(1), 178–189, 2013.

[20] V.I. Utkin, "Sliding Mode Control Design Principles and Applications to Electric Drives," IEEE Transactions on Industrial Electronics, **40**(1), 23–36, 1993.

[21] J. Zhang, W.X. Zheng, "Design of Adaptive Sliding Mode Controllers for Linear Systems via Output Feedback," IEEE Transactions on Industrial Electronics, **61**(7), 3553–3562, 2014.

[22] M. Zhihong, A.P. Paplinski, H.R. Wu, "Robust MIMO Terminal Sliding Mode Control Scheme for Rigid Robotic Manipulators," IEEE Transactions on Automatic Control, **39**(12), 2464–2469, 1994.

[23] A. Pisano, M. Rapaic, E. Usai, Sliding modes after the first decade of the 21st century, Springer-Verlag Berlin Heidelberg: 412, 2012.

[24] J. Baek, M. Jin, S. Han, "A New Adaptive Sliding Mode Control Scheme for Application to Robot Manipulators," IEEE Transactions on Industrial Electronics, **63**(6), 3628–3637, 2016, doi:10.1109/TIE.2016.2522386.
[25] Y. Shtessel, M. Taleb, F. Plestan, "A novel adaptive-gain supertwisting sliding mode controller: Methodology and Application," Automatica, **48**(5), 759–769, 2012, doi:10.1016/j.automatica.2012.02.024.
[26] C. Xia, G. Jiang, W. Chen, T. Shi, "Switching-Gain Adaptation Current Control for Brushless DC Motors," IEEE Transactions on Industrial Electronics, **63**(4), 2044–2052, 2016.
[27] M.H. Khooban, T. Niknam, F. Blaabjerg, M. Dehghani, "Free chattering hybrid sliding mode control for a class of non-linear systems : electric vehicles as a case study," IET Science, Measurement and Technology, **10**(7), 776–785, 2016, doi:10.1049/iet-smt.2016.0091.
[28] N.M. Tri, D.X. Ba, K.K. Ahn, "A gain-adaptive intelligent nonlinear control for an electrohydraulic rotary actuator," Internation Journal of Precision Engineering and Manufacturing, **19**(5), 665–673, 2018.
[29] M.H. Khooban, N. Vafamand, T. Niknam, T. Dragicevic, F. Blaabjerg, "Model Predictive Control based on T-S Fuzzy model For Electrical Vehicles Delayed Model," IET Electric Power Applications, **11**(5), 918–934, 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 912-919 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Investigation of the Effect of FBG Profiles, Temperature and Transmission Distance for Environmental Sensing & Monitoring

Muhammad Arif Riza, Yun Ii Go*

School of Engineering & Physical Sciences, Heriot-Watt University Malaysia, Putrajaya, 62200, Malaysia

ARTICLE INFO

Article history:
Received: 29 August, 2020
Accepted: 19 November, 2020
Online: 08 December, 2020

Keywords:
Apodization
Environmental
FBG
Sensor
Simulation
Temperature

ABSTRACT

Optical sensors exist in various forms and one of them is Fiber Bragg Gratings (FBG). Performance evaluation was simulated for FBG sensing. The simulation involves different grating profiles tested with complete optical sensing system. The system involves a broadband light source, FBG of various profiles and optical spectrum analyzer for data interpretation. Different profile of gratings imprinted on an FBG outputs wavelength spectra and simulated results satisfy this theory obtained from literatures. Temperature sensing was also simulated for the FBG at various temperature ranges that suit different industrial demands. At low temperatures FBG does give noticeable sensing capabilities. When reaching 400°C and above, the FBG was still capable of providing response, however when in real environments the fiber may suffer from thermal damage. FBG performance when integrated within a system with long distances of fibers were also simulated. Power attenuation was noticed at the reflectivity spectra.

1. Introduction

This paper is an extension of work originally presented in "IEEE International Conference on Sensors and Nanotechnology 2019" [1]. Environmental monitoring have been a main topic in most fields of study. Conventional sensors nowadays such as microelectronic mechanical sensors (MEMS) or probe sensors are vulnerable to electromagnetic (EM) interference and wiring complications. This phenomena prevents its use in long-range sensing applications [2]. Fiber optic sensors is a promising alternative to conventional sensors due to its compact nature, EM proofing and multiplexing capabilities [3]. The simplicity of optical based sensors allows various sensing configurations such as absorption based sensors (interferometers) and grating based sensors. Fiber Bragg Grating (FBG) sensor is an example of an optical sensor that operates by creating a periodic perturbation of light signals within the refractive index of its core. The gratings serves as wavelength filter that permits only light with wavelengths that dissatisfy the Bragg's equation to permeate through while the light waves that satisfy the Bragg equation will be reflected back. Therefore, strains acting on the gratings with a set periodicity can induce a shift in the wavelength spectra. The shift in wavelength defines the condition of the parameter being sensed [4]. Commonly, FBG sensors are used for pressure and thermal sensing. However, FBG sensors can be modified to sense other forms of parameter such as humidity by introducing a sensitive coatings onto the sensitized regions of the sensor.

Operations of FBG sensors require a source of propagated broadband light. In some cases, FBG sensors are positioned in remote areas where power input availability is scarce. Normally, the light source are powered by conventional batteries which could have a short lifespan and can be inconvenient maintenance wise. Hence, a sustainable power source is desired. A promising source of green energy is from ambient vibrations within the earth's crust [5]. The concept of vibration energy harvesting involves converting amplified vibrations at resonance into useful electrical power by means of kinetic energy distribution. Ambient vibration frequencies generally fall below 100.0 Hz. However, in civil structures, the frequency can drop to less than 10.0 Hz [6]. The acceleration level of these vibrations are very low being less than 1.0 g where g = 9.81 ms^{-2}, making it more suitable for low-power electronic application such as powering a light source.

This paper focuses on reporting simulated results of FBG tested with various different grating profiles, temperature sensing and transmission distance. Whilst simulating the FBG in different grating configurations, basis of the system which will be explained further sections will be tailored to a vibrational energy harvester power output. Since the proposed harvester design provides a low but steady power output to act as a light source, the whole FBG system will be run under low power to be simulated within various

*Corresponding Author: Yun Ii Go, Email: y.go@hw.ac.uk

conditions including thermal and attenuation. It is an opportunity to simulate how well the system of FBG can be run with sufficient power with different types of gratings which differs from most other closely similar articles involving FBG simulation [7-9].

1.1. Optical vs Conventional Sensor

Environment sensors that are still in use comes in various types. Electrical based sensor types such as the resistive probes, p-n junction diodes and MEMS sensors have been the staple sensor devices used over the years in monitoring various parameters [10]. Some sensors that have a very low power consumption of <1.0 mW and are made to be self-sustained by equipping the sensor with a solar panel harvester [11]. However, the operation these conventional sensors require electrical flow on the active sensitized regions of the sensor. This will render the sensors to be susceptible to electromagnetic interference and damage due to short circuit. Besides having a high maintenance cost, the complexity of the sensing mechanism in electrical based sensors available in market limits its capability to be embedded within a structure [12]. These drawbacks propel researchers to find alternatives such as optical fiber sensors.

Optical sensors alone can have many different configurations based on how the modification is formed on the surface, cladding, core or tip of the fiber. In addition to FBG sensors, other optical sensors include interferometers and absorption based sensors, which function based on the principal of the energy absorption spectrum and refractive index changes. However, grating based sensors possess an advantage over other optical based sensors in terms that they are more compact and ease of modification. In addition, FBG sensors have low noise, high sensitivity and a high multiplexing imprint capabilities via phase masking, which helps provide a large mapped sensing area within an environment [13].

1.2. FBG Sensing Principle

FBG are made typically from phase masking, where periodic modulations of the refractive index are imprinted within the fiber core by intense UV laser, exposed perpendicularly to a single mode optical fiber. The modulated refractive indexes within the fiber core that has different refractive index values than the core and cladding acts as the wavelength filter for the incident propagated light. Based on Figure 1, light signals at wavelengths other than the Bragg wavelength would pass through whereas light signals with wavelength equal to the Bragg wavelength that would be reflected back. Therefore, the transmitted ray that passes the grating will have the Bragg wavelength absent in the main spectrum [14]. FBGs can be serially imprinted into one fiber at multiple sensitized regions (multiplexing). This allows the FBGs to have multiple functional sensor nodes along a fiber length which is essential for area mapping.

On their own, FBGs are strain, pressure and temperature sensitive. In temperature sensing, different temperatures would impose axial strain upon the FBG fiber based on coefficient of thermal expansion of the FBG material (which would normally be silica). This would cause a change in the pattern spectrum of the reflected light, resulting in a shift in the reflected wavelength. This shift in wavelength can be calibrated to identify temperature change [15]. Detection of non-physical parameters requires the FBG to be modified by coating the sensitized region with a reversibly reactive material nanostructures that can induce mechanical strain or reaction to the FBG when exposed to a measurand [16]. Generally, it is possible for a FBGs to sense chemical parameters such as salinity concentration [8], chemical gas concentration [17] and relative humidity of the environment [18]. Ease of modification of FBG to suit various environmental sensing applications makes FBG a viable sensing device to be studied and optimized.

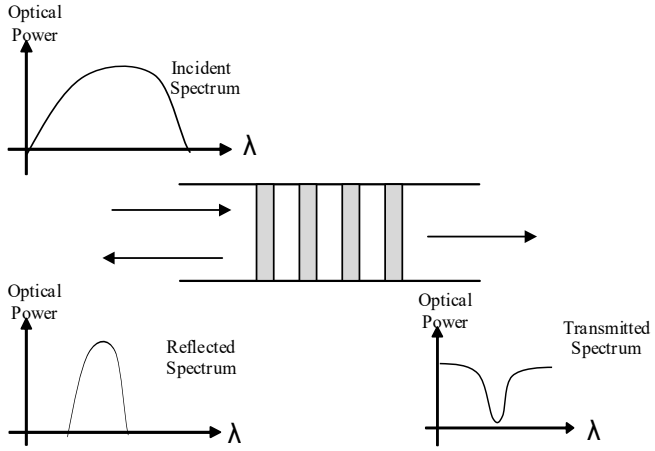


Figure 1: Principle of FBG.

2. Methodology

In this study, the behaviour of FBG fiber is simulated with different grating profiles. Different grating profiles can have an effect to the output transmission and reflectivity of the fibers. The behaviour of the FBG as a temperature sensor are simulated in temperature ranges that relates to certain industries. Connection of fibers in a long range fiber setup are also simulated to show the viability of FBG sensor array assembly in a long-range network.

2.1. Overall System Description

FBG sensors typically involve several components to be available for environmental monitoring. The following subsections explain each of the components in Figure 2 that are mandatory for FBG sensor implementation. Table 1 presents the constant parameters used in this study.

Table 1: System Setting.

Source Spec.	Optical Fibers	FBG Spec.	OSA Spec.
Broadband white light source Frequency: 193.1 THz Power: 130 dBM	Silica SMF Length: 10 cm	Grating profile: Uniform Bragg wavelength: 1550 nm IM: 0.0002	Resolution: 0.001 nm

2.2. Source

Optical based devices require sources in form of propagated photons. Light source of broad spectrum is preferred when using FBG. Broad spectrum light source examples include white light source, amplified spontaneous emission (ASE) source and broadband light source.

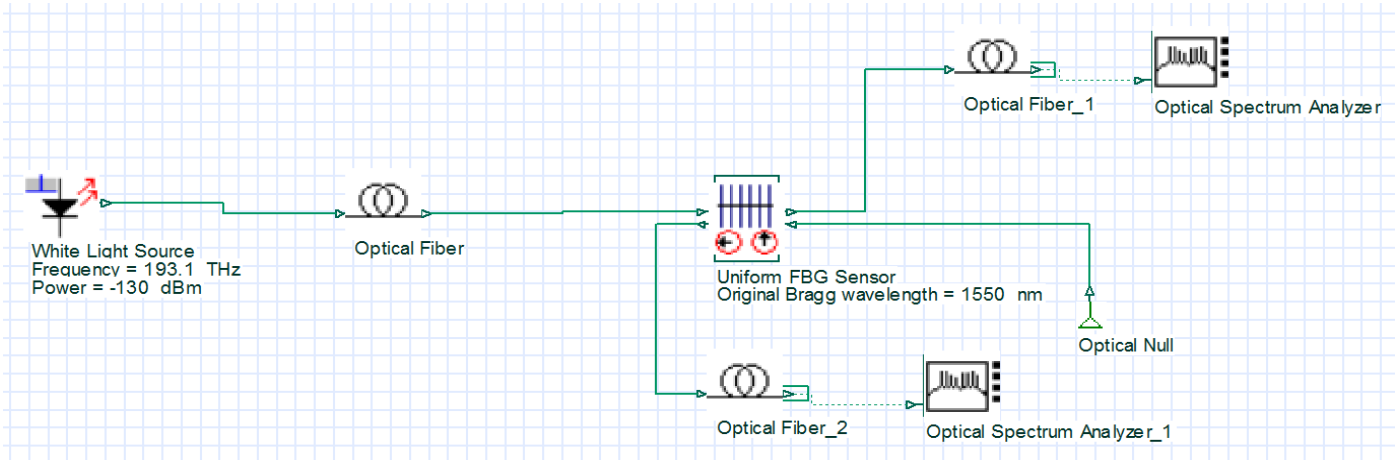


Figure 2: Arrangement of FBG setting.

2.3. FBG

The FBG component is the main component in the simulation that has variables manipulated. The Bragg grating wavelength chosen is 1550 nm since it is among the commonly used wavelength for FBG sensing. The grating length of fiber is set constant at 10 mm throughout the study. Index modulation is set at 0.0002 for the FBG and is also a controlled parameter. The grating profile will be varied in order to observe the effects of different grating profiles on the transmission and reflectivity spectra in the simulation.

2.4. OSA

Optical spectrum analyser or optical receiver is mandatory for any optical based sensors including FBG as it detects the incoming light propagation. The optical receiver then sends the received light signals to be converted into readable data. In lab-based study, OSA typically combine these functions of receiving and signal conversion into transmission and reflectivity spectrum. The OSA settings used in this simulation is with resolution of 0.001 m. The resolution value allows for more data points to be recorded brings more accurate results.

3. Results and Discussion

3.1. Effect of Grating Profiles

There are many different grating profiles of an FBG. These grating profiles may affect the spectrum of both the transmission and reflectivity. The grating profiles are made different with the periods or arrangement of the grating follows different model. Different phase mask can be used for the imprinting of different grating profiles. In this study, Uniform grating along with several other grating profiles is tested.

3.1.1. Uniform Profile

Uniform grating profiles is the most common imprinted FBG profiles that has seen its use in sensing. The uniform periodicity profile affects light modulation and follows the Braggs equation which is:

$$\lambda_B = 2n_{eff} \cdot \Lambda_G \quad (1)$$

where, λB is the Bragg wavelength, n_{eff} is the refractive index of the propagating light and ΛG is the index modulation periodicity [19]. In this study, light modulation across a uniform profile FBG has been simulated. In Figure 3 transmission and reflectivity of light propagated pass the profiles has been recorded. As predicted, the response have showed that the peak power exists on the 1550 nm wavelength. This is the wavelength at which the grating was imprinted on. Figure 2(a) showed the transmission power trough occurring and on Figure 2(b), reflectivity spectra showed peaks at the same Bragg wavelength. This simulation has demonstrated the regular uniform FBG profiles connected with regular power source. There appears to be minor side lobes occurring on Figure 2(b). These side lobes may not be favourable for FBG sensors that are tailored for high precision sensing. However, the side lobes does not hinder the uniform profiles from entering the market as region or array monitoring system.

3.1.2. Chirped Profile

A different profile than the typical uniform profile of FBG is the chirped FBG. The gratings in a chirped FBG varies in terms of periodicity along the grated regions depending on the chirp pattern on the phase mask. This produces spectrums that are different than the original, uniform and non-apodized FBGs. The wavelength spectrums produced via chirped FBGs are much broader than uniform profiled FBGs which can range between several nanometers or even thousands of nanometers depending on the chirp rate, grating length and phase mask used for writing the FBG in a fiber [20].

Based on Figure 4, chirped FBG with uniform apodization and follows similar Bragg wavelength, index modulation and grating length as table 1, has been simulated. It was seen that both Figure 4(a) and Figure 4(b) showed broader spectrum that uniform FBG. No shifts of transmission peaks and reflectivity troughs was seen which is expected. The breadth of the peaks is observed to be approximately 4 nm. Both transmission power and reflectivity shows no deterioration when chirp profile is applied without altering other parameters. Figure 4 has successfully demonstrated a chirp profile via simulation that satisfy with other studies [21, 22]. While having wider Bragg peaks in both reflectivity and transmission, chirped FBGs can have certain suitable purposes.

However, if high accuracy detection is preferred, chirped FBG may not be the best grating profile to be utilised. This is due to small shifts that occur upon detection may not be visible or detected and could lead to misinterpretation at the end signal-to-data conversion.

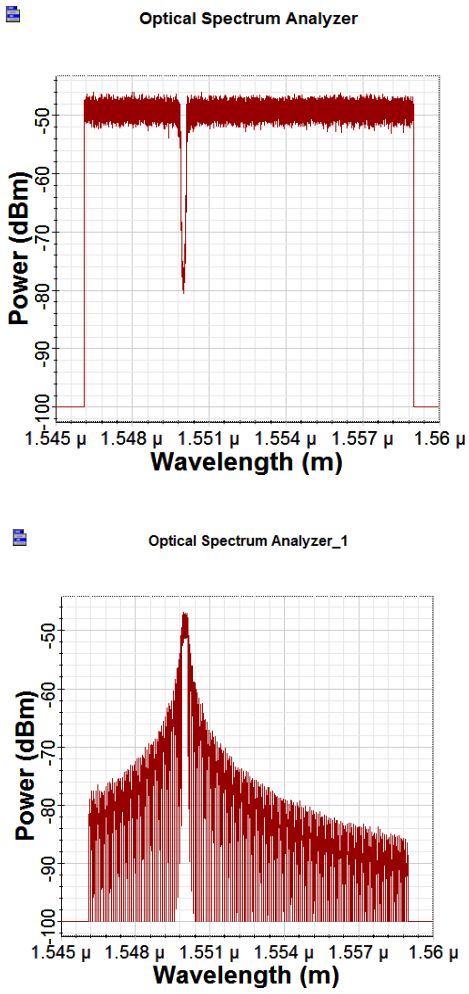
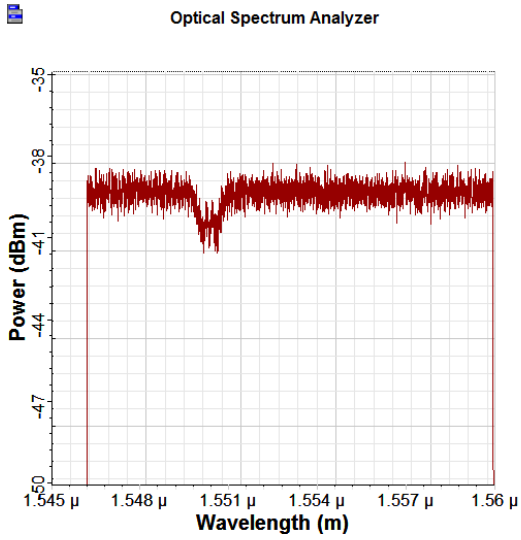
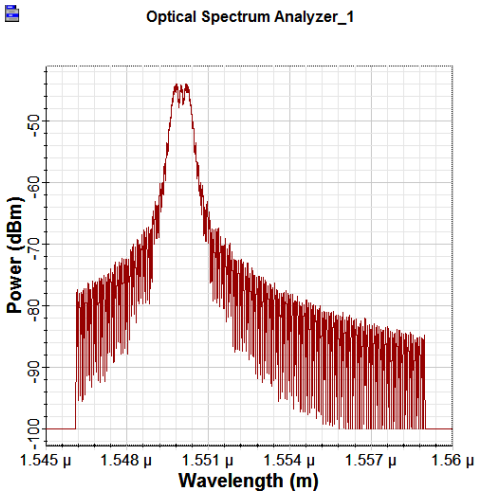


Figure 3: Optical power spectrum of uniform FBG (a) Transmission and (b) Reflectance.

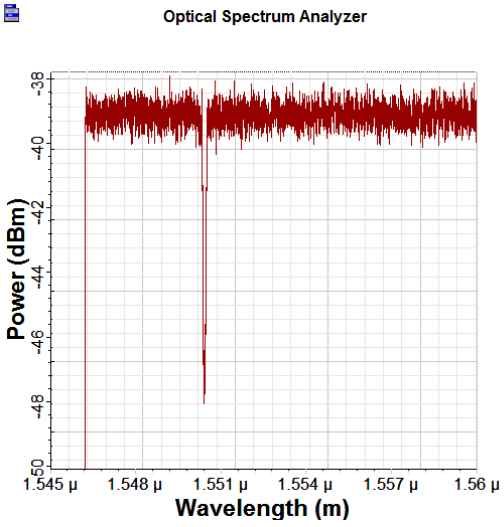


(a)

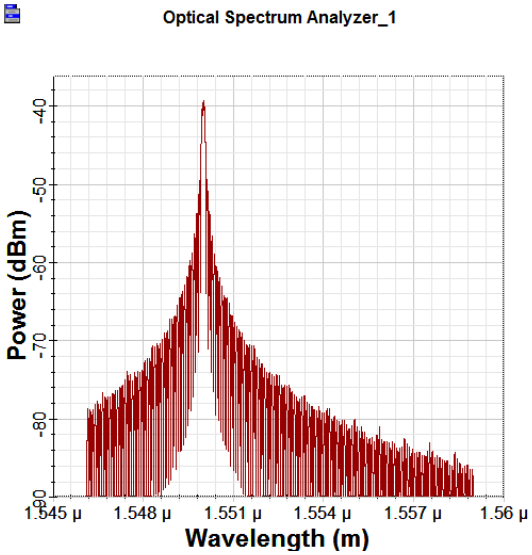


(b)

Figure 4: Optical power spectrum of linear chirped FBG (a) Transmission and (b) Reflectivity.



(a)



(b)

Figure 5: Optical power spectrum of Gaussian apodized FBG (a) Transmission and (b) Reflectivity.

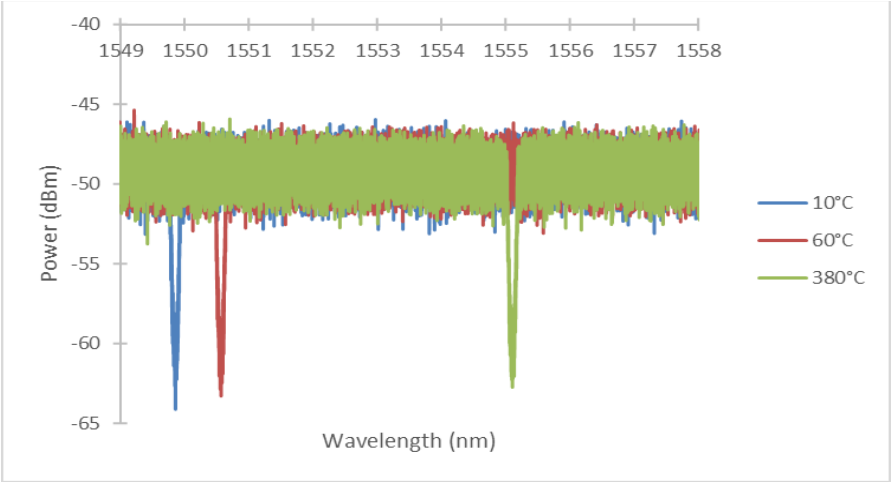
3.1.3. Gaussian Apodized Profile

Another variant of FBG profile other than chirped and uniform are apodized FBGs. These FBGs differ in terms of the way the gratings are imprinted on to the fiber. They are written in a manner that the grating intensities gradually change along the grated region until it becomes similar to the rest of the fiber. Specifically, in gausian apodized FBG, the grating intensity change along the grating length follows Gaussian modelling variant of the original Bragg wavelength equation. The emergence of this type of grating profile serves to reduce side lobes that may appear on uniform grated FBG [8]. Apodized FBG are also known to result in more narrow peaks of both reflectivity and transmission spectra. Besides gaussian model for index modulation and grating intensity determination, other models include Nuttal model and Pi (π) apodized model [23].

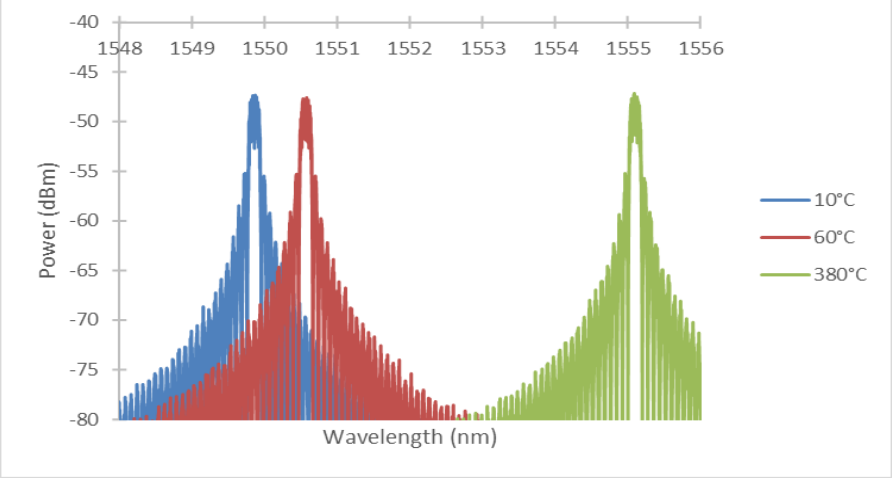
In this study, Gaussian apodized profile was simulated based on table 1 properties of the FBG. Figure 5 shows the simulated transmission and reflectivity spectra of the aforementioned FBG. Side lobe reduction occured and is observable on both of the spectra when compared to a uniform FBG. The peaks reflectivity and transmission trough are slightly narrower if compared with uniform profiles. Reflectivity and transmission power however, does suffer loss of ~10 dbm in comparison with uniform profile FBG. With the presence of side lobe reduction and narrow Bragg peak, utilization of apodized FBG is favourable for accurate sensing. Even though Gaussian apodized FBG may have lower side lobe appearance and narrow peaks, the low transmission power may be a hindrance if the FBG system is to take part in longer distances.

3.2. Effect of Temperature with FBG

FBG type sensors are used in wide variety of temperature ranges depending on the type of industries that utilize it. Simulation of FBG in different range of temperatures is shown on Figure 6. When used in oceanography for monitoring ocean temperatures, a sensor should be at least sensitive at temperature ranges of 10˚C [24]. This posed no problem for FBG sensors array as simulated in Figure 6. At higher than ambient temperatures of around 60˚C, A notable shift of the spectrum from its original wavelength of 1550 nm is detectable. At the said temperature, is observed when an FBG sensor is used in conjunction of a surgical tool to ablate cancer cells [25]. Upon higher temperature use for nuclear reactor temperature monitoring, an FBG may require modifications or special temperature proof casing. As mentioned by Mohamed, et al. [26], temperatures can reach up to 1600˚C. When using normal SMF made of silica, melting point of 400C should be taken into account. This limits the use of FBG if it is grated within a standard SMF down to only ~380˚C. While 390-400˚C is still possible to simulate, in reality, engineering constraints have to be considered for safety.

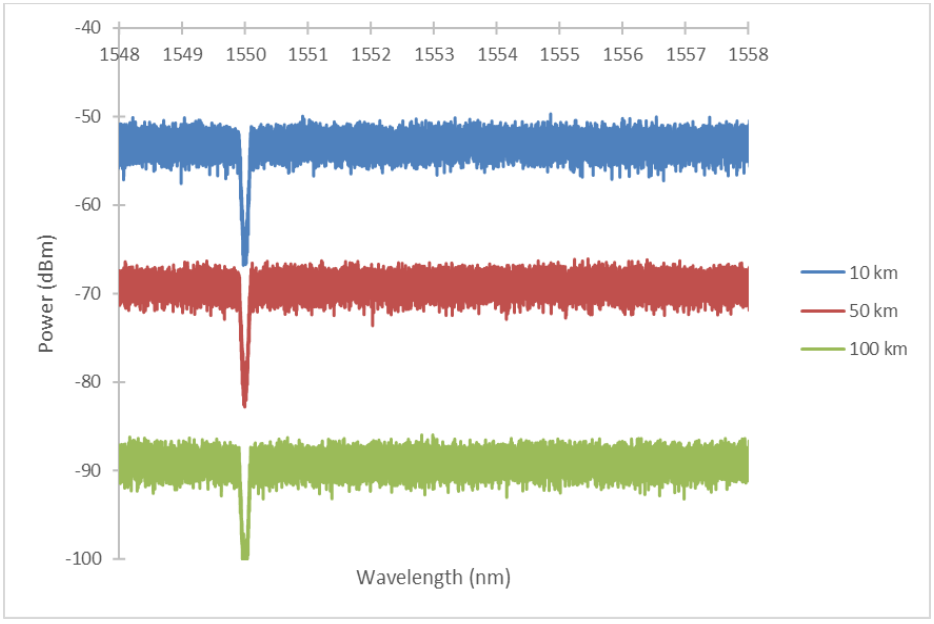


(a)

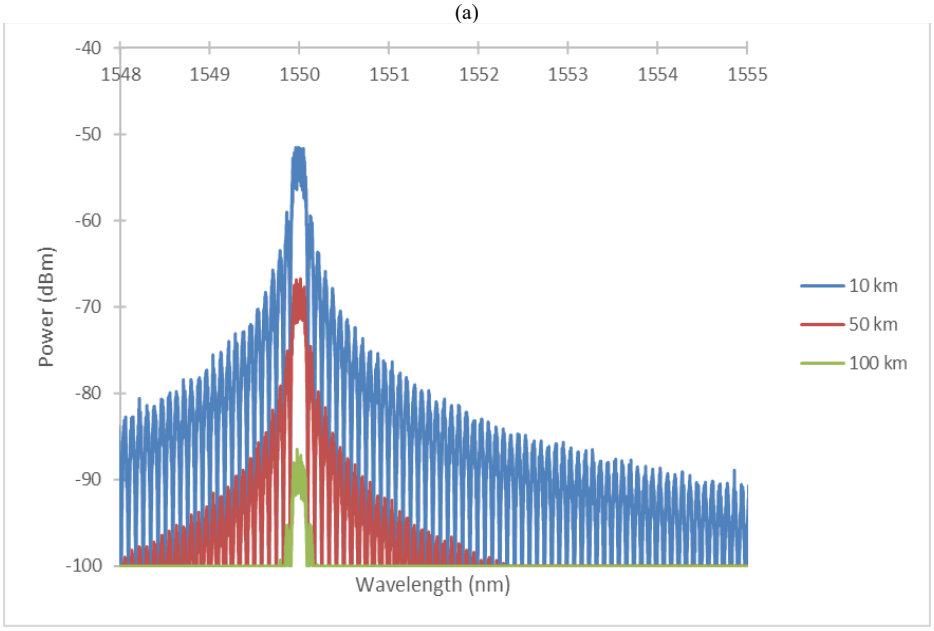


(b)

Figure 6: Optical power Spectra of FBG when exposed to various temperatures (a) Transmission and (b) Reflectivity.



(a)



(b)

Figure 7: Spectra of FBG utilized within various system distance (a) Transmission spectra. (b) Reflectivity spectra.

3.3. Effect of Transmission Power and Distance

Transmission power is at its ideal value when it is simulated or close to its ideal value when tested within lab scale lengths (<1m). When incorporating FBG within an array where the source is further than 1km in distance, power attenuation will occur. The reason for simulating this is to show that at what distance will the transmission and reflected power will deteriorate. The chosen length was based on different uses in the industry. At a distance of 10 km chosen due to most small and medium industries or premises require this distance of FBG array for environment monitoring [27]. For larger area coverage such as land and oceanic surveying applications, 50 km distance may be favourable therefore it is chosen as one of the simulation parameter for fiber distance [28]. For remote sensing in rural areas where the source and signal receiver has to be far from the area of monitoring due to power source availability, may require fiber length of up to 100 km.

Figure 7 (a) and (b) are the spectra of FBG sensor simulated within an optical circuit over various distances. The FBG used for simulation is a uniform profile variant due to it being the most stable in terms of providing low power loss due to attenuation. Furthermore, uniform FBGs are much easier to mass produce due to uniform phase mask availability. Within 10 km distance, there are hardly any power loss for both reflectivity and transmission spectra as shown on Figure 7 (a) and (b). Reflectivity on Figure 7 (b) shown significant attenuation power loss when distance of FBG optical circuit is extended to distances of 50 km and 100 km. Transmission spectra in Figure 6 (a) shown a minimal and slightly noticeable power loss. This is due to that reflected power may experience loss when reflected by the Bragg gratings due to leakage of evanescent waves occurring on FBG grated regions.

3.4. Sustainable Power Source

In environmental sensing, it is essential for light sources of sensors to have stable power input and output. While access to power may be available in established areas and urban environments, some other areas may not connected to main power grid. To remedy this issue, a light source powered by sustainable and renewable energy is favoured. A good candidate for powering light source can be a vibrational sensor which rely on the continuous minute vibration of the earth's crust [1]. In powering light source for optical sensing applications, at least 25 - 50 mW power is required. Vibrational sensors that can resonate with frequencies of <10 Hz at acceleration level (g) by <1.0 can supply enough power to FBG sensors [1]. To achieve this, vibrational sensor proposed within our previous paper [1] was equipped with rectangular clamp-free cantilever beams with attached magnet at the tips. The magnets will vibrate at the low acceleration level causing change in magnetic flux and inducing voltage in the coil. This in turn can produce electrical DC power which can be use with various low power consumption devices. By having a renewable energy source powering the FBG sensors, environmental condition mapping in remote locations is possible.

4. Conclusion and Recommendation

Several forms of FBG grating profiles has been reported via simulation of its transmission and reflectivity spectra. Temperature sensing of FBG has been simulated at various temperatures that are used in different industries ranging from as low as 10˚C to as high as 400˚C. Higher temperatures of >400˚C may require modifications that protects the FBG thermally and mechanically for certain rough applications such as monitoring greater nuclear reactor or volcanic core. When used within a high distance optical circuit, reflectivity was seen to suffer greatly with increasing distance and nearly not usable at distances spanning >100 km. Therefore, it is advisable to rely on remote optical source that is powered by renewable energy source within a FBG sensing system or array that spans greater distances. Multiple FBG of 50 km length should be optimal for use with one remote optical source. For greater coverage, multiple system of FBG and power source each with only 50 km distances between each other may be required for monitoring over a very large area. Furthermore, when an independent and self-powering optical broadband source is used, more remote areas can be monitored such as in vast open seas or subterranean regions. It has been demonstrated that FBG sensor can is capable to operate in power as low as 130 dBm but still maintains proper functions as shown in the obtained results. Although power loss due attenuation was apparent, it does not limit the sensing capability of the FBG to other environmental parameters such as temperature to some extent. This paper provided performance evaluation for FBG of various grating profiles and satisfied the theory of each grating performance.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

The funding provider of this study is the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS), No. FRGS/1/2018/TK10/HWUM/02/2.

References

[1] M. A. Riza, F. M. Foong, Y. I. Go, C. K. Thein, "Development of Sustainable FBG Sensor for Environmental Monitoring Through Vibration Energy Harvesting," in 2019 IEEE International Conference on Sensors and Nanotechnology,1-4,2019,doi: 10.1109/SENSORSNANO44414.2019.8940082.

[2] M. Amanzadeh, S. M. Aminossadati, M. S. Kizil, A. D. Rakić, "Recent developments in fibre optic shape sensing," Measurement, **128**, 119-137, 2018, doi: 10.1016/j.measurement.2018.06.034.

[3] H. Ahmad, M. T. Rahman, S. N. A. Sakeh, M. Z. A. Razak, M. Z. Zulkifli, "Humidity sensor based on microfiber resonator with reduced graphene oxide," Optik, **127**(5), 3158-3161, 2016, doi: 10.1016/j.ijleo.2015.11.184.

[4] X. Jin, S. Yuan, J. Chen, "On crack propagation monitoring by using reflection spectra of AFBG and UFBG sensors," Sensors and Actuators A: Physical, **285**, 491-500, 2019,doi: 10.1016/j.sna.2018.11.052.

[5] C. B. Williams, R. B. Yates, "Analysis Of A Micro-electric Generator For Microsystems," in Proceedings of the International Solid-State Sensors and Actuators Conference - TRANSDUCERS '95, **1**, 369-372, 1995, doi: 10.1109/SENSOR.1995.717207.

[6] M. Rhimi, N. Lajnef, "Tunable Energy Harvesting from Ambient Vibrations in Civil Structures," Journal of Energy Engineering, **138**, 185-193, 2012, doi: 10.1061/(ASCE)EY.1943-7897.0000077.

[7] A. N. Anoshkin, V. Y. Zuiko, K. A. Pelenev, P. V. Pisarev, G. S. Shipunov, "Numerical simulation for development of methodology of stress-strain state control of composite bulkhead for aviation application with the usage of FBG sensors", PNRPU Mechanics Bulletin, **4**, 47-57, 2018, doi: 10.15593/perm.mech/2018.4.04.

[8] N. A. Mohammed, T. A. Ali, M. H. Aly, "Performance optimization of apodized FBG-based temperature sensors in single and quasi-distributed DWDM systems with new and different apodization profiles," AIP Advances, **3**(12), 122125, 2013, doi: 10.1063/1.4859115.

[9] P. Xia, L.-H. Zhang, Y. Lin, "Simulation Study of Dispersion Compensation in Optical Communication Systems Based on Optisystem," Journal of Physics: Conference Series, **1187**(4), 042011, 2019, doi: 10.1088/1742-6596/1187/4/042011.

[10] G. Dastgeer, M. F. Khan, G. Nazir, A. M. Afzal, S. Aftab, B. A. Naqvi, J. Cha, K. A. Min, Y. Jamil, J. Jung, S. Hong, J. Eom, "Temperature-Dependent and Gate-Tunable Rectification in a Black Phosphorus/WS2 van der Waals Heterojunction Diode," ACS Applied Materials & Interfaces, **10**(15), 13150-13157, 2018, doi: 10.1021/acsami.8b00058.

[11] C. C. Wu, W. Y. Chuang, C. D. Wu, Y. C. Su, Y. Y. Huang, Y. J. Huang, S. Y. Peng, S. A. Yu, C. T. Lin, S. S. Lu, "A Self-Sustained Wireless Multi-Sensor Platform Integrated with Printable Organic Sensors for Indoor Environmental Monitoring," Sensors, **17**, 715, 2017, doi: 10.3390/s17040715.

[12] S. F. H. Correia, P. Antunes, E. Pecoraro, P. P. Lima, H. Varum, L. D. Carlos, R. A. S. Ferreira, and P. S. André, "Optical Fiber Relative Humidity Sensor Based on a FBG with a Di-Ureasil Coating," Sensors, **12**(7) 8847-8860, 2012 doi: 10.3390/s120708847.

[13] D. Luo, J. Ma, Z. Ibrahim, Z. Ismail, "Etched FBG coated with polyimide for simultaneous detection the salinity and temperature," Optics Communications, **392**, 218-222, 2017, doi: 10.1016/j.optcom.2016.12.068.

[14] V. G. Annamdas, "Review on Developments in Fiber Optical Sensors and Applications", International Journal of Materials Engineering, **1**(1), 2166-5389, 2012, doi: 10.5923/j.ijme.20110101.01.

[15] A. Boersma, R. Cremers, R. Jansen, "Fiber Bragg Grating Distributed Chemical Sensors," Procedia Engineering, **168**, 1245-1248, 2016, doi: https://doi.org/10.1016/j.proeng.2016.11.435.

[16] Z. L. Poole, P. R. Ohodnicki, A. Yan, Y. Lin, K. P. Chen, "Potential to Detect Hydrogen Concentration Gradients with Palladium Infused Mesoporous-Titania on D-Shaped Optical Fiber," ACS Sensors, **2**(1), 87-91, 2017, doi: 10.1021/acssensors.6b00583.

[17] S. J. Mihailov, "Fiber Bragg Grating Sensors for Harsh Environments," Sensors,**12**(2), 1898-1918, 2012, doi: 10.3390/s120201898.

[18] T. L. Yeo, T. Sun, K. T. V. Grattan, "Fibre-optic sensor technologies for humidity and moisture measurement," Sensors and Actuators A: Physical, **144**(2), 280-295, 2008, doi: 10.1016/j.sna.2008.01.017.

[19] K. O. Hill, Y. Fujii, D. C. Johnson, B. S. Kawasaki, "Photosensitivity in optical fiber waveguides: Application to reflection filter fabrication," Applied Physics Letters, **32**(10), 647-649, 1978, doi: 10.1063/1.89881.

[20] D. Tosi, "Review of Chirped Fiber Bragg Grating (CFBG) Fiber-Optic Sensors and Their Applications," Sensors, **18**(7), 2147, 2018, doi: 10.3390/s18072147.

[21] E. Vorathin, Z. M. Hafizi, A. M. Aizzuddin, M. K. A. Zaini, K. S. Lim, "Temperature-independent chirped FBG pressure transducer with high

sensitivity," Optics and Lasers in Engineering, **117**, 49-56, 2019, doi: 10.1016/j.optlaseng.2019.01.012.
[22] D. M. Wang, L. S. Wang, Y. Y. Guo, Y. C. Wang, A. B. Wang, "Key space enhancement of optical chaos secure communication: chirped FBG feedback semiconductor laser," Optics Express, **27**(3), 3065-3073, 2019, doi: 10.1364/OE.27.003065.
[23] M. A. Riza, Y. I. Go, S. W. Harun, R. R. J. Maier, "FBG Sensors for Environmental and Biochemical Applications—A Review," IEEE Sensors Journal, **20**(14), 7614-7627, 2020, doi: 10.1109/JSEN.2020.2982446.
[24] L. Wang, Y. Wang, J. Wang, F. Li, "A High Spatial Resolution FBG Sensor Array for Measuring Ocean Temperature and Depth," Photonic Sensors, **10**, 57-66, 2020, doi: 10.1007/s13320-019-0550-0.
[25] M. Jelbuldina, A. Korobeinyk, S. Korganbayev, D. Tosi, K. Dukenbayev, V. J. Inglezakis, "Real-Time Temperature Monitoring in Liver During Magnetite Nanoparticle-Enhanced Microwave Ablation With Fiber Bragg Grating Sensors: Ex Vivo Analysis," IEEE Sensors Journal, **18**(19), 8005-8011, 2018, doi: 10.1109/JSEN.2018.2865100.
[26] A. E. N. A. Mohamed, A. N. Zaki Rashed, M. Z. Elsiefy, M. M. Zaky, A. I. Elsaket, "Characterization of Fiber Bragg Grating for applications in Nuclear Research Reactors," The Egyptian International Journal of Engineering Sciences & Technology, **27**, 2019, doi: 10.21608/eijest.2019.97276.
[27] J. Nedoma, M. Fajkus, R. Martinek, "Fiber-Optic Breath Sensors: A Comparison Study," Journal of Biomimetics, Biomaterials and Biomedical Engineering,**40**,56-63,2019, doi: 10.4028/www.scientific.net/JBBBE.40.56.
[28] C. R. U. Kumari, D. Samiappan, K. R, and T. Sudhakar, "Fiber optic sensors in ocean observation: A comprehensive review," Optik, **179**, 351-360, 2019 doi: 10.1016/j.ijleo.2018.10.186.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 920-925 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Barriers and Supports in Engineering Career Development: An Exploration of First-Year Students

Rosmery Ramos-Sandoval*, Jano Ramos-Diaz

Centre for Interdisciplinary Science and Society Studies, CIICS, Universidad de Ciencias y Humanidades, Lima, Lima-39, Peru

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 18 November, 2020
Online: 08 December, 2020

Keywords:
Social Cognitive Career Theory
Barriers
Supports
Engineering student retention

ABSTRACT

Previous research has found that first-years in college is challenging due to changes in academic demands and the adaptation process. In addition to this, research on career development has found that barriers and supports may influence career interest, motivation to continue and student retention. This study explores the influence of specific supports and barriers among first-year engineering college students. To reach this goal, a case study was conducted with 425 engineering students at three universities in Lima, Peru. Based on Social Cognitive Career Theory (SCCT), we performed a Structural Equation Model (SEM) to assess effects of contextual factors such as barriers and supports on self-efficacy, coping self-efficacy, and goals. Results showed that certain supports and barriers influence self-efficacy, cope self-efficacy and goals. Effects of these factors in first-year students and possible strategies for retention in STEM careers are discussed.

1. Introduction

Science and technology are critical subjects for the advancement in societies because they create prosperity, wellbeing and more opportunities [1, 2]. Despite its increasing importance in daily life, the necessary number of science, technology, engineering and mathematics (STEM) workers is not being produced [3]. As a result, this gap has become a major problem, especially in developing countries, given the pressure to become competitiveness in the global economy. Therefore, increase and retention of STEM students, graduates, and workers have vital importance for Latin-American organizations, governments and educational institutions.

Accordingly, Peru showed an increase in higher education enrollments, however, this tend to vary significantly in specific careers. A recent report showed that between 2008 and 2019, 771 900 and 1'509 400 students were registered at national level [4]. Although, the same report showed that in 2017, more than a half (52%) were in social sciences and only 27.8% were in engineering and technology. Nonetheless, the rates of registered students in STEM fields indicated that there was a decline between 2018 and 2019 [4]. Besides, previous official reports indicated that economic issues (31%), job responsibilities (28%), and family issues (14%) [5] were the main reasons to suspend higher education in STEM. Based on the above-mentioned reports, it seems that several factors are involved in Peruvian STEM students' choice to stay or leave university. In addition to this, these rates and factors may contribute to the gap in STEM careers in Peru. As a result, this may have economic and social consequences for the country development.

Therefore, exploring how environmental and behavioral process interact in engineering career behavior may help to understand the STEM gap and first-year Peruvian students experiences in STEM careers. What is more, research on these factors is scarce in Peruvian STEM careers. Therefore, by using a survey design, the present study aims to explore how specific barriers and supports may influence self-efficacy, cope self-efficacy and goals.

1.1. Who leaves and who stays?

First-years in higher education are critical for students due to changes in study style and the time that it takes to adapt from high school to college. Previous research suggested that it is critical to focus on factors related to students retention in early years of STEM careers [6–9]. This advice is particularly relevant especially in certain areas of Peru where a lack of knowledge from parents about higher education constitute an additional challenge for students. Moreover, lack of support in first-year students may generate a biased perception of higher education that may result in withdrawal. Therefore, it is crucial to get insights based on theory and evidence to understand first-year engineering career development in the Peruvian context.

*Corresponding Author: Rosmery Ramos-Sandoval, Email: rramos@uch.edu.pe

Consequently, the literature on engineering career behaviors has highlighted several factors involved in career progression. Previous studies identified that lack of enjoyment, interest in engineering courses, demotivation and unpreparedness were the main reasons to leave engineering careers [8, 10]. In addition to this, Snyder and colleagues [7] found that self-perception, beliefs and motivations may influence competence and expectations for achievement in engineering career. Based on this, psychological factors seems to have effects in how students behave and take decisions as well as their drive to continue engineering careers.

1.2. Self-efficacy, coping self-efficacy and goals

One particular theory that has been useful in the explanation of career development is Social Cognitive Career Theory (SCCT) [11]. SCCT model hypothesizes the association between self-efficacy, outcome expectation, interest, goals, barriers and supports [11, 12]. SCCT proposed that when someone feels competent concerning a specific activity (self-efficacy), they will anticipate positive results (outcome expectation) and will develop interest in that particular domain. As a result, subjects will set goals that later will become in actions, however, these will be affected by barriers and supports. However, this study will focus in how specific barriers and supports affect certain variables of the model. For instance, several models have been constructed in order to explore in more detail specific SCCT variables [13, 14].

The term self-efficacy is defined as beliefs that people has regarding their capabilities to achieve successfully a specific task [15, 16] whereas goals is defined as student's intentions to persist in their career. The term cope self-efficacy is defined as beliefs about one's ability to manage environmental obstacles in a particular domain [12]. In this regard, previous reports have found that specific contextual factors may influence student's self-efficacy, career goals and cope self-efficacy [6, 17, 18]. In addition to this, relationships between barriers and supports may be indirectly related through coping self-efficacy [19]. Similarly, recent reports indicated that goals may affect determination to engage in career behaviors and that it is mediated by barriers and supports [11, 17, 20]. Taken together, these studies support the notion that certain individual variables of SCCT are suitable to understand career progress. Therefore, using certain core variables from SCCT seems suitable to understand in more detail how specific barriers and supports (e.g. Economics, Institutional and Family) interact with self-efficacy, cope self-efficacy and goals.

1.3. Barriers

Much of the current literature on contextual factors in career development pays particular attention to barriers and supports. Barriers refer to negative contextual influences that hinder different aspects of career progress [12, 21]. In addition to this, previous research identified that barriers are associated with career-related behaviors, engineering college student's progress and attrition [22–24]. Moreover, recent studies found that barriers associated with family, social networks and lack of social support may influence career choice and decision to stay or leave current careers [25, 26]. These studies suggest that there is an interaction between internal and contextual factors that may affect student's decision to continue engineering degrees and possible consequences such as withdrawal. Besides, specific differences in culture and context may change how these barriers are expressed in the Peruvian context. Therefore, exploring in more detail how these barriers interact could provide useful insights into Peruvian universities and possible interventions to reduce possible dropouts in engineering college students.

1.4. Support

On the other hand, support is defined as an environmental variable that facilitates the formation and pursuit of an individual's career choice [12]. Previous research has established that parents or family support have a critical role on vocational behavior and educational outcomes in undergraduate students [22]. In addition to this, peer social support was also positively associated with higher levels of aspiration and expectation in high school students [27]. Furthermore, Lent et al. [28] proposed that certain supports such as teaching staff or family are relevant in career decision-making and expectancies of career success. Therefore, by identifying specific supports, we could explore to what degree these may affect students career behavior (e.g. self-efficacy, coping self-efficacy and goals). In addition to this, by exploring these factors, Peruvian educational institutions may minimize the pre-existent barriers that exist in state engineering faculties such as the lack of resources, materials, equipment and quality standards. Moreover, it may contribute to develop strategies for education as a whole and increase student's retention.

Therefore, the present study aimed to explore the relationship between self-efficacy, cope self-efficacy and goals and how they interact with specific barriers and supports. Based on previous research [21–23, 29], the following hypotheses are proposed:

- H1. Social acceptance support have a direct effect on self-efficacy, goals, and cope self-efficacy.
- H2. Family support have a direct effect in self-efficacy, goals, and cope self-efficacy
- H3. Economic support have a direct effect on self-efficacy, goals, and cope self-efficacy.
- H4. Teacher support has a direct effect on self-efficacy, goals, and coping self-efficacy.
- H5. Social barriers have negative effects on engineering student's self-efficacy, goals, and their cope self-efficacy.
- H6. Economic barriers have negative effects on self-efficacy, goals, and cope self-efficacy.
- H7. Gender barriers have negative effects on self-efficacy, goals, and cope self-efficacy.
- H8. Teacher barriers have negative effects on self-efficacy, goal, and cope self-efficacy.
- H9. Racial barriers in have negative effects on self-efficacy, goal, and cope self-efficacy.

2. Methodology

2.1. Data collection

This research was carried out by contrasting the reference sample universe (N=40962; Male=77%; Female=23%) in students population enrolled at three university campuses in the North of Lima, Peru. This information was obtained from government public information [30]. To calculate the sample size, we used the finite population analysis formula which is considered to have a 95% confidence level. The final sample was composed of 425 engineering students (77.4% males; 22.6% females) from ten engineering fields. The mean age of participants was 21 years

(SD=4.02; Min=16; Max=38). As mentioned before, female students in this sample were low (< 23%) and this was aligned with national reports for female engineering students (25%) [30]. Before data collection, ethical standards were considered such as the right to withdrawal, inform consent and anonymity protection.

2.2. Survey characteristics

The Spanish version from the Engineering Field Questionnaire [31] was used from the original version of Social Cognitive Career Theory proposed by Lent, Brown & Hackett [11,12]. This study employed the following SCCT core variables to model the multiple relationships in the hypotheses mentioned before.

The first construct was designed to test participant's self-efficacy. This scale had 4-items with a 9-point Likert scale that ranges from 0 (no confidence at all) to 9 (complete confidence). Example items included: "How confident are you about performing well in your training over the next two courses?" To find out participants goals, 4 items were used to measure student's intention to persist. Each item was scored on a 5-point Likert scale (strongly disagree, disagree undecided, agree, strongly agree). An example item included: "I expect to remain enrolled in this course for the next year". To assess coping self-efficacy, a 7-item scale was used with a 9-point Likert scale that ranges from 0 (no confidence at all) to 9. Some of the questions were: "I feel confident to be able to continue my studies, even if I feel that the classroom environment was not welcoming".

To explore specific supports, a 5-item scale was used. These items were related to positive environmental cues that students received from friends, family, teachers as well as economic factors that facilitate their career. The response format was a 4-point Likert that ranges from 1 (No at all likely) to 5 (Extremely likely). Example items included: "During the first year, did you get financial support to continue your studies?" To identify barriers, a 6-item scale was used. This scale described obstacles that students receive during their career such as discrimination, lack of teacher support and issues with peers. The response format was in Likert format that ranges from 1 (No at all likely) to 5 (Extremely likely). Example items included: "During the first year, have you felt discouraged by the lack of social support?". This questionnaire was adapted using specific procedures to adapt psychological tests [32].The engineering field questionnaire has shown good psychometric properties and it has been used in engineering students [21,33].

2.3. Statistical analysis

Since the study aim was to confirm and explore effects from supports and barriers in first-year students, we designed a Structural Equation Model (SEM) [34]. In addition to this, data were analyzed in SPSS© IBM 23 and modelling was executed in AMOS© 21 [35].

3. Results

3.1. Independent Variables

Barriers and supports were tested as independent variables, individual scores were obtained through exploratory factor analysis (Table 1).

Table 1: Independents variables from barriers and support in first-year

Factors	Independent variable	Number of categories	Estimate
First-year Support	Social Acceptance Support	4	0.705
	Family Support		0.716
	Economic Support		0.632
	Teacher Support		0.620
First-year Barriers	Social Barriers	5	0.606
	Economic Barriers		0.420
	Teacher Barriers		0.367
	Gender Barriers		0.870
	Racial Barriers		0.843

Estimate > 0.50.

3.2. Confirmatory Factor Analysis

A confirmatory factor analysis (CFA) was implemented in order to confirm correlation relationships between the three factors proposed (self-Efficacy, goals & cope self-efficacy). A maximum likelihood (ML) estimation method was performed, providing acceptable model fit to the data: X^2= 216.806; df = 83; p<0.001; CFI = 0.963; TLI = 0.954; RMSEA [90% CI] = 0.062 (0.052; 0.072); SRMR = 0.059.

As shown in Table 2, validity measures and the internal consistency were explored through Cronbach's alpha (α), Composite reliability (CR) and the Average Variance Extracted (AVE). These were confirmed in reliability and convergent validity measures.

Table 2: Cross-construct correlation and validity measures

Factors	α	CR	AVE	Goal	Self-Efficacy	Cope Self-Efficacy
Goal	0.800	0.834	0.568	**0.754**		
Self-Efficacy	0.878	0.866	0.624	0.406	**0.790**	
Cope Self-Efficacy	0.896	0.887	0.531	0.365	0.540	**0.728**

Reliability measures: α > 0.7; CR > 0.7; Convergent validity: AVE > 0.5.

3.3. The validity of the causal structure

When CFA was confirmed, the path diagram modelled was tested in this phase. In order to determinate the predictive effects of barriers and supports on cognitive variables (self-efficacy, goals & cope self-efficacy), a SEM analysis was conducted using ML estimation. The SEM estimation including ML produced the following goodness of fit indices: X^2= 405,946; df = 193; p<0.001; CFI = 0.957; TLI = 0.938; RMSEA [90% CI] = 0.051 (0.044; 0.058); SRMR = 0.079.

The results of path relationships confirmed the model proposed (Fig. 1), indicating a significant standardized regression coefficient and a positive relationship between family supports and self-Efficacy (SE), goals (GL) and cope self-efficacy (CSE), thus, confirming H2a (β=0.192; ρ<0.05), H2b (β=0.166; ρ<0.05),

and H2c (β=0.135; ρ<0.05); at the same time, a relationship between teacher support and CSE was identified, confirmed H4c (β=0.186; ρ<0.05).

Accordingly, negative effects were found on individual variables because barriers were perceived during first year; social barriers on SE & CSE was established as a significant negative condition, confirming H5a (β=-0.185; ρ<0.05) and H5c (β=-0.129; ρ<0.05). Also, teacher barriers had negative effects on SE & CSE, confirming H8a (β=-0.161; ρ<0.05) and H8c (β=-0.300; ρ<0.001); while racial barriers had negative effects on Goals, confirming H9b (β=-0.212; ρ<0.05). In relation to economic barriers, although effects on CSE were identified as significant, the effects were not recognized as negative, thus, rejecting H6c (β=0.11; ρ<0.05).

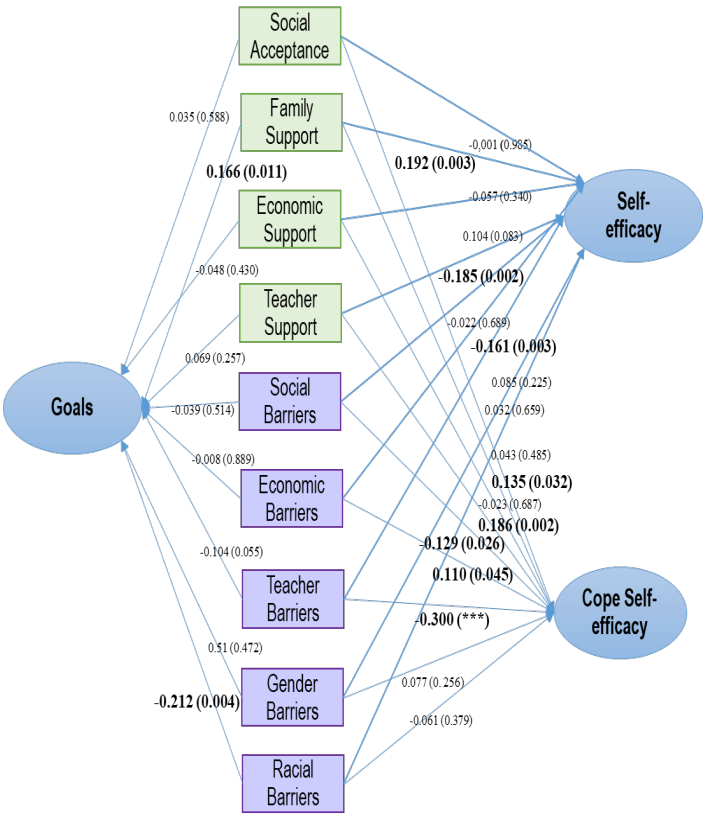


Figure 1: Path relationships confirmed the model proposed. *** p < 0.001; **p < 0.01; *p < 0.05

4. Discussions and Conclusions

The present research aimed to generate insights into contextual factors that may hinder or promote Peruvian engineering student's career progress during their first-year. The present study identified that family support was the major factor that contributes to self-efficacy, goals and cope self-efficacy. A possible explanation for this is that family support may represent acceptance among students because it guarantees economic support for other family members. This study confirms the findings from Peña-Calvo and colleagues [21], who found that family has a strong direct and indirect influence on SCCT factors in undergraduate engineering students. This finding reflects the importance of the effective role that family play in student's college life. For instance, parents may need to understand that students not only face academic challenges but also social and personal challenges (making friends, extracurricular activities, take own responsibility, time management, etc.). Future studies should assess and intervene family dynamics (e.g. authoritarian, permissive parenting style) and misconceptions that parents of first-year students may have about college life.

One interesting finding was the positive and negative effects of teacher's support on cope self-efficacy. This result may be explained by the fact that most of the students in our sample came from families who only had basic education and low income. In addition to this, Peruvian national state universities lack specialized online resources for engineering students, thus, making the role of teacher support critical for students coping and success. These conditions may limit students to find extra academic support and make the role of teachers critical to influence student's capacity to solve issues during their career. These results further support the idea that contextual factors among Peruvian students may influence the decision to continue or dropout higher education [4]. This finding also accords with earlier reports that identified that poor quality in teaching and advice influenced the decision to leave engineering careers [9]. These results have practical implications for Peruvian engineering faculties and teachers because first-year courses are challenging and there are high expectations of academic performance among students [9,10]. Teachers may have the potential to boost self-efficacy on students by providing adequate supports such as positively interpreting their failures, reduce career biases and minimize possible withdrawal.

Another interesting finding was the effect of social barriers (e.g. family, friends) on self-efficacy and coping self-efficacy. It seems possible that negative comments, a sense of not belonging to that particular group and family pressure may diminished students repertoire to cope with adverse situations as well as their judgment on their own capabilities. This result was also reported by Peña-Calvo and colleagues [21] who concluded that peer barriers is a predictor of self-efficacy and goals, which may be determinant for students career development. This finding further support the idea of Wilcox and colleagues [25] who showed the importance of family and friends support in first-year students. This finding highlights implications for student's social relationships and family dynamics. With intention or due to a lack of knowledge, family members and students may negatively influence in the decision to continue engineering degree. As a result, this situation may increase engineering career gaps in Peruvian engineering faculties. Therefore, strategies to enhance positive student's relationships and provide training to families on how to support first-year students may be beneficial.

In this study, racial discrimination had effects on goals. One possible explanation is that racial discrimination may be perceived as a long-term issue and this may affect student's determination to continue engineering career. Another possible explanation is the lack of policies related to race/ethnicity discrimination in student's universities. This finding is consistent with previous studies that found that racial discrimination influence student's decisions to leave engineering career [14, 36, 37]. Further research should be undertaken to investigate other

types of aversive interactions among students such as bullying, sexual orientation and sexual harassment.

The aim of the present research was to examine specific barriers and supports and possible effects on self-efficacy, coping-self efficacy and goals among first-year Peruvian engineering college students. This study may lay the foundation for further exploration in college dropouts, teacher support, racial discrimination and the role of peers in continuing engineering careers. For instance, further studies may explore these issues using qualitative research or longitudinal designs.

A limitation of this study is the low number of women, which may biased results. However, this gender proportion was aligned with national statistics reports [4]. The major limitation of this study is the low number of items to explore specific barriers and supports. Despite its exploratory nature, this study offers some insights into environmental factors that influence engineering career development in the Peruvian context. Further research might explore these weaknesses by using robust psychometric questionnaires and/or conducting multi-group analyses. These findings suggest several courses of action at many levels in Peruvian universities with engineering careers. For instance, policies that systematically assess these specific barriers and supports among students may be helpful to promote engineering persistence in first-year students. In addition to this, economic incentives from government agencies to universities who minimize these barriers (e.g. by creating students unions) may be useful and beneficial to students and society as a whole. In addition to this, training psychologists and tutors with SCCT framework may boost the quality of staff and student's relationships.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We acknowledge the support of the Project 104-2017-FONDECYT funded by *Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica – CONCYTEC* and *Consorcio de Investigación Económica y Social - CIES.*

References

[1] R. Atkinson, M. Mayo, Refueling the U . S . Innovation Economy : Fresh Approaches to Science, Technology, Engineering and Mathematics (STEM) Education, 2010.

[2] R.D. Atkinson, How ICT Can Restore Lagging European Productivity Growth, 2018.

[3] A. Fiszbein, C. Cosentino, B. Cumsille, El desafío del desarrollo de habilidades en América Latina: Un diagnóstico de los problemas y soluciones de política pública, Washington, DC, 2016.

[4] Superintendencia Nacional de Educación Superior Universitaria SUNEDU, II Informe Bienal sobre la realidad universitaria en el Perú, 2020.

[5] Asamblea Nacional de Rectores, Datos Estadísticos Universitarios, Lima, 2011.

[6] R.W. Lent, M.J. Miller, P.E. Smith, B.A. Watford, K. Hui, R.H. Lim, "Social cognitive model of adjustment to engineering majors: Longitudinal test across gender and race/ethnicity," Journal of Vocational Behavior, **83**(1), 22–30, 2013, doi:10.1016/j.jvb.2014.11.004.

[7] K.E. Snyder, S.M. Barr, N.B. Honken, C.M. Pittard, P.A.S. Ralston, "Navigating the First Semester: An Exploration of Short-Term Changes in Motivational Beliefs Among Engineering Undergraduates," Journal of Engineering Education, **107**(1), 11–29, 2018, doi:10.1002/jee.20187.

[8] E. Godfrey, T. Aubrey, R. King, "Who leaves and who stays? Retention and attrition in engineering education," Engineering Education 2010: Inspiring the Next Generation of Engineers, EE 2010, **0052**, 2010, doi:10.11120/ened.2010.05020026.

[9] R.M. Marra, K.A. Rodgers, D. Shen, B. Bogue, "Leaving engineering: A multi-year single institution study," Journal of Engineering Education, **101**(1), 6–27, 2012, doi:10.1002/j.2168-9830.2012.tb00039.x.

[10] B.N. Geisinger, D.R. Raman, "Why they leave: Understanding student attrition from engineering majors," International Journal of Engineering Education, **29**(4), 914–925, 2013.

[11] R.W. Lent, S.D. Brown, G. Hacket, "Toward Unifying Social Cognitive THoery of Career and Academic Interest, Choice, and Performance," Journal of Vocational Behavior, **45**, 79–122, 1994.

[12] R.W. Lent, S. Brown, G. Hackett, "Contextual supports and barriers to career choice: A social cognitive analysis," Journal of Counseling Psychology, **47**(1), 36–49, 2000, doi:http://dx.doi.org/10.1037/0022-0167.47.1.36.

[13] R.W. Lent, T.R. Morris, L.T. Penn, G.W. Ireland, "Social-cognitive predictors of career exploration and decision-making: Longitudinal test of the career self-management model," Journal of Counseling Psychology, **66**(2), 184–194, 2019, doi:10.1037/cou0000307.

[14] R.W. Lent, H. Bin Sheu, M.J. Miller, M.E. Cusick, L.T. Penn, N.N. Truong, "Predictors of science, technology, engineering, and mathematics choice options: A meta-analytic path analysis of the social-cognitive choice model by gender and race/ethnicity," Journal of Counseling Psychology, **65**(1), 17–35, 2018, doi:10.1037/cou0000243.

[15] A. Bandura, "Self-Efficacy Mechanism in Human Agency," American Psychologist, **37**(2), 122–147, 1982.

[16] A. Bandura, "Perspectives on Psychological Science Toward a Psychology of Human Agency," 2013, doi:10.1111/j.1745-6916.2006.00011.x.

[17] R.W. Lent, S.D. Brown, "On Conceptualizing and Assessing Social Cognitive Constructs in Career Research : A Measurement Guide," **14**(1), 12–35, 2006, doi:10.1177/1069072705281364.

[18] R.W. Lent, S.D. Brown, J. Schmidt, B. Brenner, H. Lyons, D. Treistman, "Relation of contextual supports and barriers to choice behavior in engineering majors: Test of alternative social cognitive models.," Journal of Counseling Psychology, **50**(4), 458–465, 2003, doi:10.1037/0022-0167.50.4.458.

[19] I. Fort, A. Murariu, "The Paths Between Gender, Barriers, Social Support, Coping Efficacy, and Educational Goals," Journal of Career Assessment, **26**(1), 68–76, 2018, doi:10.1177/1069072716679924.

[20] M.F. Shoffner, D. Newsome, C.A. Barrio Minton, C.A. Wachter Morris, "A qualitative exploration of the STEM career-related outcome expectations of young adolescents," Journal of Career Development, **42**(2), 102–116, 2015.

[21] J. Peña-Calvo, M. Inda-caro, C. Rodriguez-Menèndez, C. Fernandez-Garìa, "Perceived supports and barriers for career development for second-year STEM students," Journal of Engineering Education, **105**(2), 2016, doi:10.1002/jee.20115.

[22] J.J. Dahling, M.N. Thompson, "Contextual supports and barriers to academic choices: A policy-capturing analysis," Journal of Vocational Behavior, **77**(3), 374–382, 2010, doi:10.1016/j.jvb.2010.07.007.

[23] N.A. Fouad, G. Hackett, P.L. Smith, N. Kantamneni, M. Fitzpatrick, S. Haag, D. Spencer, "Barriers and Supports for Continuing in Mathematics and Science: Gender and Educational Level Differences," Journal of Vocational Behavior, **77**(3), 361–373, 2010, doi:10.1016/j.jvb.2010.06.004.

[24] M.C. Cadaret, P.J. Hartung, L.M. Subich, I.K. Weigold, "Stereotype threat as a barrier to women entering engineering careers," Journal of Vocational Behavior, **99**, 40–51, 2017, doi:10.1016/j.jvb.2016.12.002.

[25] P. Wilcox, S. Winn, M. Fyvie-Gauld, "'It was nothing to do with the university, it was just the people': The role of social support in the first-year experience of higher education," Studies in Higher Education, **30**(6), 707–722, 2005, doi:10.1080/03075070500340036.

[26] K.A. Tate, W. Caperton, D. Kaiser, N.T. Pruitt, H. White, E. Hall, "An Exploration of First-Generation College Students' Career Development Beliefs and Experiences," Journal of Career Development, **42**(4), 294–310, 2015, doi:10.1177/0894845314565025.

[27] M.E. Kenny, D.L. Blustein, A. Chaves, J.M. Grossman, L.A. Gallagher, "The role of perceived barriers and relational support in the educational and vocational lives of urban high school students," Journal of Counseling Psychology, **50**(2), 142, 2003.

[28] R.W. Lent, S.D. Brown, B. Brenner, S.B. Chopra, T. Davis, R. Talleyrand, V. Suthakaran, "The Role of Contextual Supports and Barriers in the Choice of Math / Science Educational Options : A Test of Social Cognitive Hypotheses," Journal of Counseling Psychology, **48**(4), 474–483, 2001, doi:10.1037//0022-0167.48.4.474.

[29] M. Inda, C. Rodríguez, J.V. Peña, "Gender differences in applying social

cognitive career theory in engineering students," Journal of Vocational Behavior, **83**(3), 346–355, 2013, doi:10.1016/j.jvb.2013.06.010.
[30] Superintendencia Nacional de Educación Superior Universitaria SUNEDU, Informe Bienal sobre la realidad universitaria en el Perú, 2018.
[31] M. Rodriguez-Menendez, M. Inda-Caro, J. Peña-Calvo, "Validacion de la teorìa cognitivo social de desarrollo de la carrera con una muestra de estudiantes de ingenierìa.," Educacion **18**(2), 257–276, 2015, doi:10.5944/educXX1.14018.
[32] J. Muñiz, P. Elosua, R.K. Hambleton, "Directrices para la traducción y adaptación de los tests: segunda edición [Guidelines for the translation and adaptation of test:]," Psicothema, **25**(2), 151–157, 2013.
[33] M. Inda-Caro, C. Rodríguez-Menéndez, J.V. Peña-Calvo, "Spanish High School Students' Interests in Technology: Applying Social Cognitive Career Theory," Journal of Career Development, **43**(4), 291–307, 2016, doi:10.1177/0894845315599253.
[34] J.F. Hair, W.C. Black, B.J. Babin, R.E. Anderson, R.L. Tatham, Multivariate data analysis, Prentice hall Upper Saddle River, NJ, 1998.
[35] B. Byrne, Structural equation modeling with AMOS: basic concepts, applications, and programming, 2010, doi:10.4324/9781410600219.
[36] L.Y. Flores, R.L. Navarro, S.R. Ali, "The State of SCCT Research in Relation to Social Class: Future Directions," Journal of Career Assessment, **25**(1), 6–23, 2017, doi:10.1177/1069072716658649.
[37] R.L. Navarro, L.Y. Flores, H.S. Lee, R. Gonzalez, "Testing a longitudinal social cognitive model of intended persistence with engineering students across gender and race/ethnicity," Journal of Vocational Behavior, **85**(1), 146–155, 2014, doi:10.1016/j.jvb.2014.05.007.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 926-936 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Prediction of Vessel Dynamic Model Parameters using Computational Fluid Dynamics Simulation

Nu'man Amri Maliky[1,*], Nanda Pratama Putra[2], Mochamad Teguh Subarkah[3], Syarif Hidayat[4]

[1]*Faculty of Mechanical and Aerospace Engineering, Aerospace Engineering, Institut Teknologi Bandung, 40132, Indonesia*

[2]*Faculty of Mechanical and Aerospace Engineering, Mechanical Engineering, Institut Teknologi Bandung, 40132, Indonesia*

[3]*School of Electrical Engineering and Informatics, Electrical Engineering, Institut Teknologi Bandung, 40132, Indonesia*

[4]*School of Electrical Engineering and Informatics, Electrical Power Engineering, Institut Teknologi Bandung, 40132, Indonesia*

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 17 November, 2020
Online: 08 December, 2020

Keywords:
Hydrodynamic Forces
Dynamic Positioning
Computational Fluid Dynamics
Control System Parameters
Autonomous Boat
Computational Simulation

ABSTRACT

Dynamic positioning, a system to maintain vessel position and heading, is a technology that is used in many vessels and being intensively used as research topic in marine engineering. In order to make this system work properly, an accurate parameters value is needed. This research focuses on finding several parameters needed in this control system, which are resistance and added mass. These parameters are identified using CFD simulation. The method has an advantage of being fairly high accuracy and has a lower cost than the experimental method. The results released from this simulation verified by several empirical methods, namely Holtrop and Mennen for resistance and Ellipsoid for added mass. Baruna Jaya III is used as an implementation object for the simulation. The simulation resulted in a small error compared to the verification method. So, this computational fluid dynamics simulation can be an alternative method for obtaining resistance and added mass values on ships.

1. Introduction

This paper is an extension of work originally presented in ICEVT 2019 [1]. Vessel is still considered as one of the most important vehicles for transportation as it is the largest vehicle for transporting objects in the world. Although it maneuvers at a slow speed, shipping transportation is the most effective mode for moving objects in large quantities and masses. In terms of cost, freight transport by sea also known to be efficient. Not only moving objects, several jobs, such as oil refineries, cable planting, system checking, and welding on the seabed, to monitoring marine areas; require underwater vehicles to facilitate. Recent research in this field has led to a vessel that is more efficient, sustainable, and reduces many human errors while working.

Research that is being intensively currently in progress is the autonomous control system, a control system that allows a vehicle to operate automatically without human intervention. One of the control systems that currently being developed is dynamic positioning. Dynamic Positioning (DP) is a computer-controlled system that automatically maintain the position and the direction of vessel using existing thrusters. It takes many components that must be considered to create a DP control system, such as: computers, sensors, drivers, etc. The determination of the hydrodynamic parameters of the vessel is considered to be important in designing the DP control system.

When talking about vessel, certainly it is necessary to also discuss hydrodynamics. Hydrodynamics is the study of fluid motion with media/substances in the form of a liquid, where the substance cannot be pressed (incompressible), which are influenced by external and internal forces. In the DP control system, there are two important parameters that must be extracted from hydrodynamics, which are drag and added mass. These hydrodynamic forces values can be obtained in several ways, including analytical, experimental, and numerical.

Computational fluid dynamics is the commonly used numerical to obtain the hydrodynamic forces that occur in a vessel. This method is used because it has a fairly high accuracy for initial predictions compared to analytics, but it doesn't require much cost than experiments. To be able to use this method, it is necessary to do some modeling from the real problem to the equations that can be understood by the computer system that will calculate the model.

*Corresponding Author: Nu'man Amri Maliky, maliky.numan@gmail.com

This paper presents the work of CFD simulation that composed into 8 sections. Section 1. Introduction, talking about the review of the CFD simulation research. Section 2. Literature Reviews, mentioning some literatures that support the work. Section 3. Materials and Methodologies, containing some materials like the system, equation, parameters, vessel dimension, and methodology to conduct the simulation. Section 4. Result and Analysis, presenting the simulation result and analyze it. Section 5. Conclusion and Future Works, presenting the conclusion of the work and review what needed to be developed on next research. Section 6. Conflict of Interests, statement from the author for submitting the paper. Section 7. Acknowledgements, gratitude from the authors to whom revise or give any critics to this work. And last Section 8. Reference, presenting the list of reference used in this work.

2. Literature Reviews

There are several literatures that serve as references and paradigms in this research and paper. Basically, this work is referring to [1], which discuss the use of computational fluid dynamics on a barge to find dynamic positioning control system parameters. This publication is the continuation and the expansion of reference [1].

Vessel modeling and control system design is explained further in [2]. This thesis discusses the equation of motion of a ship, where there are several basic parameters that affect the moving ship, namely: mass, Coriolis, and drag/resistance. These parameters will be sought from a ship design to design its dynamic positioning control system [2].

One of the parameters to obtain is added mass. Paper that discuss further about added mass identification is found on [3]. The method to design using computational fluid dynamics to obtain the added mass value is also stated on the reference. The other parameter is drag or resistance. The presence of ship resistance is caused by a combination of several factors. These factors are heavily influenced by a dimension of the ship [3].

Validation is carried out to ensure that the simulation that has been done correctly. The validation uses the empirical calculation of Holtrop and Mennen's Method described in [4]. The calculation for the parameters needed. For added mass calculation, validation uses the ellipsoid and/or strip theory method. These methods explained well in [4], [5].

For the basic theory of computational fluid dynamics and the additional theory are explained well in [6], [7].

Another reference used is the proceedings "Preliminary Computational Fluid Dynamics (CFD) Simulation of EIIB Push Barge in Shallow Water". This paper describes a method for finding the resistance of a ship using computational fluid dynamics. The formation of ship geometry to support the method discussed is described in "Analysis of Hull Resistance of Pushed Barges in Shallow Water". Then, the required mesh is discussed in "Viscous Drag Calculations for Ship Hull Geometry" [8]-[10].

This research main goal is to obtain the hydrodynamic parameters from a vessel through CFD simulation. The detailed parameters can be found from mentioned reference above which are in [1], [2]. The parameters are added mass and vessel drag, where the calculation explained in [3]. This simulation result needs to be verified with the others method. The verification can be done by a method mentioned in [4], [5]. The correct way to do a proper CFD simulation can be found in [8]-[10].

3. Materials and Methodologies

3.1. Dynamic Positioning Control System

DP System consists of the position reference system, propulsion system, and computer control at the deck room. Position reference system acting as a navigation and guidance system for DP. The sensors commonly used for this subsystem are GPS and gyro-compasses. This will inform the vessel position and heading and act as state feedback for control calculation as well.

The control calculation is conducted on the DP computer, which generates a thrust command to each thruster every defined sampling time. The signal from feedback control calculation is converted through a certain allocation algorithm. The full block diagram of the control system shown in Figure 1.

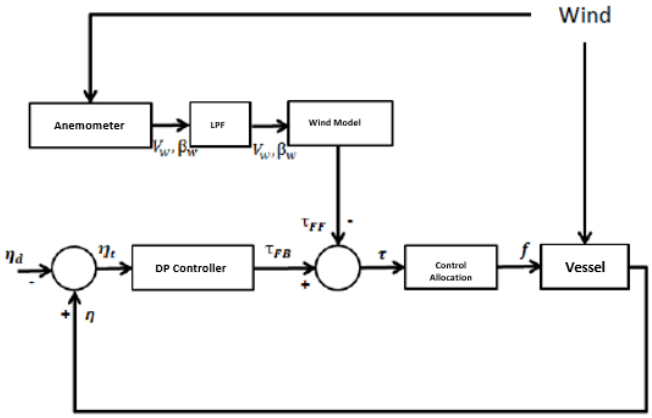


Figure 1: Dynamic Positioning Control System

A wind feedforward control is a common algorithm implemented on DP nowadays. It prevents a bigger deviation of position earlier. The wind signal is filtered with low pass filter to prevent a big oscillation of the control signal. Feedback, feedforward, control allocation, and full control system implementation are explained in [2].

In order to design or implement the controller, a vessel model is required. The model is derived from its dynamic equation of motion which consists of mass, Coriolis, and drag. The equation is can be found in equation 1 with a further detail explanation of the equation can be found in [2].

$$\mathbf{M\dot{v}} + \mathbf{C(v)\dot{v}} + \mathbf{D(v)\dot{v}} = \boldsymbol{\tau} + \boldsymbol{\tau_e} \quad (1)$$

M, C(v), and D(v) stand for mass, Coriolis, and Drag. While τ and τ_e are internal force and external force respectively. DP controls the 3 DoF motion of the vessel, surge, sway, and yaw. Equation 1 is modeled with respect to the motion so the fixed variable, mass, drag, and Coriolis are 3x3 matrix while the rest is a 3x1 matrix.

Mass and Coriolis consist of a rigid body and added mass. While there are linear and nonlinear drags. The variable of the parameters is shown in equation 2 to 10.

$$\mathbf{M} = \mathbf{M_{RB}} + \mathbf{M_{AM}} \tag{2}$$

$$\mathbf{M_{RB}} = \begin{bmatrix} m & 0 & -m\,y_g \\ 0 & m & m\,x_g \\ -m\,y_g & m\,x_g & I_z \end{bmatrix} \tag{3}$$

$$\mathbf{M_{AM}} = \begin{bmatrix} -X_{\dot{u}} & 0 & 0 \\ 0 & -Y_{\dot{v}} & -Y_{\dot{r}} \\ 0 & -N_{\dot{v}} & -N_{\dot{r}} \end{bmatrix} \tag{4}$$

$$\mathbf{C(v)} = \mathbf{C_{RB}} + \mathbf{C_{AM}} \tag{5}$$

$$\mathbf{C_{RB}} = \begin{bmatrix} 0 & 0 & -m(x_G r + v) \\ 0 & 0 & -m(y_g r - u) \\ m(x_G r + v) & m(y_g r - u) & 0 \end{bmatrix} \tag{6}$$

$$\mathbf{C_{AM}} = \begin{bmatrix} 0 & 0 & Y_{\dot{v}} v + \frac{Y_{\dot{r}} + N_{\dot{v}}}{2} r \\ 0 & 0 & -X_{\dot{u}} u \\ -Y_{\dot{v}} v - \frac{Y_{\dot{r}} + N_{\dot{v}}}{2} r & X_{\dot{u}} u & 0 \end{bmatrix} \tag{7}$$

$$\mathbf{D(v)} = \mathbf{D_{linear}} + \mathbf{D_{non-linear}} \tag{8}$$

$$\mathbf{D_{linear}} = \begin{bmatrix} X_u & 0 & 0 \\ 0 & Y_v & 0 \\ 0 & 0 & N_r \end{bmatrix} \tag{9}$$

$$\mathbf{D_{non-linear}} = \begin{bmatrix} X_{u|u|} & 0 & 0 \\ 0 & Y_{v|v|}|v| + Y_{v|r|}|r| & Y_{r|v|}|v| + Y_{r|r|}|r| \\ 0 & N_{v|v|}|v| + N_{v|r|}|r| & N_{r|v|}|v| + N_{r|r|}|r| \end{bmatrix} \tag{10}$$

where

- M_{RB} : mass of constant variable (rigid body)
- M_{AM} : mass of mass dependent variable (added mass)
- C_{RB} : coriolis of constant variable (rigid body)
- C_{AM} : coriolis of mass dependent variable (added mass)
- $X_{\dot{u}}$: surge added mass dependent to surge acceleration
- $Y_{\dot{v}}$: sway added mass dependent to sway acceleration
- $N_{\dot{r}}$: yaw added mass dependent to yaw acceleration
- $Y_{\dot{r}}$: sway added mass dependent to yaw acceleration
- $N_{\dot{v}}$: yaw added mass dependent to sway acceleration
- X_u : surge drag dependent to surge velocity
- Y_v : sway drag dependent to sway velocity
- N_r : yaw drag dependent to yaw velocity
- $X_{u|u|}$: surge drag dependent to surge velocity
- $Y_{v|v|}$: sway drag dependent to sway velocity
- $N_{r|r|}$: yaw drag dependent to yaw velocity

The rigid body of the vessel could be measured. While the added mass and drag parameters should be derived from a computational simulation through Computational Fluid Dynamics (CFD) simulation. Therefore, further hydrodynamic analysis is required to generate an exact model of the vessel.

3.2. Hydrodynamic Forces on a Ship

Hydrodynamic forces on ships can be classified into two forces, which are drag or resistance and added mass. Cited from [3], for an object which moves through a surrounding fluid, the force balance would be denoted by the equation of motion as equation 11. The separated component of hydrodynamics forces can be seen in equations 12 and 13.

$$M\mathbf{a} = \mathbf{F} - \mathbf{F_H} \tag{11}$$

where

m = Mass of the object

a = Acceleration of the object

F = Driving forces of the object

F_H = Hydrodynamic forces acting on the object

$$\mathbf{F_H} = \mathbf{F_D} + \mathbf{F_A} \tag{12}$$

$$\mathbf{F_A} = m_A\mathbf{a} \tag{13}$$

where

F_D = Drag forces acting on the object

F_A = Hydrodynamic reaction forces

m_A = Proportionality constant (added mass)

a = Acceleration of the object

The main idea was to accelerate the fluid that surrounds the body and measures the forces that are acting on the surface of the body. By subtracting the drag F_D from the total forces, we could obtain the hydrodynamic reaction forces F_A that is proportional to the acceleration of the fluid (i.e. the body). [3]

- Resistance of a ship

As the ship moves through the water, the ship would experience opposing forces which are often referred to as drag or resistance. The presence of the drag is caused by a combination of several factors. Some major components of the total drag can be denoted as written in equation 14.

$$R_T = R_V + R_W + R_{AA} \tag{14}$$

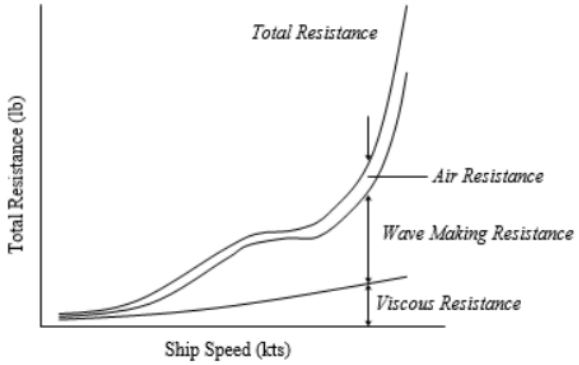
where

R_T = Total hull resistance

R_V = Viscous (friction) resistance

R_W = Wave making resistance

R_{AA} = Air resistance caused by ship moving through calm air



As portrayed in Figure 2, the effect of the wave making resistance tends to dominates as the speed of the vessel increases.

In order to quantify this effect, a dimensionless parameter, "Froude number" (F_n) is used. The Froude Number defined as equation 15.

$$F_n = V/\sqrt{gL} \quad (15)$$

where

V = Velocity

g = Gravity

L = L_{PP}, L_{BP} or L_{WL} of the ship

Sketch of L_{PP}, L_{BP} and L_{WL} of ship can be seen on Figure 3.

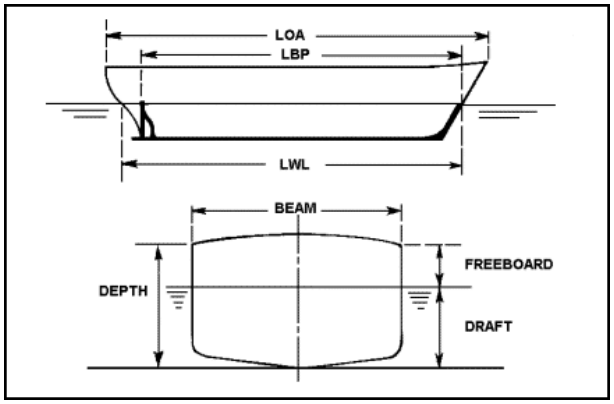


Figure 3: Ship Parameter

L_{PP} or L_{BP} is the length between perpendicular, a longitudinal distance between the forward and aft perpendiculars, and L_{WL} is the length of the hull that is intersecting with the surface of the water. The relation between this Froude number and the effect of the generated wave can be seen in Figure 4. Typically, at service speed, the effect of surface dynamics on the vessel cannot be ignored.

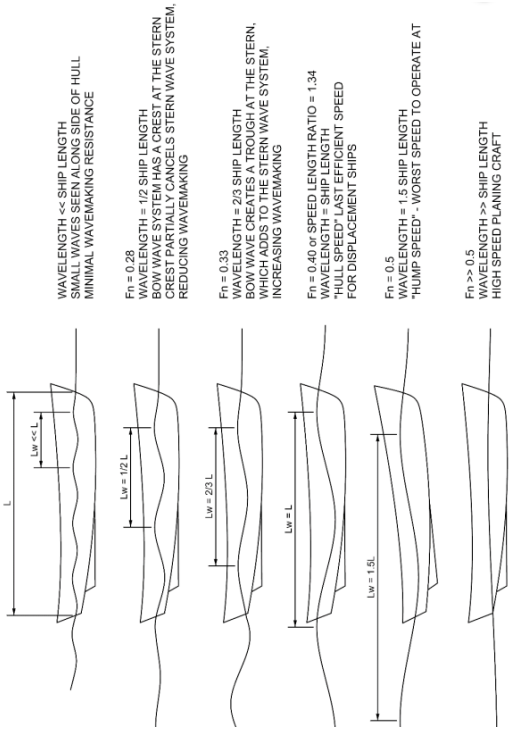


Figure 4: The Effect of Vessel's Speed on Generated Wavelength

The drag or resistance of the ship can be calculated using analytical, computational, and experimental method. Each method has its own advantages and disadvantages. One of the analytical methods that commonly used is Holtrop and Mennen's method. This method used a regression analysis of a vast range of model tests and trial data. However, recommended that, in order to obtain reasonable results from Holtrop's method, the ship should fulfill the following criteria, written in equation 16 to 18. [4]

$$F_n \leq 0.45 \quad (16)$$

$$0.55 \leq C_p \leq 0.85 \quad (17)$$

$$3.9 \leq L/B \leq 9.5 \quad (18)$$

where

F_n = Froude number

C_P = Prismatic coefficient

L/B = Length to breadth ratio

The prismatic coefficient is a dimensionless number from the ratio of the volume of the ship to the volume of the prism, a sketch of the picture can be seen in Figure 5. The equation can be written as in equation 19. [4]

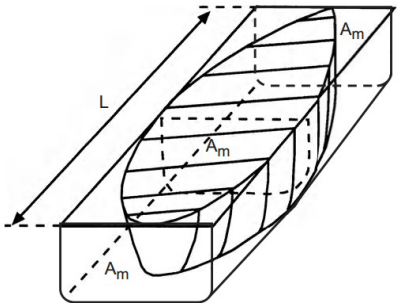


Figure 5: The Prismatic Coefficient

$$C_p = \frac{V}{A_m L} \quad (19)$$

where

V = Submerge volume of the ship

A_m = Maximum cross-sectional submerge area of the ship

L = L_{PP}, L_{BP} or L_{WL} of the ship

Holtrop method computes the total drag by dividing it into several components that can be calculated easily. Equation 20 provides the components contained in the Holtrop calculation. The calculation for each component is shown in equation 21 to 44. [4]

$$R_T = (1 + k)R_F + R_{APP} + R_W + R_B + R_{TR} + R_A + R_{AA} \quad (20)$$

where

R_F = frictional resistance

R_{APP} = appendage resistance

R_A = model–ship correlation resistance

R_W = wave resistance

R_B = resistance due to bulbous bow near the water surface

R_{TR} = pressure resistance due to immersed transom

R_{AA} = air resistance

$$R_F = \frac{1}{2}\rho v^2 S C_F \quad (21)$$

$$C_F = \frac{0.075}{[\log_{10}(Re)-2]^2} \quad (22)$$

$$Re = \frac{\rho v L_{WL}}{\mu} \quad (23)$$

$$k = -0.07 + 0.487118\, c_{14}\, [a_1\, a_2\, a_3\, a_4\, a_5] \quad (24)$$

$$a_1 = \left(\frac{B}{L_{WL}}\right)^{1.06806} \quad (25)$$

$$a_2 = \left(\frac{T}{L_{WL}}\right)^{0.46106} \quad (26)$$

$$a_3 = \left(\frac{L_{WL}}{L_R}\right)^{0.121563} \quad (27)$$

$$a_4 = \left(\frac{{L_{WL}}^3}{V}\right)^{0.36486} \quad (28)$$

$$a_5 = (1 - C_P)^{-0.604247} \quad (29)$$

$$R_{APP} = \frac{1}{2}\rho v^2 (1 + k_2)_{eq} C_F \sum_i S_{APP_i} + \sum R_{TH} \quad (30)$$

$$(1 + k_2)_{eq} = \frac{\sum_i (1+k_{2i}) S_{APP_i}}{\sum_i S_{APP_i}} \quad (31)$$

$$R_{TH} = \rho v^2 \pi d^2 C_{D_{TH}} \quad (32)$$

$$R_W(Fr) = \begin{cases} R_{W_a}(Fr) & \text{if } Fr \le 0.4 \\ R_{W_{a,b}}(Fr) & \text{if } 0.4 < Fr \le 0.55 \\ R_{W_b}(Fr) & \text{if } Fr \ge 0.55 \end{cases} \quad (33)$$

$$R_{W_a} = c_1 c_2 c_5 \rho g V \exp\left[m_1 Fr^d + m_4 \cos(\lambda Fr^{-2})\right] \quad (34)$$

$$R_{W_b} = c_{17} c_2 c_5 \rho g V \exp\left[m_3 Fr^d + m_4 \cos(\lambda Fr^{-2})\right] \quad (35)$$

$$R_{W_{a,b}} = R_{W_a}(0.4) + \frac{20Fr-8}{3}\left[R_{W_b}(0.55) - R_{W_a}(0.4)\right] \quad (36)$$

$$R_B = 0.11\rho g\left(\sqrt{A_{BT}}\right)^3 \frac{{Fr_i}^3}{1+{Fr_i}^2} e^{-3.0{P_B}^{-2}} \quad (37)$$

$$P_B = 0.56\frac{\sqrt{A_{BT}}}{T_F - 1.5h_B + h_F} \quad (38)$$

$$Fr_i = \frac{v}{\sqrt{g(T_F - h_B - 0.25\sqrt{A_{BT}} + h_F + h_W)}} \quad (39)$$

$$R_{TR} = \frac{1}{2}\rho v^2 A_T c_6 \quad (40)$$

$$R_A = \frac{1}{2}\rho v^2 (C_A + \Delta C_A)[S + \sum S_{APP}] \quad (41)$$

$$C_A = 0.00546(L_{WL} + 100)^{-0.16} - 0.002 + 0.003\sqrt{\frac{L_{WL}}{7.5}}\, {C_B}^4 c_2 (0.04 - c_4) \quad (42)$$

$$\Delta C_A = \begin{cases} 0 & \text{if } k_s = 150\ \mu m \\ \dfrac{0.105 k_s^{\frac{1}{3}} - 0.005579}{L_{WL}^{\frac{1}{3}}} & \text{if } k_s > 150\ \mu m \end{cases} \quad (43)$$

$$R_{AA} = \frac{1}{2}\rho v^2 A_V C_{DA} \quad (44)$$

where

ρ = density of the fluid

μ = viscosity of the fluid

v = ship velocity

L_{WL} = length in waterline

B = beam

T = mean draft

T_A = draft at aft perpendicular

T_F = draft at forward perpendicular

V = volumetric displacement

C_P = prismatic coefficient (based on L_{WL})

C_M = midship section coefficient

C_{WP} = waterplane area coefficient

C_{stern} = stern shape parameter

c_i, m_i = specific coefficient, detail explanation in [4]

A_V = area of ship and cargo above waterline

A_T = immersed transom area

A_{BT} = transverse area of bulbous bow

h_B = height of center of A_{BT} above basis

D = propeller diameter

S = wetted area of the hull

S_{APP} = wetted surface of appendages

- Added mass of a ship

As the vessel accelerates through the surrounding fluid, the vessel would experience an opposing force that is proportional to the magnitude of acceleration. Recalling the general equation of motion for the vessel is on equation 1, The matrix **M** could be separated into 2 components; the inertia matrix of the rigid body, and the added mass, can be seen in equation 2 to 4.

The computational fluid dynamics simulation can be a solution for obtaining the value of added mass. Another approach to calculate the added mass is explained in [5]. The paper compared the ellipsoid and Lewis transformation method to the experimental value and gives recommendations on which method to use.

The equivalent ellipsoid method is used to estimate the amount of added mass on the ship by modeling the ship as an ellipsoid form. The modeling sketch can be seen in Figure 6. Thus, the added mass value in the surge direction can be calculated using equation 45 to equation 48. [5]

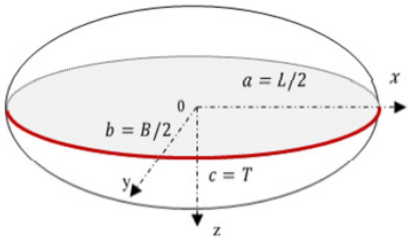


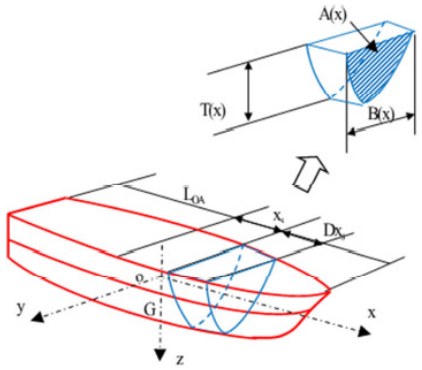
Figure 6: The ship is modeled as an ellipsoid

$$m_{11} = mk_{11} \quad (45)$$

$$k_{11} = \frac{A_0}{2-A_0} \quad (46)$$

$$A_0 = \frac{2(1-e^2)}{e^3}\left(\frac{1}{2}\ln\left(\frac{1+e}{1-e}\right) - e\right) \quad (47)$$

$$e = \sqrt{1-\frac{b^2}{a^2}} \quad (48)$$



Meanwhile, the strip theory method divides the ship into several pieces that resemble the hull shape of the ship. The sketch of this method can be seen in Figure 7. Due to the complexity of each hull cut to shape on the ship, the Lewis transform is used to simplify calculations. However, to simulate the added mass of ships in the surge direction, it is recommended to use the ellipsoid method. [5]

3.3. Computational Fluid Dynamics

Computational Fluid Dynamics is a method that uses numerical analysis and data structures to analyze and solve problems involving fluid flow. Computers are used to carry out the calculations needed to simulate fluid flow and fluid interactions with surfaces determined by boundary conditions. Analysis using the CFD method requires some modeling which is the conversion of physical data in the real world to a property understood by the software. Therefore, a technical approach is needed for each variable to be analyzed. Figure 8 shows the process of the computational fluid dynamics method [6].

Turbulence in CFD can be modeled with several governing equations. Each model has advantages and disadvantages. The choice of turbulent models is determined by the flow phenomenon to be simulated or analyzed. For obtaining the resistance and added mass, this research is using k-epsilon for the turbulence model. This model was chosen because it has the advantage of being robust and good for getting initial values based on what ANSYS has written. It also tends to require less computer memory [7].

Because the effect of surface dynamics on the vessel cannot be ignored, in order to obtain the whole value of drag force on the ship, it is necessary to conduct a two-phase simulation. The two-phase simulation consists of water on the bottom and air on the top of the computational domain. To perform a two-phase simulation, a computational domain that can capture the phenomena is needed. A suitable domain for the simulation described in [8], the sketch can be seen in the Figure 9.

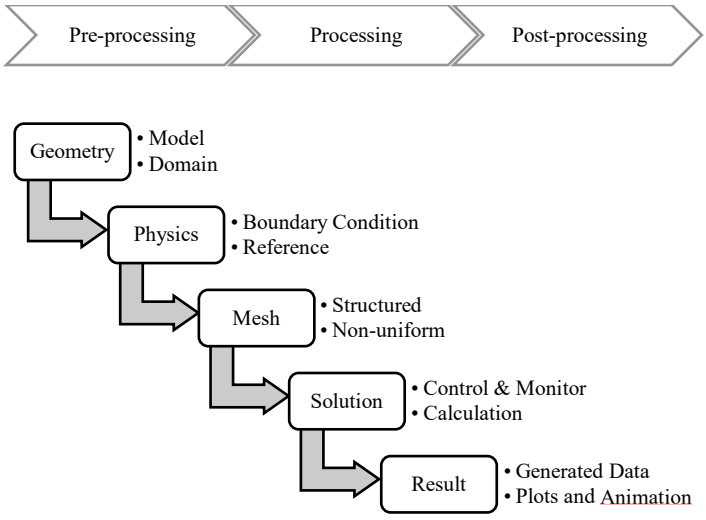
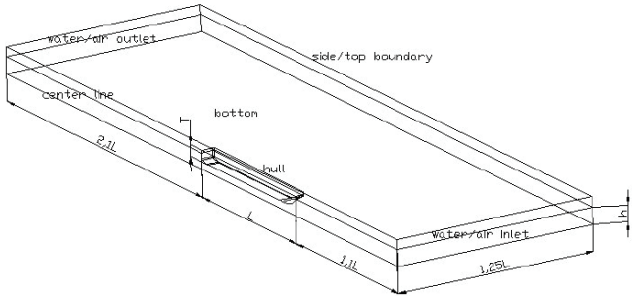
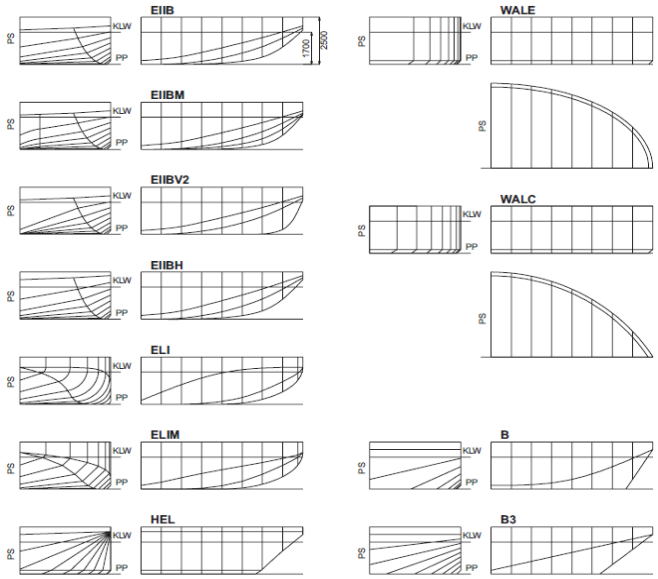


Figure 8: CFD Simulation Flowchart



The model of the ship should also be created in a way that would reflect the original ship. Oversimplification of the hull shape may result in a significant difference in the CFD results. For simulations in shallow water, the results might be significantly different from the ones that are conducted in the deep-water computational domain. As a rule of thumb, for barge ship, there exist several types of ship hulls discussed in [9]. These types of hulls can be seen on Figure 10.



For free surface flows, the meshing process needs to be able to discretize the boundary layer and the free surface. Scaling and the usage of structured mesh might be used to save computational time [8]. The computational domain for two-phase simulation could be created based on flat-plate boundary layer theory and setting the parameter $y^+ = 1$. Y^+ can be approached by using equation 49 [10].

$$y^+ = 0.172 \left(\frac{y}{L}\right) Re^{0.9} \quad (49)$$

where

y = Distance from wall surface

L = Body length

Re = Reynold numbers

3.4. Methodology

The methodology is divided into two parts: obtaining the value of resistance and added mass. Although using the same method, computational fluid dynamics, these two components have quite different steps for obtaining them because of the unique characteristics of each component. Flowcharts can be seen in Figure 11 and Figure 12 respectively.

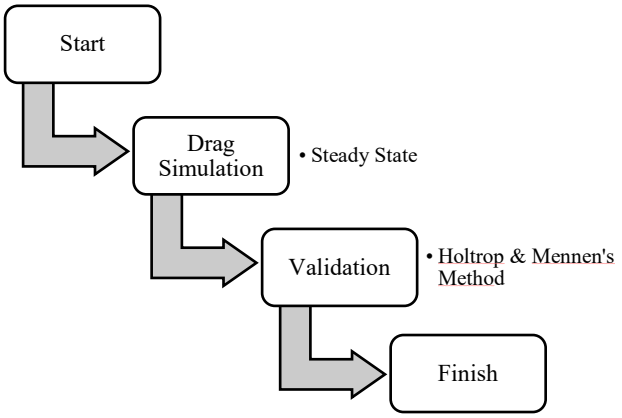


Figure 11: Drag Simulation Flowchart [8]

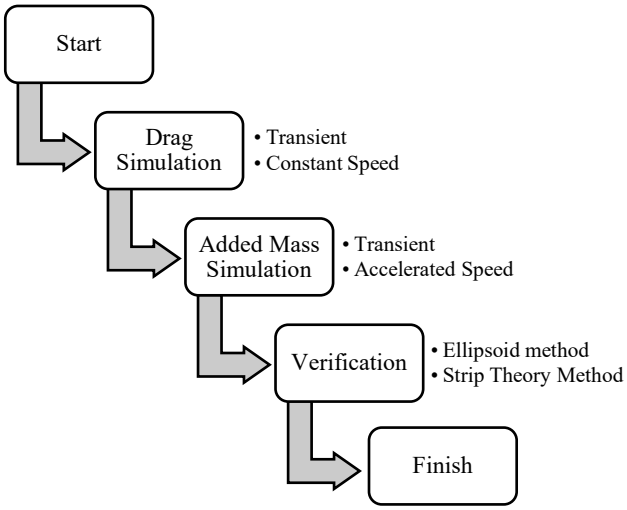


Figure 12: Added Mass Simulation Flowchart [3]

The CFD was conducted by using ANSYS FLUENT software. Both obtaining value of resistance and added mass use the same simulation settings and boundary conditions. Simulation settings and boundary conditions are listed in Table 1 and Table 2 respectively. The boundary condition for resistance is similar to the one that is listed in [8] and for added mass is similar to [3]. For the calculation of added mass, it is necessary to define the speed change of the ship in the simulation.

Table 1: Simulation Settings

Parameters	Settings
Solver	Pressure based, steady/transient
Model	Multiphase, VOF
Vol. Fraction Parameters	Open channel flow - implicit
Interpolation method	Presto! QUICK

Table 2: Boundary Condition

Region	Settings
Ship	Wall
Top/Side/Bottom	Wall
Inlet	Pressure inlet, open channel, vessel speed, free-surface and bottom level defined.
Outlet	Pressure outlet, open channel, free-surface and bottom level defined.
Symmetry	Symmetry

Air (FLUENT database) and sea water (user defined, $\rho = 1025$ kg/m^3, $\mu = 0.00109$ Pas) were used as the fluids in the analysis. The choice of viscous model such as k-epsilon or SST would significantly affect the result as demonstrated in [8]. As the symmetry method were used in the analysis, the results obtained to describe forces must be multiplied by two for the full-scale result. The value of coefficient might result in wrong amount if the references values were not defined. The simulation should be iterated until convergence criteria are achieved. Typically, the criteria for convergence are as follow.

- RMS residual values are below 10^{-4}
- The solution imbalance is less than 1%
- Quantity of interest shows a steady value

To make sure that the result obtained is mesh independent, the simulation above should be repeated several times with multiple number of mesh. The solution is considered to be mesh independent if the increase on number of mesh doesn't affect the result by significant values.

The result should be validated with data in order to make sure that the simulation reflect the actual situation. The easiest method would to compare the results with experimental data obtained from the vessel. However, the experimental data is unavailable, so, this research will use Holtrop and Mennen's method [4] for the validation of the drag and approximation for added mass.

3.5. Implementation

This particular research used ship called Baruna Jaya III as the geometric model and calculation references. Baruna Jaya III is one of the former TNI-AL warships built in 1990. This ship is used for mapping tasks for areas in Indonesia. Figure 13 shows the photo of the Baruna Jaya III ship. Technical drawings of Baruna Jaya III can be seen in Figure 14 to Figure 17.

Figure 13: Baruna Jaya III: Ship Photo

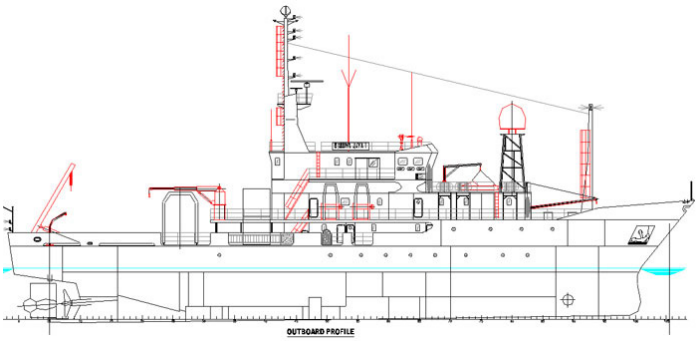


Figure 14: Baruna Jaya III: Outboard Profile

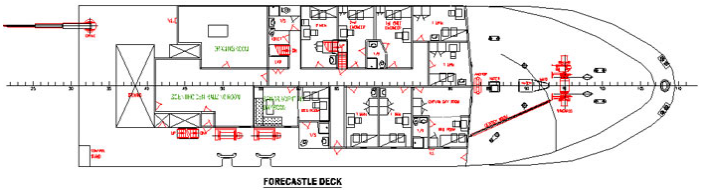


Figure 15: Baruna Jaya III: Wheelhouse deck dan Forecastel deck

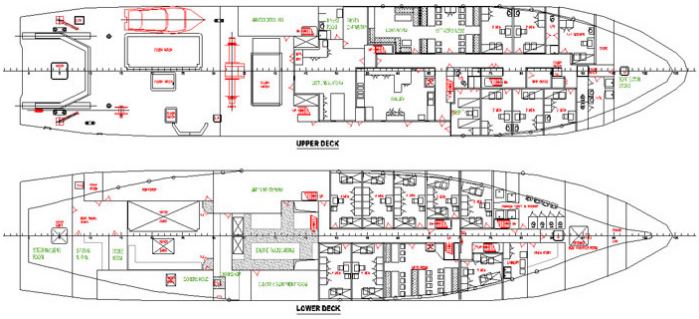


Figure 16: Baruna Jaya III: Upper deck dan lower deck

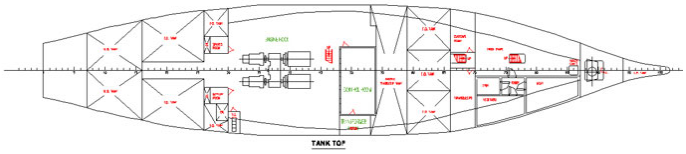


Figure 17: Baruna Jaya III: Tank Top

Table 3: Baruna Jaya III Dimension

Dimension	Value
LOA	60.40 m
LBP	52.39 m
Breadth	11.60 m
Draft Mean	5.70 m
Speed	8 knots

The pictures and dimensions listed in Table 3 will be used as references for modeling the ship to be carried out the simulation and calculating the analytical solution.

Geometry

The geometry of the ship was modelled to be able to represent its original shape while considering the approximation as discussed earlier. The geometry of the ship made must be able to represent its original shape. Therefore, the approximation discussed earlier is needed. Figure 19 shows the CAD of the Baruna Jaya Ship which has been approximated from its original form and ready for further processing. Detailed approximations are required on the bottom of the vessel which is subject to the seawater phase. This section needs to be quite detailed because even a few changes will affect the final result.

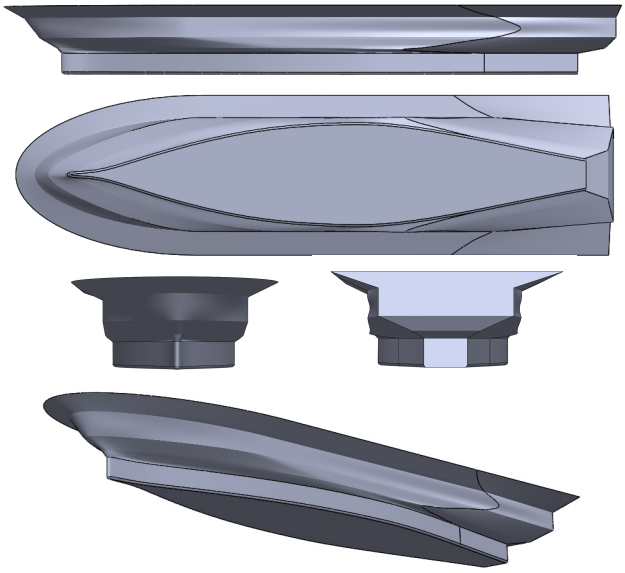
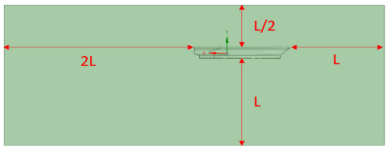
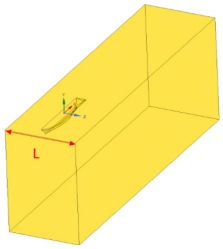


Figure 18: CAD for CFD Simulation of Baruna Jaya III

Meshing

After obtaining a suitable geometry to represent the ship, the next process is meshing the ship's far-field. The computational domain must be defined here. Figure 19 gives the computational domain used by this implementation.





Additional geometry in the form of blocks was added in the computational domain to form a body of influence in the computing domain. This is done to form a body of influence in the computing domain. Thus, the mesh will be smoother on that part. The computational domain for body of influence can be seen on Figure 20.

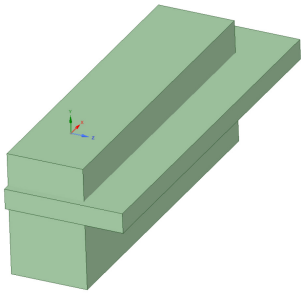


Figure 20: Computational Domain for Baruna Jaya III

The meshing process or discretization of the model is done using face sizing on the surface area of the ship with the dimensions of the mesh is 0.84 m. Meanwhile, for other computing domains, body sizing with a size of 1.2 m is used with the choice of the body of influence activated. The total number of mesh is 1,701,030 elements. The result mesh can be seen on Figure 21 and Figure 22.

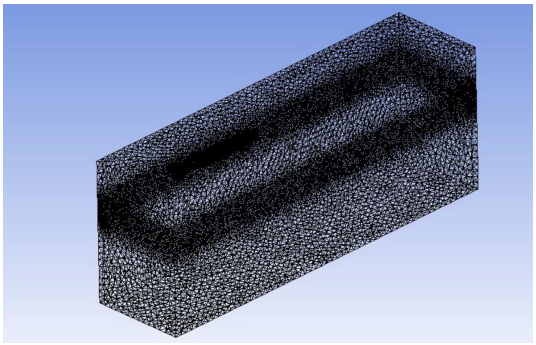


Figure 21: Mesh for Baruna Jaya III

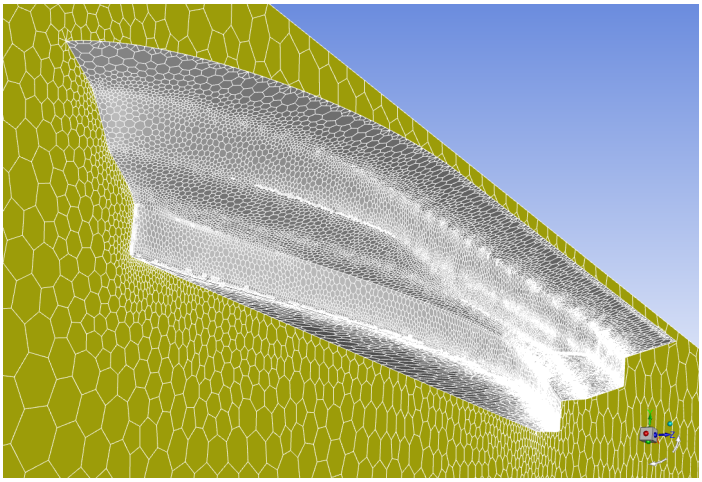


Figure 22: Symmetry and Hull Mesh of Baruna Jaya III

Simulation Setting

As mentioned before, simulation settings will follow [8] for resistance and [3] for added mass. Time setting should be steady for resistance and transient for added mass. Figure 23 shows the simulation settings, the other settings will follow the default one.

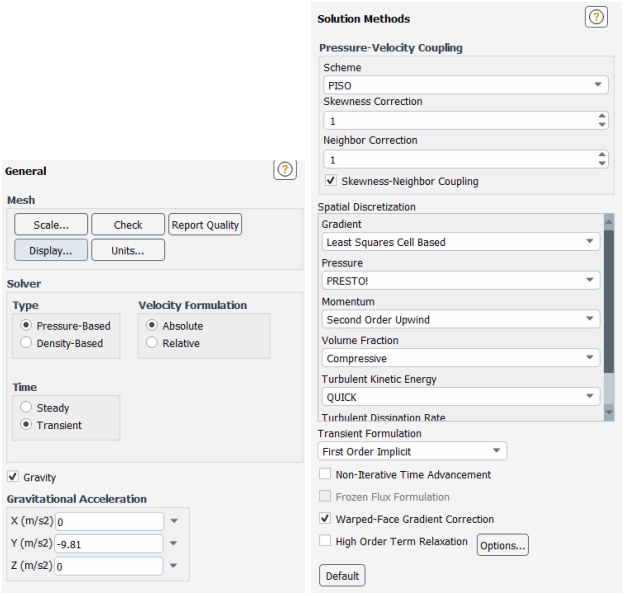


Figure 23: Simulation Settings on Fluent: General and Solution Method

The multiphase model used is the volume of fluid (VOF). The formulation is set to implicit form and the open channel flow option is activated. For the turbulence model, the simulation uses Realizable k-epsilon. These settings can be seen on Figure 24.

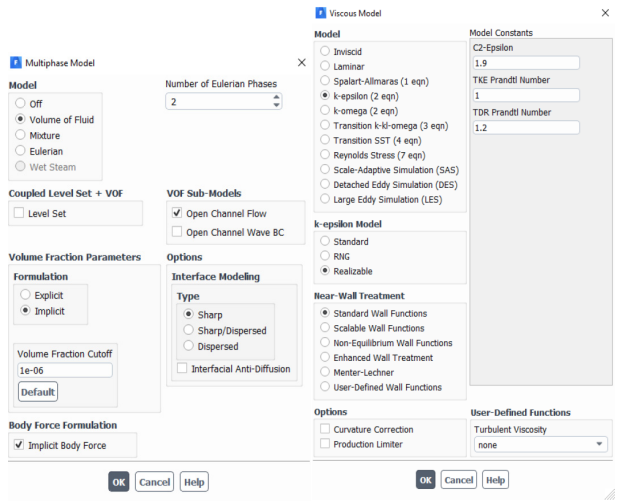


Figure 24: Simulation Settings on Fluent: Multiphase and Turbulence Model

The inlet portion of the computational domain is defined as the pressure inlet. In the inlet, the open channel option is activated and defined based on the entry speed and the location of the bottom and free surface in the compute domain. Based on measurements in the model, -62.5 m refers to the lower part of the computational domain and 2.2 m is equivalent to a draft height of 5.7 m. The input settings are listed in Figure 25.

The simulation is carried out until the convergence criterion is reached. Based on previous experiments, the simulation will converge before 900 iterations. The timestep is set to automatic with a scale factor of 0.2. The convergence criterion is set to obtain loose convergence, or the value of each residue is below 10^{-4}.

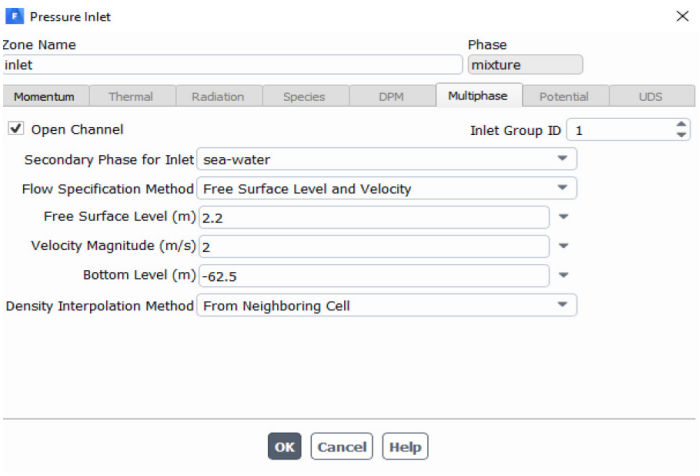


Figure 25: Simulation Settings on Fluent: Inlet Input

For the added mass, there is one simulation setting that has to be defined, which is the function of velocity magnitude to get the effect of added mass. The simulation setting should be defined as Figure 26.

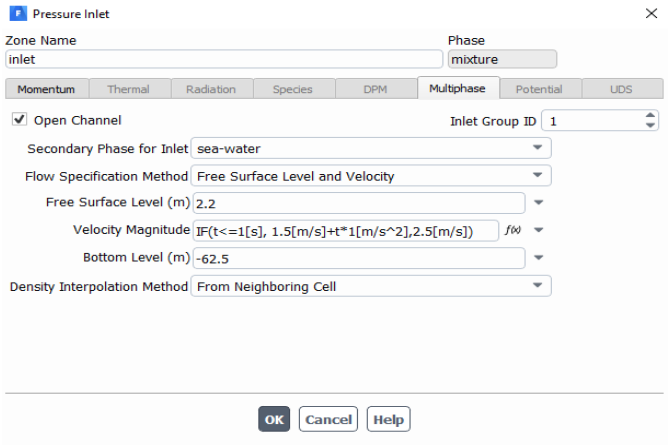


Figure 26: Simulation Setting Addition for Added Mass

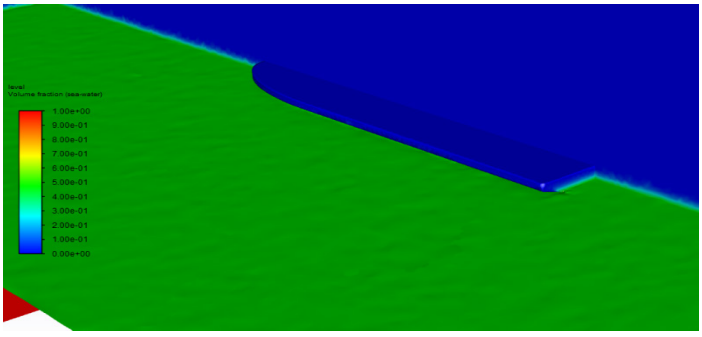


Figure 27: Visualization of CFD Simulation

4. Result and Analysis

The results obtained from the simulations that have been carried out are the value of drag in Newton units and the value of added mass in kilograms which is derived from the force by the acceleration. Visualization of both can be seen in Figure 27.

The validation of the drag simulation results on the ship is done by comparing the values obtained from CFD, with the theoretical values proposed by Holtrop and Mennen. The geometric parameters of the ship are calculated first to check whether this method is valid for use on the ship that is tested. The geometric model parameters of the ship that obtained from the CAD can be seen in Table 4 and the prerequisite geometry table can be seen in Table 5.

Table 4: Geometric Model Parameters

Parameter	Symbol	Value	Unit
Waterline Length	L_{WL}	57.576	m
Length Overall	L_{OA}	60.4	m
Beam/Breadth	B	12.379	m
Depth	d	6.5	m
Draft/Draught	T	5.7	m
Displacement	∇	2580.207	m^3
Wetted Surface Area	S	1170.01	m^2
Cross Section Area	A_M	61.23	m^2
Immersed Transom Area	A_T	38.819	m^2

Table 5: The Geometric Prerequisites for the Holtrop and Mennen Method

Parameter	Symbol	Value
Froude Number	F_n	0.084
Prismatic Coefficient	C_p	0.732
Length to Breadth Ratio	L/B	4.651

Because the values obtained are based on the calculations in Table 5 in correspondence with the criteria, Holtrop and Mennen's method is suitable for use as a verification method. Calculation of Froude number used equation 15 and prismatic coefficient using equation 19. The ship parameters used for this calculation use the specifications previously given.

In this verification, the components that are considered to have a large contribution to the total drag on the ship are drag due to friction, drag due to wave generation, and additional pressure drag due to the submerged transom area of the ship. The other components are considered to have no significant effect on the total drag, so they are assumed to be zero. Based on the calculation method written in [4], equation 21 to 44, it is found that the value of each component is listed in Table 6.

Table 6: Resistance Component of Baruna Jaya III

Resistance Components	Value
Friction	7003.9 N
Wave Generation	0.012 N
Submerge Transom Area	14830.2 N
Total	21834.14 N

For added mass, validation uses the ellipsoid method. The calculation can be done by substituting the ship dimension parameters into equations 19 to 20. From these calculations, the results are shown in Table 7.

Table 7: Surge Added Mass (Ellipsoid Method)

Parameter	Symbol	Value	Unit
Ellipticity	e	0.9850246	
	A_0	0.0907254	
	k_{11}	0.0475182	
Surge Added Mass	m_{11}	123547.40	kg

Recaps for the simulation results of resistance and added mass data can be seen in Table 8 and Table 9, combined with the method values for validation.

Table 8: Result for Drag Simulation

Method for Resistance	Result
CFD Simulation	21030.37 N
Holtrop and Mennen	21834.14 N
Error	-3.681%

Table 9: Result for Added Mass Simulation

Method for Added Mass	Result
CFD Simulation	115719.91 kg
Ellipsoid	123547.40 kg
Error	-6.764%

The simulations carried out can capture the phenomena that occur well and have a value that is quite close to the analytical results. The error has a value below 5% for resistance or drag simulation and has 6% error for added mass simulation. It means this CFD simulation method is accurate in analyzing a ship, especially Baruna Jaya III.

The difference between the simulation and analytical results could occur due to modelling and analytical result value. Error in modelling could happen due to simplification of the original shape that was modelled into 3D Model, the convergence of the mesh, or the simplified equations used. If the modeling of the CFD simulation is the problem, it is necessary to check the step-by-step modeling to be able to trace what went wrong. In terms of analytical result value, further comparison with experimental results must be done do verify the results. Further research should be taken to solve this problem.

5. Conclusion and Future Works

Drag and added mass estimation using computational fluid dynamics has been carried out and gives a drag simulation result of 21030.37 N, verified using the Holtrop and Mennen method, and added mass value of 115719.91 kg, verified by the Ellipsoid method. These estimated values have a very low error compared to the verification value. This means that the CFD simulation can be an alternative method to obtain drag and added mass. The methodology can be used to capture the hydrodynamic forces that occur on the ship, especially Baruna Jaya III, and complete the parameters required for a dynamic positioning control system. With a cost lower than the experimental method, computational fluid dynamics could be considered as the best option for getting the hydrodynamic parameters of the dynamic positioning control system.

In the future, it is necessary to investigate the added mass method in the CFD simulation further so that it can have a value closer to the analytical solution. Validation also needs to be compared with the experimental results so that values are more convincing because they are compared with actual values. The values of drag and added mass that occurs in the sway and yaw direction of the ship also will be further investigated in order to complete all the parameters needed to create a dynamic positioning control system.

Conflict of Interest

The authors: Nu'man Amri Maliky, Nanda Pratama Putra, Mochamad Teguh Subarkah, and Syarif Hidayat; declare that we have no conflict of interest for writing this paper to be submitted to ASTES Journal: Special Issue Paper (Special Issue on Multidisciplinary Innovation in Engg Science & Tech).

Acknowledgement

The authors are very grateful to all reviewers who have checked and criticized the papers that have been written to become a great paper in this journal. Thank you also to all ASTES committees, especially those who take care of the Special Issue Paper (Special Issue on Multidisciplinary Innovation in Engg Science & Tech) section who has invited us to submit our paper to the journal here.

References

[1] N.A. Maliky, M. Teguh Subarkah, S. Hidayat, "Added Mass and Drag Prediction using CFD Fluent Simulation for an Autonomous Barge Parameters," in ICEVT 2019 - Proceeding: 6th International Conference on Electric Vehicular Technology, 2019, doi:10.1109/ICEVT48285.2019.8993988.

[2] H. Qu, E.I. Sarda, I.R. Bertaska, K.D. Von Ellenrieder, "Wind feed-forward control of a USV," in MTS/IEEE OCEANS 2015 - Genova: Discovering Sustainable Ocean Energy for a New World, 2015, doi:10.1109/OCEANS-Genova.2015.7271438.

[3] V. Mishra, S. Vengadesan, S.K. Bhattacharyya, "Translational added mass of axisymmetric underwater vehicles with forward speed using computational fluid dynamics," Journal of Ship Research, 2011.

[4] L. Birk, Holtrop and Mennen's Method, 2019, doi:10.1002/9781119191575.ch50.

[5] D.T. Sen, T.C. Vinh, "Determination of Added Mass and Inertia Moment of Marine Ships Moving in 6 Degrees of Freedom," International Journal of Transportation Engineering and Technology, 2016.

[6] J. Anderson, "Computational Fluid Dynamics: The Basics with Applications. 1995," McGrawhill Inc, 1995.

[7] ANSYS, "ANSYS - Turbulence Modelling and the Law of the Wall: Tutorial," ANSYS User Manual, 2014.

[8] P. Beneš, R. Kollárik, "Preliminary Computational Fluid Dynamics (CFD) Simulation of EIIB Push Barge in Shallow Water," Scientific Proceedings Faculty of Mechanical Engineering STU in Bratislava, 2012, doi:10.2478/v10228-011-0012-1.

[9] T. Tabaczek, J. Kulczyk, M. Zawiślak, "Analysis of hull resistance of pushed barges in shallow water," Polish Maritime Research, 2007, doi:10.2478/v10012-007-0002-4.

[10] C. Lin, S. Percival, E.H. Gotimer, Viscous Drag Calculations for Ship Hull Geometry, Virginia, 1995.

[11] A. Fluent, "ANSYS Fluent 12.0 user's guide," Ansys Inc, 2009.

[12] A. Fluent, "Ansys Fluent Theory Guide," ANSYS Inc., USA, 2013.

[13] C. Chas, R. Ferreiro García, "Introduction to ship dynamic positioning systems," Journal of Maritime Research, 2008.

[14] J.D. Anderson, Fundamentals of Aerodynamics SI, 2011.

[15] D.L. Logan, A First Course in Finite Element Method, Biomedicine, 2004.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 937-943 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Design and Implementation of DFT Technique to Verify LBIST at RTL Level

Nagaraj Vannal[1,2,*], Saroja V Siddamal[3]

[1]*Department of ECE, BVBCET Hubballi, VTU Belgaum,590018, India*

[2]*Technical Lead (Level-2), ASIC Division, Einfochips (an ARROW company) Pvt. Ltd Bangalore, 560095, India*

[3]*School of ECE, KLE Tech University, Hubballi,580031, India*

ARTICLE INFO

Article history:
Received: 17 August, 2020
Accepted: 04 October, 2020
Online: 08 December, 2020

Keywords:
DFT (Design for Testability)
LBIST (Logic Built in Self-Test)
SoC (System on Chip)
ECO (Engineering Change Orders)
Netlist
RTL (Register-Transfer Level)
ATPG (Automatic Test Pattern Generator)

ABSTRACT

According to IEC 61805 and ISO 26262 standards requirement inclusion of LBIST (Logic Built in Self-Test) became mandatory to achieve safety critical application such as automotive field. In such systems, once device is switched ON LBIST (Logic Built in Self-Test) is activated and testing of digital logic is performed. After safety subsystem says that the LBIST passed, the SoC (System on Chip) moves into the functional mode otherwise, the SoC moves into a safe state. In this entire start-up sequence the LBIST interacts extensively with the safety sub-system of the SoC. Startup sequence remains un-verified at RTL (Register-Transfer Level) leading to painful ECOs (Engineering Change Orders) and post Silicon issues in some cases. LBIST verification can only run if scan chains are present in design which is not the case at RTL. The paper describes design of a Design-for-Testability (DFT) technique to enable LBIST based system verification with different test approaches at RTL which eliminates the possibility of ECOs by catching most of the issues at RTL level. Simulation results are demonstrating the feasibility of the approach with emphasizing the benefits obtained on significant computational modules.

1. Introduction

LBIST became alternative approach for ATPG (Automatic Test Pattern Generation) due to increased complexity and test cost requirement for functional blocks [1]-[3]. Electronic component general safety-critical applications [4] are referenced in IEC 61805 [5] and Automotive Safety Integrity Level (ASIL) for automotive domain is referenced in ISO 26262 standard [6].

Latent faults (LFs) detection for automotive systems using power on self-test (POST) based on logic built in self-test (LBIST) is executed during engine start -up to test safety – critical devices before staring any functional operations [7]. POST is having limited test application time in automotive systems [8]. Getting required LF coverage (>90%) within less test application time is primary requirement [9].

the pass/fail comparison of the unit [10], [11]. PRPG (Pseudo Random Pattern Generator) is used for generating input stimulus for the design, and MISR (Multiple Input Signature Register) to capture the compact the response. Test patterns which is supplied to internal scan chains are output of PRPG block and response is the input to the MISR compaction. Signature is the compacted response for that cycle of operation [12], [13]. The LBIST implementation is as shown in Figure 1.

LBIST is based on 'scan' methodology to test digital logic [14], this suffers from limitation – The design cycle will be delayed by the time the netlist is ready with Scan insertion, X-bounding [15] and Test-point insertion done [16] to perform LBIST verification. Performing LBIST patterns gate-level simulations early in the design cycle will take weeks to complete since execution is performed in series and are large number [17].

In this paper, we propose an innovative strategy for the in-field testing of automotive devices. Methodology to perform parallel verification of sub-modules of System on Chip (SoC) is described. Section 2 describes proposed design methodology to per-form Logic BIST verification at RTL level with dummy netlist. Section 3 describes implementation details such as scan chain insertion steps, dumpy netlist creation and direct mode entry. Simulation results with debugging analysis details are discussed in section 4 and in section 5 conclusion of the work is described.

*Corresponding Author: Nagaraj Vannal, Email: nagaraj.vannal@gmail.com

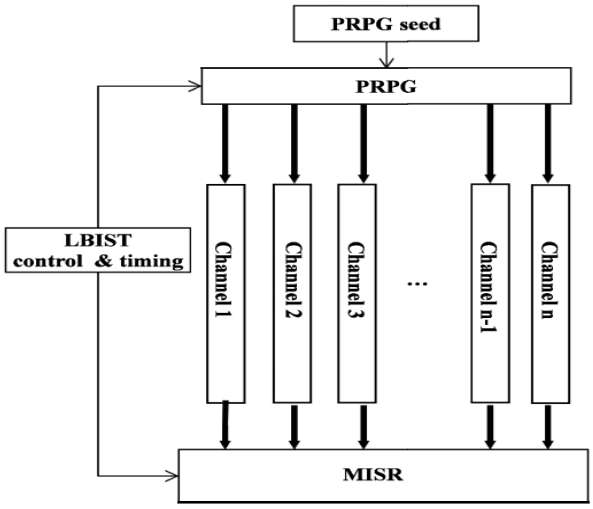


Figure 1: Standard LBIST Implementation

2. Proposed Design Methodology

Logic Built-In Self-Test (LBIST) is implemented by several LBIST controllers which operate independently on different partitions of the device as shown in Figure 2. This is needed to fulfil safety requirements about the independence and diversity of replicated IP, helps in physical flow and helps to avoid exceeding power limits. For applying traditional external scan patterns, LBIST controllers are put into trans-parent mode. Typically scan chains used for LBIST are shorter than scan chains used for compressed ATPG (Automatic Test Pattern Generation). Therefore, in LBIST transparency mode, the short LBIST chains are concatenated during to longer pro-duction scan chains which then are connected to the EDT (Embedded Deterministic Test) scan compression IP.

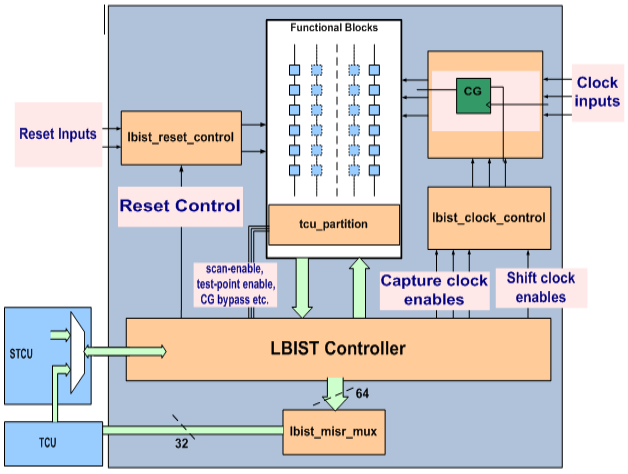


Figure 2: LBIST Architecture

LBIST controller IP itself will become part of the scan chains used for external scan patterns. Certain registers may not change value during LBIST operation but should be part of regular ATPG scan. These registers are put into one or more separate scan chains which are not shifted during LBIST. During LBIST their values are not influenced by shifting. These special scan chains are only used for ATPG. These registers will be identified during implementation. Each of the LBIST controllers is connected independently to the STCU (Self-Test Control Unit). The STCU can either start all LBIST controllers at the same time in parallel or in a sequence defined by the STCU. Only for debug reasons it is possible to access and operate each LBIST controller separately. This is done via JTAG (Joint Test Action Group) and a path through TCU (Test Control Unit) to the selected LBIST controller.

2.1. LBIST Partitioning

According to the safety concept eight LBIST partitions are implemented. They are numbered A0, A1, B0, B1, C0, C1, P0, P1. Each LBIST partition is contained in a separate LBIST island or lake as shown in Table 1. List of blocks where LBIST implementation is exempted:

- JTAG Controller
- TCU
- IO_top level ((Input-Output) / Pad ring
- PMU (Power Management Unit)/ PMU-Controller
- STCU
- PLL (Phase Locked Loop)
- XOSC (Oscillator)

Table 1: Characteristic data of LBIST partitions

	Characteristics				
Partition	**Size**	**LBIST chains**	**PRPG size**	**Clock domains**	**Chain overhead**
A0	77K	1600	46	System, tck	10%
A1	60K	1100	36	System, tck	10%
B0	24K	500	24	System, tck	10%
B1	60K	1100	36	System, tck	10%
C0	45K	1000	34	System, tck	10%
C1	60K	1100	36	System, tck	10%
P0	26K	600	26	System, tck	10%
P1	60K	1100	36	System, tck	10%

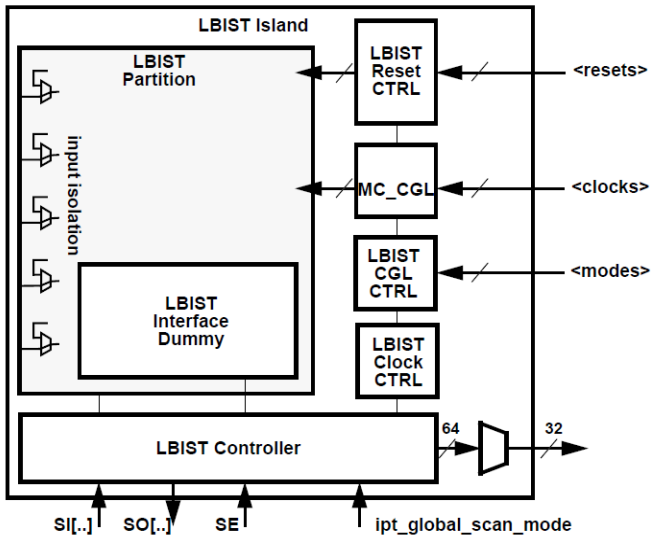
2.2. LBIST Integration

The design is split into several LBIST islands, where each LBIST island contains the actual LBIST partition, i.e. the logic to be covered by LBIST, a dedicated LBIST controller, dedicated clock and reset control blocks as shown in Figure 3. LBIST Controller:

Essential features of the LBIST controller IP used are:

- External clock control.
- Serial and parallel interface (to simplify communication, both functional and test/debug).
- Neg-edge lockup registers after PRPG and before MISR (to simplify LBIST scan chain timing)

With the introduction of lockup registers after PRPG and before MISR all LBIST scan chains must begin and end with a pos-edge flop. With this architecture all LBIST clock domains with less than half a cycle clock skew can be handled safely.



LBIST Controller Hierarchy: The controller needs to be generated with the following parameters:

set rtl hierarchy -Clk_mux Out -external_clk_enables
set lbist Controller -programmable_shift_rate ON
set shift_rate Divisor 8

This generates the LBIST controller IP in a hierarchy. The LBIST clock controller is entirely based on clock enable rather than the clocks generated by the LBIST IP in order not to generate a new clock path. Figure 4 gives an overview of the generated Controller IP and its hierarchy.

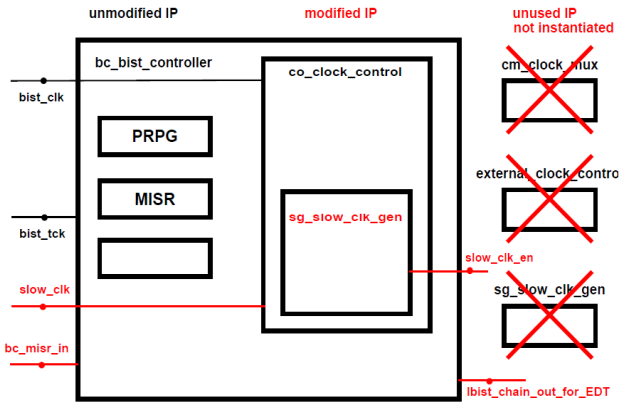


Figure 4: LBIST Controller IP Hierarchy

2.3. LBIST Controller Clocks

The LBIST controller operates on three different clocks, the LBIST engine clocks, bist_clk, and slow_clk, as well as the STCU communication clock bist_tck. The LBIST engine clocks bist_clk and slow_clk must fulfill the following requirements:

- The clock source for 'bist_clk' and 'slow_clk' must be identical.
- The clock source for 'bist_clk' and 'slow_clk' must be the fastest clock used inside the LBIST partition.
- The clock source for 'bist_clk' and 'slow_clk' must be the clock with the highest clock insertion delay used inside the LBIST partition.

2.4. LBIST Interface Dummy

The LBIST interface dummy propagates a standard LBIST interface into the LBIST partition, to ensure, that all relevant signals are present and preserved through synthesis and scan insertion. The LBIST interface dummy combines the following interfaces / features:

- LBIST controller IP interface
- LBIST clock interface
- LBIST clock gating control ('ipt_cg_bypass' and 'ipt_se_gatedclk')
- Test points that can be connected by the physical team to improve coverage if required
- Production Scan: Launch-Off-Shift (LOS) per clock domain and blocking of LBIST control signals for at-speed scan.

The LBIST interface dummy is protected during synthesis because connections to the module are created during scan chain stitching and automatic clock gate insertion only.

2.5. LBIST Interface Dummy

For direct execution of LBIST, the optional parallel interface of Mentor's LBIST controller is used. It provides the following interface signals:

- bist_run_input- triggers LBIST run
- bist_done - flags a finished LBIST run
- misr_value - vector output providing parallel read access to MISR value

Debug and diagnosis require access to LBIST control registers. It is done through LBIST's serial interface. Writing and reading control registers happens in a JTAG-like serial mode. The LBIST serial debug interface is used by the STCU to con-figure LBIST run parameter, such as pattern count. Full debug and diagnosis are performed through the TCU only. The interface is based on a JTAG-like protocol with shift- and update-DR states. Both TCU and STCU share the interface to the LBIST controller IP, where the muxing between TCU and STCU is performed inside the STCU controlled by 'tcu_lbist_direct_ctrl'. The actual TCU access is performed though the TCU external TDR mechanism based on the register selection through the LBIST TDR (LTDR). Expected LBIST signatures are kept in the Non-Volatile Memory (NVM). Through STCU programming the required signature is fetched from the NVM for comparison.

2.6. LBIST Clock and Reset Control

LBIST Clock Control: All LBIST clocks are controlled through the Magic Carpet Clock Gating Logic (MC_CGL) associated with each LBIST partition. The clock control by the MC_CGL is purely based on clock enables rather than the clocks generated by the LBIST IP. Figure 3 shows the modules relevant for LBIST clock control for one LBIST partition 'C0'.

The LBIST controller, as well as the blocks lbist_clock_control, lbist_cgl_control, and cgl_lake exist local to all eight LBIST partitions. The LBIST controller with its modified co_clock_control block is connected to the main lbist_clock_control module, that combines the shift and capture enable signals. The lbist_cgl_control block switches between the clock enables generated by the TCU and by the LBIST controllers depending on the mode setting. The clock gate inside the MC_CGL will generate the slow bist clock based on the enable provided by the LBIST IP. The LBIST clock control will provide punch out clocks, not 50% duty cycle clocks.

Slow Clock Generation: The LBIST slow clock ('slow_clock') is a punch out clock divided down from bist_clk (== fast_clk) using a programmable clock divider inside the 'sg_slow_clock_gen'. It is used to control several key LBIST functions, such as PRPG, MISR, pattern counter etc. The same slow clock divider is used to program the shift clock frequency. The default divider ratio of the 'sg_slow_clk_gen' module is set to 8 but can be reprogrammed via the LBIST control registers.

LBIST Reset Control: To control all resets within the LBIST partition, all resets must be routed through the LBIST reset control block. The block allows reset control for LBIST, while the system resets are not gated. This is to make sure, that a system reset is not blocked during device startup.

LBIST reset requirements are:

- All asynchronous resets of an LBIST partition must be routed through the LBIST reset control block.
- All functional resets must be in off state during LBIST execution.
- No LBIST partition output to MC_RGM/MC_RSL must have an influence on self-test / LBIST execution.

The original reset control logic provided by the LBIST controller IP is an active-high reset (clock) signal multiplexed into the functional reset path. With the non-blocking reset control, where the functional reset is not blocked by LBIST a new reset control is introduced replacing the original one. To pre-vent reset glitches from propagating into the LBIST partition in scan mode, the LBIST reset control must be bypassed by a scan mode reset. The non- blocking reset control dictates, that all functional resets are off during LBIST execution, otherwise the LBIST MISR becomes invalid.

2.7. Making the LBIST Partitions LBIST Ready

The interface of the LBIST partitions need special attention to ensure that there is no X poisoning into the LBIST partition. All critical signals causing problems in the standard X-bounding flow will be X-bound in RTL. Mentor Testkompress fault simulation is used to identify missing X-bounds. The decision on whether an input is to be X-bounded is taken based on the partition input signal categories shown in Table 2.

The X-bounds inserted in RTL will instantiate a MUX controlled by lbist_en from the LBIST controller IP. Whenever possible the partition outputs are used as inputs to the MUX. This increases coverage as it provides toggling and not constant inputs to the MUXes, while at the same time adding observability to the partition outputs.

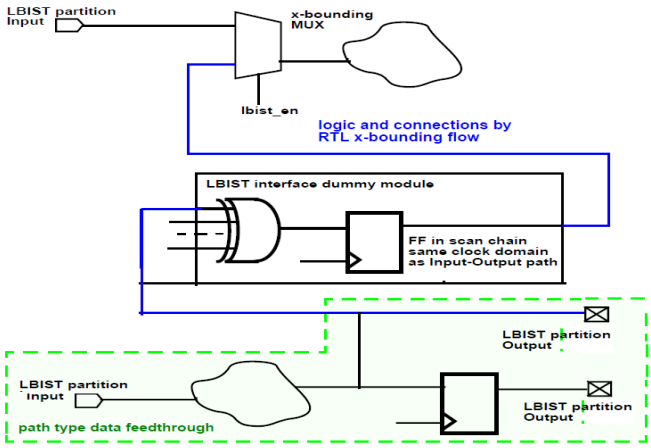


Figure 5: X-bounding to avoid multi-cycle path generation

Figure 5 shows a concept used to avoid multi-cycle path generation on known critical input – output combination. The LBIST interface dummy module will include several XOR gate FF combinations as shown below. The RTL X-bounding flow uses and connects these structures. In a later stage the automatic RTL x-bounding flow might also include instantiation of these structures instead of using predefined ones.

Table 2: LBIST partition input categories

Category	LBIST Control	Bound
Reset input	Via LBIST reset control	No
Clock input	Via LBIST clock control	No
Clock gate control	Via LBIST interface dummy	No
Common Digital Input		Yes
Selected Tied Digital Input		Yes
Unconnected Digital Input	None required	No
Digital Input from TCU – no influence on LBIST		yes
Digital Input from TCU – influence on LBIST	Constant from TCU	No
Bidirectional	NA	No
Analog/Special Nets	All hard IP in LBIST mode	No
LBIST Controls	From LBIST control IP	No

3. Implementation Details

LBIST scan chain insertion and concatenation to top level LBIST chains including lockup-FF insertion followed these basic steps in the DFT and physical flow.

1. DFT flow - Generated EDT controllers everything pos-edge based w/o lockup-FFs
2. DFT flow - Generated LBIST Controller and connected internal LBIST chains at the BIST controller to form ATPG scan chains. Modified LBIST controller to provide ATPG chain outputs with LBIST controller interface. Connecting ATPG chain inputs and outputs of LBIST controller to EDT controllers using IPXACT flow. (Assigned fixed value to EDT controller pins for remaining chains outside LBIST to protect them in synthesis)
3. Physical flow -One step scan insertion: Building LBIST chains, connecting them to the LBIST controller. Building ATPG chains for logic outside LBIST, connect-ed them to the

EDT controller using the pins with fixed value assigned by IPXACT.

4. Physical flow - Post processing step to insert lockup FFs, including required lockup-FFs in concatenated LBIST chains.
5. Checked for missing lockup-FFs on

a. change of clock name or clock polarity in scan cell report - DFT flow

b. clock hierarchy change - Physical flow

Figure 6 depicts concatenation of four LBIST scan chains to form one production chain with input and output available at the LBIST controller boundary for hook up to the EDT controller. Also shows where lockup FFs are placed by the physical flow.

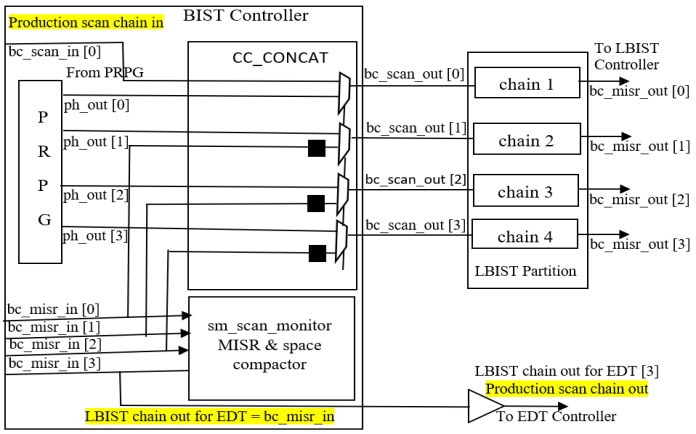


Figure 6: LBIST chain concatenation and lockup latch insertion

3.1. LBIST Dummy Netlist Integration at RTL

Integrating a dummy netlist in the design at RTL level with the same number of scan chains, scan clocks and scan resets is done. This dummy netlist should only be visible to the verification environment, in the LBIST testcases expect the MISR calculated for the dummy netlist.

Flow: Generate a dummy netlist (with scan chains) → Generate the MISR value for the same dummy netlist →Integrate this dummy netlist into the actual design RTL → run the testcases for verifying the complete start-up sequence. Details of each step are described below:

Step1: Generate a dummy netlist using "skeleton design creation".

create_skeleton_design -o lbist_for_verif -i skeleton_design.txt

Step2: Generate MISR for dummy netlist by doing fault simulation using LBIST Architect

*lbistarchitect -fault_simulation *
*lbist_for_verif.v -verilog *
*-lib skeleton.fslib -nogui *
-dofile lbist_for_verif.dofile

Step3: Integrate the dummy netlist into the lbist partition using "ifdef"

Step4: Perform RTL simulations for LBIST based start-up sequence using the com-pilation argument:

+define+LBIST_DUMMY_SIM

3.2. LBIST Direct Control Mode

For LBIST debug purposes there is a 'LBIST direct control' mode, where the different LBIST partitions can be run stand alone or in parallel. Which LBIST partitions are run, as well as LBIST parameters like pattern count, MISR start value etc., are programmed via JTAG. The LBIST Direct Test Mode Entry uses the LBIST Test Data Register (LTDR) to select the LBIST controllers, start the LBIST run. The following sequence shown in Table 3 is used in the Teskompress setup as well is in the Verilog stimuli for LBIST direct mode verification and pattern generation.

Table 3: LBIST direct mode entry

Step	Description
1	Common steps to TCU based mode entry
2	Load and update the Generic TCR
3	Load and update the RGMTDR to disable gating reset phase 3.
4	Load and update the CLK TDR to program the PLL0 clock source for LBIST execution
5	Set LBIST direct control: Load and update LTDR LBIST direct mode: tcu_lbist_testmode=1; tcu_lbist_direct_control=1;
6	Turn TEST mode on to be able to program LBIST Controllers
7	Program LBIST controllers Select the LBIST controllers: Load and update the LTDR: ipt_lbist_sel[4:0]=<controller id>Program controller settings like pattern start/end counter using the steps in Table V.
8	Deassert RESET and apply EXTAL clock pulses to get rgm state machine to stabilize.
9	Start LBIST run on selected controllers: Load and update LTDRipt_lbist_sel[3:0]=<selected controllers> and ipt_lbist_run=1.
10	2 cycles for LBIST finder PRPG and MISR recognition
11	Deassert RESET_B

The following Table 4 describes the sequence of Writing to / Reading from an LBIST controller register in LBIST direct test mode and Table 5 describes the sequence to read out the MISR values after completion of an LBIST run in LBIST direct test mode.

Table 4: Writing to / reading from an LBIST controller register

Step	Description
1	Write register address to LBIST controller dselect register: a. Load and update the LTDR to select dselect register: ipt_lbist_reg_sel=1 b. Send JTAG instruction 6'd5 to access external LBIST controller register. Write value to LBIST controller dselect register. E.g. '0110' to map pattern count end value into data register.
2	Write value to / read value from an LBIST controller data register a. Load and update the LTDR to select data register: ipt_lbist_reg_sel=0. b. Send JTAG instruction 6'd5 to access external LBIST controller register. Write value to / read value from LBIST controller data register

Table 5: Reading out MISR values

Step	Description
1	Read the LTDR to observe LBIST done flags lbist_done [3:0]
2	Turn off EXTAL before switching from bist_clk to TCK
3	Load and update LTDR: ipt_lbist_run=0. This switches from bist_clk to TCK
4	Read MISR registers a. Select the LBIST controller: Load and update the LTDR: ipt_lbist_sel[3:0] b. Read the MISR value using the steps in Table V.

4. LBIST verification results at RTL

Verification of all different partitions are carried out with four different testcases listed below running parallelly with respective MISRs generated during LBIST Fault simulation using LBIST Architect tool [18].

1. Testcase with PLL clock
2. Testcase with EXTAL clock
3. Scan chain mask decoder
4. MISR exposed on PADs

4.1. Testcase with PLL clock frequency

LBIST controller logic/integration reset and clocks of the respective partitions, mode entries are verified at lbist_fast_clk = 320MHz & lbist_slow_clk = 80 MHz frequencies. Figure 7 shows one of the issues identified for a partition.

- MISR mismatching is observed after pattern count value 237, expected MISR value is 415B3579F4B1C0DA, simulation result MISR value is DE2F87E1C111F4D1.
- Analysis: Capture window for pattern count value after 237 is as below but from waveform we can see along with below 4 clk_domain clk_domain_58 is also pulsed which is causing the MISR mismatch.

clk_domain_1 → pulse '10000000'
clk_domain_18 → pulse '00100000'
clk_domain_39 → pulse '00001000'
clk_domain_30 → pulse '00000010'

Figure 7: MISR mismatch issue caught during RTL verification

4.2. Testcase with EXTAL clock Frequency

LBIST controller logic/Integration reset and clocks of the respective partitions, mode entries are verified at EXTAL = 100MHz frequencies. As shown in Figure 8 MISR is observed after pattern count value 260, expected MISR value is B8D5_F506_668B_F517, simulation result MISR value is B8D5_F506_668B_F517.

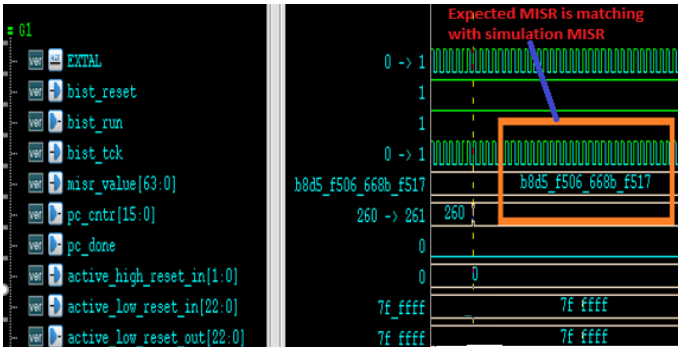


Figure 8: LBIST Verification with EXTAL clock frequency 100MHz

4.3. Scan Chain Mask Decoder Testcase

LBIST controller logic/Integration, reset and clocks of the respective partitions, verified at lbist_fast_clk = 320MHz & lbist_slow_clk = 80 MHz frequencies. As shown in the Figure 8 chain_0 is masked for every LBIST partition and pattern lbist_si[980:0] toggles for mask decoder lbist_si[980:1] and lbist_si[0] will be static 1.

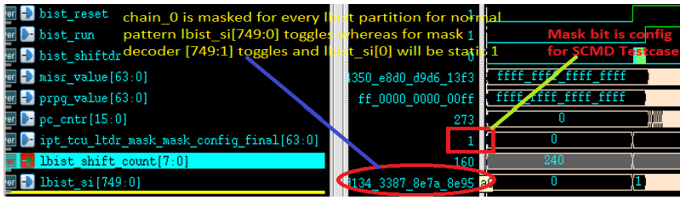


Figure 9: Scan chain mask decoder testcase result

4.4. MISR Expose at PAD Testcase

LBIST controller logic/Integration, reset and clocks of the respective partitions, verified at lbist_fast_clk = 320MHz & lbist_slow_clk = 80 MHz frequencies along with MISR value is exposed at PAD during end of the pattern count using lbist_misr_word_sel bit to expose 1st 32 -bit and later 32 -bit values respectively. As show in Figure 10 MISR value DCCF_7DA6_225C_9279 at program counter value 272 is exposed on PAD pins using lbist_misr_word_sel (0/1).

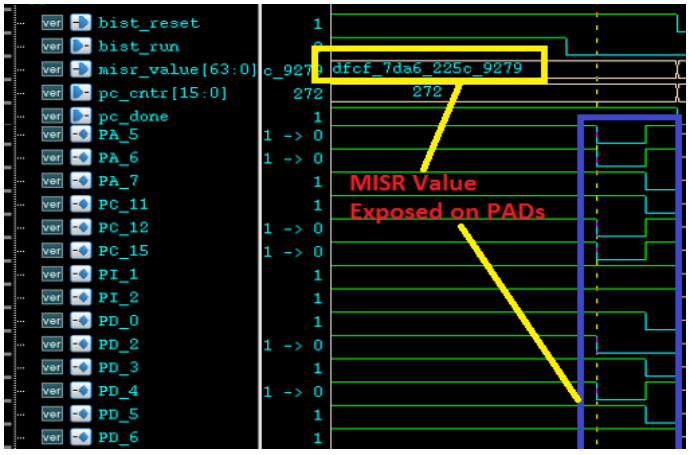


Figure 10: MISR exposed at PAD testcase result

5. Conclusion

The proposed methodology for verifying LBIST implementations at RTL in SoCs is not a replacement for gate

level verification of LBIST implementations, with this approach a significant effort and design cycle time can be reduced by moving the verification process early into the SoC design cycle. The results are proven with measurable improvements on real design where the flow/methodology was implemented and helped in identifying the bugs & fixing them before logic freeze, thereby avoiding ECOs.

References

[1] Andal Jayalakshmi and Tan Ewe Cheong, "A methodology for LBIST logic diagnosis in high volume manufacturing," 2012 4th Asia Symposium on Quality Electronic Design (ASQED), Penang, 249-253, 2012, doi: 10.1109/ACQED.2012.6320510.

[2] K. Butler, "ATPG versus Logic BIST -Now and in the Future," in 2013 IEEE International Test Conference (ITC), Baltimore, Maryland, 1181, 2001. doi: 10.1109/ITC.2001.10048

[3] P. Maxwell, I. Hartanto and L. Bentz, "Comparing functional and structural tests," Proceedings International Test Conference 2000 (IEEE Cat. No.00CH37159), Atlantic City, NJ, USA, 400-407, 2000, doi: 10.1109/TEST.2000.894231.

[4] S. Carbonara, P. Bernardi and M. Restifo, "A Hybrid In-Field Self-Test Technique for SoCs," 2019 14th International Conference on Design & Technology of Integrated Systems In Nanoscale Era (DTIS), Mykonos, Greece, 1-6, 2019, doi: 10.1109/DTIS.2019.8735075..

[5] International Standard - IEC61508 - Functional safety of electrical/electronic/programmable electronic safety-related systems, International Electrotechnical Commission, 2010.

[6] ISO Standard 26262, Part 5, "Road vehicles— functional safety," 24 May 2016.

[7] Y. Sato et al., "A circuit failure prediction mechanism (DART) for high field reliability," 2009 IEEE 8th International Conference on ASIC, Changsha, Hunan, 581-584, 2009, doi: 10.1109/ASICON.2009.5351352.

[8] H. Iwata and J. Matsushima, "Multi-configuration Scan Structure for Various Purposes," 2016 IEEE 25th Asian Test Symposium (ATS), Hiroshima, 131-131, 2016, doi: 10.1109/ATS.2016.32.

[9] S. Wang, Y. Higami, H. Takahashi, H. Iwata and J. Matsushima, "Automotive Functional Safety Assurance by POST with Sequential Observation," in IEEE Design & Test, **35**(3), 39-45, June 2018, doi: 10.1109/MDAT.2018.2799801.

[10] Wu-Tung Cheng, Manish Sharma, T. Rinderknecht, Liyang Lai and C. Hill, "Signature based diagnosis for logic BIST," 2007 IEEE International Test Conference, Santa Clara, CA, 1-9, 2007, doi: 10.1109/TEST.2007.4437703.

[11] M. Abramovici, M. A. Breuer, and A. D. Friedman, "Digital System Testing and Testable Design", AT&T Bell Laboratories and W. H. Freeman and Company, 1990.

[12] H. D. Schnurmann, E. Lindbloom and R. G. Carpenter, "The Weighted Random Test-Pattern Generator," in IEEE Transactions on Computers, C-**24**(7), 695-700, July 1975, doi: 10.1109/T-C.1975.224290.

[13] G. Kiefer and H. -. Wunderlich, "Using BIST control for pattern generation," Proceedings International Test Conference 1997, Washington, DC, USA, 347-355, 1997, doi: 10.1109/TEST.1997.639636.

[14] Y. Maeda, J. Matsushima and R. Press, "Automotive IC On-line Test Techniques and the Application of Deterministic ATPG-Based Runtime Test," 2017 IEEE 26th Asian Test Symposium (ATS), Taipei, 237-241, 2017, doi: 10.1109/ATS.2017.52.

[15] M. Sadi, G. K. Contreras, J. Chen, L. Winemberg and M. Tehranipoor, "Design of Reliable SoCs With BIST Hardware and Machine Learning," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, **25**(11), 3237-3250, Nov. 2017, doi: 10.1109/TVLSI.2017.2734685.

[16] P. Wohl, J. A. Waicukauski, G. A. Maston and J. E. Colburn, "XLBIST: X-Tolerant Logic BIST," 2018 IEEE International Test Conference (ITC), Phoenix, AZ, USA, 1-9, 2018, doi: 10.1109/TEST.2018.8624738.

[17] E. Moghaddam, N. Mukherjee, J. Rajski, J. Solecki, J. Tyszer and J. Zawada, "Logic BIST With Capture-Per-Clock Hybrid Test Points," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, **38**(6), 1028-1041, June 2019, doi: 10.1109/TCAD.2018.2834441.

[18] LBIST Architect process guide Manual Mentor Graphics, Software Version 2018.1.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 944-950 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Empathy in Nursing Students that do the Non-Medical Internship in Three Universities in Lima, 2019

Liseth Acuña-Medina[1], Yumira Arias-Quispe[1], Yackeline Espeza-Veláquez[1], Brian Meneses-Claudio[2], Hernan Matta-Solis[1], Eduardo Matta-Solis[3]

[1]*Faculty of Health Sciences, Universidad de Ciencias y Humanidades, Lima, 15314, Perú*

[2]*Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, Lima, 15314, Perú*

[3]*Research and Intellectual Creativity Direction, Universidad María Auxiliadora, Lima, 15408, Perú*

ARTICLE INFO

Article history:
Received: 03 August, 2020
Accepted: 19 November, 2020
Online: 08 December, 2020

Keywords:
Empathy
Students
Nursing internship
Non-medical internship

ABSTRACT

Empathy is an attribute of importance between the nursing professional and the patient, where it influences the recovery of their health. Several studies carried out have shown worrying results where they reflect a great decrease in empathy during the formative academic years. The objective of the study is to determine empathy in nursing interns from three university institutions in Lima, 2019. The study is a quantitative approach, with a non-experimental, prospective, descriptive and cross-sectional design. The application of the Jefferson Medical Empathic Scale instrument for students in Spanish version S is proposed, where the data was entered into a matrix in the Microsoft Excel 2013 program, then the IBM SPSS Statistics version 24.0 is subsequently analyzed obtaining the percentage of empathy of the internal, also take actions to improve these variables. 216 nursing interns from three universities in Lima participated. 51.4% are from Universidad Norbert Wiener, 89.4% (n = 193) were women, 56.4% (n = 123) are from the ninth cycle, 42.4% (n = 91) they study and work, 58 nursing interns from Universidad Norbert Wiener representing 52.3% have a medium level, 34 nursing interns from the Universidad de Ciencias y Humanidades, representing 30.6% have a medium level and finally 19 nursing interns from the Universidad Privada del Norte representing 17.1% have a medium level. Universidad Norbert Wiener obtained a higher score prevailing the medium level.

1. Introduction

The definition of empathy over time has been modified and expanded, coming to recognize it more than a static trait, also a state or ability that flows, being susceptible to change according to personal experiences and cultural interventions [1]. Empathy in the nurse-patient relationship should be used on a day-to-day basis, since it may affect their work of caring for patients, the decrease of it may lead to the loss of ability that allows better communication in the opinion of the patient since this is considered a relevant factor in the recovery process [2].

The comparison of the three universities is very important since it is going to differ in the type of teaching that each university offers to nursing interns, in turn the ability of nursing interns to function in the clinical field, where His training will stand out depending on each intern willing to improve and be a successful professional.

These circumstances may be due to the educational system because it must be oriented to the service of society and the individual, which should be based on the needs and demands that are important for development and competent from the professional point of view [3].

To carry out a relationship between nurse and patient, health personnel provide empathetic treatment to the patient to not worse the situation of the patient's disease, thereby also giving the staff the development of attitudes, capacities and skills, giving a care of quality, that's why, different causes of the deficit are visualized in its application.

The health personnel with the longest service time and high hierarchy level are those that present the greatest decrease in the level of empathy [4].

*Corresponding Author: Brian Meneses-Claudio, Mr., +51 1 950159924 & bmeneses@uch.edu.pe

In a study on health students, "Evaluation of empathy in human medicine students from the private university of Peru" [5], they found that those students who think that their specialty is linked to Patient Care presented greater empathy with 66.9%, the same as those who reported having a professional model in patient contact with 72%.

In a research work [6] studies carried out in nursing and medical students, indicate that students have found variations in empathy in terms of specialty and clinical competence, being worrying because some of them reflect a decrease in empathy over the years of training, the score is from 0.0 to 140.0; the women with 118.7 obtained a higher level of empathy score than the men with 110.4 points, differences that were statistically significant.

In the research work "Concepts on empathy: development of an instrument for formative evaluation based on a teaching experience with pediatric residents of the Universidad de Antioquía" [7], indicates that a deficit in student training on the basis and practice of empathy, because the participants had misconceptions about empathy, where they adopted universal criteria for all their patients, believing that they had to make the patient's wishes come true, make the patient or family feel that are important in the decisions of health personnel, among others.

Currently, professionals who can work normally in the socio-labor context that allows them to function interpersonally are required, the variable empathy is essential in the deployment of a professional in different medicine and social fields that its performs, its importance is appreciated in the development of professional skills [8].

The objective of the research work is to determine empathy in nursing interns from three University Institutions in Lima, 2019 according to their three dimensions of compassionate care, perspective taking and putting themselves in the other's place. The implementation of empathy on a daily basis will help, also the mental state of the patient and the health personnel, to recognize and identify mental problems that could be going unnoticed, being a national research priority 2015-2021 established by the Instituto Nacional de Salud [9].

The technique to be used is the survey that will be through the questionnaire and data collection instrument the Jefferson Scale of Empathy (JSE) for students in Spanish version S, created by the Jefferson group, led by Mohammadreza Hojat, consisting of 20 evaluated questions on a 7-point Likert-type scale, the score ranges from a minimum of 20 to a maximum of 140 points, there are 10 items written in a positive sense and 10 in a negative sense, whose items explore the dimensions of perspective taking, compassionate care and ability to put yourself in the other's place. Permission has been offered to use the Jefferson Empathy Scale in 85 countries. The JSE has been translated into 56 languages[10]. After this, the data will be entered into a matrix in Microsoft Excel 2010, and then exported to the IBM SPSS Statistics Base 22 statistical package, in which it will be analyzed.

The following research work is structured as follows: In section II, the development of the used data processing will be presented. In section III, the results of the tables and figures of sociodemographic agreement, institution and dimensions will be shown. In section IV, it will present the discussions of the research work and finally in section V, the conclusions, as well as the future work that is intended to be wanted with the research work.

2. Methodology

2.1. Research type and design

The present research is of a quantitative approach because the variables were measured, statistical data was used generalizing the hypothesis, the methodological design is a non-experimental, prospective research because according to the planning of data collection, they are primary, cross-sectional because the measure is non-experimental, it was for a single occasion, descriptive because it consists of a single variable.

2.2. Population and sample

The study population is made up of 216 nursing interns from three private universities in Lima, enrolled in 9th and 10th semesters, who carry out their pre-professional practices and are in contact with the patient, during the period of 2019, according to the Population type are non-probabilistic type, selected according to the inclusion and exclusion criteria.

Inclusion criteria

University students of both sexes who are studying nursing at the time of data collection, nursing interns enrolled during the academic semester 2019 who are taking the 9th and 10th semesters who voluntarily participate in the interview or fill out the data collection questionnaire, nursing students who agree to sign the informed consent.

Exclusion criteria

Nursing students who are from the 1st semester to the 8th semester and nursing interns who have not signed the informed consent to carry out the study.

2.3. Technique and instruments

The technique used was the survey, in the form of the Jefferson Scale of Empathy in Spanish for students, containing general student data such as: sex, age and year of training that will be used to obtain information of the students who participate in the survey, specially designed to measure empathy in the context of patient care, whose items consist of 3 dimensions of perspective taking, compassionate care and ability to put yourself in the patient's place. It consists of 20 items which are evaluated on a 7-point Likert-type scale ranging from: Strongly Agree (7 points) to Strongly Disagree (1 point). The score is from a minimum of 20 points to a maximum of 140 points, the highest scores correspond to a high level of empathy. There are 10 items written in a positive sense and 10 in a negative sense (inverted score) [10].

"Jefferson Scale of Empathy" has been validated in many countries around the world, including Latin American countries and previously applied in Peru, it has construct validity; convergent, parts that comprise are related to each other; and discriminating, its parts are not related to topics different to empathy. In addition, it has the Chi X^2 that is 0.5 in its various versions and translations, which is a reliable measurement scale and each of the questions are correlated with each other and with the entire instrument [11].

2.4. Data collection procedure

- An authorization was presented to the Nursing faculty for the execution of the project in order to obtain the respective permission.
- Subsequently, the nursing faculty provided facilities for the execution of this research, also will provide a cover letter.
- Once obtained, it was presented to teachers and delegates of the ninth and tenth semesters informing the objectives of the project and the informed consent, in order to obtain their acceptance, support and collaboration in the application of the instrument.
- The evaluators met teachers and delegates of ninth and tenth semester to discuss the date and time of application of the instrument.

2.5. Application of data collection instruments

- The evaluators attended the classrooms of each semester, ninth and tenth, on coordinated dates and times.
- The students were informed of the importance of their participation, a brief explanation about the instrument.
- The informed consent form was distributed to be signed for the student and then the questionnaire.
- The questionnaire development lasted approximately 10 to 15 minutes.
- Then the instrument was collected, verifying that all questions are answered.
- In closing, the students were thanked for their participation and collaboration in the research.

As shown in Figure 1, nursing interns were presented and explained how to carry out the survey. The protection aspects of the research participants and bioethical principles were taken into account, likewise informed consent was applied after clear information given to the participants, the application of the instrument is important, it will help especially in the mental state of the patient and the health personnel themselves, to recognize and identify mental problems that could be going unnoticed, being a national priority, which asks them to collaborate in providing with certain data that will allow to get the study objective .

Figure 1: Teacher permission and approval to apply the survey to students.

After obtaining the approval of the teacher of the classroom, the survey was applied, where advice was given before and during the application because questions were raised by the students, most of them were just to clarify questions, as shown in Figure 2.

Figure 2: Advice before and during the survey application

2.6. Data analysis and processing techniques

The data collection process was carried out during May 2020, the data was initially entered into a data matrix designed in the Microsoft Excel 2010 program, then exported to the IBM SPSS Statistics version 22.0 program. There, they were analyzed, obtaining tables and graphs that were later described and interpreted in the results and discussion sections, respectively.

For the statistical analysis of each main variable involved in the study, descriptive statistical tests such as absolute frequencies, relative frequencies, and measures of central tendency were applied, in addition to some correlational tests to verify the study hypothesis.

3. Results

A summary table of the surveys carried out following the guidelines corresponding to the research work will be shown, in the study participants were 216 nursing interns of 9th and 10th semester from the three University Institutions, in the following table, it will present results of the sociodemographic data of the study participants.

Table 1: Sociodemographic data of empathy in nursing interns from three University Institutions in Lima, 2019 (N = 216)

Participant Information	Total	
	N	%
Total	**216**	**100**
University		
Universidad de Ciencias y Humanidades	64	29,6
Universidad Norbert Wiener	111	51,4
Universidad Privada del Norte	41	19,0
Gender		
Female	193	89,4
Male	23	10,6
Study semester		
9th Semester	123	56,9
10th Semester	93	43,1
Occupation Status		
Study	61	28,2
Study and eventually work	64	29,6
Study and work	91	42,1

Regarding the university, 111 nursing interns representing 51.4% are from Universidad Norbert Wiener, 64 nursing interns representing 29.6% are from the Universidad de Ciencias y Humanidades, and finally 41 nursing interns representing 19% are from the Universidad Privada del Norte.

Regarding gender, the female sex predominated with 193 students representing 89.4%, followed by the male with 23 students representing 10.6%.

Regarding the study semester of nursing interns, the ninth semester is 123 representing 56.9%, and the tenth semester, 93, representing 43.1%.

Regarding the occupation status of nursing students, 91 with (42.1%) cases study and work, followed by 64 with (29.6%) cases study and eventually work and 61 with (28.2%) cases only study.

Table 2: Empathy in nursing interns from three University Institutions in Lima, 2019 (N = 216)

			Universidad de Ciencias y Humanidades	Universidad Norbert Wiener	Universidad Privada del Norte
Empathy	**Low**	N	0	0	1
		%	0,0	0,0	100,0
	Medium	N	34	58	19
		%	30,6	52,3	17,1
	High	N	30	53	21
		%	28,8	51,0	21,2

In Table 2, it can see Empathy in nursing interns from three University Institutions in Lima, 58 nursing interns from Universidad Norbert Wiener representing 52.3% have a medium level, 34 nursing interns from the Universidad de Ciencias y Humanidades, representing 30.6% have a medium level and finally 19 nursing interns from the Universidad Privada del Norte representing 17.1% have a medium level. Universidad Norbert Wiener obtained a higher score prevailing the medium level.

Table 3: Empathy in relation to gender of nursing interns from three universities in Lima, 2019 (N = 216)

		Empathy			
		Low	Medium	High	Total
Student gender	Female Count	0	98	95	193
	% within Student's gender	0,0%	50,8%	49,2%	100,0%
	Male Count	1	13	9	23
	% within Student's gender	4,3%	56,5%	39,1%	100,0%
Total	Count	1	111	104	216
	% within Student's gender	0,5%	51,4%	48,1%	100,0%

Chi-square tests

	Value	df	Asymptotic significance (bilateral)
Pearson's Chi-square	8,958[a]	2	,011
Likelihood ratio	5,069	2	,079
Linear by linear association	1,649	1	,199
N of valid cases	216		

a. 2 cells (33.3%) have expected a count less than 5. The minimum expected count is ,11.

In Table 3, empathy is related according to sex, where women present a higher level of empathy (49.2%) than in men (39.1%), in which it was determined with the chi-square test of Pearson (X^2). The level of significance of the test obtained a value of 0.11 (p> 0.05) (X^2 = 8.958; d.f = 2). Therefore, an association hypothesis is not rejected, for which there are statistical data that verify the relationship between empathy and the sex of nursing interns from three universities in Lima.

In Table 4, we can observe the relationship between the dimension of perspective taking with the Universities of the nursing interns, in which it was determined with Pearson's chi-square test (X^2). The level of significance of the test obtained a value of 0.95 (p> 0.05) (X^2 = 3.512; d.f = 4). Therefore, an association hypothesis is not rejected, for which there are statistical data that verify the relationship between the perspective taking dimension and the universities of the nursing interns in the research. By which, we can interpret that the interns of the Universidad Nolbert Wiener have a high perspective taking (79.3%).

Table 4: Empathy in its perspective-taking dimension in nursing interns from three universities in Lima, 2019 (N = 216)

			Institution			
			Universidad de Ciencias y Humanidades	Universidad Nolbert Wiener	Universidad Privada del Norte	Total
Perspective taking	Low	Count	1	3	1	5
		% within Institution	1,5%	2,7%	2,5%	2,3%
	Medium	Count	19	20	8	47
		% within Institution	29,7%	18,0%	19,5%	21,8%
	High	Count	44	88	32	164
		% within Institution	68,8%	79,3%	78,0%	75,9%
Total		Count	64	111	41	216
		% within Institution	100,0%	100,0%	100,0%	100,0%

Chi-square tests

	Value	df	Asymptotic significance (bilateral)
Pearson's Chi-square	3,512[a]	4	,476
Likelihood ratio	3,398	4	,494
Linear by linear association	,981	1	,322
N of valid cases	216		

a. 3 cells (33,3%) have expected a count less than 5. The minimum expected count is ,95.

Table 5: Empathy in its compassionate care dimension in nursing interns from three universities in Lima, 2019 (N = 216)

			Institution			
			Universidad de Ciencias y Humanidades	Universidad Nolbert Wiener	Universidad Privada del Norte	Total
Beware with compassion	Low	Count	8	16	6	30
		% within Institution	12,5%	14,5%	14,6%	13,9%
	Medium	Count	26	54	16	96
		% within Institution	40,6%	48,6%	39,0%	44,4%
	High	Count	30	41	19	90
		% within Institution	46,9%	36,9%	46,4%	41,7%
Total		Count	64	111	41	216
		% within Institution	100,0%	100,0%	100,0%	100,0%

Chi-square tests

	Value	df	Asymptotic significance (bilateral)
Pearson's Chi-square	2,278[a]	4	,685
Likelihood ratio	2,285	4	,684
Linear by linear association	,138	1	,710
N of valid cases	216		

a. 0 cells (0,0%) have expected a count less than 5. The minimum expected count is 5,69.

In Table 5, we can observe the relationship between the dimension of compassionate care with the Universities of nursing interns, in which it was determined with Pearson's chi-square test (X^2). The level of significance of the test obtained a value of 5.69 ($p > 0.05$) ($X^2 = 2.278$; d.f = 4). Therefore, it is not possible to reject an association hypothesis, in which there is no statistical evidence in a significant relationship in the compassionate care dimension with the universities of nursing interns. By which we can interpret that the interns of the Universidad de Ciencias y Humanidades have a high compassionate care (46.9%).

Table 6: Empathy in its dimension, ability to put oneself in the place of others in nursing interns from three universities in Lima, 2019 (N = 216)

			Institution			
			Universidad de Ciencias y Humanidades	Universidad Nolbert Wiener	Universidad Privada del Norte	Total
Ability to put oneself in the place of others.	Low	Count	17	41	12	70
		% within Institution	26,5%	36,9%	29,2%	32,4%
	Medium	Count	41	54	25	120
		% within Institution	64,1%	48,6%	61,0%	55,6%
	High	Count	6	16	4	26
		% within Institution	9,4%	14,5%	9,8%	12,0%
Total		Count	64	111	41	216
		% within Institution	100,0%	100,0%	100,0%	100,0%

Chi-square tests

	Value	df	Asymptotic significance (bilateral)

Pearson's Chi-square	4,581[a]	4	,333
Likelihood ratio	4,607	4	,330
Linear by linear association	,068	1	,794
N of valid cases	216		

a. 1 cells (11,1%) have expected a count less than 5. The minimum expected count is 4,94.

In Table 6, we can observe the relationship between the dimension ability to put oneself in the place of others with the Universities of nursing interns, in which it was determined with Pearson's chi-square test (X^2). The level of significance of the test obtained a value of 4.94 ($p > 0.05$) ($X^2 = 4.581$; $d.f = 4$). Therefore, an association hypothesis cannot be rejected. In which, there is no statistical evidence in a significant relationship in the compassionate care dimension with the universities of nursing interns. By which it can be interpreted that the interns of the Universidad de Ciencias y Humanidades have a medium level of empathy (64.1%).

This dimension was the most affected, it makes understand that being nursing interns where the work of a professional is being carried out having a great responsibility, the first thing the hospital prioritize is the technological or the material and forget the emotional, the feelings; this makes understand that during academic training, it does not have a good empathetic training where it must be emotionally prepared and not as professional competitiveness among University institutions.

These results will be used as antecedents for future researchers to give greater emphasis to humanized treatment and improve the level of empathy at first base that are the students.

4. Dicussion

In the present study, an increase in medium empathy was found in nursing students where the level of empathy occurred more in women than in men, this is because women, because they are more emotionally sensitive, feel a more empathic concern, personal discomfort where concern for the care of the child, adult and older adults perceive themselves based on their understandings and at the same time on non-verbal language. Our findings are similar to the results of the authors M. Montilva and her collaborators in their study "Empathy according to the Jefferson scale in students of Medicine and Nursing in Venezuela", the average score of the EMed was 120.3, and the EEnf was 117.9; found a non-significant difference in the score with a higher level of empathy in women than in men [6]. The presence of empathy in the Peruvian population of three universities was 111, also have a medium empathy.

According to the study, empathy in nursing students according to their compassionate care dimension, it found with a high frequency, regarding the author M. Mejía and her collaborators in their study "Empathic behavior in science students of the health from Universidad de los Andes ", it is observed that there is not as much relationship since the students of the Universidad de los Andes are less compassionate [12].

In the found findings, it can observe that empathy in nursing students who carry out the non-medical internship, in its dimension perspective taking is high, which is related to the study of the author R. Lorch and his collaborators "a phenomenological study of nature of empathy ", which leads to the importance of developing empathy as an aspect of emotional competence, this study is related to the understanding that nursing students perceive and experience [13].

Another findings of empathy in health students, elaborated by J. Huancaya and his collaborators in their study "Evaluation of empathy in human medicine students of the private university of Peru", they found that those students who think their specialty was linked to patient care, they showed greater empathy, like those who reported having a professional role in patient contact, a higher percentage of the level of empathy in women was also found [5].

In the study by J. Blanco and collaborators in their study "Medical empathy in residents and medicine tutors of family and community. The vision of the professional and the patient "found that the health personnel, who have the longest service time, presented low levels of empathy compared to other professionals [14]. Regarding nursing students, it is observed that the majority have medium empathy, comparing with the mentioned study, the difference between them was that in the non-medical internship, there is a decrease in empathy.

It can be seen that there is a similar relationship in terms of gender, empathic levels in women are high, which indicates that women are superior to men.

5. Conclusions

It is concluded that empathy in nursing interns from three university institutions in Lima, 2019, medium empathy predominates. Regarding the dimensions of the three universities, the most affected dimension was putting themselves in the place of others the low level predominates, according to their compassionate care dimension, the medium level empathy predominates and, in the perspective, taking dimension the high level predominates.

In the bibliographic reviews of the national and international scientific databases, it was determined that there is not an adequate number of national scientific articles that involve this type of subject: Empathy in nursing interns and the comparison between different universities. Knowing that these students of the new generation must be prepared not only with scientific aspects but also with the human part. For this reason, the present research wants to complete this research gap.

The study contributes to the development and consolidation of the topic of research in empathy, and also introduces the concept of empathy in nursing interns through an instrument, which its new version is adapted to the field of nursing-patient. The results will contribute to the improvement of the humanized treatment of the patient.

Conflicts of Interest

The authors declare no conflict of interest

References

[1] J. McKinnon, "In their shoes: An ontological perspective on empathy in

nursing practice.," Journal of Clinical Nursing, **27**(21), 3882–3893, 2018, doi:10.1111/jocn.14610.

[2] J. Álvarez, J. Sáchica, J. Villalba, "Percepción de los pacientes acerca de la empatía de las enfermeras en Monterrey (México).," Revista Española De Comunicación En Salud, **9**(1), 46, 2018, doi:10.20318/recs.2018.4252.

[3] Navarro G., Vaccarezza G., Gracia M., Catalán R., Construcción de conocimiento en educación superior., 2015.

[4] M. Neumann, F. Edelhäuser, D. Tauschel, M. Fischer, M. Wirtz, C. Woopen, A. Haramati, C. Scheffer, "Empathy decline and its reasons: A systematic review of studies with medical students and residents.," Academic Medicine, **86**(8), 996–1009, 2011, doi:10.1097/ACM.0b013e318221e615.

[5] J. Huarcaya, M. Rivera, "Evaluación de la empatía en alumnos de medicina humana de una universidad privada del Perú.," Educacion Medica, **20**, 100–107, 2019, doi:10.1016/j.edumed.2018.05.008.

[6] M. Montilva, M. García, A. Torres, M. Puertas, E. Zapata, "Empatía según la escala de Jefferson en estudiantes de Medicina y Enfermería en Venezuela.," Investigación En Educación Médica, **4**(16), 223–228, 2015, doi:10.1016/j.riem.2015.04.006.

[7] O. Salazar, L. Dupuy, "Concepciones sobre empatía: desarrollo de un instrumento para la evaluación formativa a partir de una experiencia docente con residentes de pediatría de la Universidad de Antioquia.," Revista de La Fundación Educación Médica, **20**(5), 227, 2017, doi:10.33588/fem.205.910.

[8] L. Loyola, S. Landgrave, E. Ponce, A. González, "Nivel de empatía en médicos del servicio de urgencias de un hospital público de la ciudad de México.," Revista Atención Familiar, **22**(2), 43–45, 2015.

[9] Instituto Nacional de Salud, Prioridades de Investigación en Salud en el Perú 2019-2023., 2019.

[10] A. Callhan, J. Gonnella, V. PharmD, J. Veloski, M. Hojat, M. LaNove, Center for Research in Medical Education and Health Care., Jefferson Philadelphia University Thomas Jefferson University, 1–39, 2018.

[11] M. Hojat, Empatía en el cuidado del paciente: Antecedentes, desarrollo, medición y resultados., 1–318, 2012.

[12] M. Mejía, J. Poveda, V. Díaz, "Comportamiento empático en los estudiantes de las ciencias de la salud. Universidad de Los Andes.," Fermentum, **23**(67), 203–220, 2013.

[13] R. Lorch, P. Maras, R. Ruiz, L. Verschaffel, "Descenso de empatía en estudiantes de enfermería y análisis de posibles factores implicados.," Psicología Educativa, **20**, 53–60, 2014.

[14] J. Blanco, C. Valcárcel, M. Guerra, E. Ortigosa, F. García, F. Caballero, L. Cabello, D. Monge, "Empatía médica en residentes y tutores de medicina familiar y comunitaria. La visión del profesional y del paciente.," Atención Primaria, **52**(3), 185–192, 2020, doi:10.1016/j.aprim.2018.04.009.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 951-958 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Investigating the Optical Behavior of Single/Multi-Dimensional Photonic Crystal Structures for Photovoltaic Applications

Gehad Ali Alsayed, Zahraa Ismail, Sameh O. Abdellatif *

FabLab, Centre for Emerging Learning Technologies (CELT) and Electrical Engineering Department, Faculty of Engineering, the British University in Egypt (BUE), Cairo, 11387, Egypt

ARTICLE INFO

Article history:
Received: 28 September, 2020
Accepted: 25 November, 2020
Online: 08 December, 2020

Keywords:
Third generation solar cells
Light trapping structures
Optical modeling
Distributed Bragg reflectors
Optimized Photonic crystals
Experimental fabrication
Optical Characterization

ABSTRACT

The results investigated in this work are toward the optimization of the photonic crystal structures in 1D and 2D scale. One-dimensional distributed Bragg reflectors (DBRs) have demonstrated substantial potential in various optoelectronic applications, due to the observed tunable optical band-gap. Herein, the use of DBRs in light trapping solar cells was simulated and validated, representing its effect as a back reflector structure. In terms of the layer thickness, material selection and number of layer, the optimized DBR structure was modeled and evaluated with respect to previously published numerical and experimental data. The proposed model is capable of designing photonic crystal structures with tunable band-gap varies from 400 nm to 700 nm while controlling the pass-band in both Visible and Near Infra-red regions. On the other hand, 2D grating structure has been simulated where the transmission spectra under various design dimensions have been investigated. Finally, thin film deposition is utilized for experimental validation to our proposed optical model.

1. Introduction

This paper is an extension of the work originally presented at the IEEE Conference on Power Electronics and Renewable Energy (CPERE) in 2019 [1].

Nowadays, new techniques of solar cells have shown a perfect solution to provide low cost alternative compared with conventional methods that used in solar cells [2-10]. These methods depend on using low cost active layer with less than 1 μm thickness and also minimizing the fabrication cost [2, 11-14]. By observing the literature, the reported structures showed low absorption in the active layer [15]. Based on that, the researchers began to find method to avoid the limitation in the absorption layer by using light trapping structures [8, 16, 17] (cf. Figure1). The light trapping structures can be divided into two main classifications, either by function or composition of the structure. The composition of the light trapping structure can be in the form of photonic crystal (PCs) [18] or plasmonics structures [1, 16, 17, 19-23]. The fabrication of PCs structures depends on the contrast between dielectric materials and semiconductor materials (refractive index contrast) which influences the optical properties of the structure [24]. This paper will discuss 1D PCs [25] as well as 2D grating structure for solar cells utilization

Additionally, other new generations of solar cells, namely dye sensitized solar cells and perovskites solar cells, showed enhanced performance with embedding two-dimensional colloidal crystal structures [26-29]. Such colloidal crystal structures are not only acting to enhance the photon capturing of the active layer of solar cells, but can be integrated for decorative purposes [29]. Over and above, light trapping techniques can have a positive role while harvesting diffused light in underwater applications [30, 31]

Light trapping structures can be applied in new technologies of solar cells, by trapping the light in the active layer through embedding nano-structures either above or below the active layer [8]. The technique of applying light trapping in the active layer depends on minimizing the reflected light in the upper layer by using anti-reflecting coating (ARC) while maximizing the reflected light in the back layer. The angle of reflection in the back reflector must exceed the critical angle, targeting total internal reflection in the active layer, which attributes to the absorption [8] (see Figure 1).

Additionally, other new generations of solar cells, namely dye sensitized solar cells and perovskites solar cells, showed enhanced

*Corresponding Author: Sameh O. Abdellatif, British University in Egypt (BUE), Cairo, 11387, Egypt, +201222194077, sameh.osama@bue.edu.eg

performance with embedding two-dimensional colloidal crystal structures [26-29]. Such colloidal crystal structures are not only acting to enhance the photon capturing of the active layer of solar cells, but can be integrated for decorative purposes [29]. Over and above, light trapping techniques can have a positive role while harvesting diffused light in underwater applications [30, 31].

The demonstrated optical model introduces an advanced optimization tool for 1D/2D PCs, seeking for maximum absorption in the active layer utilized in new generation of solar cells. Using Matlab scripting, the developed tool was essentially deployed and proved using a free optical simulator platform called "Open-Filter" [32], while absorption profiles were simulated numerically using Comsol multi-physics. The optimization procedure was applied on particular parameters such as the number of layers constricting the PC structure, the difference in refractive index and the layer thickness. The output results showed complete consistency with the literature data previously measured [20] and simulated [8, 17].

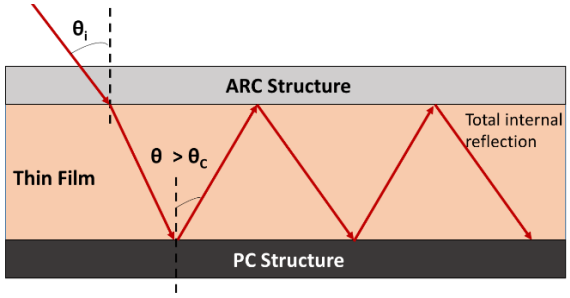


Figure 1: Schematic diagram of sandwiched light trapping structure for solar cells.

2. Optimization Model

2.1. Analytical modeling

The thin film optical modeling represented by characteristics matrix analytical method [33] is utilized in this paper. This method depends on Maxwell's equations which describe by electric field that cross continuously from medium to another. When the incidence of layer was given by electric and magnetic field, the field of both electric and magnetic can derived by matrix that given above [33]:

$$\begin{bmatrix} E_1 \\ H_1 \end{bmatrix} = M \begin{bmatrix} E_2 \\ H_2 \end{bmatrix} \quad (1)$$

where the intensity of electric and magnetic fields in the input medium are E_1 and H_1, while the electric and magnetic fields in the output medium are E_2 and H_2. Herein, M is given by:

$$M = \begin{bmatrix} \cos k_o h & i \sin k_o h/\gamma_1 \\ \gamma_1 i \sin k_o h & \cos k_o h \end{bmatrix} \quad (2)$$

and γ_1 is given by:

$$\gamma_1 = \sqrt{\frac{\varepsilon_o}{\mu_o}} \frac{n_1}{\cos \theta_i} \quad (3)$$

where h is thickness of thin film, k_o represented as propagation coefficient ($k_o = 2\pi/\lambda$), ε_o and μ_o are represented as air permittivity and permeability, refractive index of material called n_1, angle of incidence called θ_i.

A concatenation of layers can be referred to as a system matrix, represented by the multiplication of individual layer matrices. This matrix used to calculate the overall coefficients of reflection and transmission. Open-filters [32] is used as open source of analytic thin film optical modeling where we can validate our characterization matrix results.

Distributed Bragg Reflector (DBR) is a one-dimensional structure of simple photonic crystal. Optical band-gap is formed for Distributed Bragg Reflector as observed by solving Maxwell's equation (more detailed analytical solution for 1D and 2D photonic crystals can be accessed in [16]). The angle of incident and polarization should directly affect the width of band-gap. By incrementing the incidence angle to oblique, the TE modes band-gap becomes wider, while the band-gap of TM mode becomes narrower. This finding leads us to the independence of the band-gap from the polarizations and angle of incidence, as presented later in the discussion section. The photonic band-gap would exist, when contrast of dielectric occurred. In this manner, DBR could be shaped by layers of silicon with thickness 550 nm (n = 3.48) and SiO_2 of thickness 550 nm (n =1.46). The Lorentzian-Drude (LD) model introduced in [9] is used, which explain the absorption and dispersion of silicon as a wavelength dependent function.

2.2. Material modeling

The coefficient of dispersion should be calculated to modal dispersive and absorptive materials. Herein, Lorentzian-Drude (LD) model through FDTD simulations are enrolled [34]. The LD modal divided into two models. The Drude model is defined as the first model that, due to intra-band transitions, directly affects background absorption and dispersion. The second modal that represents dispersive media by multiplying resonant frequencies is the Lorentz model.

The model of semi-quantum reflects the electronic inter-band transitions. A material using the LD model's complex permittivity could be expressed by:

$$\varepsilon(\omega) = \varepsilon_{\bowtie} + \sum_n \frac{\sigma_n \omega_n^2}{\omega_n^2 - \omega^2 - i\omega\Gamma_n} \quad (4)$$

where $\varepsilon_{\bowtie}$ is the instantaneous dielectric response (DC), Γ_n and ω_n are the LD coefficients and σ_n is a position function that defines the power of the nth resonance. Figure 2 shows the optical model for the DBR structure to be optimized. The optimization process is considering the number of DBR layers, reflective index (n_1– n_2), and the thickness of the layer (h_1, h_2).

2.3. Structure optical modeling using FDTD

The finite difference time-domain (FDTD) algorithm is used to make simulation of light interaction with various structure of layers [35]. The (FDTD) algorithm is considered as a standard numerical technique in classical electromagnetism. To simulate (FDTD) algorithm, open source algorithm is used called MEEP [36].The simulated volume is reflected on the required simulation time and RAM that required to make simulation on MEEP. A runtime of nearly 10 ns per pixel per time step of time is needed with RAM of almost 100 bytes per pixel [37].

MEEP was used to implement simulation of reflection and transmission of light by using different structure of light trapping. 10 μm × 10 μm is volume that used in simulation with unit step and resolution of 10 nm and 10 respectively. The wavelength is set as 550 nm to simulate sun irradiance for TM and TE modes. Figure 3, shows the modal with two detectors integrated for the calculation of transmitted and reflected field.

The maximum intensity of AM1.5G is determined to be 1000 W/m^2. Two identical Gaussian light sources with a width from 200 nm to 1500 nm and a mean wavelength of 550 nm, were utilized to represent TE and TM modes. It is important to model and integrate in MEEP the wavelength dependent variation in the real and imaginary parts of the refractive index for different materials used. Here, we integrate the published LD model to illustrate the complex refractive indices using the Lorentzian-Drude (LD) coefficients for the chosen materials with specific focus on Si as a semiconductor material. The compatibility of the LD model with MEEP is considered as the main reason for selecting it [38].

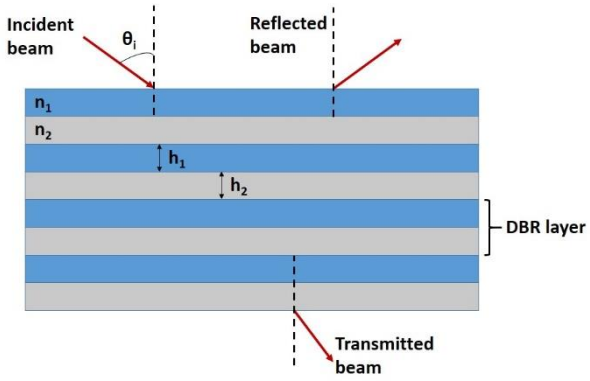


Figure 2: Schematic diagram of the structure of the DBR under test.

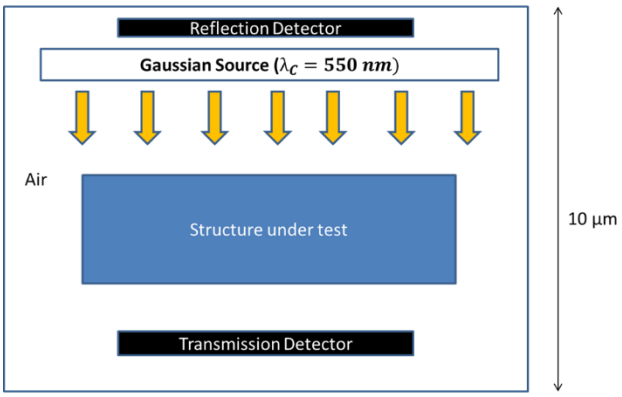


Figure 3: Schematic diagram used for the calculation of transmission and reflection spectra for the MEEP model.

2.4. 2D optical modeling using Comsol Multi-physics

Using the commercial simulator Comsol Multiphysics 5.0 [37], optoelectronic modeling is carried out. Comsol utilize finite element method (FEM) solver such that a tetrahedral mesh is selected based on the minimum feature length in the geometry under investigation. Various 2D optical structures devices could be simulated by optical and electromagnetic libraries and are modeled by the continuity, Poisson, drift-diffusion and Maxwell's equations assuming 2D solution. The targeted solar cells with light trapping structure are described as a p-i-n junction [39]. Consideration is also provided to the bulk, surface recombination, and contact effects. The primary governing equations are given by [8]:

$$-\nabla.\left(\frac{\varepsilon_o \varepsilon_r}{q}\nabla V\right) = -n + p + N_D(1 - f_D) - N_A(f_A) + \sum_j N_{tj}(\delta_j - f_{tj}) \quad (5)$$

$$\nabla.J_n - \sum_j R_n^{tj} - R_{sp} - R_{st} - R_{av} + G_{opt}(t) = \frac{\partial n}{\partial t} + N_D\frac{\partial f_D}{\partial t} \quad (6)$$

$$\nabla.J_p - \sum_j R_p^{tj} - R_{sp} - R_{st} - R_{av} + G_{opt}(t) = \frac{\partial p}{\partial t} + N_A\frac{\partial f_A}{\partial t} \quad (7)$$

where p and n are the excess hole concentration and minority-electron concentration with respect to equilibrium, N_A and N_D are the doping levels of acceptors and donors, J_p and J_n are the current density for holes and electrons, R_p and R_n represent the bulk recombination of holes and electrons, G_{opt} is the term for optical generation, R_{sp} represents the recombination of surfaces, and R_{av} and R_{st} are trapping recombination. The required optoelectronic parameters per each layer are captured from the simulation tool library and the optical generation term is calculated using an Matlab based optical model, taking into confederation a doping level of 10^{18} cm^{-3} in both the n and p sides [10].

Herein, the optical generation parameter G_{opt} plays a critical role in investigating the optical impact of the proposed photons management structure in enhancing the performance of the solar cell. The expected enhancement can be attributed to the elongation of the optical path length inside the active materials under simulation. This elongation is directly reflected on the generation term.

3. Simulation Results

3.1. Optimizing number of layer in DBR structure

Si-SiO$_2$ DBR structures have been simulated following the characteristics matrix analytical approach defined in the section two. Based on our previous work, these DBR structures were taken into account as an initial model to be optimized [8]. Firstly, the influence of the DBR layers on the presence of the optical band-gap was originally discussed in Figure 4. The spectra showing the reflectance are plotted and the structures of two, four and six DBR layers have been simulated. Here, the ratio between the the reflected light and the incident light is termed as the optical reflectivity (dimensionless unit), cf. Figures 2 and 3.

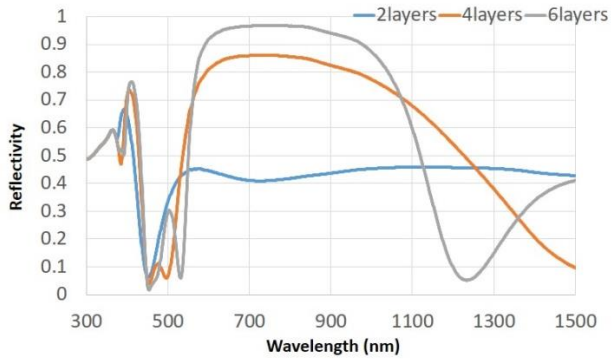


Figure 4: Si-SiO2 Reflectivity of DBR for different layer numbers.

It is evident from Figure 4 that the occurrence of the optical band-gap is associated to the number of layers constructing the one dimensional photonic crystal structure. 550 nm optical band-gap

(from 550 nm to 1100 nm) was formed with reflectively of around 0.95, while utilizing six DBR layers structures. In contrast, attention should be given to the trade in parameters concerned to the overall DBR dimensions and to the fabrication procedure sophistications. For instance, a sharper optical band-gap is anticipated by increasing the number of layers, more than six, however the complicity and fabrication cost will increase as well. Thus, the optimal structure in terms of layer count for such Si-SiO_2 DBR structure is the six layer DBR structure. It is worth noting that this layer number is not a global optimum, as the observation of the optical band-gap depends not only on the DBR layers count, but also plays the refractive index contrast plays an critical role.

3.2. DBR thickness optimization

Following the first optimization stage demonstrated in section 3.1, Figure 5 shows the influence of the thickness variance on reflectivity. The DBR structure of six layers optimized in section 3.1 is used. To reduce the variables included in the optimization problem under investigation, the thickness of both layers, SiO_2 and Si, are considered equal. Figure 5 examined 550 nm, 250 nm, and 125 nm reflectivity spectra. The nominated wavelengths are selected to fit the peak wavelength for the input source at 550 nm. Consequently, λ_o , $\lambda_o/2$ and $\lambda_o/4$ thicknesses have been simulated. As shown in Figure 4, the presence of the photonic-gap is associated at thin film thickness of $\lambda_o/4$. This is attributed to the maximum local point at which the maximum reflection occurs. The thin film thickness is nominated to be a 1/4 of the optical wavelength where maximum reflection is observed [14]. In view of the projection of a quarter wavelength layer of Si with a refractive index of n_1 and light on a wavelength layer of λ_o, the thickness h_1, which induces maximum reflection, is determined by [14]:

$$h_1 = \frac{\lambda_o}{4\, n_1} \quad (8)$$

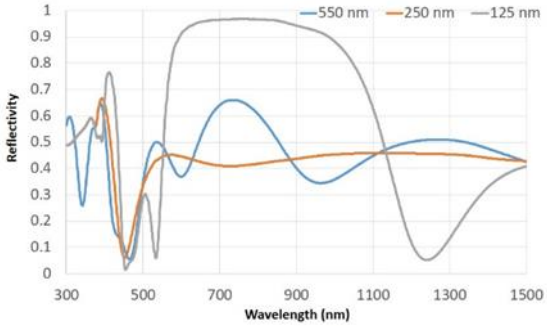


Figure 5: Six Si-SiO2 DBR reflectivity layers with various layer thicknesses.

3.3. Refractive index contrast optimization

Finally, Titanium dioxide is integrated as a direct substitution for Si, in order to examine the impact of refractive index contrast variation on the photonic-gap. TiO_2 is chosen as an easily prepared, low-cost and efficient n-type semiconductor material, knowing that we have already investigated its refractive index experimentally in [28]. Manual as well as automated Dr Blade screen printing can be utilized, as a sol-gel method, to deposit thin film of TiO_2 layer in micrometer range as previously discussed in [28]. In such context, diluted paste with isopropanol can be managed to tune the thin film thickness [27]. The large variation between the refractive index of TiO_2, nearly 1.78, and that of Si, around 3.9 [26], negatively contribute to the detected width of the band-gap. This narrowing can be observed in Figure 6, as a reduced, less sharp optical band-gap with maintaining the number of layers in both DBRs under test.

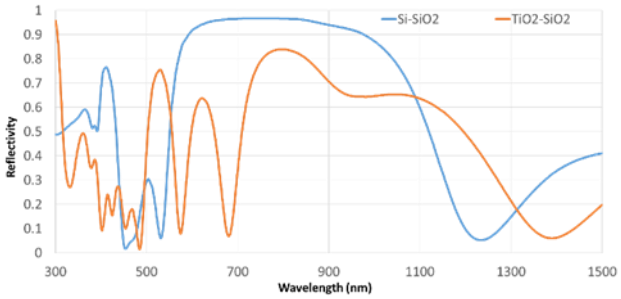


Figure 6: 125 nm, six layers Si-SiO2 and TiO2-SiO2 DBR reflectivity.

3.4. Absorption profile simulation

Towards more optical investigation in the DBR structure optimized in the previous versions, Comsol Multi-physics was utilized. Herein, six layers Si-SiO_2 DBR modeled in the previous section were simulated. The electric field components are used to describe the absorption profile following the Poyting vector method illustrated in [8]. The demonstrated optical absorption profile shown in figure 7, at 550 nm, illustrates the trapping functionality of the DBR structure to trap light inside the active material seeking for optical path length elongation. Accordingly an enhancement of nearly 38% is recorded in the optical generation term. The simulation has been conducted over an optical wavelength range from 200 nm (UV-range) to 2000 nm (IR-range) with a peak at 550 nm.

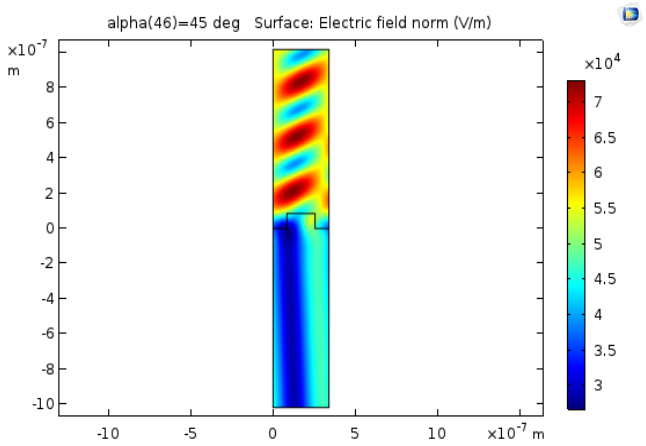


Figure 7: 125 nm, six layers Si-SiO_2 absorption profile.

3.5. Two-dimensional grating light trapping structure

As shown in figure 8a, a 2D grating is characterized as structures that are periodic with a periodicity L in two-dimensions, (e.g. x-direction and y-direction), and are uniform in only one dimension (z-direction). We consider the propagation of incident light in the xz-plane. Since the structure considered is two-dimensional, the reaction of the structures can be separately considered for the s-polarization, which has its electric field polarized along the y-direction, and for the p-polarization, which, as shown in figure 8b, has its magnetic field polarized along y-direction.

With a periodicity L of 1 µm, SiO_2 thickness d_2 of 100 nm, and an aspect ratio (a/b) of 1 (figure 8b), a 2D Si/SiO_2 grating structure is simulated. The minimum x-direction dimension b = 500 nm is used to allow such a device to be fabricated in an area of 200 µm x 200 µm. Furthermore, another experiment was carried out in order to simulate the same structure but with aspect ratio of 2.4. The reflectance of this structure is shown in Figure 9 in the near infrared region with an optical band-gap. With respect to the DBR structure, this shifting in the optical band gap structure is due to the scaling up of the minimum dimension length from 50 nm to 500 nm.

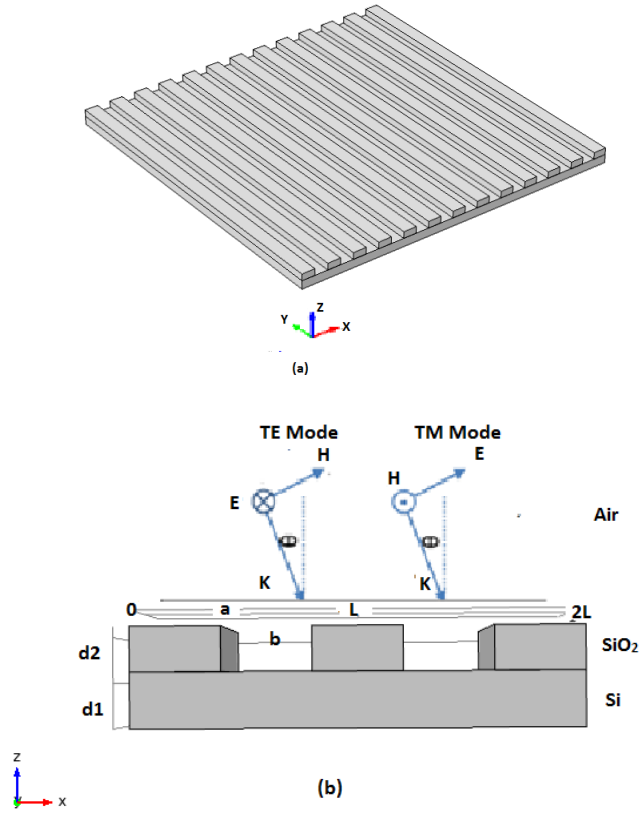


Figure 8: (a) 3D schematic for the structures of 1D grating. (b) X-Z cross section for 1D grating structure.

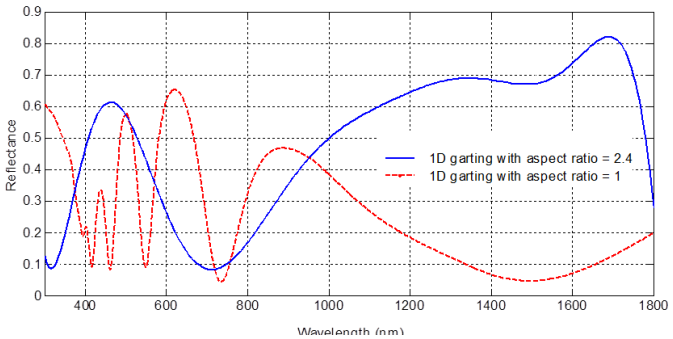


Figure 9: 1D grating structure with aspect ratio of 1 (red curve) and 2.4 (blue curve).

4. Fabrication Process

As mentioned in the introduction, one of our targets here is to fabricate thin film to validate our proposed optical model. Herein, thermal evaporator and spin coater are suggested to be used for this process, while UV-Vis spectrometer is used for the optical characterization purposes.

The thermal evaporator consists of different subsystems. One of the subsystems is the vacuum chamber, the control valves, the vacuum pumps are integrated to satisfy the vacuum itself, and the gauges is utilized to evaluate the pressure level in the chamber. Another obvious subsystem is the filament and the current source required to heat the filament and to control its temperature. Another subsystem measures the thickness of the deposited film so that we can know when the targets are evaporating and when the film deposition is complete. A more trivial subsystem is the framework used physically to hold the substrates and to control whether the evaporated material reaches the substrates.

The Spin coater is equipment enrolled in deposing thin films to flat surfaces substrates. A coating material, usually called suspension, is drooped in the center of the sample before starting the spinning process. The sample is then rotated either in one step or multi-steps with a certain rotating speed and acceleration. This rotation produces the thin film with the aid of centrifugal force.

The speed and the acceleration chosen play an important role in adjusting the thickness of the produced film. The used suspension is usually chosen to be easily evaporated and volatile. It is also worth to mention that not only the speed and acceleration of rotation tune the thickness but also the viscosity and concentration of the suspension. However, in principle, the faster the rotation speeds, the thinner the produced film. Typically, spin coater are used in deposing thin films of few nanometers thick to micrometers range such as the photoresist materials.

For optical characterization, the UV-Vis spectrometer used to scan the ultraviolet to near infrared regions including the visible portion of spectrum. The basic setup of the UV-VIS spectrometer is consisting of a light source which includes two lamps. The first is a tungsten-halogen lamp for the visible and the near infrared (NIR) region. For the UV region, the deuterium lamp is used A NIR and a visible light detector are used in detecting light. Finally, fibers optics cables are used to connect the source detectors and sample holder (cf. Figure 10).

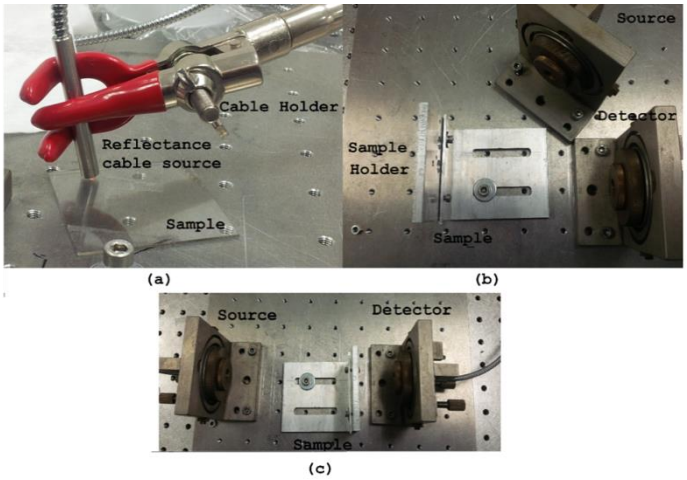


Figure 10: (a) Measuring reflectance using three terminals fibres cable. (b) Measuring reflectance using source/detector setup. (c) Measuring the transmission

5. Experiential validation

In order to validate order optical proposed model, experimental validation take place using thin film based samples. A thermal

evaporation process is used to evaporate a 70 nm of Ag on a highly P-doped Si wafer. The highly P-doped Si wafer is chosen to maintain its optical characteristics during the fabrication process. The thermal evaporation process is done under a current rate of 12% where the tooling factor was previously determined using a dektak profilometer and it was found to be 0.551. The output evaporation rate is 0.1 nm/sec. Then a characterization process takes place using UV-VIS spectrometer to measure the reflectance of the sample in the desired portion of the spectrum as shown in figure 10. The effect of adding the collimating lens is also considered here showing a more accordance with the simulation results than the without collimating lens results. The simulation results shown here are based on our optical numerical model described in the previous section.

The results shown in figure 11 is not only considered as a new validation step to our optical model but it can also be considered as a validation to the new suggested measurement setup of the UV-VIS spectrometer with adding collimating lens. The results show that the adjusted lens doesn't cause any undesirable effect in the wavelength portion of interest making it possible to use such a setup in measuring the reflectance of small area samples.

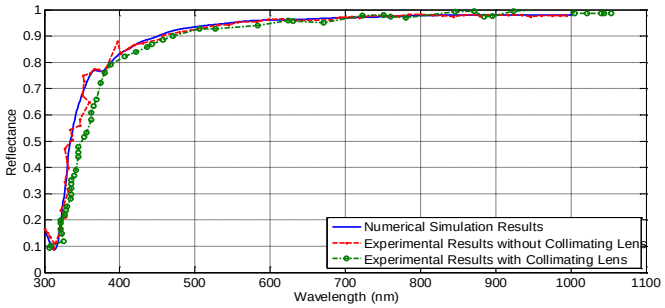


Figure 11: Comparison between the simulated and the measured reflectance of a 70 nm Ag thin film.

A 40 nm thin film of SiO_2 is deposited on a Si <100> wafer using thermal evaporator. The tooling factor is calculated and it was 0.1812 while the current rate is set to be 35%. It is observed that the current needed to evaporate SiO_2 is much higher than that used for Ag, this is due to the high evaporating temperature of SiO_2 with respect to that of Ag. The evaporation rate is varying around 0.01 nm/sec. The Reflectance of the thin film is measured using the UV-VIS spectrometer as observed in figure 12. The figure shows a high accordance between the numerical simulated and the measured reflectance. It was found that due to the high evaporating temperature of SiO_2 (about 1800 ^oC); it is not efficient to evaporate thin film of SiO_2 using thermal evaporation technique. As deposing thin film of SiO_2 must be done by more than one iteration, due to the very low evaporating rate, which causes a poor uniformity of the surface.

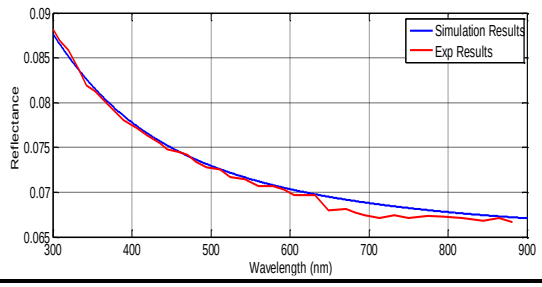


Figure 12: Comparison between the simulated and the measured reflectance of a 40 nm SiO_2 thin film.

It's worth to mention here that the optical validation process implemented on two sample of thin films behaving as optical filter showed perfect agreement with the data captured from our proposed optical model. However, the validation process demonstrated in this paper is limited to thin films samples due to constrain associated with photonic crystals fabrication. Accordingly, we consider the validation step with fabricated photonic crystal structure as a future extension of the current work.

6. Discussions

After the iterations used to optimize DBR structure the output for three steps of optimization that shown in section 3 and also take in consideration the previous measurement in [8], shown in figure 13. By our requested of optimization modal, it is shown the suitable match. Our represented structure is suggested structure to back reflector of thin film of solar cells. This model of structure lead to optical band-gap in near infrared and far visible region, (from 800 nm to 1100 nm) the low absorption occurred at the layer of Si. We attribute such extended band-gap to the refractive index contrast with respect to the DBR shown in figure 5. This structure showed below describe the performance of non-absorbed photons which cannot escape from active layer, after that the light trapping occurred inside material. It is worth to mention that the height (y-axis reflectance value) is inversely proportional to the optical band gap-width [16]. The higher the reflectance the lower the band-gap.

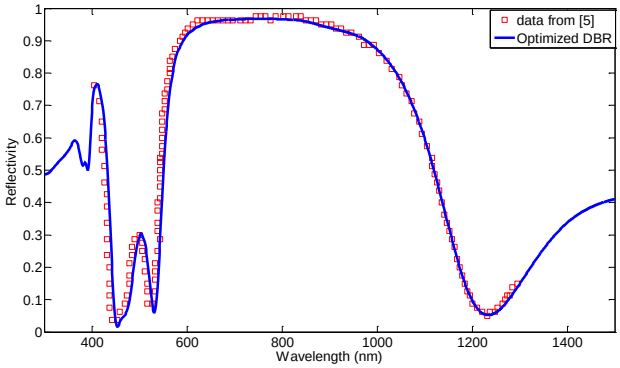


Figure 13: 125 nm, six layers Si-SiO_2 DBR verification with previously published data in [5].

A primary downside within the proposed silicon-slica DBR structure is that the nonappearance of metal which may act as a counter electrode. Consequently, another altered DBR structure with Indium tin-oxide (ITO) rather than Si can be suggested. Though, the proposed planned structure will lack of a condensed optical band-gap because of the less index of refraction distinction with the ITO-SiO_2 DBR with respect to the optimized Si-SiO_2 DBR structure. Additionally, the proposed ITO-SiO_2 DBR structure will show more refined fabrication process, particularly within the deposition of SiO_2 layer the thin ITO layer. Regularly the SiO_2 deposition is achieved by chemical vapor deposition (CVD) at moderately high temperature. Such testimony conditions might influence the ITO electrical properties yet as morphology.

Fundamentally, Nano-lithography technology can be essentially used to fabricate higher dimensions (2D and 3D) photonic crystal structure. In this manner, more refined manufacturing processes and costs would be expected. This specifically influences the cost of delivered energy (Wh/$) by TFSCs. An inclusive analysis showing the effect of the 1D, 2D, and 3D PCs structure in terms of cost versus increase in conversion

efficiency can be carried out as a future task. In principle, alternative low-absorption solar cells can be upgraded using the same light trapping process.

7. Conclusion

The optimization in DBR structure that represented in the paper depends on thickness of layers, refractive index contrast and number of layers. The optimization of DBR structure implement in two step, firstly analytical implantation by MATLAB and secondly verification by using Open-filter. In order to simulate the solar spectrum (AM1.5G), Gaussian sources and FDTD were used.

By using 6 layers of Si-SiO_2 of DBR as result of optimization, the band-gap occurred in 550 nm. This band-gap represents a massive potential in Si as reflected near-infrared of photonics and this reflection avoid light from escaping from material. After comparing between previous data that published and our simulation results, the suitable match was chosen. On another hand, TiO_2 was examined as alternative of Si to observe the effect of changing refractive index in the optical band-gap. At the end the effect of adding metal layer of ITO on Si as substitution was discussed in this paper and released the effect of it on the optical band-gap. Additionally, 2D grating structure was also examined based on the developed optical mode, where transmission spectra were simulated. Fabricated thin films were utilized to validate the optical model while a perfect matching was recorded.

The proposed optical model has showed a great potential in designing optimal 1D and 2D photonic crystal structures with a tunable optical band-gap varying from 400 nm to 700 nm capable of integrating in various solar cells applications. Moreover, the study investigates the optical absorption profile for various photonic crystals seeking for minimum parasitic absorption.

Acknowledgment

In this work, the authors would like to acknowledge the contribution and support of the STDF. As part of the STDF Project entitled, “Mesostructured Based Solar Cells for Smart Building Applications”, Project ID#33502. Additionally, the authors would like to thank the Centre for Emerging Learning Technology (CELT), directed by Prof. Hani Ghali, in BUE for providing all the fabrication facilities needed.

References

[1] S. O. Abdellatif and G. A. Alsayed, "Optimizing 1D Photonic Crystal Structures for Thin Film Solar Cells," in 2019 IEEE Conference on Power Electronics and Renewable Energy (CPERE), 333-337, 2019, doi:10.1109/CPERE45374.2019.8980046.

[2] M. Mujahid, C. Chen, W. Hu, Z.-K. Wang, and Y. Duan, "Progress of High-Throughput and Low-Cost Flexible Perovskite Solar Cells," Solar RRL, 1900556, 2020, doi:10.1002/solr.201900556.

[3] M. Sajadnia, S. Dehghani, Z. Noraeepoor, and M. H. Sheikhi, "Highly improvement in efficiency of Cu (In, Ga) Se2 thin film solar cells," World Journal of Engineering, **17**(4), 527–533 2020, doi:10.1108/WJE-02-2020-0068.

[4] S. Amiri and S. Dehghani, "Design and Simulation of Single-Junction and Multi-junction Thin-Film Solar Cells Based on Copper Tin Sulfide," Journal of Electronic Materials, **49**, 5895–5902, 2020, doi.org/10.1007/s11664-020-08382-6.

[5] F. Sadegh, S. Akin, M. Moghadam, V. Mirkhani, M. A. Ruiz–Preciado, Z. Wang, et al., "Highly efficient, stable and Hysteresis–less planar perovskite solar cell based on chemical bath treated Zn_2SnO_4 electron transport layer," Nano Energy, 105038, 2020, doi:10.1016/j.nanoen.2020.105038.

[6] S. Liu, J. Yuan, W. Deng, M. Luo, Y. Xie, Q. Liang, et al., "High-efficiency organic solar cells with low non-radiative recombination loss and low energetic disorder," Nature Photonics, **14**, 300–305, 2020, doi.org/10.1038/s41566-019-0573-5.

[7] F.-I. Lai, J.-F. Yang, Y.-C. Hsu, and S.-Y. Kuo, "Enhanced omnidirectional light harvesting in dye-sensitized solar cells with periodic ZnO nanoflower photoelectrodes," Journal of Colloid and Interface Science, **562**, 63-70, 2020, doi:10.1016/j.jcis.2019.12.003.

[8] S. Abdellatif, K. Kirah, R. Ghannam, A. Khalil, and W. Anis, "Enhancing the absorption capabilities of thin-film solar cells using sandwiched light trapping structures," Applied optics, **54**, 5534-5541, 2015, doi:10.1364/AO.54.005534.

[9] S. Abdellatif, R. Ghannam, and A. Khalil, "Simulating the dispersive behavior of semiconductors using the Lorentzian–Drude model for photovoltaic devices," Applied optics, **53**, 3294-3300, 2014.

[10] S. Abdellatif and K. Kirah, "Numerical modeling and simulation for a radial pin nanowire photovoltaic device," Energy Procedia, **36**, 488-491, 2013, doi:10.1364/AO.53.003294.

[11] M. M. Hassan, A. Sahbel, S. O. Abdellatif, K. A. Kirah, and H. A. Ghali, "Toward low-cost, stable, and uniform high-power LED array for solar cells characterization," in New Concepts in Solar and Thermal Radiation Conversion III, 114960Q, 2020, doi:10.1117/12.2572874.

[12] G. R. Li and X. P. Gao, "Low-cost counter-electrode materials for dye-sensitized and perovskite solar cells," Advanced Materials, **32**, 1806478, 2020, doi:10.1002/adma.201806478.

[13] B. Pashaei, S. Bellani, H. Shahroosvand, and F. Bonaccorso, "Molecularly engineered hole-transport material for low-cost perovskite solar cells," Chemical Science, **11**, 2429-2439, 2020, doi:10.1039/c9sc05694g.

[14] E. G. Durmusoglu, G. S. Selopal, M. Mohammadnezhad, H. Zhang, P. Dagtepe, D. Barba, et al., "Low-Cost, Air-Processed Quantum Dot Solar Cells via Diffusion-Controlled Synthesis," ACS Applied Materials & Interfaces, **12**, 32, 36301–36310, 2020 doi:10.1021/acsami.0c06694.

[15] J. H. Werner, J. Mattheis, and U. Rau, "Efficiency limitations of polycrystalline thin film solar cells: case of Cu (In, Ga) Se2," Thin Solid Films, **480**, 399-409, 2005, doi:10.1016/j.tsf.2004.11.052.

[16] R. Biswasa and D. Zhoua, "Enhancing Light-trapping and Efficiency of Solar Cells with Photonic Crystals," in Materials Research Sociaty Spring Meeting - Symposium A – Amorphous and Polycrystalline Thin-Film Silicon Science and Technology, Germany, 0989-A03-02, 2007, doi:10.1557/PROC-0989-A03-02.

[17] S. Abdellatif, K. Kirah, R. Ghannam, A. Khalil, and W. Anis, "Comprehensive study of various light trapping techniques used for sandwiched thin film solar cell structures," in Physics, Simulation, and Photonic Engineering of Photovoltaic Devices VII, 10527152018, doi:10.1117/12.2291613.

[18] J. Hou, M. Li, and Y. Song, "Patterned colloidal photonic crystals," Angewandte Chemie International Edition, **57**, 2544-2553, 2017, doi:10.1002/anie.201704752.

[19] W. Zhu, R. Esteban, A. G. Borisov, J. J. Baumberg, P. Nordlander, H. J. Lezec, et al., "Quantum mechanical effects in plasmonic structures with subnanometre gaps," Nature communications, **7**, 1-14, 2016, doi:10.1038/ncomms11495 .

[20] M. Mariano, P. Mantilla-Pérez, P. Romero-Gómez, A. Martínez-Otero, X. Elias, R. Betancur, et al., "One-Dimensional Photonic Crystals for Light Management in Organic Solar Cells," in Organic and Hybrid Photonic Crystals, ed: Springer, 303-320, 2015, doi:10.1007/978-3-319-16580-6_13.

[21] F. Enrichi, A. Quandt, and G. C. Righini, "Plasmonic enhanced solar cells: Summary of possible strategies and recent results," Renewable and Sustainable Energy Reviews, **82**, 2433-2439, 2018, doi: 10.1016/j.rser.2017.08.094.

[22] C. Zhao, P. Tong, L. Shi, T. Ji, F. Zhu, Y. Hao, et al., "Performance improvement in inverted organic solar cells by incorporating core-shell SiO_2@ Au plasmonic structures," in 9th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Micro-and Nano-Optics, Catenary Optics, and Subwavelength Electromagnetics, 108400N, 2019, doi:10.1117/12.2510899.

[23] R. S. Moakhar, S. Gholipour, S. Masudy-Panah, A. Seza, A. Mehdikhani, N. Riahi-Noori, et al., "Recent Advances in Plasmonic Perovskite Solar Cells," Advanced Science, **7**, 1902448, 2020, doi:10.1002/advs.201902448.

[24] J. D. Joannopoulos, P. R. Villeneuve, and S. Fan, "Photonic crystals: putting a new twist on light," Nature, **386**, 143-149, 1997, doi:10.1038/386143a0.

[25] A. Fu, H. Gao, P. Petrov, and P. Yang, "Widely tunable distributed Bragg reflectors integrated into nanowire waveguides," Nano letters, **15**, 6909-6913, 2015, doi:10.1021/acs.nanolett.5b02839.

[26] S.-P. Yu, J. A. Muniz, C.-L. Hung, and H. Kimble, "Two-dimensional photonic crystals for engineering atom–light interactions," Proceedings of

the National Academy of Sciences, **116**, 12743-12751, 2019, doi: 0.1073/pnas.1822110116.
[27] S. Abdellatif, S. Josten, P. Sharifi, K. Kirah, R. Ghannam, A. S. Khalil, et al., "Optical investigation of porous TiO_2 in mesostructured solar cells" 10526: SPIE, 2018, doi:10.1117/12.2288376.
[28] S. Abdellatif, P. Sharifi, K. Kirah, R. Ghannam, A. Khalil, D. Erni, et al., "Refractive index and scattering of porous TiO_2 films," Microporous and Mesoporous Materials, **264**, 84-91, 2018, doi: 10.1016/j.micromeso.2018.01.011.
[29] S. O. Abdellatif, S. Josten, A. S. Khalil, D. Erni, and F. Marlow, "Transparency and Diffused Light Efficiency of Dye-Sensitized Solar Cells: Tuning and a New Figure of Merit," IEEE Journal of Photovoltaics, **10**, 522-530, 2020, doi:0.1109/JPHOTOV.2020.2965399.
[30] S. Kamal, G. M. Al-sayyad, R. Abdelmoteleb, M. Abdellatif, and S. O. Abdellatif, "Submerged Solar Energy Harvesters Performance for Underwater Applications," in 2019 International Conference on Innovative Trends in Computer Engineering (ITCE), 444-449, 2019, doi:10.1109/ITCE.2019.8646459..
[31] M. M. Abdellatif, S. M. Maher, and M. Ghazal, "Implementation of a low cost, solar charged RF modem for underwater wireless sensor networks," International Journal on Smart Sensing and Intelligent Systems, **13**, 1, 1-11, doi:10.21307/ijssis-2020-015.
[32] S Larouche and L Martinu, "OpenFilters: open-source software for the design, optimization, and synthesis of optical filters," Appl. Opt. **47**, C219-C230, 2008, doi:10.1364/AO.47.00C219.
[33] E. Hecht, OPTICS, Fourth eddition. Adelphi university, San Francisco, Boston, New York, 2002.
[34] A. D. Rakić, A. B. Djurišić, J. M. Elazar, and M. L. Majewski, "Optical properties of metallic films for vertical-cavity optoelectronic devices," Applied optics, 37, 5271-5283, 1998, doi:10.1364/AO.37.005271.
[35] K. Kunz and R. Luebbers, Finite differnce time domain method for electromagnatics. United State of America: CRC Press, 1993.
[36] J. D. Joannopoulos, "MEEP," ed: Massachusetts Institute of Technology, https://meep.readthedocs.io/en/latest/, accessed October 2020.
[37] A. F. Oskooi, D. Roundy, M. Ibanescu, P. ermel, J. D. Joannopoulos, and S. G. Johnson, "Meep: A flexible free-software package for electromagnetic simulations by the FDTD method," Computer Physics Communications, ELSEVER, **181**, 687–702, 2010, doi:10.1016/j.cpc.2009.11.008.
[38] "Reference solar spectral irradiance: Air mass 1.5," ed: National Renewable Energy Lab.
[39] C. Multiphysics, "version 4.2," Heat Transfer Module, https://www.comsol.com/products/4.2/, accessed October 2020.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 959-966 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Standalone Operation of Modified Seven-Level Packed U-Cell Inverter for Solar Photovoltaic System

Kishan Bhushan Sahay[1,*], Pankaj Kumar Singh[1], Rakesh Maurya[2]

[1]*Department of Electrical Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur,273010, India*

[2]*Department of Electrical Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, 395007, India*

ARTICLE INFO

Article history:
Received: 29 October, 2020
Accepted: 16 November, 2020
Online: 08 December, 2020

Keywords:
Cascaded H-Bridge (CHB)
Modified Packed U-Cell (MPUC)
Packed U-Cell (PUC)
Solar Photovoltaic System

ABSTRACT

In this paper, a modified configuration of Single-Phase Seven-Level Packed U-Cell (PUC) Multilevel Inverter for solar photovoltaic system is presented & investigated for standalone operation. The Seven-Level MPUC Inverter comprises of six semiconductor switches & two DC links which generates seven voltage levels at the inverter output. The maximum amplitude of inverter output voltage is more than that of available maximum DC link value (i.e. sum of two DC link values). The voltage levels generated in the MPUC inverter is more with reduced number of components counts as compare to conventional multilevel inverter topologies like Cascaded H-Bridge (CHB) and Neutral Point Clamped (NPC). The proposed inverter finds application in PV System where green power is derived from two PV panels with different power rating and voltage rating connected to DC links. Several design considerations viz. the RMS value of inverter output voltage, switching frequency & voltage rating of semiconductor switches are taken into consideration to prove the advantages of the MPUC inverter. The simulation results of the MPUC inverter shows the appropriate operation of multilevel inverter.

1. Introduction

Owing to growing number of consumers in addition to large scale power industries, the power grid has experienced large energy demand from last decade. Traditional bulky transformers are significantly replaced by power electronic devices because a lot of research work is happening in power electronic field these days resulting in advancement of semiconductor device technology. Power converters are utilized as a conversion medium for transferring green power either for grid or a standalone load. Economic pricing of power semiconductor devices makes it beneficial to produce and allow them in competitive market. Recently, the use of a greater number of switches in the converter causes no significant rise in pricing. This is the reason that the traditional two-level converters having high harmonics and power losses are being interchanged by the multilevel inverters having less switching frequency [1, 2].

Now, the focus of researchers has moved towards developing various topologies of multilevel inverters and their control strategies. So, the foremost attention is to reduce the component counts in these inverters. Fewer the component counts, lower the cost and power loss [3]-[5]. The two most popular commercially available multilevel inverter (MLI) topologies are CHB and NPC inverter topologies [6].

Recently, the flooding of green power into the grid has become a prime concern [7]. Most of the countries have initiated to implement the solar power panels at broad scale to cater the local energy need. With the use of power electronic converters, the DC power is converted into AC power which is further transferred to the grid or a standalone load.

Thus, the improvement in conversion efficiency and reduction in power loss of these converters are tackled in several reports [8-9]. Conversion of DC voltage of PV system into AC waveform with low harmonics to be used by the grid or standalone load with high conversion efficiency and small size filtering requirement can be fulfilled by single-phase multilevel inverters (MLI) [10].

The key problem with most of the topologies are to use numerous isolated DC sources which makes it difficult for its practical implementation. An isolated DC source can be made by an AC transformer and diode bridge rectifier which is a costly affair and become bulky for manufacturing. Consequently, topologies having single DC source i.e. NPC (Neutral Point

*Corresponding Author: Kishan Bhushan Sahay, kishansahay16@gmail.com

Clamped) inverter and five-level PUC (PUC5) inverter [11] are commercially more viable than other new topologies. Nevertheless, some topologies e.g. CHB are utilized for very high-power applications owing to necessity of large number of DC sources.

In latest survey, seven-level PUC inverter topology was introduced which has many advantages like low component counts and numerous output voltage levels [12]. The key problem with PUC inverter is its maximum output voltage level which is actually the maximum available DC link value of the topology. In recent times, a seven-level modified PUC inverter topology has been presented which provides an alternative for those applications where several Photovoltaic panels are available for connecting it to isolated DC links. In [13], the authors demonstrated that Photovoltaic panels having different voltage rating & power rating can be coupled with two DC link of modified PUC inverter topology.

In this paper, modification of Seven Level Packed U-Cell inverter is done for solar PV systems. MPUC inverter topology is similar to CHB inverter having unequal DC sources where two cells produces seven voltage levels. The introduced inverter has a smaller number of semiconductor switches than the CHB inverter with equivalent performance. In MPUC inverter, two unequal DC sources (i.e. two PV panels with different power rating) are required to generate seven output voltage levels by means of reduced harmonic and uniformly distributed voltage levels. The maximum inverter output voltage level is sum of amplitudes of two dc sources. So, authors call it as Modified PUC Inverter (MPUC).

The seven-level PUC inverter is elaborated in section II. Subsequently, seven-level MPUC topology is described in section III. The rms value of seven level inverter output voltage for standalone operation is discussed in section IV. The results obtained from simulation of MPUC inverter has been given in section V. Conclusion is specified under section VI.

2. Seven-Level Packed U-Cell Inverter

The template is used to PUC inverter topology was first introduced by Al-Haddad in 2010 & later it was advanced by Hani Vahedi in 2015 [14]. PUC inverter is among the most promising multilevel inverters which takes advantages of both Flying Capacitor Converter (FCC) & CHB inverter. This topology is known as Packed U-Cell as the shape of each inverter unit is in U form. In Each U Cell, it contains two semiconductor switches & one DC voltage source. PUC topology work only with one DC source & a smaller number of active and passive components. Six semiconductor switches with anti-parallel diodes are used to achieve seven voltage levels which is evident from figure 1.

These 6 switches are divided into two legs having three switches in each leg. In one leg, there are three main switches (Q_1, Q_2 & Q_3) and in another leg, there are three complimentary switches of main switch ($Q_{1'}$, $Q_{2'}$ & $Q_{3'}$).

Various voltage levels are achieved either through DC voltage sources available or through their series opposition provided the amplitude of second DC source (V_2) is equal to one third of amplitude of first DC voltage source (V_1) i.e. $V_1 = 3V_2$. Thus, inverter output voltage has seven-voltage levels i.e. 0, $\pm V_2$, $\pm 2V_2$, and $\pm 3V_2$. From the Table I, it can be concluded that there are 3 positive voltage levels, 3 negative voltage levels and 2 zero voltage levels (viz. state 4 and 5). All eight switching states are shown graphically in Figure 2. Zero voltage level can be achieved only when the load is either shorted through main switches or through complimentary switches. Redundant zero voltage states are utilized by appropriate means just for reduction in switching frequency at the time when there are changeover in between positive & negative half cycles of AC signal.

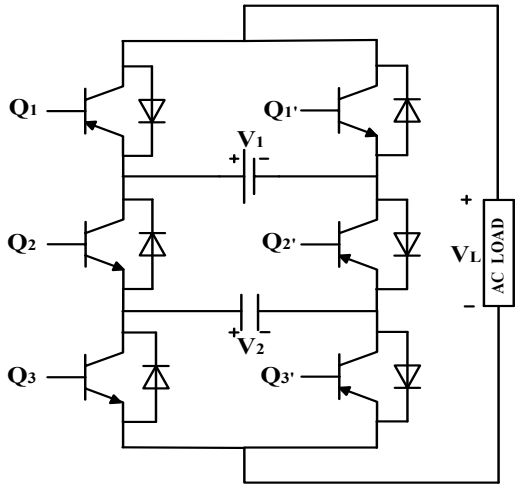


Figure 1: Packed U-Cell (PUC) Inverter Topology

Table 1: Complete Switching States & Voltage Levels

Switching States	Q_1	Q_2	Q_3	$Q_{1'}$	$Q_{2'}$	$Q_{3'}$	$V_{inv} = V_L$
1	1	0	0	0	1	1	V_1
2	1	0	1	0	1	0	$V_1 - V_2$
3	1	1	0	0	0	1	V_2
4	1	1	1	0	0	0	0
5	0	0	0	1	1	1	0
6	0	0	1	1	1	0	$-V_2$
7	0	1	0	1	0	1	$V_2 - V_1$
8	0	1	1	1	0	0	$-V_1$

From above switching states Table I, it can be concluded that irrespective of change in switching frequency and carrier frequency of adopted PWM technique, the switches Q2 & Q2' operate at line frequency (i.e. 50 Hz). This results in significant reduction of switching losses. Thus, the switches produce flow of current in the DC bus resulting in the use of a series diode for blocking the reverse voltage. To overcome this problem, an energy storage element i.e. Capacitor is used as a second DC source. The voltage across capacitor must be one third to that of V1 which can be achieved by various voltage balancing techniques. The current through the capacitor is in positive and negative parts which shows that there is charging and discharging in the capacitor [15]-[19]. Thereby achieving the desired voltage level for generating constant desired inverter output level. The biggest issue in using PUC inverter topology is its low maximum voltage level i.e. the maximum amplitude of the PUC inverter topology is maximum value of amplitude of the two voltage sources. Thus, the PUC inverter can't generate voltage more than the maximum value of the two voltage sources. So, it doesn't get application in medium or high-power applications [20]-[23].

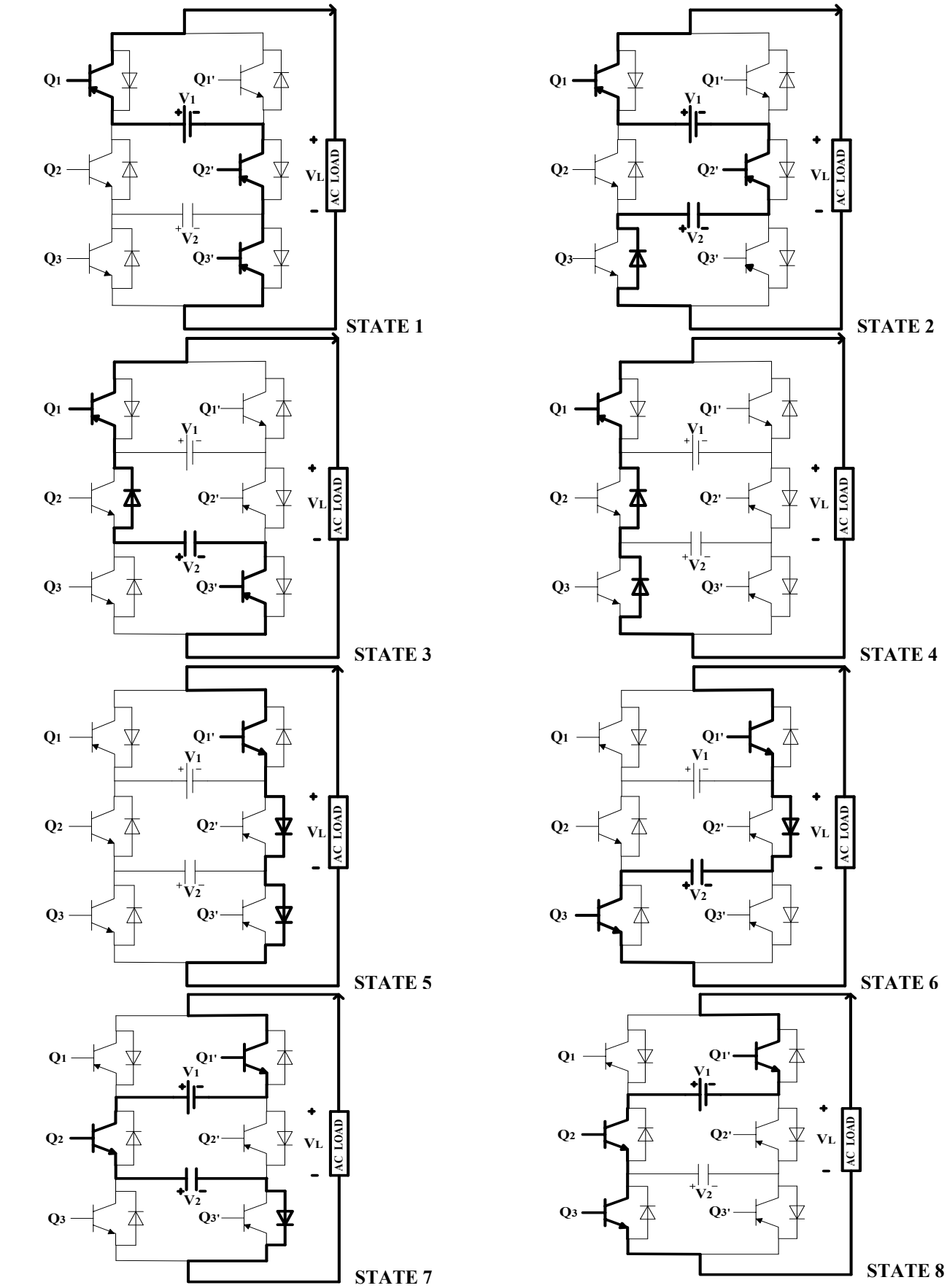


Figure 2: Switching States for PUC Inverter Topology

3. Seven-Level Modified Packed U-Cell Inverter

Due to recent advancement in solar panels and its economic cost, it becomes feasible to use number of PV panels as DC sources. Hence multilevel inverters using large number of DC sources are justified economically. The maximum output voltage in PUC inverter is not more than that of maximum DC voltage source magnitude. To solve this issue, the MPUC inverter topology is proposed as shown in figure 3. The DC source (V_2) is attached in reverse direction when compared to PUC inverter. So, the lower switches viz. Q_3 & $Q_{3'}$ have also been reversed to hinder current through diodes. Thus, with appropriate use of gate pulses, the path of current flow can be allowed or hindered. Various Switching states & output voltage levels for MPUC inverter is shown in Table II. From table II, it can be concluded that maximum inverter output voltage of MPUC inverter is nothing but total addition of two DC voltage sources i.e. V_1+V_2 which can be obtained by joining two voltage sources in series. Thus, we are getting more voltage amplitude by this proposed topology. It is evident that seven uniform voltage level is produced by MPUC inverter provided the second DC source (V_2) amplitude is two times the first DC source (V_1) amplitude i.e. $V_2 = 2V_1$. So, the output voltage levels have following levels: 0, $\pm V_1$, $\pm 2V_1$, $\pm 3V_1$. Thus, the maximum output voltage amplitude will be $3V_1$ while maximum DC source is equal to $2V_1$. The voltage level $3V_1$ is obtained by joining two DC voltage source V_1 & V_2 in series which appear on load. The voltage level $2V_1$ can be achieved by making second DC voltage source V_2 to be appear on the load. The voltage level V_1 can be achieved by making first voltage source V_1 to be appear on the load [24]-[26].

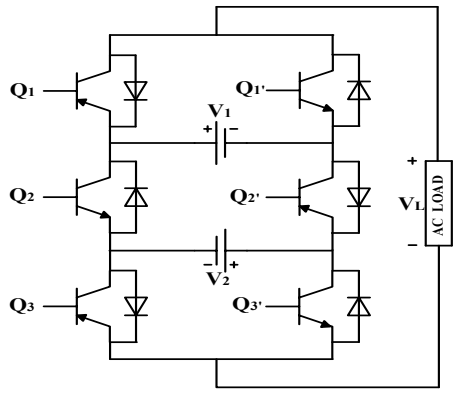


Figure 3: MPUC Inverter Topology

Table 2: Complete Switching States & Voltage Levels

Switching States	Q_1	Q_2	Q_3	$Q_{1'}$	$Q_{2'}$	$Q_{3'}$	$V_{inv} = V_L$
1	1	0	0	0	1	1	$V_1 + V_2$
2	1	0	1	0	1	0	V_2
3	1	1	0	0	0	1	V_1
4	1	1	1	0	0	0	0
5	0	0	0	1	1	1	0
6	0	0	1	1	1	0	$- V_1$
7	0	1	0	1	0	1	$- V_2$
8	0	1	1	1	0	0	$- V_1 - V_2$

Eight switching states are presented graphically in figure 4. Here are two redundant zero voltage states viz. state 4 and 5 which is used in the reduction of switching frequency. When zero voltage level is required to be generated at the output, the modulation technique selects the appropriate state in state 4 and 5 to give gate pulse to switches which gives less change in Turn-On & Turn-Off mode of operation resulting in significant reduction in switching frequency. From the switch states shown in table II, it can be concluded very clearly that irrespective of change in switching & carrier frequency, the two switches viz. Q_2 & $Q_{2'}$ operate at the line frequency i.e. 50 Hz.

4. Design Consideration of Standalone Operation

A mathematical analysis has been done for the standalone application of MPUC inverter supplying power to single-phase load. Here we aim to determine general expression for output voltage waveform of seven-level MPUC inverter based on figure 5.

The general expression for determining RMS voltage is given by:

$$V_r = \sqrt{\frac{1}{T}\int_0^T v_0^{\,2}(t)\,dt} \tag{1}$$

where Vr is RMS voltage, v_0 is time function of voltage and T is time period of the periodic function.

Thus equation (1) can be rewritten as follows:

$$V_r = \sqrt{\frac{1}{2\pi}\int_0^{2\pi} v_0^{\,2}(t)\,dt}$$

$$= \sqrt{\frac{1}{2\pi}\left[4\int_{\theta_1}^{\theta_2} V_1^{\,2}\,d\omega t + 4\int_{\theta_2}^{\theta_3} V_2^{\,2}\,d\omega t + 2\int_{\theta_3}^{\theta_4}\left(V_1+V_2\right)^2 d\omega t\right]}$$

$$= \sqrt{\frac{1}{2\pi}\left[4V_1^{\,2}\alpha + 4V_2^{\,2}\beta + 2\left(V_1+V_2\right)^2\gamma\right]} \tag{2}$$

Since $V_2 = 2V_1$

$$V_r = \sqrt{\frac{1}{2\pi}\left[4V_1^{\,2}\alpha + 4\left(2V_1\right)^2\beta + 2\left(3V_1\right)^2\gamma\right]}$$

$$= V_1\sqrt{\frac{2\alpha + 8\beta + 9\gamma}{\pi}} \tag{3}$$

where α , β and γ are angles at which reference signal crosses voltage levels as fixed lines between the carriers. The RMS value of output voltage can also be determined in term of modulation index (m_a) and maximum output voltage level ($V_{max} = 2V_1$) which is given as follows:

$$V_r = 0.725 \times m_a \times V_{max} \tag{4}$$

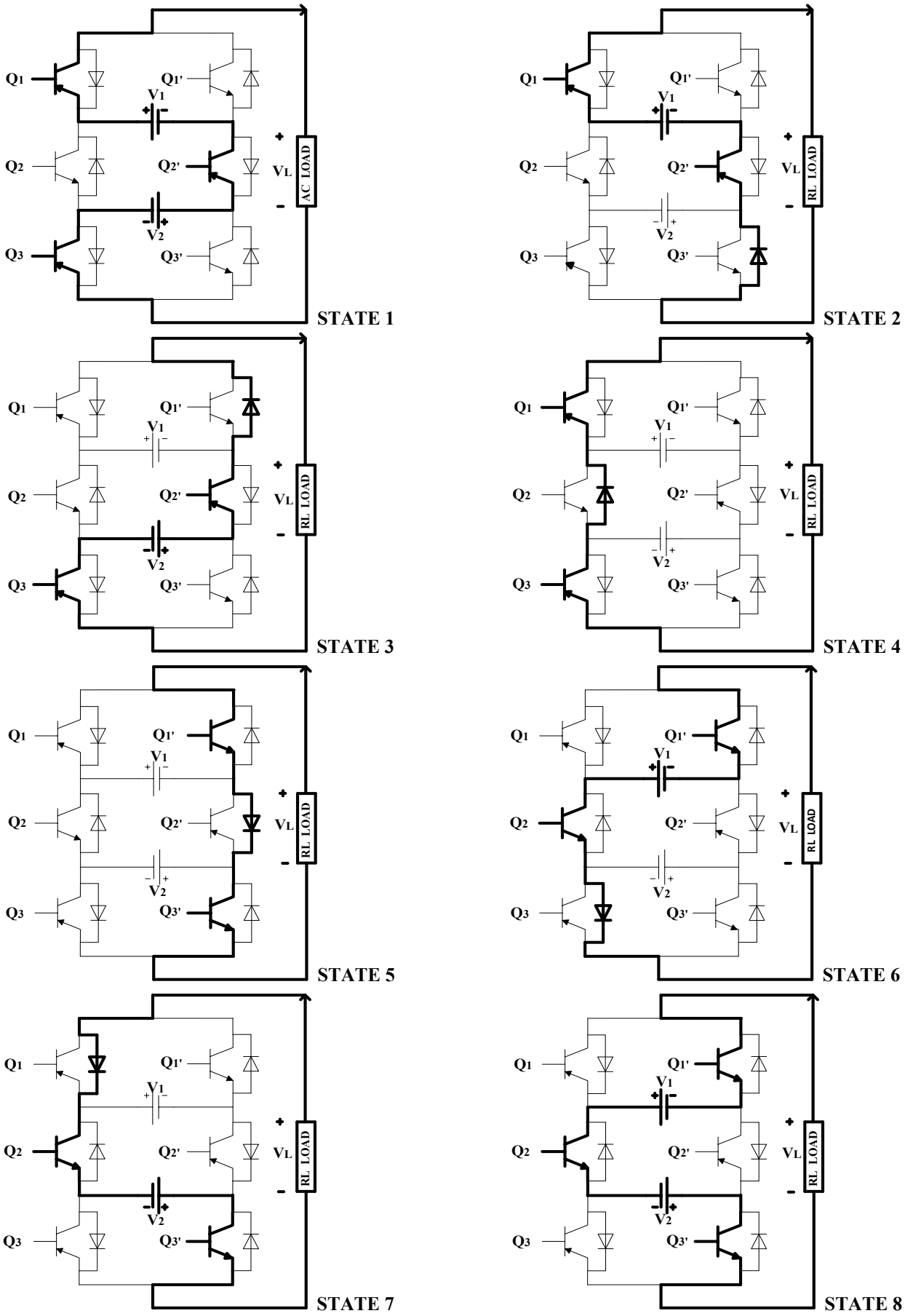


Figure 4. Switching States for MPUC Inverter

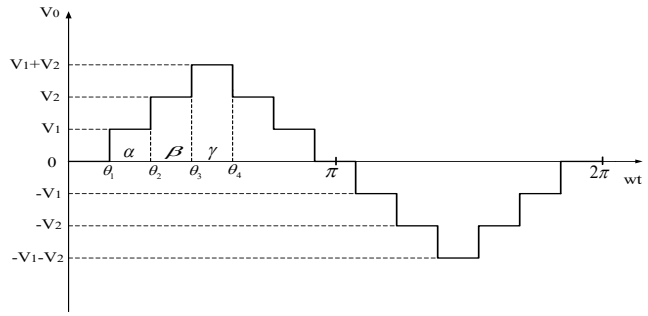


Figure 5: Seven-Level Inverter Output Voltage Waveform

5. Simulation & Results

The introduced MPUC inverter topology is simulated in standalone mode of operation to validate the above analysis. Six IGBT switches are taken in the simulation model. The switching algorithm adopted is elucidated in [16] where four level shifted carriers are used for modulating the reference signal. Thus, gate pulses are given to all the six pulses. Table III gives complete view of simulation parameters.

Here two DC sources viz. V_1 and V_2 are joined to different DC links of PV panels having different power & voltage rating and standalone load has been taken at the output. At higher switching frequency, passive components size reduces significantly.

Here the MPUC inverter is connected to RL load because most of the AC loads are inductive in nature. The parameters of RL load is specified into table III and results obtained from simulation has been given in figure 6.

The inverter output voltage waveform i.e. V_{inv} has seven uniform voltage levels having maximum of 120 Volt i.e. the total addition of two DC sources. The uniformly distributed & symmetric structure of inverter output voltage & current shows proper selection of modulation technique. THD of inverter output voltage waveform with no use of any filter is 23.96 % which is evident from figure 7. THD for output current waveform is 1.24 % which is evident from figure 8. Thus, the current THD is less than 5 % which is as per IEEE STD 519. Thus, the filtering requirement will be small and economical for MPUC inverter applications especially renewable energy.

For determining switching frequency of MPUC inverter topology, the inverter output voltage as well as gate pulses for switches Q_1, Q_2 & Q_3 has been shown for one time period of output voltage which is given in figure 9. Based on number of switching pulses for one cycle, we can determine the switching frequency of MPUC inverter switches which is given in table IV. Since each switch in one U cell operate in complimentary, so both the switches in a U cell will have same frequency. During switching time, the upper switches viz. Q_1 & Q_1, have to block voltage V_1 & lower switches viz. Q_3 & Q_3, have to block voltage V_2. So, the voltage rating of upper switches and lower switches are V_1 and V_2, respectively. The two middle switches viz. Q_2 & Q_2, have to suffer higher voltage than other switches i.e. V_1+V_2. So, the voltage rating of two middle switches are V_1+V_2.

The middle switches operate online frequency i.e. 50 Hz. So middle switches operating online frequency and high power are placed into middle U cell. Thus, a comparative study based on component counts is performed among popular Cascaded H-Bridge (CHB) inverter & MPUC inverter. The main limitation of MPUC inverter is that the unequal shading of Photovoltaic panels may cause deterioration of power balance between two DC links which may lose voltage control.

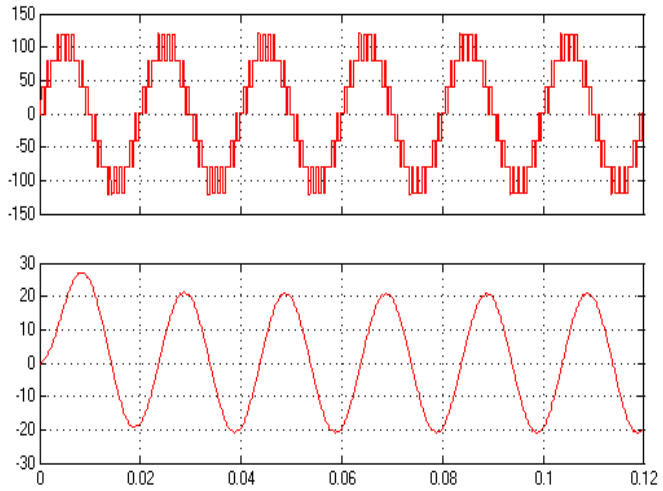


Figure 6: MPUC Inverter Output Voltage & Current Waveform

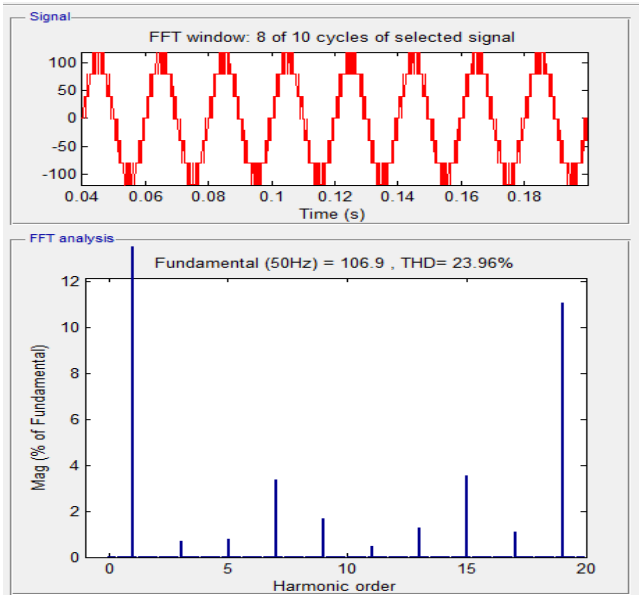


Figure 7: FFT Analysis for Inverter Output Voltage

Table 3: Parameters Used in Simulation

Parameters	Values
First DC Source (V_1)	40 Volt
Second DC Source (V_2)	80 Volt
Frequency (f)	50 Hz
Switching Frequency (f_{PWM})	1 KHz
Resistance (R)	2 ohm
Inductance (L)	15 mH

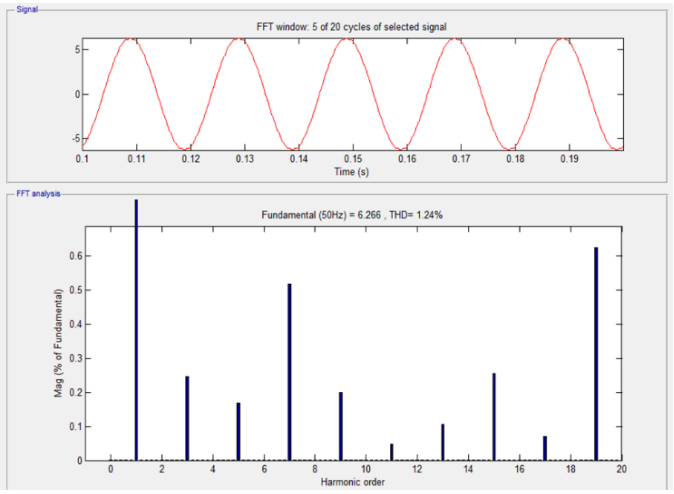


Figure 8: FFT Analysis for Inverter Output Current

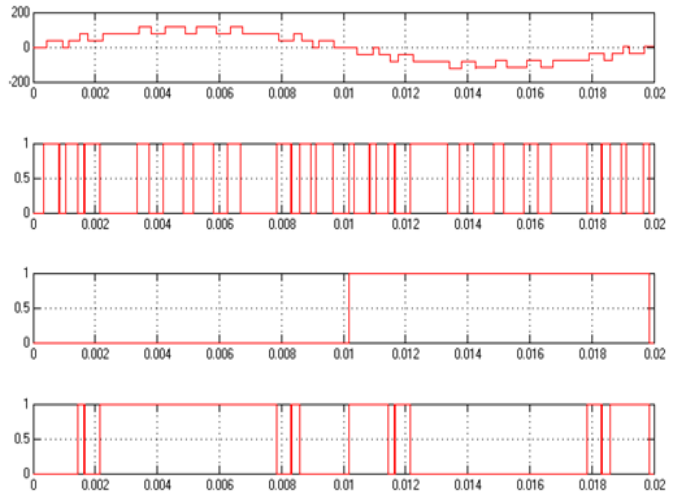


Figure 9: One Complete Cycle of (i) Inverter Output Voltage (ii) Q_1 Gate Pulse (iii) Q_2 Gate Pulse (iv) Q_3 Gate Pulse

Table 4: Switching Frequency of Inverter Switches

Q_1 & Q_1'	1250 Hz
Q_2 & Q_2'	50 Hz
Q_3 & Q_3'	500 Hz

Table 5: Comparative Study of Component Counts

Topology	Number of Switches	DC Sources
7-level CHB (Equal DC Sources)	12	3
7-level CHB (Unequal DC Sources)	8	2
7-level MPUC Inverter	6	2

6. Conclusion

In this paper, the modified form of Packed U-Cell inverter topology has been presented and validated on MATLAB Simulink. The proposed MPUC topology can generate seven levels of inverter output voltage with acceptable harmonic contents. Unlike PUC topology, the proposed topology is capable to produce maximum output voltage level which is more than the available maximum voltage source. In this topology, the DC source amplitudes gets added so that it can supply more power at the output. The switching algorithm adopted in MPUC inverter is such that it can operate at less switching frequency. Furthermore, this topology is targeted for PV applications for delivering power from different voltage rating PV panels. To show this, simulation results are given to verify THD for inverter output voltage & output current. Likewise, the comparison of MPUC inverter & CHB inverter proved the suitability of proposed inverter for advantages like fewer component counts, economical manufacturing cost and compact package owing to reduced filter size.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] D. Heydt, M.V. West, "Directions of esearch on Electric Power Quality," IEEE Trans. on Power Delivery, **8**(1), 429-436, 1993, doi: 10.1109/61.180365.

[2] S. Bakran, "Comparison of Modular Multilevel Converter based HV DC-DC-converters,"18th EPE'16 ECCE Europe, EISBN:978-9-0758-1524-5, 1-10, 2016 doi : 10.1109/EPE.2016.7695259.

[3] V. Al-Haddad, L. Rahmani, "Cascaded Multilevel Inverter with Multicarrier PWM technique and Voltage Balancing Feature," IEEE 23rd ISIE, Istanbul, 2155-2160, 2014, doi : 10.1109/ISIE.2014.6864951.

[4] Abolhasani, "Modular multipulse rectifier transformers in symmetrical cascaded H-bridge medium voltage drives," IEEE Tran. on PE, **27**(2), 698-705, 2012, doi: 10.1109/TPEL.2011.2161593.

[5] R. Bernet et al., "A Survey on Neutral-Point Clamped Inverters,"IEEE Trans. on IE, **57**(7), 2219-2230, 2010, doi: 10.1109/TIE.2009.2032430.

[6] S. Vahedi et al., "Modified Selective Harmonic Elimination Employed in Four-leg NPC Inverters," IEEE IECON, Dallas, TX, USA, Electronic ISBN:978-1-4799-4032-5, 2014, doi: 10.1109/IECON.2014.7049291.

[7] B. Singh, "Solar PV Array Fed Brushless DC Motor Driven Water Pump,"IEEE 6th ICPS, New Delhi, EISBN:978-1-5090-0128-6 2016, 1-5, doi: 10.1109/ICPES.2016.7584057.

[8] Enerdata, global energy statistical yearbook, 2015 [Online].

[9] L.B. Hiscock, "National survey report of pv power application in Canada," International Energy Agency,2013.

[10] J. Fahimi, "Multi-port power electric interface for renewable energy sources,"24th Annual IEEE Applied Power Electronics Conference and Exposition, Washington DC, 2009, doi: 10.1109/APEC.2009.4802680.

[11] V. Al-Haddad, "Multilevel electric power converter," US Patent 20110280052, 2011.

[12] V. Al-Haddad, "Real-Time Implementation of a Seven Level Packed U-Cell Inverter with a Low-Switching-Frequency Voltage Regulator," IEEE Trans. on Power Electr., **31**(8), 5967-5973, 2016, doi: 10.1109/TPEL.2015.2490221.

[13] V. Hani et al., "PUC converter review: Topology, Control and Applications," 41st Annual Conference of the IEEE Industrial Electronics Society, 4334-4339, 2015, doi: 10.1109/IECON.2015.7392774.

[14] J.I. Metri et al., " Real-Time Implementation of Model-Predictive Control on Seven-Level Packed U-Cell Inverter", IEEE Transactions on Industrial Electronics , **63**(7) ,4180 - 4186, 2016, doi: 10.1109/TIE.2016.2542133.

[15] S. Arazm et al., " Phase-shift modulation technique for 5-level packed U-cell (PUC5) inverter ", IEEE CPE-Powereng, 2018, doi: 10.1109/CPE.2018.8372577.

[16] V. Hani et al., "Modified seven level pack U-cell inverter for photovoltaic applications" IEEE Journal of Emerging and Selected Topics in Power Electronics, **6**(3), 1508-1516, 2018, doi: 10.1109/JESTPE.2018.2821663.

[17] S. Najafi, "Standalone operation of modified seven-level packed U-cell (MPUC) single-phase inverter" MDPI, **8**(3) , 2019, doi: 10.3390/electronics8030268.

[18] R. Vivekanand et al., "Improved Solar PV Cell Matlab Simulation Model and Performance Analysis," i-Manager's Journal on Electrical Engineering,

9(3), 12-19, 2016, doi: 10.26634/jee.9.3.4883

[19] S. Suneet et al., "A low cost power quality management tool for different load types,"IEEE 6th International Conference on Power Systems, Electronic 2016, doi: 10.1109/ICPES.2016.7584045.

[20] S. Deeksha et al., "Basic concepts of superconducting fault current limiter," IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems, ISBN:978-1-4673-8587-9,2016, doi: 10.1109/ICPEICES.2016.7853069.

[21] S. Bhandari et al., "Optimization Techniques In Modern Times & Their Applications", EISBN: 978-1-5386-2317-6, 6th IEEE iEECON2018 Krabi Thailand, 2018, doi: 10.1109/IEECON.2018.8712308.

[22] S.K. Bhushan et al., "Implementation of MPPT Technique in PV Array for a Varying Load by Modeling and Simulation,"International Electrical Engineering Congress (iEECON), 2018, doi: 10.1109/IEECON.2018.8712313.

[23] P. Shubham, S.K. Bhushan, "Modeling of Solar Energy Grid Integration System Using Typhoon HIL," International Electrical Engineering Congress (iEECON), 2018, doi: 10.1109/IEECON.2018.8712253 .

[24] S.K. Bhushan, M. Goel. "Grid Integrated Solar Photovoltaic Array Power Plant Modeling and Simulation," International Conference and Utility Exhibition on Green Energy for Sustainable Development, EISBN:978-9-7482-5799-0, 2018, doi: 10.23919/ICUE-GESD.2018.8635688.

[25] S.K. Bhushan et al., "Modeling and Simulation of Grid Connected 1 MW Solar PV Power Plant." 2018 International Conference and Utility Exhibition on Green Energy for Sustainable Development (ICUE). IEEE, 2018, doi: 10.23919/ICUE-GESD.2018.8635745.

[26] R. Vivekanand, K. Sahay. "Power Quality Improvement of Grid Connected PV System Using Dynamic Voltage Restorer," i-Manager's Journal on Power Systems Engineering, **4**(1), 5-13, 2016, doi: 10.26634/jps.4.1.5966.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 967-971 (2020)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Hand Gesture Classification using Inaudible Sound with Ensemble Method

Jinwon Cheon[1], Sunwoong Choi[*,2]

[1]*Department of Security Enhanced Smart Air Mobility, Kookmin University, Seoul, 02707, South Korea*

[2]*School of Electrical Engineering, Kookmin University, Seoul, 02707, South Korea*

ARTICLE INFO

Article history:
Received: 07 September, 2020
Accepted: 02 November, 2020
Online: 08 December, 2020

Keywords:
Hand gesture classification
Short-Time Fourier Transform (STFT)
Ensemble

ABSTRACT

Recognizing the human behavior and gesture has become important due to the increasing use of wearable devices. This study classifies hand gestures by creating sound in the inaudible frequency range from a smartphone and analyzing the reflected signals. We convert the sound using Short-Time Fourier Transform to magnitude and phase. We trained two types of data on Convolutional Neural Network model. And then we propose a method applying soft voting, an ensemble technique, to improve classification accuracy taking the average of two models' result. In this paper, the classification accuracy of the Mag model is 96.0% and the classification accuracy of the Phase model is 90.0% for 8 hand gestures. While the ensemble model showed 96.88%, which is better than Mag and Phase models.

1 Introduction

With the advancement of IT technology, the use of wearable devices such as smartwatches and IoT-based devices is becoming more popular. However, as devices become smaller for portability, there are some limitations, including the difficulty of controlling these devices using buttons or touches. To date, lots of studies are being conducted to overcome this limitation. Most studies collect and recognize data from sensors like cameras and controllers [1]–[4]. However, it is a fatal flaw that requires a sensor or product to recognize movement. Therefore, in this study, we propose a hand gesture classification using a smartphone without a separate sensor.

When the smartphone's speaker makes an inaudible sound, the microphone records this sound. A subject performs a hand gesture while recording is in progress, the signal hit by the subject's hand and reflected changes in frequency domain because of the Doppler effect. There are several methods to analyze a signal into time and frequency domain, among them, we analyzed the signal using the Short-Time Fourier Transform (STFT) [5]. This is because STFT converts the data into a 2D grid-like topology with time and frequency domains. It can be easy to input in 2D Convolutional Neural Network (CNN) [6]. In this way, the sound signal is transformed into magnitude and phase. We propose a method to increase the classification accuracy of the independently trained magnitude and phase CNN models by applying soft voting, an ensemble technique. As a result, the proposed method showed a classification accuracy of 96.88% for 8 hand gestures.

The rest of this paper is organized as follows. It introduces the hand recognition and classification research using ultrasound and ensemble technique in Chapter 2, describes the method proposed in Chapter 3, evaluates its performance in Chapter 4, and finally concludes this paper in Chapter 5.

2 Related Work

In this chapter, we briefly introduce how hand gesture recognition and classification are being used, and studies using ultrasonic signals and studies to improve the performance of a model using multiple CNN models are described.

First, the current status of research on hand gesture recognition and classification is introduced [7]. Data for hand gesture recognition and classification are collected from various sensors, for example, mount-based sensors, multi-touch screen sensors, vision-based sensors, and ultrasonic-based sensors. The collected data are trained on machine learning models used in various fields, such as human-robot interaction, virtual manipulation, sign language, and gaming.

Second, studies on recognizing or classifying hand motions using ultrasonic signals will be described. There is a study called

[*]Corresponding Author: Sunwoong Choi, Kookmin University, School of Electrical Engineering, 82-2-910-4416 & schoi@kookmin.ac.kr

FingerIO that uses smartphone and smartwatch to find fingers using the microphone and speaker [8]. In this study, they use sound in a inaudible frequency range and track the fingers' position by Orthogonal Frequency Division Multiplexing (OFDM). Other study proposes a deviceless gesture tracking scheme. This is called LLAP (Low-Latency Acoustic Phase) [9]. In this study, they use microphones and speakers that built-in on mobile devices to track hands and fingers. This study measures the phase changes that is formed by movement and converts the changes into the distance of the movement. There is a paper that classifies hand gestures using two smartphones [10]. One used as a microphone, while the other used as a speaker. In this paper, they used STFT. These window size is 500, and overlapped size is 475. The frequency resolution is 2,048. As a result, their proposed CNN model classified 8 hand gestures with a classification accuracy of 87.75%.

Finally, studies using multiple CNN models will be described. These two papers [11, 12] explained ensemble method. In paper [12], the performance of the model for classifying hand gestures was improved using the ensemble technique. They ensembled three types of classifiers: a polynomial classifier, a multi-layer perceptron, and a support vector regression. So, they obtained median recognition rates per moment in time of more than 86%. Another paper [13] is a paper that improves performance by fusion of GoogLeNet, VGGNet, and ResNet. The performance was improved by extracting feature layers (100-dimensional and 40-dimensional fully connected and softmax) to each model and attaching them. This paper implemented a model that recognizes the motions of a person in an image, and performance improved in order when only one of the three models mentioned above was used, when two were used, and when all three were used.

3 Proposed Method

Our system can be divided into three steps (data acquisition, data pre-processing and model training) like Figure 1. First, we collect sound data using our own application. Then, the data are pre-processed by STFT on MATLAB. Lastly, We train the pre-processed data on the proposed CNN model.

Figure 1: Overall system architecture.

3.1 Data Acquisition

We collected data with a smartphone application developed in Android Studio. This application produces a single frequency of 20 kHz on the smartphone's speaker. At the same time, the application starts recording on the smartphone's microphone. The speaker and microphone are activated for 3 seconds, and a subject makes a hand gesture within that time at a distance of 1–5 cm from the smartphone. The application stores the sound reflected by gestures. When collecting data, only one subject in the office performing hand gestures to reduce the noise.

3.2 Data Pre-processing

We have used STFT to convert sound data because STFT informs the change of frequency over time. A window size of 5,000, an overlap size of 4,750, and a frequency resolution of 2,048 are detailed information of STFT we used. The microphone recorded for 3 seconds, but the speaker only produced up to 2 seconds. So, we cut off the last second of the data. We also trimmed the data around 20 kHz to suppress the background noise and reduce the number of data. As shown in the Figure 2, the raw data graph represents amplitude value only, however the STFT spectrogram represents the signal intensity using color. STFT data contain magnitude and phase parts. Figure 3 shows magnitude of STFT data, and Figure 4 shows phase of STFT data. So, we pre-processed the sound data to magnitude and phase data.

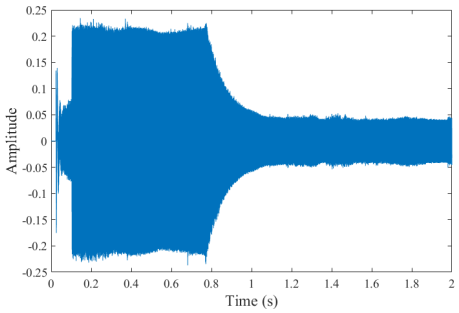


Figure 2: Recorded raw data amplitude graph.

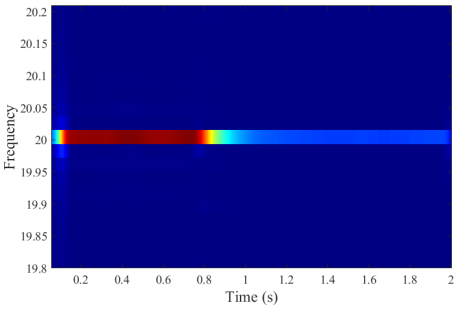


Figure 3: STFT spectrogram (magnitude).

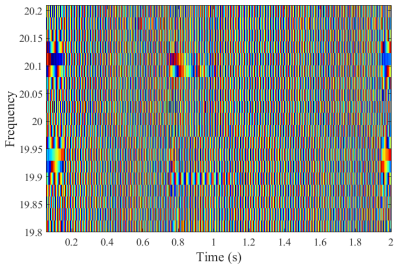


Figure 4: STFT spectrogram (phase).

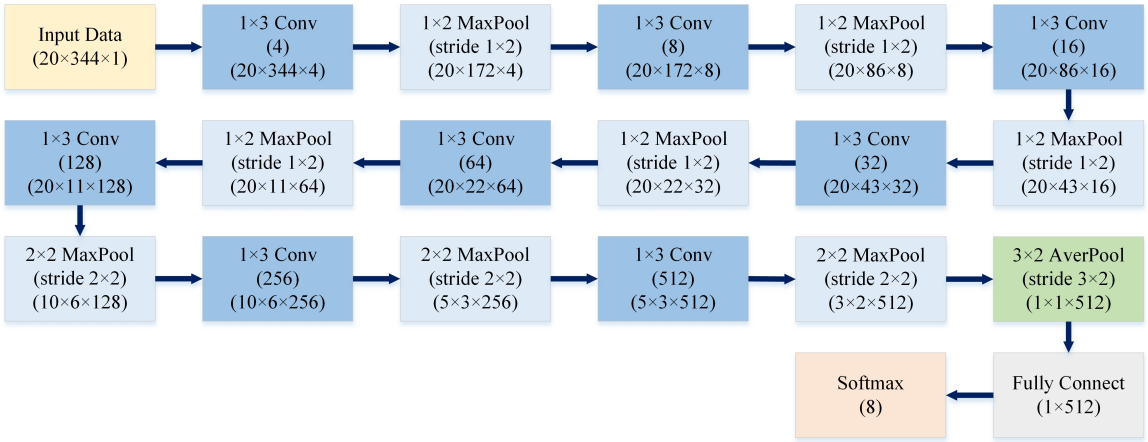


Figure 5: The structure of Mag and Phase models.

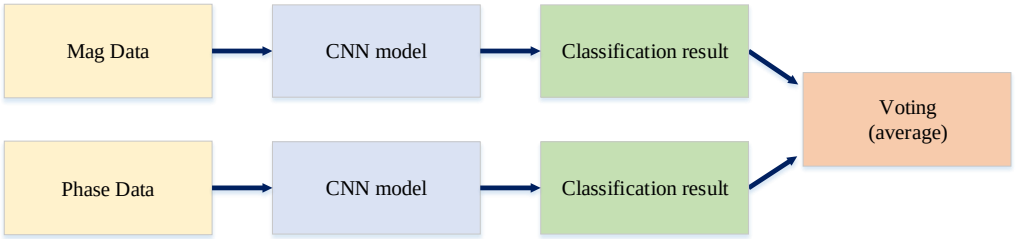


Figure 6: The structure of ensemble model.

3.3 Model Training

We propose two CNN models which called Magnitude (Mag) model and Phase model. Mag and Phase models have the same architecture. The models trained with one part of pre-processed data (magnitude or phase). The data were trimmed to the specific frequency range of 19.8–20.2 kHz through STFT. The models are largely composed of 5 parts: convolution layers, max pool layers, an average pool layer, a fully connected layer and a softmax function. There are 8 pairs of a convolutional layer and a max pool layer in the models to extract data feature. The filter size of the convolutional layer is 1×3, and the filter size of the max pool layer is 1×2 and 2×2. The output of pairs of the convolutional layer and the max pool layer passes through an average pool layer of 3×2 and then through a fully connected layer. Figure 5 shows the structure of our CNN models.

After that, the models were ensembled by taking the average of the softmax function outputs (soft voting). When Mag and Phase models train, the ensemble model not affected training but is only used when testing. Two results of softmax function are calculated by inputting magnitude data and phase data for one hand gesture into each model. The size of the softmax result of each model is the same as the number of classes, and a new softmax result is created by averaging two softmax results corresponding to each hand gesture. The operation is as (1). In this equation, S means the result of softmax funcion.

$$S^{ensemble} = \frac{S^{mag} + S^{phase}}{2} \quad (1)$$

Figure 6 shows how to progress the ensemble model. The magnitude part of the STFT is input to the Mag model and the phase part is input to the Phase model to calculate the average of the softmax matrix outputs from each model. The class that has the largest value among the values of the average matrix is the predicted class of the ensemble model.

4 Evaluation

4.1 Experiment Setup

In the experiment, we have collected the data to TG&Co.'s LUNA smartphone in an office where only one subject performed hand gestures. The application was developed in the Android Studio environment. The UI of developed application is as Figure 7. The smartphone was placed on a flat desk and recording started when the REC-START button was pressed. The application activated the speaker to produce a single frequency at 20 kHz and the microphone to start recording. After 3 seconds, recording stopped automatically.

The saved sound data were transferred from the smartphone to the computer. The data were converted to STFT on MATLAB R2019b. The models were implemented using the Tensorflow 1.6.0, and Python 3.4.8 for running the Tensorflow. Training and test were performed on a GPU server which operates CentOS 7.5 and uses Intel Xeon Silver 4114 CPU. The learning progressed about 4,000 epochs. Besides, a 5-fold cross validation technique was used to increase the reliability of the models.

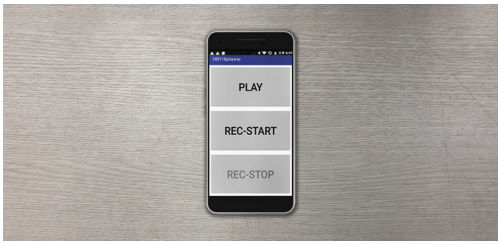


Figure 7: UI of developed application.

4.2 Gesture Dataset

In this study, 8 hand gestures were classified. Figure 8 shows 8 hand gestures that were used as a dataset. The subject gestured at a distance of 1–5 cm from the phone without touching the screen. This action was performed within 3 seconds and repeated 100 times per gesture. Through this process, we have collected a total of 800 data. The data were divided into training data and test data at a ratio of 8:2.

- Do nothing : Do not move the hand while recording.
- Block : Block the microphone with palm while recording.
- Move : With the palm open, record perpendicular to the screen. Place the hand at the beginning of the arrow and move it in the direction of the arrow while recording (to right, to bottom, to left, to top).
- Draw : Stretch only index finger and draw on the screen once in the direction of the arrow (circle and triangle).

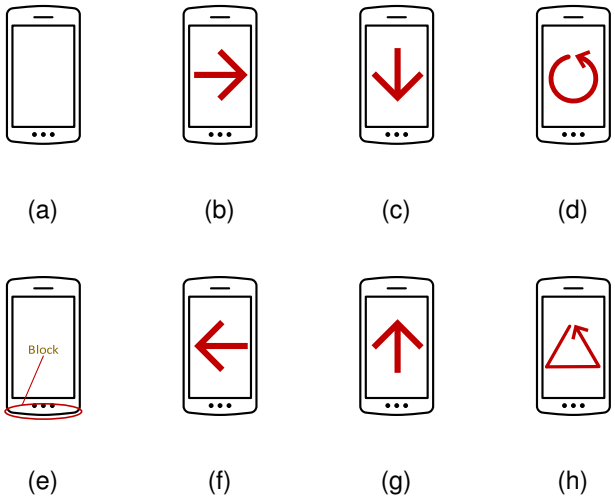


Figure 8: Used hand gestures.

4.3 Experiment Result

When performing STFT, the STFT result value also changes depending on the window size, so that we evaluated this. We tested the classification accuracy by increasing the window size from 2,000 to 8,000 in steps of 1000 using only magnitude data. In all window sizes, the frequency resolution and overlap size were unified to 2,048 and 95% during the evaluation. The classification accuracy by window size is shown in Figure 9. Therefore, the evaluations of the Phase and ensemble model were proceeded when window size is 5,000.

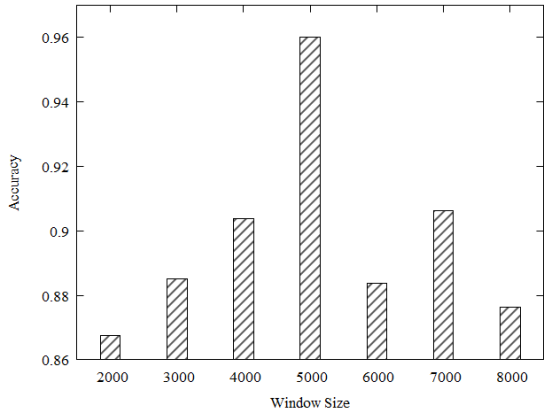


Figure 9: Classification accuracy by window size.

In conclusion, we presented the ensemble model that classified 8 hand gestures as 96.88% classification accuracy. We estimated Mag, Phase, and the ensemble model when window size is 5,000. Mag model showed 96.0% classification accuracy, and Phase model showed 90.0%, slightly lower than Mag model. However, when these two models were ensembled with our method, the performance improved, showing a classification accuracy of 96.88%. Not only classification accuracy but also various evaluation indicators were tested. Figure 10 shows precision, recall, F1-score, and accuracy of each model. It can be seen that the ensemble model better than the two models in all evaluation indicators.

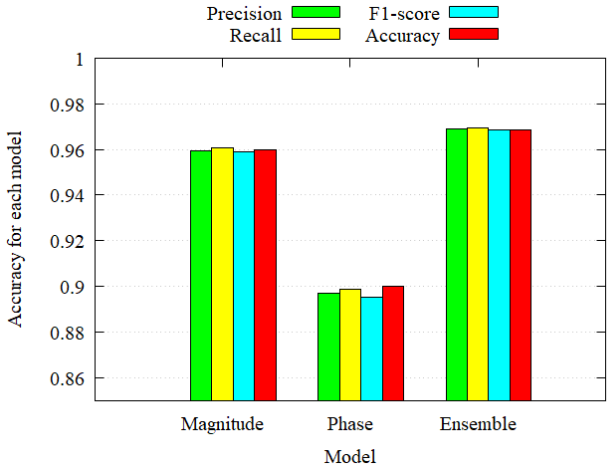


Figure 10: The performance comparison of each model.

We plotted the confusion matrices to also check the accuracy corresponding to each class. The confusion matrices are shown in figures 11– 13. Comparing the confusion matrices, it can be seen that the Mag and Phase models differ in the class that predicts well. For example, in Mag model, 'do nothing' gesture shows the highest accuracy, but not in Phase model. Since the accuracy of Mag model

is higher than that of Phase model, the ensemble model shows a similar pattern to Mag model. Also, almost classes of the ensemble model show better accuracy than Mag and Phase models.

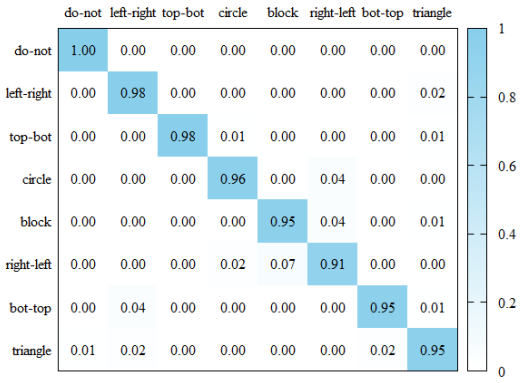


Figure 11: The confusion matrix of Mag model.

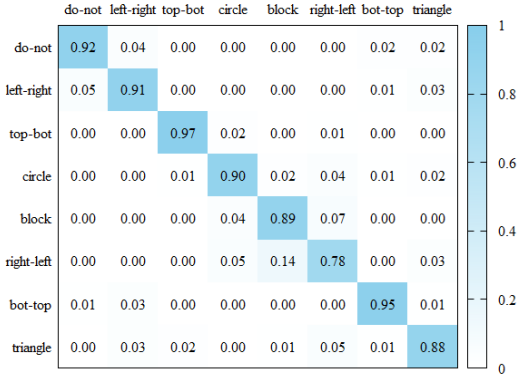


Figure 12: The confusion matrix of Phase model.

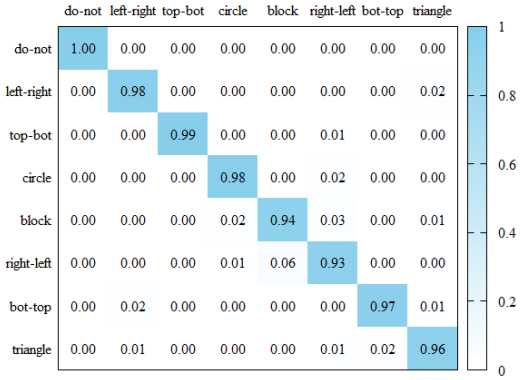


Figure 13: The confusion matrix of the ensemble model.

5 Conclusion

This study classified hand gestures using inaudible sound using a smartphone's speaker and microphone. We used our own application to produce and record 20 kHz sound while gesturing. The sound data were converted to STFT as magnitude and phase data. These data used for training and test in our CNN models. We proposed Mag and Phase models, and also suggested ensemble method using soft voting to improve classification accuracy. As a result, our model showed 96.88% classification accuracy for 8 hand gestures.

Acknowledgment This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIP) (No.2016R1A5A1012966).

References

[1] G. Zhu, L. Zhang, P. Shen, J. Song, "Multimodal gesture recognition using 3-D convolution and convolutional LSTM," IEEE Access, **5**, 4517–4524, 2017, doi:10.1109/ACCESS.2017.2684186.

[2] W. Lu, Z. Tong, J. Chu, "Dynamic hand gesture recognition with Leap Motion controller," IEEE Signal Processing Letters, **23**(9), 1188–1192, 2016, doi:10.1109/LSP.2016.2590470.

[3] T. Schlömer, B. Poppinga, N. Henze, S. Boll, "Gesture recognition with a Wii controller," in Proceedings of the 2nd International Conference on Tangible and Embedded Interaction, 11–14, 2008, doi:10.1145/1347390.1347395.

[4] M. Sathiyanarayanan, S. Rajan, "MYO armband for physiotherapy healthcare: A case study using gesture recognition application," in 2016 8th International Conference on Communication Systems and Networks (COMSNETS), 1–6, 2016, doi:10.1109/COMSNETS.2016.7439933.

[5] R. Schafer, L. Rabiner, "Design and simulation of a speech analysis-synthesis system based on short-time Fourier analysis," IEEE Transactions on Audio and Electroacoustics, **21**(3), 165–174, 1973, doi:10.1109/TAU.1973.1162474.

[6] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," Nature, **521**(7553), 436–444, 2015, doi:10.1038/nature14539.

[7] H. Cheng, L. Yang, Z. Liu, "Survey on 3D hand gesture recognition," IEEE Transactions on Circuits and Systems for Video Technology, **26**(9), 1659–1673, 2016, doi:10.1109/TCSVT.2015.2469551.

[8] R. Nandakumar, V. Iyer, D. Tan, S. Gollakota, "FingerIO: Using active sonar for fine-grained finger tracking," in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 1515–1525, 2016, doi:10.1145/2858036.2858580.

[9] W. Wang, A. X. Liu, K. Sun, "Device-free gesture tracking using acoustic signals," in Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, 82–94, 2016, doi:10.1145/2973750.2973764.

[10] J. Kim, S. Choi, "Hand gesture classification based on nonaudible sound using convolutional neural network," Journal of Sensors, **2019**, 1084841, 2019, doi:10.1155/2019/1084841.

[11] T. G. Dietterich, "Ensemble methods in machine learning," in Multiple Classifier Systems, 1–15, 2000, doi:10.1007/3-540-45014-9_1.

[12] J. Schumacher, D. Sakič, A. Grumpe, G. A. Fink, C. Wöhler, "Active learning of ensemble classifiers for gesture recognition," in Pattern Recognition, 498–507, 2012, doi:10.1007/978-3-642-32717-9_50.

[13] Y. Lavinia, H. H. Vo, A. Verma, "Fusion based deep cnn for improved large-scale image action recognition," in 2016 IEEE International Symposium on Multimedia (ISM), 609–614, 2016, doi:10.1109/ISM.2016.0131.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 972-980 (2020)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Docker-C2A : Cost-Aware Autoscaler of Docker Containers for Microservices-based Applications

Mohamed Hedi Fourati[*,1], Soumaya Marzouk[1], Mohamed Jmaiel[1,2], Tom Guérout[3]

[1] *ReDCAD Laboratory, National Engineering School of Sfax, University of Sfax, Sfax, 3038, Tunisia*

[2] *Digital Research Center of Sfax, University of Sfax, Sfax, 3021, Tunisia*

[3] *Laboratory for Analysis and Architecture of Systems, University of Toulouse, Toulouse, 31031, France*

ARTICLE INFO

Article history:
Received: 14 September, 2020
Accepted: 26 November, 2020
Online: 08 December, 2020

Keywords:
Cloud Computing
Microservices
Docker
Kubernetes
Autoscaling
PSO algorithm

ABSTRACT

*This article proposes a cost-aware autoscaler for microservices-based applications deployed with docker containers. This autoscaler decreases the cost of the application deployment as it reduces computing resources. In elastic treatment, microservice resources are scaled when the used metric as the central processing unit (CPU) exceeds the threshold. In case of threshold exceeding, an autoscaler adds many instances of docker containers in order to satisfy the need of the application. In many studies, the autoscaler adds many containers without selecting the appropriate microservices for scaling and without in advance calculation of the adequate number of containers. This may lead to allocating additional resources to inappropriate microservices and a non optimal number of containers. For this reason, we propose our autoscaler "*Docker-C2A*" which identifies the adequate microservices to add resources. It also calculates the optimal number of needed containers. "*Docker-C2A*" analyses the state of the application, uses the execution history and uses a Particle Swarm Optimization (PSO) algorithm to identify the adequate microservices for scaling resources and to determine the optimal number of containers. As a result, "*Docker-C2A*" helps to reduce computing resources and to save extra costs. Experimental measurements were conducted on a microservices-based application as a concrete use-case demonstrating the effectiveness of our proposed solution.*

1 Introduction

Microservices is an architectural style of development consisting of a collection of small independent and loosely coupled components [2]. Nowadays, several companies like Amazon, Linkedin, Spotify and Netflix have migrated towards microservice architecture in order to increase the efficiency and the scalability of computing resources.

Currently, the most of microservices-based applications are deployed in docker containers [3] and orchestrated with kubernetes tool [4]. Docker is a container technology designed for creating, deploying and running applications using containers. The latter allow developers to package up an application with all needed parts such as libraries and deploy it as one package. Each instance of a microservice is a docker container. They are orchestrated and deployed in a cluster of VMs (Virtual Machines) configured by kubernetes, an open-source container orchestration system created by Google for automation of application deployment, scaling and management.

When there is a rise on the workload, microservices resources are overloaded and thus additional resources should be allocated. One of the most important stakes in Cloud environment is to minimize computing resources so as to reduce the deployment cost of the application. We can resolve this issue and optimize computing resources using autoscaling techniques. Over-provisioning of computing resources leads to allocating unused resources. Whereas, under-provisioning causes performance degradation of the application. Thus, an autoscaling technique should identify the needed resources for the application in order to allocate quasi-exact needed computing resources.

In literature, few autoscalers are proposed for containers and microservices-based applications, and most existing solutions have the same treatment as kubernetes autoscaler. The autoscaler of kubernetes is called Horizontal Pod Autoscaler HPA [5]. HPA scales the number of containers based on CPU usage. CPU threshold is

[*]Corresponding Author: Mohamed Hedi Fourati, ReDCAD Laboratory Sfax, Tunisia. Email : mohamed-hedi.fourati@redcad.org
This paper is an extension of work originally presented in [1].

specified by the user. When the application receives high workload due to the increase of requests number, many microservices are overloaded and the CPU threshold is exceeded. Consequently, kubernetes HPA adds one container instance to each overloaded microservice. If the threshold is still exceeded, the autoscaler continues adding other containers until reaching the normal state of the CPU or the maximum number of containers. This autoscaling process has several issues that allocate many resources and thus charge a high deployment cost.

The first issue is the usage of CPU metric. In fact, using CPU metric in autoscaling is not effective and may lead to allocating unnecessary resources. Indeed, it was proved in [6] that a high CPU usage means that the container instance is fully utilized, but it can still provide acceptable response time without adding more resources. So, it is possible to execute further requests and to have acceptable response time even if we have high CPU and high workload. This is also why many elastic solutions tend to use response time metric instead of CPU usage to benchmark their autoscaler. Moreover, it is highly recommended to use response time metric instead of resource metrics as CPU when launching autoscaling actions.

The second issue is that existing autoscalers do not select the adequate microservices for adding resources. In fact, the microservices of an application are in relationship, and an issue in a given microservice may be propagated to other ones. Similarly, if this issue is resolved for a microservice, it is automatically resolved for related microservices. Thus, an autoscaler should select and find the eligible microservices for scaling. To do that, the autoscaler should analyze the whole application with an overview of all microservices. Also, it should use the execution history of the application to find eligible microservices.

The third issue is that existing autoscalers do not calculate the needed number of containers for each microservice and do not react fast. For example, the autoscaler of kubernetes adds containers one by one until reaching the normal state. In this case, the normal state is recovered after a period of time, while in Cloud applications, the time constraint is critical and fast scaling actions are requisite. Then, the autoscaler should calculate and estimate the needed number of containers for each microservice from the beginning to launch fast actions. To do this, the autoscaler should analyze the application and its microservices to estimate the adequate number of containers for each microservice. Moreover, it should consider the execution history and the architecture of the application as the relationship between microservices to estimate the adequate number of containers. As a result, the autoscaler calculates precisely the needed number of container and launches fast actions. This leads to optimize computing resources and deployment cost.

This paper proposes "Docker-C2A" a **C**ost **A**ware **A**utoscaler for microservices-based applications deployed with docker containers. This autoscaler focuses on existing issues in order to optimize computing resources and reduce deployment costs. Firstly, "Docker-C2A" is based on response time metric instead of CPU metric. Also, it selects eligible microservices for scaling and calculates the optimal number of containers to be added to each microservice. To do this, during autoscaling treatment, "Docker-C2A" considers the architecture of the application and the relationship between microservices. In addition, it uses the execution history of the application and executes a PSO algorithm. As a result, "Docker-C2A" selects appropriate microservices for scaling and calculates the needed number of containers for each microservice. Experimental results conducted on a concrete use case demonstrates that we can satisfy high workload with minimum computing resources by allocating a small number of containers. Thus, "Docker-C2A" optimizes computing resources and reduces the deployment cost. And that is the main purpose of this autoscaler.

This paper is organized as follows: Section 2 presents the related work. Section 3 presents the motivating example of our autoscaler. Section 4 presents in details Docker-C2A autoscaler. Section 5 details the PSO algorithm in Docker-C2A. Section 6 discusses the experimentations. We conclude the paper in Section 7.

2 Related Work

Autoscaling is dedicated for optimizing resource allocation in order to avoid allocating unused resources and to reduce the cost of deployment. Few studies were proposed for the autoscaling of microservices-based applications focusing on containers. As it was explained in section 1, these solutions have many issues. In fact, most of these solutions focus on CPU metric. Also, the resource allocation is not accurate as many resources are allocated without selecting the adequate microservices for scaling and without previous calculating the needed amount of resources. This may lead to allocating unused and unnecessary resources and thus charging the user extra costs. In this section we review these studies and explain the shortcomings of each one.

In [7], the author designed a container-based autoscaling policy with a mathematical model dedicated for Iot applications based on containers and microservices. This autoscaler uses two main metrics, number of requests and resource related metrics as CPU or memory usage. Also, it uses the history of the application to take adequate actions for similar cases recorded in the history. After each scaling action, the autoscaler waits for a cooling period defined by the user to avoid launching other scaling actions until the application becomes stable. The main objective of this autoscaler is to minimize computing resources and the deployment costs. Using the history, this autoscaler calculates and estimates the optimal number of containers used to scale each microservice. Although this solution estimates the needed resources for allocation, it does not select the eligible components for scaling.

In [8], the author presented a message queue as a specific use case in IoT context. They considered metrics related to message queue mi- croservices. Two classes of microservices were considered, compute-intensive and I/O-intensive. In [9], the author aimed to adjust computing resources to the incoming load. In scaling process, a new container is added when the response time exceeds 1000 ms in a 15 s window. A container can be removed if it uses less than 10% of cpu. These two studies neither select containers and components for scaling nor calculate the adequate quantity of resources to be allocated.

In [10], the author proposed a genetic algorithm for resources allocation and elasticity management dedicated for microservices-based applications deployed on a containerized environment. The main objectives of the proposed solution is the container allocation

in VMs and elastic treatment of containers. This study models the whole system such as applications, microservices, the relationship between microservices, the required resources of each microservice, containers, the needed resources, physical machines and the network. Then, it optimizes this model using a meta heuristic genetic algorithm to get optimal results for resource allocation of containers and the number of containers for each microservice. This autoscaler considers the architecture of the application and the relationship between microservices in scaling process. However, it does not calculate the needed amount of resources for scaling.

In [11], the researchers developed an autoscaler called Microscaler for microservices-based applications which identifies the scaling-needed services using a customized metric called Service Power. When detecting a response time violation, Microscaler launches the process of detecting services that should be scaled using the customized metric Service Power. This metric is based on response time and used to identify the microservices that need to be scaled with additional resources. Besides, this autoscaler calculates the adequate number of containers to be added for each selected microservice. This autoscaler uses a Bayesian Optimization (BO) and an heuristic approach to determine the optimal number of instances for selected services. Although this autoscaler selects the appropriate microservices to be scaled and calculates the number of additional containers, it is strongly related to applications based on service mesh that make this approach not applicable in other contexts.

In this paper, we propose "Docker-C2A" autoscaler for microservices-based applications using docker containers. We focus on docker container technology because it is the most used in the context of containerization and microservices-based applications. This autoscaler is based on response time metric. Also, "Docker-C2A" selects eligible microservices for scaling and calculates the optimal amount of resources to be allocated. This autoscaler uses the execution history of the application and a PSO heuristic algorithm. The execution history helps to determine which of application microservices need to be scaled. Then, the PSO algorithm calculates the optimal number of containers to be added to each selected microservice.

"Docker-C2A" can be distinguished over existing solutions as it allocates the optimal amount of resources to the adequate microservices. "Docker-C2A" is a fast and accurate autoscaler that gives more precise scaling actions. This leads to optimizing application resources and deployment costs.

3 Motivating Example

To illustrate existing issues, we detail the behavior of kubernetes autoscaler on a concrete microservices-based application. We take the example of kubernetes autoscaler as most existing autoscalers have the same treatment. In fact, we consider Bookinfo [12] a microservices-based application. Figure 1 illustrates Bookinfo application composed of four microservices.

This application displays the data related to each book. It is similar to a unique catalog entry of an online book store. It displays book details as ISBN, number of pages, and book reviews. Bookinfo application is composed of four separate microservices: **Productpage** microservice calls details and reviews microservices to get the book data. **Details** microservice gives the book information. **Reviews** microservice contains the book reviews and calls ratings microservice. **Ratings** microservice contains ranking formation deducted from the book review. Reviews microservice has three versions as three types of instance : **Reviews v1** does not call Rating microservice. **Reviews v2** invokes Rating microservice and displays rating stars as 1 to 5 black stars. **Reviews v3** invokes Rating microservice and displays rating stars as 1 to 5 red stars. This application is deployed with docker containers and deployed on a kubernetes cluster.

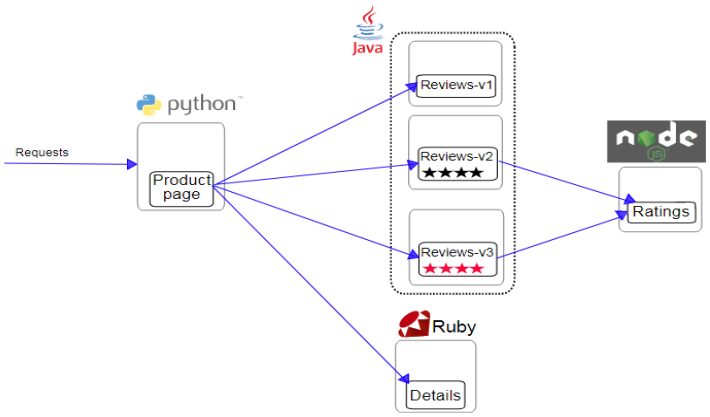


Figure 1: Bookinfo Application

Kubernetes is the most known and efficient orchestrator for docker containers. In the following, we use the pod concept of kubernetes. A pod in kubernetes is considered as a single docker container of a microservice. A microservice is composed by a set of pods (a set of containers). The elasticity controller of kubernetes is called Horizontal Pod Autoscaler HPA [5] and is based on CPU metric. It adds pods when the CPU threshold is exceeded. CPU threshold is specified by the user. If the maximum number of pods (containers) is reached, HPA autoscaler cannot add any more pods.

Many types of workload can be launched in this application. For example we can manage a workload that routes all incoming requests to Reviews v3 of microservice Review. Another example of workload routes incoming requests in 50% to Reviews v2 and 50% to Reviews v3. Technically, it is difficult to route requests flow to a specific container of a microservice. However, it is easy to manage the request flow with Istio [13] service. Istio is based on service mesh which facilitates the creation and management of the network of deployed services. Istio allows to manage network communication between microservices as well as the traffic of requests. In fact, it helps to create configuration rules for traffic routing in order to control the traffic flow between microservices.

In this paper we consider a workload that routes all incoming requests of microservice review to reviews-v3 instances. This workload launches a flow of 70 Req/s. It is not a very high workload but it saturates resources and generates overloaded microservices. This section presents the behavior of kubernetes HPA autoscaler using this workload.

Figure 2 shows the behavior of kubernetes HPA autoscaler using the presented workload with a request flow of 70 Req/s. Many pods were saturated and Kubernetes HPA launches autoscaling ac-

tions. Kuberenetes HPA adds 4 pods for productpage microservice resulting in a total of 5 pods. It adds also 1 pod to reviews-v3 microservice. From t=70, we have a total of 12 pods of reviews-v3. Ratings and details microservices were not saturated and we have no additional pods. The curves of ratings and details are superimposed as they have 1 pod all the time. The total number of pods required for this workload is 19 pods as 5 pods for productpage, 12 pods for reviews-v3, 1 pod for ratings and 1 pod for details. This autoscaler is based on CPU. If the CPU usage of a microservice exceeds threshold, the autoscaler adds many pods until recovering of the normal state.

However, as explained in section 1, using CPU usage in autoscaling treatment is not effective and may lead to allocating unnecessary resources. Therefore, we should also analyze the behavior of kubernetes autoscaler towards the response time. The red curve in Figure 2 illustrates the evolution of the response time in kubernetes autoscaler.

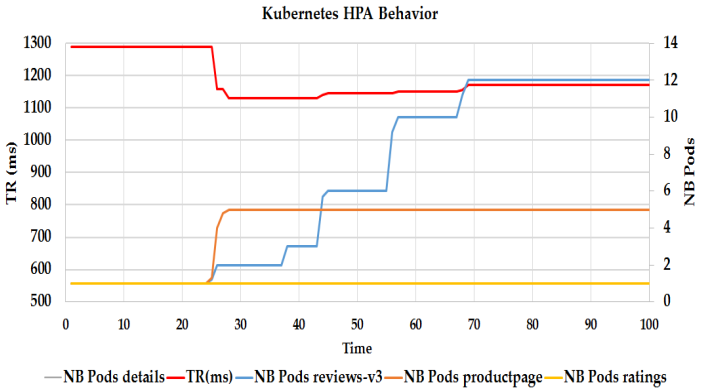


Figure 2: Kubernetes HPA Autoscaler behavior

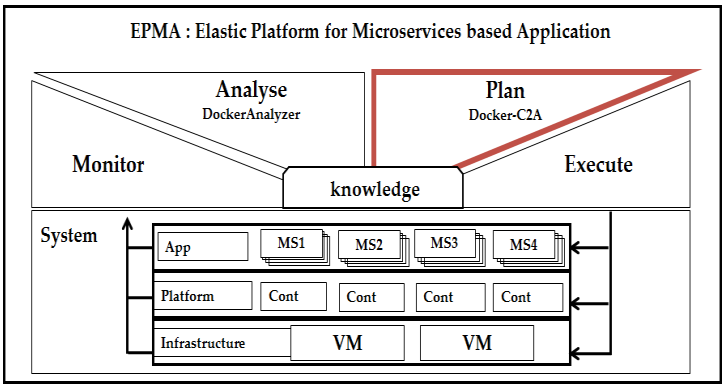


Figure 3: EPMA : Elastic Platform for Microservices based Application : Autoscaling Engine

Approximately, the response time starts with 1300 milliseconds (ms). At time t=25, we have 5 pods for productpage microservice and 2 pods for reviews-v3 microservice. We can notice a considerable evolution of the response time with approximately 1100 milliseconds. From t=37 to t=100, kubernetes autoscaler adds pods for the microservice reviews-v3 until reaching 12 pods. Following this addition, we notice no improvement in response time, but we can see a slight increase starting from t=43. It is concluded that during this interval of time, kubernetes autoscaler added 10 pods for reviews-v3 microservice, while the response time is not improved. So the added pods for reviews-v3 are not useful and kubernetes allocates unnecessary resources without any benefit.

With our autoscaler Docker-C2A, based on response time metric, we aim at optimizing the addition of resources and avoid allocating unnecessary resources. Docker-C2A uses PSO heuristic algorithm and the execution history of the application to optimize resource allocation. Using Docker-C2A, we can satisfy the same workload used in kubernetes example with a very small number of pods and with a better application performance.

4 Docker-C2A Autoscaler

4.1 EPMA platform

Docker-C2A is part of our EPMA platform (Elastic Platform for Microservices based Applications) illustrated in Figure 3. This platform is based on the autonomic MAPE-K loop proposed by IBM [14]. MAPE-K loop is composed of four components that share the Knowledge: Monitor, Analyze, Plan and Execute. EPMA manages the entire system with all levels: VM level, container level and microservice level. The analyze component DockerAnalyzer was previously presented [1] while this paper details the plan component Docker-C2A.

The monitor component collects data from different levels: VM, container and Microservice. This component collects metrics related to resources as CPU usage and memory usage and metrics related to the application performance as the response time. These parameters are filtered and correlated in order to determine symptoms that should be analyzed. As example of a symptom, an overload state caused by the CPU violation. If the monitor detects a symptom, DockerAnalyzer, the analyze component, is invoked to check whether this symptom is caused by a normal behavior as resource saturation or by an abnormal behavior as VM problem. If an abnormal behavior is detected, the analyzer had to check the root cause of the problem. Then, the plan component Docker-C2A generates the appropriate change plan describing the desired set of changes, and deliver the change plan to the execute component. The execute component applies the change plan and executes adequate actions for each layer. These four components share the knowledge containing particular data as policies and symptoms. This paper explains deeply the plan component Docker-C2A of EPMA platform.

Docker-C2A is based on the execution history of the application and a PSO heuristic algorithm. When the application state needs additional resources, Docker-C2A selects the best microservices for scaling and calculates the optimal amount of needed resources for each microservice.

4.2 Execution History

The main purpose of the execution history is to identify the microservices that should be scaled, and those that should not be scaled as there is no need to add resources. To do that, we attribute a score value for each microservice. The score value is a number varying from 0 to 5 depending on the utility of adding resources to the

microservice. For example, if we have a microservice that the addition of pods (containers) improves greatly the response time, the attributed score will be 5. If we have a microservice where adding pods will not improve the response time, the attributed score will be 0. As a result of the execution history, we generate the score value for each microservice. This helps to scale the appropriate microservices.

We account for the score attribution using bookinfo application with the same workload explained in section 3 which routes the incoming traffic of reviews microservice to reviews-v3 instances and uses a flow of 70 Req/s. Bookinfo application is composed of 4 microservices. We present in the following an extract from the history related to two microservices in order to demonstrate the score attribution.

NB Pods Productpage	NB Pods Reviews-v3	NB Pods Ratings	NB Pods Details	Response Time (ms)
1	1	1	1	1289
5	1	1	1	1134

Figure 4: An extract of Productpage history

In Figure 4, we change the number of pods in microservice productpage and we keep the same number of pods for other microservices. In the second row in this table of Figure 4, we have 5 pods of productpage and the response time has been improved considerably compared to the first line with 1 pod. The response time decreased effectively with 155 milliseconds. So requests are much faster with 5 pods than if we had 1 pod of productpage. In this case, we notice that adding pods to productpage microservice improves greatly the response time. We attribute the high score with value 5 for this microservice.

NB Pods Productpage	NB Pods Reviews-v3	NB Pods Ratings	NB Pods Details	Response Time (ms)
1	1	1	1	1289
1	5	1	1	1292

Figure 5: An extract of Reviews-v3 history

In Figure 5, we change the number of pods in microservice reviews-v3 and we keep the same number of pods for other microservices. There is no remarkable difference in response time between 1 and 5 pods. There is a very small increase in response time with 3 milliseconds, which is not profitable when we add 4 pods. In this case, we notice that adding pods to reviews-v3 microservice has no effect on the response time. Then, there is no change in the result. So we attribute the lowest score to reviews-v3 with value 0.

With the same process, we analyze the evolution of the response time when we add pods to ratings and details microservices. In fact, there is a weak improvement of the response time when we add pods to these two microservices. Then, we attribute the score value of 1 for ratings and details microservices while adding pods improves the results slightly.

As a conclusion for the execution history, we can admit that adding more resources to productpage microservice can effectively improve the response time of the application. Adding pods to reviews-v3 microservice has no effect on the response time. So adding pods to this microservice is unnecessary. Adding pods to ratings and details microservices improves the results slightly. They are not privileged in autoscaling treatment.

We illustrate the score values for each microservice in Figure 6.

	Microservice Index	Score value of scaling
productpage microservice →	0	5
reviews-v3 microservice →	1	0
ratings microservice →	2	1
details microservice →	3	1

Figure 6: Score value for each microservice

The execution history is very helpful when adding resources in scaling actions. In fact, it helps to select the appropriate microservices for scaling which leads to optimize resource allocation. After selecting eligible microservices for scaling, Docker-C2A calculates the optimal number of pods to be added to each microservice. To do this, Docker-C2A uses PSO algorithm to estimate the needed amount of resources for each microservice.

4.3 PSO Algorithm

Particle Swarm Optimization (PSO) is a computational technique based on the behavior of herds of animals that optimizes a given problem. It tries iteratively to improve candidate solutions with regard to a given quality measure. It was developed by Kennedy and Eberhart [15] in 1995 and widely used and researched ever since.

In PSO algorithm we define a particle which is analogous to a bird in a flock of birds. The flock of birds forms the search space. The movement of each particle is guided by the velocity vector. Each particle moves according to its best position pbest and the best position of the swarm gbest.

In this algorithm, we use a fitness function to measure the performance of each particle. In each iteration, the algorithm changes the velocity of each particle towards pbest and gbest positions. The algorithm stops when a stopping criterion is met or the number of iterations is reached. The velocity and positions equations are represented with Eqs. (1) and (2), respectively. The pseudo-code of the algorithm is presented in Algorithm 1.

$$V_i^{k+1} = wV_i^k + c_1 r_1 (pbest_i - X_i^k) + c_2 r_2 (gbest - X_i^k) \quad (1)$$

$$X_{i+1}^k = X_i^k + V_i^{k+1} \quad (2)$$

where

- w : inertia weight

- c_1 and c_2 : acceleration coefficients
- X_i^k : position of particle i at iteration k
- X_i^{k+1} : position of particle i at iteration k+1
- V_i^k : velocity of particle i at iteration k
- V_i^{k+1} : velocity of particle i at iteration k+1
- $pbest_i$: best position of particle i
- $gbest$: best position of the swarm

Algorithm 1 PSO Algorithm

```
Initialize a population of particles with random values positions and velocities from D dimensions in the search space
while Termination condition or Number of iterations not reached do
    for Each particle i do
        Update the velocity of the particle using Equation 1
        Update the position X_i of the particle using Equation 2
        Calculate the fitness value f(X_i)
        if f(X_i) < f(pbest_i) then
            pbest_i ← X_i
        end if
        if f(X_i) < f(gbest) then
            gbest ← X_i
        end if
    end for
end while
```

5 Docker-C2A : PSO Algorithm

Docker-C2A uses PSO algorithm to calculate the optimal number of pods (containers) to be added to each microservice. Two main objects should be defined in PSO algorithm: particle and fitness function.

5.1 Particle definition in Docker-C2A

In PSO algorithm, a particle represents a solution candidate of the problem. In our context, a solution is a particle that contains the adequate number of pods for each microservice. The dimension of the particle is defined by the number of microservices of the application. We consider the same example and workload explained in section 3. In this example, we use bookinfo application composed by 4 microservices: productpage, reviews-v3, ratings and details. In this case we define particles of 4 dimensions. Each microservice is referenced by an index. For example, we consider the productpage microservice with index 0, reviews-v3 with index 1, ratings with index 2 and details with index 3. We illustrate the particle model for bookinfo application in Figure 7.

	Particle_i	
	Microservice Index	Number of Pods
productpage microservice →	0	Nb_pods_ms_0
reviews-v3 microservice →	1	Nb_pods_ms_1
ratings microservice →	2	Nb_pods_ms_2
details microservice →	3	Nb_pods_ms_3

Figure 7: Particle model of bookinfo application

In Figure 8. below we present an example of two particles. Particle_0 contains 3 pods for productpage microservice, 1 pod for reviews-v3 and ratings microservices, 3 pods for details microservice. Particle_1 contains 1 pod for productpage microservice, 5 pod for reviews-v3 microservice, 1 pod for ratings details microservices.

Particle_0	
Microservice Index	Number of Pods
0	3
1	1
2	1
3	3

Particle_1	
Microservice Index	Number of Pods
0	1
1	5
2	1
3	1

Figure 8: Example of Particles

The PSO algorithm 1 evaluates each particle using the fitness function. The fitness value helps to decide about the best and which one is better than another by searching pbest and gbest positions as described in the algorithm.

5.2 Fitness function in Docker-C2A

The fitness function helps to evaluate particles and update pbest and gbest positions. We aim to minimize the fitness value to get the best solution. In Docker-C2A, the fitness function depends on the score based on the execution history as depicted in section 4.2 and the cost of the particle. We aim to maximize the score of particles in order to improve the response time and to minimize the cost of the particles to save deployments charges.

The fitness value of each particle P_i is calculated following equation 3 which is the cost of the particle divided by the score of the particle.

$$fitness(P_i) = \frac{Cost(P_i)}{Score(P_i)} \tag{3}$$

If we reduce the cost and increase the score then we minimize the fitness value and obtain a better solution. The particle with the lowest fitness value is the best particle, while it has a low cost with high score. The high score reflects a great improvement of the response time.

The score value of the particle is calculated following equation 4 which is the sum of the score of each microservice MS_k multiplied by the number of pods of the microservice.

$$Score(P_i) = \sum Score(MS_k) * NB\_Pods(MS_k) \tag{4}$$

The cost of the particle is calculated according to equation 5 which is the sum of the cost unit of each microservice MS_k multiplied by the number of Pods of the microservice.

$$Cost(P_i) = \sum Cost(MS_k) * NB\_Pods(MS_k) \quad (5)$$

5.3 PSO Example

In this section, we explain a concrete example of executing PSO algorithm. We consider the bookinfo application with the same workload presented in section 3 which routes requests of reviews microservice to reviews-v3 pods. In this example, we use 3 particles and 3 iterations. We follow the algorithm 1.

	Best position of the swarm gbest	Fitness value of gbest position
After Particle Initialization	Particle_1 Microservice Index / Number of Pods 0 / 2 1 / 2 2 / 2 3 / 1	Fitness = 0,538
After Iteration 0	Particle_0 Microservice Index / Number of Pods 0 / 3 1 / 2 2 / 2 3 / 1	Fitness = 0,444
After Iteration 1	Particle_0 Microservice Index / Number of Pods 0 / 4 1 / 2 2 / 2 3 / 1	Fitness = 0,391
After Iteration 2	Particle_0 Microservice Index / Number of Pods 0 / 5 1 / 2 2 / 2 3 / 1	Fitness = 0,357

Figure 9: Summarize of the execution treatment

In the first step, the algorithm initializes the number of pods needed to the 3 particles randomly. Initially, the best position of each particle pbest is the particle itself. The fitness value of each particle is calculated following the equation 3. When we calculate the cost in fitness function, we consider that all pods have the same cost as 1 unit of cost. In the first step, the fitness values of particles 0, 1 and 2 are ”0.75”, ”0.538” and ”0.714” respectively. The particle with the lowest fitness value is the best particle which is Particle_1 with a fitness value of ”0.538”. PSO algorithm selects Particle_1 as the best particle having the best position gbest in the swarm.

In each iteration, the algorithm calculates the fitness value for each particle, selects the gbest position and moves particles to new positions in order to ameliorate particle positions with better fitness values.

Figure 9. summarizes the execution treatment in this example. This figure illustrates the best position gbest after each iteration and its fitness value. We remember that we use 3 iterations in this example. It is clear that the fitness value of the best solution gbest is improved in each iteration, as the purpose of the algorithm is to minimize this value. In fact, if we minimize the fitness value, we guarantee that the particle position has the minimum cost with significant gain in response time.

At the beginning, the fitness value of gbest was ”0.538”, this value was decreased after each iteration, so our algorithm is improving the solution correctly with the decrease of fitness value. The lowest value of fitness ”0.357” is presented in the final iteration and it is the output of the algorithm. This particle has 5 pods for microservice 0 which is productpage, 2 pods for microservices 1 and 2 which are reviews and ratings microservices respectively, and 1 pod for microservice 3 which is details microservice.

In the next section, we discuss the results of our autoscaler Docker-C2A based on PSO algorithm and execution history using this example and we discuss the difference and the contribution compared to Kubernetes HPA autoscaler.

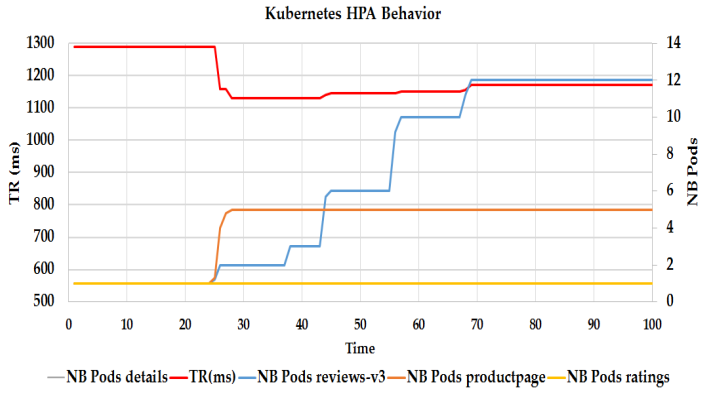
6 Experimentations

In this section, we discuss the impact of our autoscaler Docker-C2A compared to Kubernetes HPA autoscaler.

In our experimentations, we use a cluster Kubernetes composed by 4 virtual machines (VMs). Each VM runs on Ubuntu OS and contains 4 cores CPU, 8GB memory and 100 GB for disk storage. We use the same microservices-based application bookinfo detailed in section 3 composed by 4 microservices: productpage, reviews, ratings and details. Each microservice is deployed in Kubernetes cluster as a deployment containing a set of similar pods. We use the same workload used previously which routes the incoming traffic of reviews microservice to reviews-v3 instances with a flow of 70 Req/s.

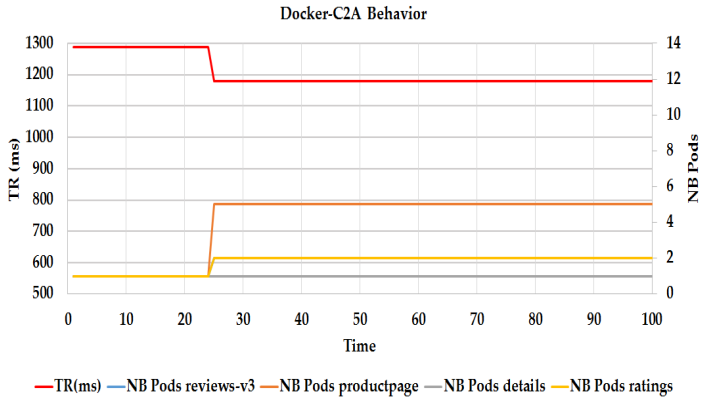
Figure 10 illustrates the difference between Kubernetes HPA and Docker-C2A behaviors. At the time t=24, we launch these two autoscalers. Kubernetes behavior is illustrated in Figure 10a. We detailed the behavior of Kubernetes HPA in section 3 using the workload of bookinfo application. When Kubernetes HPA detects a high load it waits 15 seconds to launch autoscaling actions. In each period of 15 seconds, it calculates and launches the needed number of pods. Every period, it updates the number of needed pods and monitor the situation. Kubernetes HPA stops adding pods when the state of the application becomes stable or it reaches the maximum number of pods. As a result, HPA autoscaler launches 5 pods for productpage microservice, 12 pods for reviews-v3 microservice and 1 pod for both details and ratings microservices. So for this workload Kubernetes HPA launches 19 pods.

Whereas, with Docker-C2A autoscaler, based on the execution history and the PSO algorithm, we can manage the same workload with reduced number of pods. Docker-C2A behavior is illustrated in Figure 10b. When it detects a high workload, it waits 15 seconds and launches the action plan issued by PSO algorithm. In this example, the output of PSO algorithm is presented in Figure 9. The output is the gbest particle of the last iteration. This particle contains 5 pods for productpage microservice, 2 pods for both reviews-v3 and ratings microservices and 1 pod for details microservice. Docker-C2A



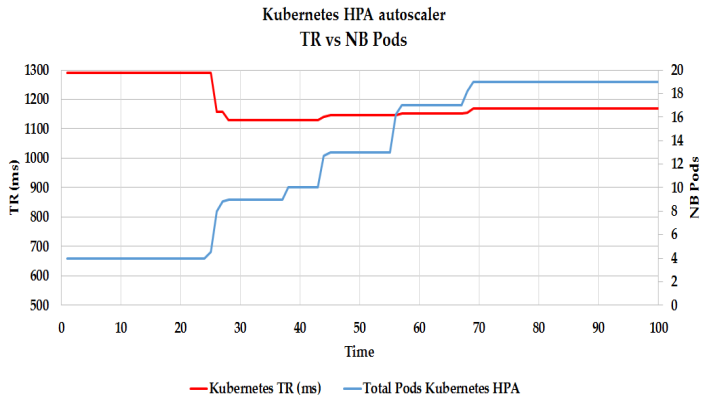
a

(a) Kubernetes HPA Autoscaler behavior



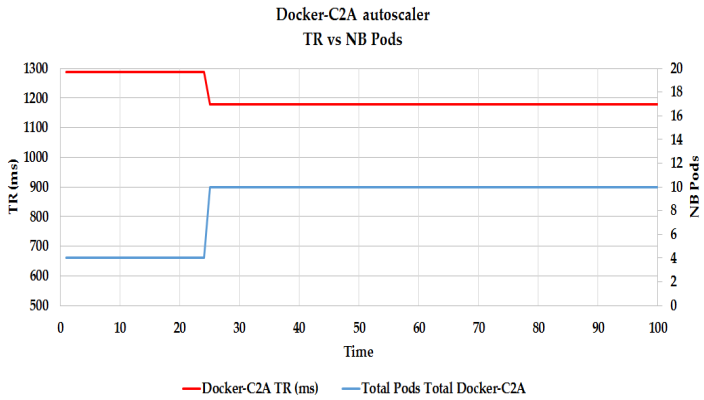
b

(b) Docker-C2A autoscaler behavior



c

(c) Kubernetes HPA Response Time vs NB of Pods



d

(d) Docker-C2A Response Time vs NB of Pods

Figure 10: Kubernetes HPA autoscaler vs Docker-C2A autoscaler

launches in total 10 pods.

If we consider the response time of requests which is presented with red curves in Figures 10a and 10b, we notice that the response time is almost the same in Kubernetes or Docker-C2A autoscalers. While Docker-C2A launches only 10 pods which is 9 pods less than Kubernetes HPA that launches 19 pods. We summarize the number of pods and the response time of these two autoscalers in Figure 11. The response time of requests with Docker-C2A is 9 milliseconds higher than the response time with Kubernetes. This difference is negligible and is not considerable as the difference is appreciated if it is in the order of tens or even hundreds of milliseconds. So, we have certainly reached our goal with Docker-C2A using the execution history and the PSO algorithm which gives an optimized amount of resources with good performance.

Autoscaler	NB Pods Productpage	NB Pods Reviews-v3	NB Pods Ratings	NB Pods Details	Total NB of Pods	Response Time (ms)
Kubernetes HPA	5	12	1	1	19	1170
Docker-C2A	5	2	2	1	10	1179

Figure 11: Kubernetes HPA autoscaler vs Docker-C2A autoscaler

Also, Docker-C2A takes in average 60 milliseconds to generate the decision of scaling, which is a short period of time that does not disturb the elastic treatment. However, Docker-C2A uses a high amount of resources of about 1.2 cores of CPU to process the PSO algorithm. The high resource usage is the cost of problem optimization and among the weak points of our autoscaler. Moreover, to run our algorithm, it is recommended to execute up to 4 iterations as a maximum. Indeed, in our context, PSO algorithm generates erroneous results from 5 iterations.

7 Conclusion and Future Works

In this paper, we presented our autoscaler Docker-C2A a cost aware autoscaler for microservices-based applications deployed with docker containers. Docker-C2A is based on response time metric. It uses the execution history of the application and PSO algorithm. Docker-C2A selects appropriate microservices for scaling and calculates the optimal number of containers for each microservice. We clearly presented the effectiveness of our autoscaler compared to Kubernetes HPA autoscaler using bookinfo application as a real case study. We compared with Kubernetes autoscaler as most of existing autoscalers have the same issues. They neither select appropriate microservices for scaling nor calculate the optimal number of additional containers. As a result, Docker-C2A optimizes computing resources and reduces deployment cost. And that is the main purpose of our autoscaler. However, there are some

limitations of Docker-C2A. In fact, it consumes a high amount of resources to generate the solution. This fact is caused by the PSO algorithm which requires high resources and this is the case of several optimization algorithms.

In future work, we improved Docker-C2A to consume less resources, either by proposing another particle definition in PSO algorithm, or by integrating another optimization algorithm instead of PSO that consumes less resources and generates good results. Also, in future work, we improved the execution history using other performance metrics and more execution data in order to analyze the application behavior in each workload. This greatly helps to estimate the needed resources with more precision and allows better optimization of computing resources.

References

[1] M. H. Fourati, S. Marzouk, K. Drira, M. Jmaiel, "DOCKERANALYZER : Towards Fine Grained Resource Elasticity for Microservices-Based Applications Deployed with Docker," in 2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), IEEE, 2019, doi:10.1109/pdcat46702.2019.00049.

[2] J. Lewis, M. Fowler, "Microservices," https://martinfowler.com/articles/microservices.html, march 2014.

[3] "Docker," https://www.docker.com/.

[4] "Kubernetes," https://kubernetes.io/, june 2019.

[5] "Kubernetes Autoscaler," https://kubernetes.io/docs/tasks/run-application/horizontal-pod-autoscale/, june 2019.

[6] G. Yu, P. Chen, Z. Zheng, "Microscaler: Automatic Scaling for Microservices with an Online Learning Approach," in 2019 IEEE International Conference on Web Services (ICWS), IEEE, 2019, doi:10.1109/icws.2019.00023.

[7] F. Zhang, X. Tang, X. Li, S. U. Khan, Z. Li, "Quantifying cloud elasticity with container-based autoscaling," Future Generation Computer Systems, **98**, 672–681, 2019, doi:10.1016/j.future.2018.09.009.

[8] M. Gotin, F. Lösch, R. Heinrich, R. Reussner, "Investigating Performance Metrics for Scaling Microservices in CloudIoT-Environments," in Proceedings of the 2018 ACM/SPEC International Conference on Performance Engineering, ACM, 2018, doi:10.1145/3184407.3184430.

[9] G. Toffetti, S. Brunner, M. Blöchlinger, J. Spillner, T. M. Bohnert, "Self-managing cloud-native applications: Design, implementation, and experience," Future Generation Computer Systems, **72**, 165–179, 2017, doi:10.1016/j.future.2016.09.002.

[10] C. Guerrero, I. Lera, C. Juiz, "Genetic Algorithm for Multi-Objective Optimization of Container Allocation in Cloud Architecture," Journal of Grid Computing, **16**(1), 113–135, 2017, doi:10.1007/s10723-017-9419-x.

[11] G. Yu, P. Chen, Z. Zheng, "Microscaler: Automatic Scaling for Microservices with an Online Learning Approach," in 2019 IEEE International Conference on Web Services (ICWS), IEEE, 2019, doi:10.1109/icws.2019.00023.

[12] "Kubernetes Autoscaler," https://istio.io/latest/docs/examples/bookinfo/, august 2020.

[13] "Istio," https://istio.io/.

[14] IBM, "An architectural blueprint for autonomic computing," 2005.

[15] J. Kennedy, R. Eberhart, "Particle swarm optimization," in Proceedings of ICNN'95 - International Conference on Neural Networks, IEEE, doi:10.1109/icnn.1995.488968.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 981-987 (2020)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

NPC five level inverter using SVPWM for Grid-Connected Hybrid Wind-Photovoltaic Generation System

Elamri Oumaymah[*,1], Oukassi Abdellah[1], Bouhali Omar[2], El Bahir Lhoussain[1]

[1]*Systems and Applications Engineering Laboratory (LISA), National School of Applied Sciences, Cadi Ayyad University, Marrakech, 40000, Morocco*

[2]*Mechatronic Laboratory, Jijel University, Jijel, 18000, Algeria*

ARTICLE INFO

Article history:
Received: 16 September, 2020
Accepted: 26 November, 2020
Online: 08 December, 2020

Keywords:
Wind power generation
Solar power
PMSG
Five-level NPC inverter
SVPWM
Grid injection

ABSTRACT

This paper covers a new topology, a synchronous wind turbine generator, and a solar photovoltaic generator. The Permanent Magnet Synchronous Generator is linked to the grid by back-to-back voltage source converters (BtB VSC), consisting of a two-stage rectifier and a five-stage Neutral Point Clamped inverter, operated by the Space Vector Pulse Width Modulation technique. A solar photovoltaic system has a direct interface with the capacitor of the DC link of the BtB VSCs, with no additional DC/DC conversion stage, while the efficiency of the system is maximized. This work features a Perturb and Observe algorithm Maximum Power Point Tracking for to extract the optimum power from both wind turbine and solar PV generator. The application of Space Vector Pulse Width Modulation control scheme is used for harmonics compensation THD. A proportional integral regulator is used to regulate the DC link voltage. A prototype is tested under various conditions, wind velocity variations as well as under several photovoltaic solar isolations. The present work has been treated using Matlab/Simulink. Simulation results proved efficiency provided by this controlling approach: reduced distortion of harmonic and increased power factor.

1 Introduction

Today, the major energy sources used for generating electricity are fossil fuel resources [1]. To reduce the dependency on fossils based energy it necessary for power generation to move towards the renewable resources including solar energy, wind power, hydroelectricity, and biomass. In literature, most of the distributed production systems are uniquely focused on one type of sustainable resource, e.g., a wind power like in [2] or solar power like in [3].

The combination and the penetration of the renewable resources, face considerable challenges due to growing power industry. The injection into the grid of the energy yielded by wind and solar photovoltaic generators, is proven to be one of the technological solution [4, 5]. Their combination offers additional benefits such as increasing the overall operational efficiency and power reliability, optimal use of lands resources, and improves the capital investments [5]. Grid-connected wind/Photovoltaic (PV) cogeneration is not largely discussed [6]–[7], whereas in literature, most of wind-PV hybrid systems are proposed for standalone off-grid applications [5]. The optimization method of generation of power using hybrid system used for battery charging is developed in [8], to increase the stability and efficiency. In [9], the author have proposed a small scale hybrid solar- wind power generation system to reduce Electricity shortage in Egypt.

Much research was conducted to track maximum power point (MPP) of the wind using methods based on Perturb and Observe (P&O) algorithm [6], neural network technique, or other approaches based on fuzzy logic. Due to the nonlinear voltage-current properties of PV solar panels, the necessity of MPP tracking has become apparent. MPP tracking schemescan be accounted for as perturb and observe (P&O), incremental conductance method, Fractional Open-Circuit Voltage, Neural Network, Fuzzy Logic Control, Current Sweep [10]. The combination of multilevel inverter and renewable energy sources are advantageous. In [11], the authors have proposed a hybrid cascaded nine level inverters for photovoltaic applications. The proposed system in [12] are based on Maximum Power Point Tracking (MPPT) combined with multilevel inverter to achieve the maximum power. In [6], the system comprises back-to-back voltage source converters with no extra dc/dc conversion stages to connect wind and photovoltaic generators to the power grid. Moreover, the control of the currents on the machine- and grid- side are carried out using hysteresis regulators, introducing higher harmonic contents

[*]Corresponding Author: Elamri Oumaymah, Email: oumaymahelamri@gmail.com

from the variable switching frequency. The objective in [13] is to realize a photovoltaic system with a MPPT linked to the 11 KV grid by employing a new control technique in order that it is possible to transfer maximum active power from the PV to the grid without injecting harmonics. The study proposes in [14, 15] a control strategy for a single-phase, three-phase photovoltaic system connected to a distribution network. The proposed technology is based on the modification of the conventional control in current mode to adjust the power factor of the photovoltaic system and control the voltage of the converter link.

This paper extends the work originally presented in ICEIT'4 [2], where the modeling and simulation of a wind turbine string, based on a vector-controlled five-level inverter, was presented. Its objective was the optimization of the energy produced, adjust the DC link and control power transferred into the grid. Motivated by advantages of hybrid production systems, presented paper covers a new topology to interface both is Permanent Magnet Synchronous Generator (PMSG)-based wind energy and PV generators with power grid. Note a blocking diode is series connected with each PV string to avoid reverse flow of current. Two P&O algorithms, namely P&O Wind and P&O PV, are used to optimize the energy extracted by turbine and solar system, respectively. Note that perturbations could impact correct operation of the receivers attached to the grid if the voltage output wave is not sinusoidal. To avoid this problem, we have implemented in this chain a two-stage rectifier followed by a five-stage Neutral Point Clamped (NPC) inverter regulated by the Space Vector Pulse Width Modulation (SVPWM) technique, without additional DC/DC conversion stages[16]. The converter control strategy is detailed in which the presented study was treated using Matlab/Simulink. Simulation results proved efficiency provided by this controlling approach: decrease distortion of harmonic and increased power factor.

2 Wind Photovoltaic Generator Side Model

2.1 Wind turbine model

Mechanical energy absorbed by a wind turbine is defined by [2]:

$$P_m = \frac{1}{2}.\rho.C_p(\lambda,\beta).V_v^3.R^2.\pi \tag{1}$$

where: C_p is the coefficient of the rotor that is a non-linear function of the tip speed ratio λ and the pitch angle of blades β, R and V_v are the radius of blades and the wind speed, respectively.

2.2 PMSG model

PMSG in the dq marker is modelised using these equations [2]:

$$V_d = -R_s.I_d - \frac{d\varphi_d}{dt} + \omega.(L_{sq}.I_q + \varphi_f) \tag{2}$$

$$Vq = -R_s.I_q - \frac{d\varphi_q}{dt} - \omega.(L_{sd}.I_d + \varphi_f) \tag{3}$$

The mechanical equation is as shown below:

$$J.\frac{d\Omega}{dt} + f_c.\Omega = C_r - C_m \tag{4}$$

Where: R_s and L_s are stator-winding resistance and inductance, respectively, I_d and I_q are the stator currents, Ω is Rotational speed of the PMSG, φ is stator phase magnetic flux, f_c and J are the friction coefficient and motor inertin, respectively. φ_f: the flux generated by the permanent magnets across stator windings, C_m is the mechanical torque .

2.3 Photovoltaic Generator

The model of the PV generator is presented in Appendix whereas it is modeled as following:

$$V_{pv} = -R_{dc}.I_{pv} - \frac{d\,I_{pv}}{dt}.L_{dc} + Vdc \tag{5}$$

The I_{op}-V_{op} characteristics of the PV array, there is an optimal operating point (I_{op}-V_{op}) corresponding to the maximum generated PV power, at any solar irradiance level.

3 Control Strategy

The suggested wind-PV system appliances two appropriate control techniques for machine-side Voltage Source Rectifier Control (VSRC) and grid-side Voltage Source Inverter Control (VSIC).The proposed system operates efficiently despite solar insolation changes and wind speed variability. The next section deals with control strategies.

3.1 Voltage Source Rectifier Control (VSRC)

Figure 1 visualizes the VSRC schema. Momentary torque is regulated through the application of the vector control schema. Vector Control technology treats the technique of control of AC machines. This control is based on eliminate the coupling between the inductor and the induced by dissociating the static current into two components (i_{sd},i_{sq}). So that, the reference direct axis current (i_{sd}) is controlled relative to desired flux, while reference quadrature axis current (i_{sq}) is adjusted according to the desired torque. Which allows for a function comparable to a dc machine, with separate excitement [4].

The Voltage Source Rectifier (VSR) equivalent circuit is presented in Figure 1. The control of the VSR is split over two parts. The first is the direct and quadrature reference current generation (i_{sdref} and i_{sqref}) and generation of switching signals for VSR.

3.1.1 Generation of i_{sdref} and i_{sqref}

In proposed system, the electromagnetic torque is controlled by controlling the current i_{sqref} while the current i_{sdref} is set to zero.

For each wind speed, the speed of the rotor has only one value for maximum wind power. Using MPPT P&O algorithm, the reference speed Ω_{ref} is generated.

Speed is regulated by the proportional integral (PI) speed controller, as illustrated in Figure 1, and its output is considered as i_{sqref}.

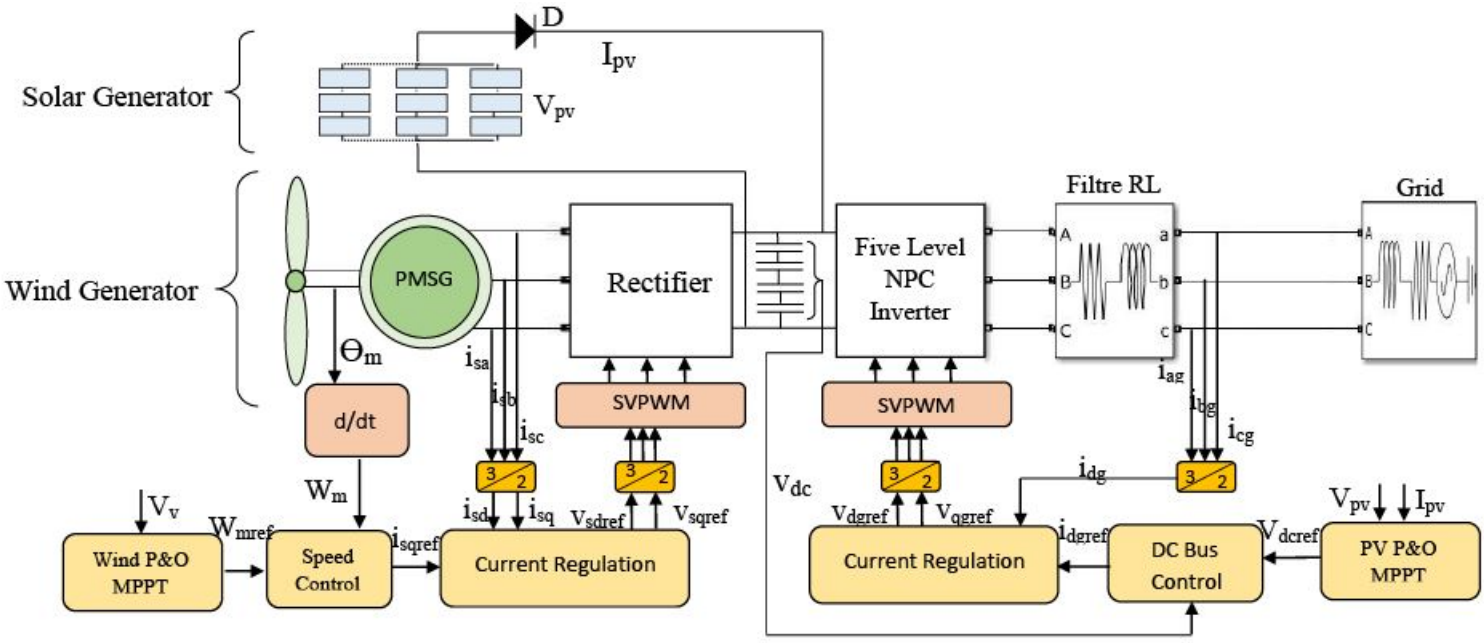


Figure 1: The proposed Wind-PV System Configuration

$$\begin{cases} i_{sdref} = 0 \\ i_{sqref} = \frac{2.Te_{ref}}{3.p.\varphi} \\ Te_{ref} = K_p.(\Omega_{ref} - \Omega) + K_i.\int(\Omega_{ref} - \Omega)\,dt \end{cases} \quad (6)$$

With: K_p and K_i are the proportion and integral coefficients of PI speed controller.

3.1.2 Generation of switching signals for VSR

The PI current controller is employed to ensure stator currents generated by PMSG correspond to references in (6), as shown in Figure 1 and (7).

$$\begin{cases} V_{sdref} = K_p.(i_{sdref} - i_{sd}) + K_i.\int(i_{sdref} - i_{sd})\,dt - \\ \Omega.L_s.i_{sq} \\ V_{sqref} = K_p.(i_{sqref} - i_{sq}) + K_i.\int(i_{sqref} - i_{sq})\,dt + \\ \Omega.L_s.i_{sd} + \Omega.\varphi \end{cases} \quad (7)$$

With: K_p and K_i are proportion and integral coefficients of PI current controller.

Three phase reference stator voltages (V_{saref}, V_{sbref}, V_{scref}) are generated from V_{sdref} and V_{sqref} using inverse Parks transform. They fed the SVPWM controller.

Voltage reference is determined using:

$$V_{ref} = \frac{2}{3}(V_{saref} + e^{-j2\frac{\pi}{3}}V_{sbref} + e^{-j4\frac{\pi}{3}}V_{scref}) \quad (8)$$

The rectifier delivers eight states, consisting of two zero vectors and six active vectors. The complex diagram is split by the active vectors in six sectors with 60 degrees forming each hexagon, with voltage reference vector corresponding to a pair of two active adjacent vectors (T_a, T_b) and a zero vector (T_0) of commutation time T_s [17].

Commutation times are determined by the equations below:

$$\begin{aligned} T_a &= T_s.m.sin(\frac{\pi}{3} - \theta) \\ T_b &= T_s.m.sin(\theta) \\ T_0 &= 1 - T_a - T_b \end{aligned} \quad (9)$$

$$\text{With: } m = \frac{(\sqrt{3})}{V_{dc}}|V_{ref}| \text{ modulation index} \quad (10)$$

3.2 Voltage Source Inverter Control (VSIC)

Figure 9 shows the Voltage Source Inverter (VSI) commutation configuration, uses the SVPWM strategy. VSI control is structured in two sections. The first section is the active power component i_{dgref} generation, the second is the VSI switching signal generation.

3.2.1 Generation of i_{dgref} and i_{qgref}

VSI system is regulated by using vector control scheme in which a phase locked loop (PLL) is applied to converter to synchronize it with power grid. Active and reactive power regulation are expressed by:

$$\begin{cases} P = 1.5(V_{dg}.i_{dg} + V_{qg}.i_{qg}) \\ Q = 1.5(V_{qg}.i_{dg} - V_{dg}.i_{qg}) \end{cases} \quad (11)$$

This control has the advantage of decoupling active and reactive control, whereas to regulate DC bus voltage V_{dc}, current i_{dg} is used and to regulate reactive power transited Q, current in quadrature i_{qg} is used. .

$$\begin{cases} P = 1.5V_{dg}.i_{dg} \\ Q = -1.5V_{dg}.i_{qg} \end{cases} \quad (12)$$

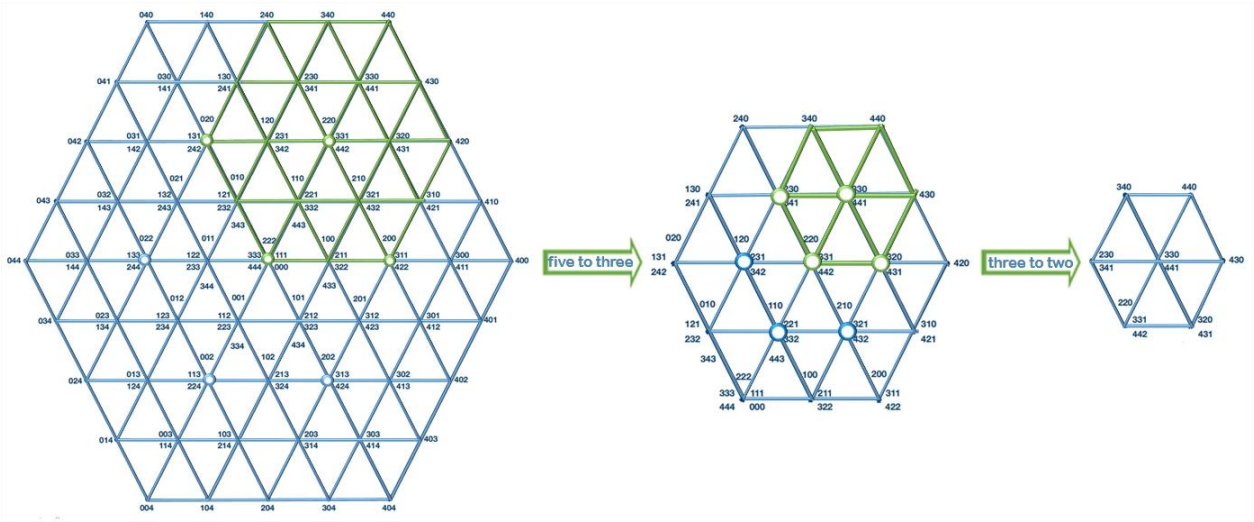


Figure 2: Decomposition of the five-level inverter vector diagram

We can therefore, through the imposition of reference currents, provide referencing for active power P_{ref} and reactive power Q_{ref} fixed in zero that maintains a unity power factor (PF), we have:

$$\begin{cases} i_{dgref} = \frac{P_{ref}.V_{dg} - Q_{ref}.V_{qg}}{V_{dg}^2 + V_{qg}^2} \\ i_{qgref} = \frac{P_{ref}.V_{qg} + Q_{ref}.V_{dg}}{V_{dg}^2 + V_{qg}^2} \end{cases} \quad (13)$$

The DC link reference (V_{dcref}) is obtained by the PV P&O MPPT algorithm as shown in figure 1, corresponding to its maximum PV power output voltage at various irradiation levels. The regulation of DC bus voltage is performed by PI regulator and its output is considered as the active power component i_{dgref}. The i_{dgref} is given as,

$$i_{dgref} = K_p.(V_{dcref} - V_{dc}) + K_i.\int (V_{dcref} - V_{dc})\,\mathrm{d}t \quad (14)$$

With: K_p and K_i are the proportion and integral coefficients of DC-bus voltage controller PI, respectively.

3.2.2 *Generation of switching signals for VSI*

The i_{dgref} and i_{qgref} are compared with i_{dg} and i_{qg}, respectivly, to obtain the currents error and are given as,

$$\begin{cases} i_{dge} = i_{dgref} - i_{dg} \\ i_{qge} = i_{qgref} - i_{qg} \end{cases} \quad (15)$$

The currents error are fed to PI controllers and the output of decoupled controller is given as:

$$\begin{cases} V_{dgref} = -K_i.\int i_{dge}\,\mathrm{d}t - K_p.i_{dge} + \\ \Omega_g.L_f.i_{qg} + v_{dg} \\ V_{qgref} = -K_i.\int i_{qge}\,\mathrm{d}t - K_p.i_{qge} + \\ \Omega_g.L_f.i_{dg} + v_{qg} \end{cases} \quad (16)$$

The three phase reference grid voltages (V_{agref}, V_{bgref}, V_{cgref}) are generated from V_{dgref} and V_{qgref} using inverse Parks transform.

This approach is founded on geometrically splitting of a vector diagram to several hexagons that represent low-level vector diagrams, as shown in figure 2 [18].

To attain this objective, two actions need to be completed:
- Depending of voltage reference vector position V_{ref}, one of six referenced hexagons is chosen;
- V_{ref} is moving to center of the hexagon selected in the first action.
Following these actions two times would reduce the five-level inverter's vector diagram to the two-level inverter's vector diagram. Next, we will employ the classical method of vector modulation of the two-level inverter, discussed in the section 3.1.2.

4 Simulation results

For the hybrid system, a simulation model shown at figure 1 has been modulated with Matlab/Simulink® in order to validate and evaluate system performance. Simulation parameters are as shown in the appendix A. Penetration of wind and PV energy is studied under different weather conditions.As shown in Figure 3 and 4, the wind speed increases from 20 to 8, and finally drops to 5 m/s, at t= 0.3 and 0.7s, respectively. Solar irradiance level increases from 400 to 1000, than drops at 500, and increases to 300 W/m, at t= 0.3, 0.5, and 0.7s, respectively.

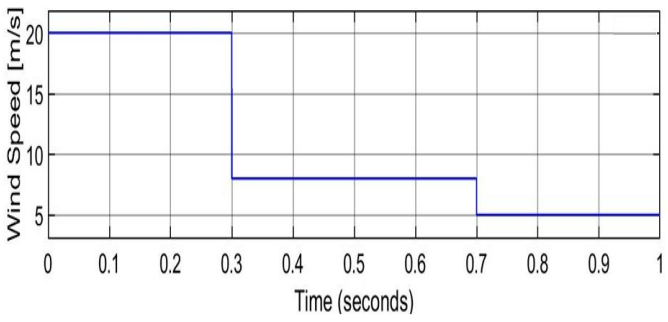


Figure 3: Cp(λ,β)curves typical for different inclination angles

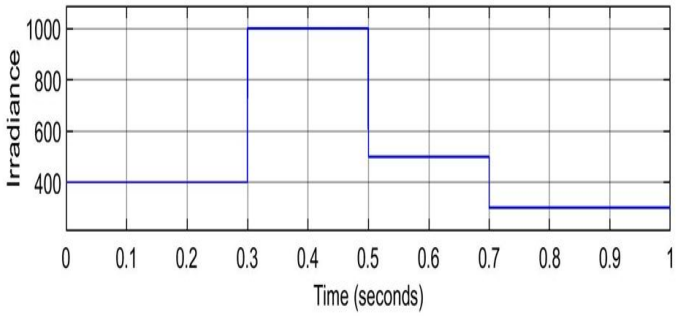


Figure 4: I_{pv} V_{pv} photovoltaic characteristics at different irradiation levels

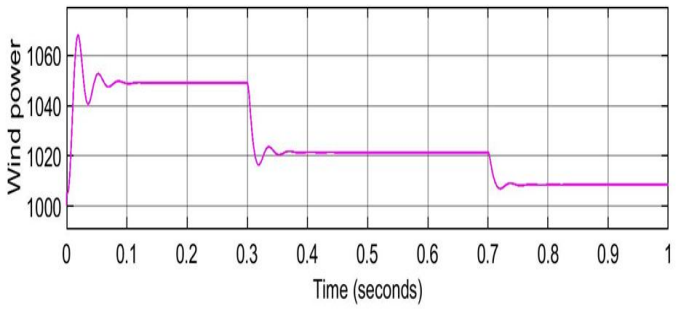


Figure 7: wind power

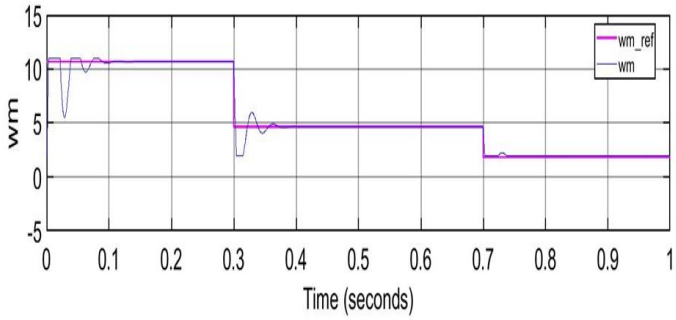


Figure 5: Generator rotation speed

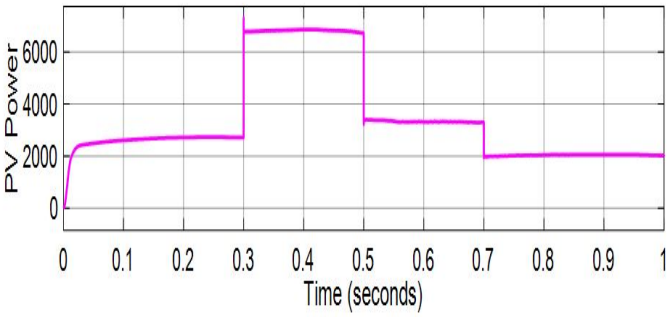


Figure 8: PV power

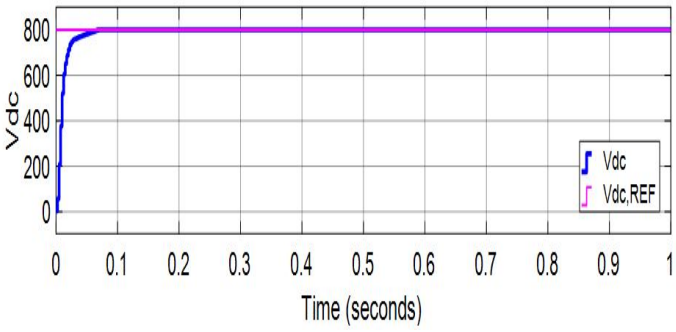


Figure 6: DC-link voltage

Optimal Ω_{ref} and V_{dcref} are generated by Wind and PV MPPT P&O, respectively. As depicted in Figure 5 and 6, the corresponding variables are then regulated.

Under these conditions, Ω_{ref} and V_{dcref} are both refracted on the corresponding PV and wind powers, as shown at figures 7 and 8.

Three-phase grid voltages before filtering are shown in figure 9, representing five-level inverter output voltage waveform obtained from capacitors. Economically, it is not interesting to use a rectifier for each capacitor, so it is necessary to ensure the balance of the continuous bus, as depicted in Figure 10 .

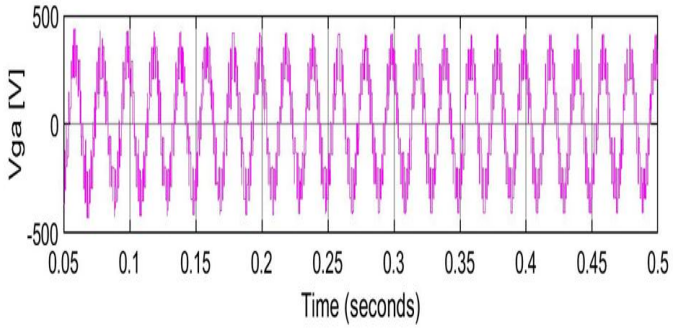


Figure 9: Output voltages of the three-phase inverter at five levels

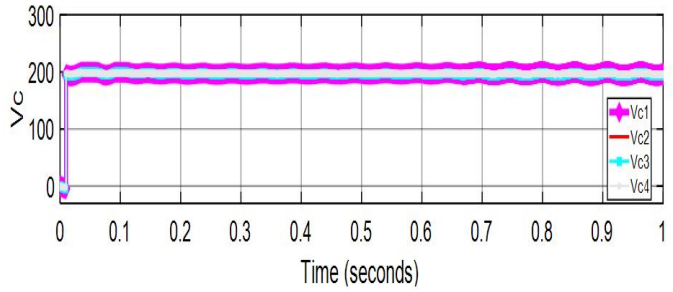


Figure 10: Capacitor voltage

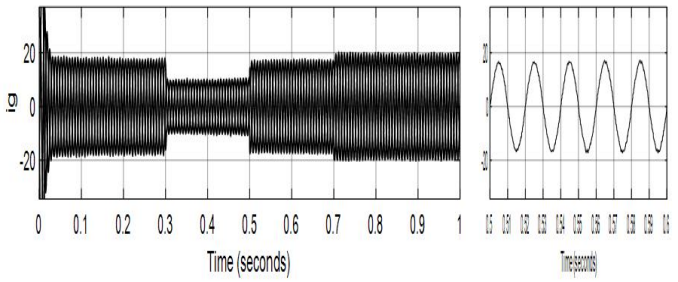


Figure 11: Grid current

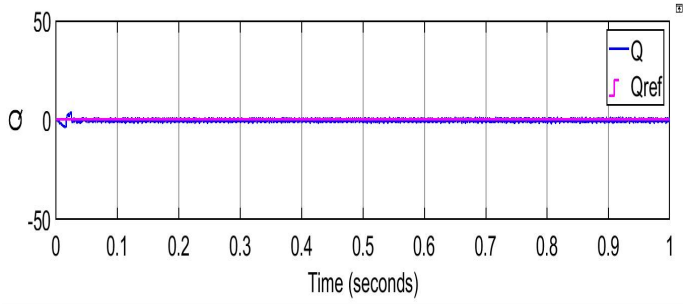


Figure 12: Reactive power

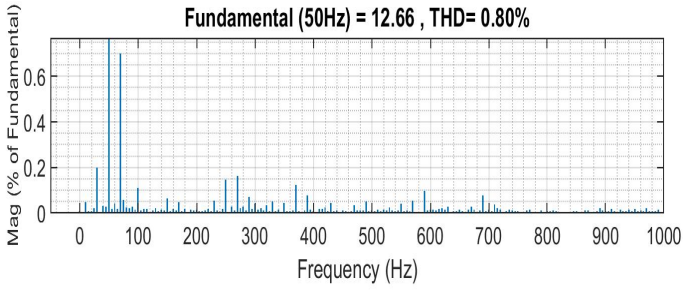


Figure 13: THD of grid currents injected

The AC current injected into grid is indicated in Figure 11. During entire operating range, the reactive power is set to zero, then a unity PF is maintained, as shown in in Figure 12.

The figure 13 illustrates the THD currents injected in the grid. According to THD analysis above, the total harmonic distortion at Fc=1KHz given THD=0.80%. Therefore, as the number of levels is increased, THDs are minimized in comparison with classical multi-level inverters [19].

5 Conclusion

In this paper, the modeling and simulation of wind-power hybrid systems connected to the grid, using vector control systems were presented.A hybrid system based on wind and photovoltaic energy was set up with a five-level inverter. The VSRC's purpose consists of extracting the maximum wind power, and the VSIC is responsible for extracting maximum PV power, and balance of dc-link capacitor in order to transfer produced power to utility-grid. This way, reactive power is fixed at zero, then a unity Power Factor PF is maintained. The proposed system has enhanced efficiency and reliability via a hybrid renewable energy system, the maximum power is extracted from independent MPPT using P&O algorithm, and reduces current THD. The results of the simulations showed the correspondence of the quantities assessed with the reference quantities and allowed the validation of the mathematical models.

References

[1] A. Demirbas, "Global Renewable Energy Projections," Energy Sources, Part B: Economics, Planning, and Policy, **4**(2), 212–224, 2009, doi:10.1080/15567240701620499.

[2] E. Oumaymah, O. Abdellah, E. A. Mustapha, "The Injection of Wind Power into a Grid Using a Multi-Level Inverter Controlled by SVPWM, booktitle = 2020 International Conference on Electrical and Information Technologies (ICEIT)," 1–6, IEEE, Rabat, Morocco, 2020, doi:10.1109/ICEIT48248.2020.9113201.

[3] L. Wang, D. Zhang, J. Duan, J. Li, "Design and Research of High Voltage Power Conversion System for Space Solar Power Station," in 2018 IEEE International Power Electronics and Application Conference and Exposition (PEAC), 1–5, IEEE, Shenzhen, 2018, doi:10.1109/PEAC.2018.8590543.

[4] F. Chishti, S. Murshid, B. Singh, "LMMN-Based Adaptive Control for Power Quality Improvement of Grid Intertie Wind–PV System," IEEE Transactions on Industrial Informatics, **15**(9), 4900–4912, 2019, doi:10.1109/TII.2019.2897165.

[5] F. Chishti, S. Murshid, B. Singh, "An Intelligent Control for Synchronous Generator Based Wind-PV-Battery Microgrid System," in 2018 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), 1–6, IEEE, Chennai, India, 2018, doi:10.1109/PEDES.2018.8707685.

[6] A. A. A. Radwan, Y. A.-R. I. Mohamed, "Grid-Connected Wind-Solar Cogeneration Using Back-to-Back Voltage-Source Converters," IEEE Transactions on Sustainable Energy, **11**(1), 315–325, 2020, doi:10.1109/TSTE.2019.2890828.

[7] S. Z. Hassan, H. Li, T. Kamal, M. Awais, "Stand-alone/grid-tied wind power system with battery/supercapacitor hybrid energy storage," in 2015 International Conference on Emerging Technologies (ICET), 1–6, IEEE, 2015.

[8] R. S. S. Singh, M. Abbod, W. Balachandran, "A Design Scheme of Control/Optimization System for Hybrid Solar — Wind and Battery Energy Storages System," in 2016 51st International Universities Power Engineering Conference (UPEC), 1–6, IEEE, Coimbra, 2016, doi:10.1109/UPEC.2016.8114093.

[9] I. Elsayed, I. Nassar, F. Mostafa, "Optimization and Economic Evaluation of Small Scale Hybrid Solar/Wind Power for Remote Areas in Egypt," in 2017 Nineteenth International Middle East Power Systems Conference (MEPCON), 25–30, IEEE, Cairo, 2017, doi:10.1109/MEPCON.2017.8301158.

[10] K. Saidi, M. Maamoun, M. Bounekhla, "Simulation and analysis of variable step size P&O MPPT algorithm for photovoltaic power control," in 2017 International Conference on Green Energy Conversion Systems (GECS), 1–4, IEEE, Hammamet, Tunisia, 2017, doi:10.1109/GECS.2017.8066265.

[11] N. Kalaiarasi, S. S. Dash, S. Paramasivam, R. Bayindir, "Hybrid Cascaded Nine Level Inverter Using Dspace Controller for Standalone Photovoltaic Applications," in 2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA), 745–750, IEEE, San Diego, CA, 2017, doi:10.1109/ICRERA.2017.8191159.

[12] B. S. N. Raj, R. Siddharth, S. M. Shyni, "MPPT with Bi-Directional DC-DC Converter and Multi-Level Inverter for Grid Connected Hybrid System," in 2017 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC), 773–777, IEEE, Melmaruvathur, 2017, doi:10.1109/ICCPEIC.2017.8290467.

[13] P. K. Hota, B. Panda, B. Panda, "Fault Analysis of Grid Connected Photovoltaic System," American Journal of Electrical Power and Energy Systems, **5**(4), 35, 2016, doi:10.11648/j.epes.20160504.12.

[14] A. Yazdani, P. Dash, "A Control Methodology and Characterization of Dynamics for a Photovoltaic (PV) System Interfaced With a Distribution Network," IEEE Transactions on Power Delivery, **24**(3), 1538–1551, 2009, doi: 10.1109/TPWRD.2009.2016632.

[15] P. P. Dash, A. Yazdani, "A Mathematical Model and Performance Evaluation for a Single-Stage Grid-Connected Photovoltaic (PV) System," International Journal of Emerging Electric Power Systems, **9**(6), 2008, doi: 10.2202/1553-779X.2033.

[16] A. Ghosh, F. Shahnia, "Applications of Power Electronic Devices in Distribution Systems," in J. A. Martinez-Velasco, editor, Transient Analysis of Power Systems, 248–279, John Wiley & Sons, Ltd, Chichester, UK, 2014, doi:10.1002/9781118694190.ch7.

[17] M. M. Irfan, P. H. K. Prasad, P. V. Rao, "Simulation of five-level five-phase SVPWM voltage source inverter," in 2010 International Conference on Power, Control and Embedded Systems, 1–5, IEEE, Allahabad, 2010, doi: 10.1109/ICPCES.2010.5698659.

[18] K. Himour, K. Ghedamsi, E. M. Berkouk, "New alternative of design and control of grid connected PV-Storage systems with the five level diode clamped inverter," in 2013 Conférence Internationale des Energies Renouvelables (CIER'13), volume 2, 10, 2013.

[19] Yaosuo Xue, Liuchen Chang, "Closed-loop SPWM control for grid-connected buck-boost inverters," in 2004 IEEE 35th Annual Power Electronics Specialists Conference (IEEE Cat. No.04CH37551), 3366–3371, IEEE, Aachen, Germany, 2004, doi:10.1109/PESC.2004.1355070.

A Appendix

The system Parameters: Wind turbine: $p = 122\ Kg/\ m^3$, $R = 412\ m$, $J = 0.0002\ Kg.m$; PMSG: P= 2 kW, $R_s = 12\ m$, $L_{sq} = 0.0125\ mH$, $L_{sd} = 0.0057\ mH$,P = 4; Solar PV :$N_s = 40$, $N_p = 2$, $I_{phc} = 3.45$ A, $G_{ref} = 1000\ W/m^2$, Open circuit voltage = 20.1 V, Saturation diode current = 4.05e-7 A, Reference temperature = 24 $°C$; DC link: $U_c = 50\ V$, $C = 2200\ pF$; Grid connection: $L_f = 1\ mH$, $R_f = 0.1\ \Omega$; Grid: U= 338 V and f =50 Hz.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 988-994 (2020)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Mobile Money Wallet Attack Resistance using ID-based Signcryption Cryptosystem with Equality Test

Seth Alornyo[*,1], Mustapha Adamu Mohammed[1,2], Francis Botchey [1,3], Collinson Colin M. Agbesi [1]

[1]*Koforidua Technical University, Computer Science Department, KTU, Koforidua, Ghana*

[2]*Kwame Nkrumah University of Science and Technology, Computer Science Department, KNUST, Kumasi, Ghana*

[3]*University of Electronic Science and Technology of China, School of Information and Software Engineering, UESTC, Chengdu, China*

ARTICLE INFO

Article history:
Received: 18 September, 2020
Accepted: 17 November, 2020
Online: 08 December, 2020

Keywords:
ID-based encryption
Signcryption
Equality test

ABSTRACT

This paper is an extension of a research work presented at ICSIoT 2019. An attack continuum against the insider attack in mobile money security in Ghana using a witness based cryptographic method proposed by Alornyo et al. resisted the service provider from peddling with users data for economic gains. Our improved scheme achieves a simultaneous benefit of digital signature in public key encryption (PKE). The adoption of signcryption cryptosystem in our scheme achieved a desired security property of EUF-CMA using the random oracle model.

1 Introduction

Ghana and other African countries have seen a tremendous growth in mobile money service patronage in recent times. The use of the mobile money services platform requires the use of a less-resource constraint mobile devices that do not require access to internet or mobile app for mobile payment. Thus, any mobile device that does not have access to internet but can access the mobile money service provider cellualr network can perform financial transactions and other utility payments. This and many other reasons has lead to the high patronage of mobile money services in Ghana. However, the mobile money system is not immune to data forgery and re-play attacks. The system model of our scheme is depicted in 1

Our research work is an extension of a publication presented at ICSIoT 2019. The construction presented at ICSIoT 2019 [1] considered the service provider as the adversary who could launch the insider attack. The service provider had access to users token information and transaction details and was possible for the service provider to peddle with users data via the following:

- The mechant receives valid transaction details such as transaction ID, td_{ID} and finds out the content M_1 as well as the token information td_{ID} from the transaction details TSD.
- The adversary (insider) attains the ciphertext CT_1 of a guessed message $M_1^{'}$ and token tdk_{ID}.
- The adversary checks if the test $Test(CT_1, TSD, td_{ID}^{'}) = 1$

In the above scenario, the insider wins the game if it succeeds in deriving the token information from the transaction details. However, the scheme utilized the witness based cryptographic primitive with an added pairing operation to resist the insider attack continuum.

In spite of their construction, there were a possible attack during data transmission such as forgery and re-play attacks in mobile money payment systems. To solve the problem, we propose a scheme using a signcryption cryptographic primitive to achieve the simultanoeus benefit of digital signature and PKE in mobile money systems deployed in Ghana (West Africa). This approach resist data forgery and re-play attacks.

The first generic construction of identity-based signcryption was unveiled [2]. Their work fulfilled a dual function of digital signature and public key cryptosystem (PKE). Other traditional approaches that employed signature-then-encrypt had a high cost computations as compared to [2]. However, the scheme [2] adopted data encapsulation method instead of key encapsulation method [3]–[6] to achieve confidentiality and unforgeability instantiation in the standard model.

Signcryption scheme unveiled in [7] did not propose a secu-

[*]Corresponding Author: Seth Alornyo, Koforidua Technical University, Computer Science Department, +233 554936729 & sabigseth@ktu.edu.gh

rity proof even though their construction were based on discrete logarithm assumptions. In view of this, several research in signcryption and signature schemes have been unveiled [8]–[15] with other functional extensions [16]–[19]. In 2011, a survey of identity-based signcryption cryptosystem was examined by Li et al.[20] to examine the security models as well as a comparative study of their security properties and efficiency. Analysis of other flavours of signcryption constructions [4], [11]–[13], [21]–[24], threshold signcryption [25]–[30], proxy signcryptions [31]–[37] and ring signcryption [38]–[42] have been proposed and constructed. Subsequently, Xiong et al. [43] did a cryptanalysis on the scheme [21]. The security notion of CPA in their work was diffused by Xiong et al. [43]. Hence, [43] showed that their scheme did not achieve chosen plaintext attack (CPA) security as they claimed.

Due to the need to resist the cloud server from peddling users outsourced data, efficient signcryption cryptosystem was unveiled in [44]. According to him, some existing signcryption schemes in [45]–[55] had certain lapses such as lack of data integrity, authentication and non-repudiation. Hence the need to construct a scheme to deny the insider adversary in the cloud from data peddling and modification for economic gains became paramount. Furthermore, the emergence of distributed computing and interconnected systems propelled [56] to unveiled a signcryption scheme in heterogeneous systems. Although in [57], the author earlier discussed the above problem, they could not construct a scheme to solve the problem of signcryption in heterogeneous systems. However, the constructions in [57, 58] and [59] achieved insider and outsider attack resistant respectively.

Recently, secure identity-based cryptosystem has been unveiled in [60]. Their security improvement was based on certain proposed signcryption algorithms [11]–[14], [61] constructed using the random oracle, and as well as schemes designed using the standard model [21, 23, 24, 62] with certain deficiencies such as IND-CCA2 and existential unforgeable chosen message attack (EUF-CMA). However, an attack was launched in [63] to unveil a new functional secure identity-based signcryption cryptosystem in [60]. A scheme to curtail an attack continuum in mobile money wallet system in Ghana to prevent message forgeability and re-play attack is still an open problem.

1.1 Our Contribution

By considering that the integrity, authentication and non-repudiation of the data in mobile money wallet system in Ghana, we proposed the mobile money wallet attack resistant scheme using the ID-based signcryption cryptographic primitive with equality test (known as MWAR-ID-SET) to achieve a simultaneous benefit of digital signature with public key encryption, and with equality test. Concretely, the formal definition, security model and concrete construction of MWAR-ID-SET are proposed in this paper. Further, our proposed scheme was shown to achieve the security property of existential unforgeable chosen message attack (EUF- CMA) by using the formal security proof.

1.2 Paper Organization

The rest of this work is organized as follows; In Section 2, our scheme outlines preliminaries for the construction and formulate the definitions of MWAR-ID-SET. In Section 3, the definitions of our scheme are outlined, section 4 outlines the security model of MWAR-ID-SET. Section 5 details the construction of our scheme, and proof the security in Section 6. Section 7 compares our work with existing schemes. Section 8 concludes our work.

2 Preliminaries

Definition 1: Bilinear Map. Let G and G_T be two multiplicative cyclic groups of prime order p. Suppose that q is agenerator of G. A bilinear map $e : G \times G \rightarrow G_T$ satisfies the following properties:

1. Bilinearity: For any $g \in G$, and $b \in Z_p$, $e(g^a, g^b) = e(g, g)^{ab}$.
2. Non-degenerate: $e : (g, g) \neq 1$.
3. Computable: There is an efficient algorithm to compute $e(g, g)$ for any $g \in G$.

Definition 2: Bilinear Diffie-Hellman (BDH) problem. Let G and G_T be two groups of prime order q. Let $e : G \times G \rightarrow G_T$ be an admissible bilinear map and let q be a generator of G. The BDH problem in (q, G, G_T, e) is as follows: Given (q, q^a, q^b, q^c), for random $c, d, f \in Z_p^*$, for any randomized algorithm. A computes the value $e(q, q)^{cdf} \in G_T$ with advantage:
$ADV_A^{BDH} Pr[A(q, q^c, q^d, q^f) = e(q, q)^{cdf}]$.

We say that the BDH assumption holds if for any polynomial-time algorithm A, its advantage ADV_A^{BDH} is negligible.

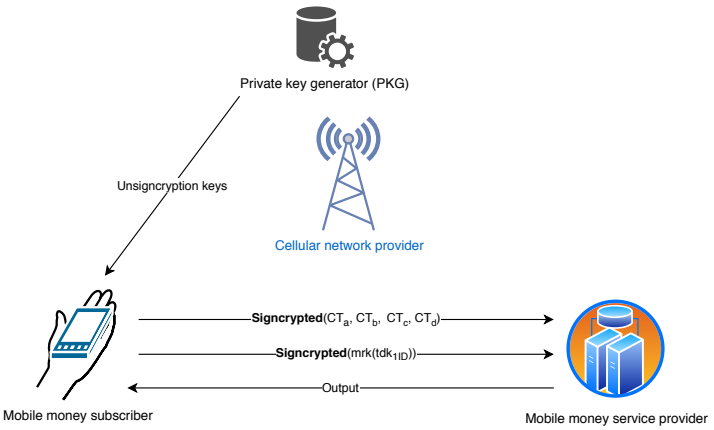


Figure 1: System model of our scheme

3 Definitions

In this section, the formal definition of our proposed scheme is outlined. MWAR-ID-SET achieves a formal security property of EUF-CMA. The construction specifies six(6) steps: $Setup$, $MW - Extract$, $TokenGen$, $MW - Signcrypt$, $MW - Unsigncrypt$, $MW - Test$. M_T and C_T are considered as the plaintext space and ciphertext space respectively.

1. $Setup$: The construction on input a security parameter k, output public parameter K with MSK as master secret key.

2. $MW - Extract$: The scheme on input MSK, $ID \in \{0,1\}$ arbitrary and gives out a message recovery key mrk corresponding to an identity.

3. $TokenGen$: With the input message recovery key mrk, arbitrary $ID \in \{0,1\}^*$ and it return token tdk_1 corresponding to an identity (ID_1).

4. $MW - Signcrypt$: The scheme on input $ID \in \{0,1\}^*$, random witness $w \in W$, a chosen plaintext $m_1 \in M_1$, and output ciphertext $CT_1 = (x_1, c_1)$ where $x \in X$ from generated witness $WInsGen(w) = x$ with the relation R satisfied [1].

5. $MW - Unsigncrypt$: On input the ciphertext CT_1, message recovery key mrk with a random chosen witness $w \in W$, the plaintext $m_1 \in M_1$ is uncovered provided CT_1 is deemed as a valid ciphertext derived from a witness relation R.

6. $MW - Test$: The scheme on input the ciphertext $CT_A \in CT_1$ with its corresponding receiver ID_A derived from the token tdk_{1_A} with its corresponding ID_A, the ciphertext $CT_A \in CT_1$ with its corresponding receiver ID_B derived from the token tdk_1 with its corresponding ID_B. The scheme respond 1, provided CT_A and CT_B have same message. Otherwise it respond as $\perp$.

4 Security Model

We assume $\sqcap = \{Setup, MW - Extract, TokenGen, MW - Signcrypt, MW-Unsigncrypt, MW-Test\}$ as the scheme and polynomial time algorithm adversary A. $MWAR - ID - SET$ achieves two main security notion of $IND - CCA2$ and $EUF - CMA$. However, our scheme adds the security notion of ID-based indistinquishability to IND-CCA2 and is coined as IND-ID-CCA2, similarly presented in [55] via the standard model..

1. $Setup$: The challenger A executes the security parameter k and outdoors K with a randomly chosen witness $w \in W$ and generates $x \in X$ of the relation R. It gives out the relation R to the adversary A.

2. Phase 1: The mobile merchant adversary A then issues $(\wedge_1, \wedge_2, ..., \wedge_{n-1})$. It is assumed that such query is of:

   - $MW-Query(ID_i)$: The merchant executes $H(.)$ and generates mrk_i which corresponding to ID_i as the identity. The recovered mrk_i is forwarded to A.
   - $TokenGen(ID_i)$: The merchant executes the $MW - Unsigncryption$ to generates tdk_i via the witness relation R. The algorithm then forwards tdk_i to A.

3. $MW - Challenge$: When the phase comes to an end, two messages (m_1, m_2) with equal-length and ID^* is submitted by A to the challenger and wishes to be challenged. But (m_1, m_2) were both not issued during signcryption and ID^* was never extracted during phase 1 section. Given this, the challenger randomly chooses $t \in \{0,1\}$ and returns $CT_1 = MW - Signcrypt(m_t, ID^*, w^*)$. Again, a challenge ciphertext token $mrk^* = (ID^*, x^*)$ is issued by the execution of $TokenGen$ phase $mrk^* \leftarrow mrk(tdk, m_t, x^*)$ and it output mrk^* to A.

4. $MW - Unsigncrypt$: The merchant executes an unsigncryption algorithm to unsigncrypt CT_1. The merchant obtains mrk_1 which corresponds to a public key of ID_i. The plaintext m_i is forwarded to A.

5. Phase 2: The merchant adversary issue the query $(\wedge_1, \wedge_2, ..., \wedge_n)$. The query is of the form:

   - $MW - Query$. A similar response as in phase 1 is given because $ID_i \neq ID^*$.
   - $TokenGen(ID_1)$. Given $x \neq x^*$, the merchant respond same as in phase 1.

6. Result. The adversary A does a guess v' of v. Since $v' = v$, then the adversary will win the game.

   The adversary advantage in breaking the scheme is noted as: $ADV_{MWAR-ID-SET}(k) = Pr[v' = v] - \frac{1}{2}$ as a negligible probability.

5 Construction

A detailed construction of our proposed scheme is outlined as follows:

1. On input a secured parameter k, the algorithm output public parameter K, with MSK as master secret key.

   - Two multiplicative group G and G_T are generated by the system with the same order of lemgth θ bits with a bilinear map $e : G \times G \rightarrow G_T$. The group generater $P \in G$ is selected.
   - The algorithm exploits keyed permutation $F : \{0,1\}^s \times \{0,1\}^n \rightarrow Z_p^*$ with a positive interger $D = k(i)$ and $L = b(i)$, it then activate a random value r_1 from $\{0,1\}^L$. Message authentication code (MAC), $MAC = GSV$. Thus, generate, sign, verify (GSV). After executing $G(i)$, it obtains r_2. It then set master token key as $MTK_1 = (r_1, r_2)$.
   - A hash functions $H_a : \{0,1\}^\tau \rightarrow Z_p^*$, $H_b : \{0,1\}^* \rightarrow G$, $H_c : A \times G \times G_T \rightarrow \{0,1\}^{\tau + r_1}$, where r_1 represent random numbers and τ as length of message. $(\tau_1, \tau_2) \in Z_p^2$ is randomly chosen and $R_v = P^{\tau_1}$, $R_m = P^{\tau_2}$. The paramrter $K = (A, \tau, G, G_T, P, R_v, R_m, MAC, H_a, H_b, H_c)$ is made public (published)

2. $MW-Extract$: With an $ID \in \{0,1\}$ as string, the system computes $Q_{ID} = H_b(ID) \in G$ and private key $mrk_{ID} = (Q_{ID}^{\tau_1}, Q_{ID}^{\tau_2})$. It should therefore be noted that (τ_1, τ_2) are secret value randomly chosen by the algorithm.

3. $TokenGen$: The algorithm with an input string $ID \in \{0,1\}^*$, the computation $Q_{ID} = H_b(ID) \in G$ is executed and the token $tdk_{1_{ID}} = (Q_{ID}^{\tau_2})$ is generated.

4. $MW - Signcrypt$: The algorithm with an input public parameter K, string ID, it executes $Q_{ID} = H_b(ID) \in G$ and a **signcryption** $m_1 \in G$ is triggered by choosing two random values $(q_1, q_2) \in Z_p^*$. The ciphertext is set as $CT_1 = (CT_a, CT_b, CT_c, CT_d)$ as :

$$CT_a = H_b(e(Q_{ID}, R_m)^{q_1}) \cdot m_1^{q_1}, \qquad CT_b = P^{q_1}$$

$$CT_c = P^{q_2} \quad CT_d = (m_1 \| w) \oplus H_b(CT_a \| CT_b \| Z \| e(Q_{ID}, R_v)^{q_2}).$$

The signature $Z \leftarrow S(r_2, CT_c)$ is used to signcrypt the ciphertext of the employed MAC. The signcrypted tag Z is used to verify signcrypted ciphertext CT_c

5. $MW - Unsigncrypt$: To **unsigncrypt**, the algorithm with an input the ciphertext CT_1, private **unsigncryption** key $mrk = (Q_{ID}^{\tau_1}, Q_{ID}^{\tau_2})$ with $CT_1 = (CT_a, CT_b, CT_c, CT_d)$ corresponding to an identity ID. The algorithm executes $(m_1 \| w^{'}) = CT_b \oplus H_b(CT_a \| CT_b \| CT_c \| e(CT_c, Q_{ID}^{\tau_1}))$. The algorithm do a check of $CT_a = (m_1^{'}) \| x^{'}$ and $CT_b = P^{q_1^{'}}$ to determine whether they are equal. If they are equal, the algorithm returns m_1, contrarily, it returns $\perp$.

6. $MW - Test$: With a given **signcrypted** ciphertext CT_{1_A} with trapdoor tdk_{1_A} and a different **signcrypted** ciphertext CT_{1_B} with a trapdoor tdk_{1_B}. The algorithm determines whether $m_{1_A} = m_{1_B}$ is equal or not. The algorithm does this by executing:

$$TD_A = \frac{CT_{a_A}}{H_b(e(CT_{a_A}, R_{m_A}))}, \quad TD_B = \frac{CT_{a_B}}{H_b(e(CT_{a_B}, R_{m_B}))}.$$

If the above equation holds, the algorithm then output 1 on success and 0 on failure.

Construction Correctness.
We assume that $CT_1 = (CT_a, CT_b, CT_c, CT_d)$.). Test algorithm executes:

$$CT_{1_A} = \frac{CT_{a_A}}{H_b(e(CT_{a_A}, R_{m_A}))}, \quad CT_{1_B} = \frac{CT_{a_B}}{H_b(e(CT_{a_B}, R_{m_B}))}$$

$$CT_{1_A} = \frac{H_b(e(Q_{ID}, P^{\tau_2})^{q_1}) \cdot m_{1_A}^{q_1}}{H_b(e(Q_{ID}, P^{\tau_2})^{q_1})}, CT_{1_B} = \frac{H_b(e(Q_{ID}, P^{\tau_2})^{q_1}) \cdot m_{1_B}^{q_1}}{H_b(e(Q_{ID}, P^{\tau_2})^{q_1})}$$

$$CT_{1_A} = m_{1_A}^{q_1}, \quad CT_{1_B} = m_{1_B}^{q_1}$$

Hence, $m_{1_A}^{q_1} = m_{1_B}^{q_1}$

From the above, the algorithm output 1 on success and 0 on failure.

Therefore:

$e(CT_{b_A}, R_{m_B}) = e(CT_{b_A}, R_{m_B})$.

$$e(CT_{b_A}, R_{m_B}) = (P^{q_1}, P^{\tau_2}) = e(P, P)^{q_1 \tau_2},$$

$$e(CT_{b_A}, R_{m_B}) = (P^{q_1}, P^{\tau_2}) = e(P, P)^{q_1 \tau_2}.$$

This implies that if, $m_{1_A} = m_{1_B}$, then

$e(CT_{b_A} = R_{m_B}) = e(CT_{b_A} = R_{m_B})$.

6 Security Analysis

In this section, we consider a formal security property of IND-CCA2 and EUF-CMA [60]. Our scheme adds the notion of ID-based indistinquishability to IND-CCA2 and referred to as IND-ID-CCA2.

6.1 IND-CCA2 Security

Our MWAR-ID-SET is $(SET_\varepsilon, t_s, q_{ks}, q_{ns}, q_{us})$-IND-CCA2 secure if $(\varepsilon_{mdbdh}, t_s) - mDBHDH$ assumption holds. Thus, H_1 and H_2 serves as (ε_{H_1}) and (ε_{h_2}) are both collision resistant hash functions, such that:

$$\varepsilon_{SET} \leq \varepsilon_{mdbdh} + \varepsilon_{H_1} + \varepsilon_{H_2} + \frac{q_{ks}+q_{us}+3}{p} + \frac{q_{ns}}{p^2}.$$

Where t_s refers to index period time, q_{ks} refers to number of extraction key queries, q_{ns} represent number of signcryption queries and q_{us} represents number of unsigncryption queries. Therefore, the security analysis with a collission resistant hash function proves our scheme secured.

6.2 EUF-CMA Unforgeability

Proof Theorem: This section outlines the security proof of unforgeability against adaptive CMA derived from the security constructions in Chow[41] ID-based **signcryption** cryptosystem. Thus, it is expected that the adversary can forge a ciphertextof a message m_1, if the assumption $CT_1 = (CT_a, CT_b, CT_c, CT_d)$ corresponding to a user identity ID holds. Thus, $CT_d = (m_1 \| w) \oplus H_3(CT_a \| CT_b \| CT_c \| e(Q_{ID}, R_v))^{q_2}$ is regarded as the signature of the message $m_1 \| w$, where $e(Q_{ID}, R_v)^{q_2}$ is regarded as the pairing of a corresponding user with secret **signcryption** key R_v. It is however noted that the difficulty of CDH problem makes the scheme unforgeable via the random oracle model.

6.3 Security Analysis of Token Key

Further details on the token security analysis can be accessed in [1]. The token security analysis experiment to our scheme is defined as: $EXP_{MWAR-ID-SET_A}^{IND-ID-CCA2}(k)$.

With a security parameter k, a master token key $MTK_1 = (r_1, r_2)$ and A adversary against token security. According to [1], the adversary A win the game if $b^{'} = b$, which shows that the output of the experiment is 1 on success and 0 on failure. Adversary A advantage in the experiment is defined as:

$Adv_{MWAR-ID-SET}^{IND-ID-CCA2}(w) = |Pr[Exp_{MWAR-ID-SET}^{IND-ID-CCA2}] - \frac{1}{2}|$

However, the probability for the adversary to win the game is negligible, hence proves our construction secured.

7 Comparison

A security strength comparison computations with existing **signcryption** schemes are outlined in Table 1. Constructions in ID-based **Signcryption** cryptosystem in [21, 23, 60, 64] are compared to with respect to security strength. The parameters for our comparison includes group multiplication, group exponentiation, inverse computations, pairing operation, test for equality, and support for EUF-CMA.

Table 1: The performance computational cost and Communication overheads

Scheme	G_{Mult}	G_{Exp_a}	$G_{T_{Mult}}$	$G_{T_{Exp_a}}$	$G_{T_{Inv}}$	Pr	IND-CCA2	EUF-CMA	ET
[21]	$2n_{u_a} + 2n_{m_a} + 1$	3	5	1	1	7(+2)	⊠	√	No
[23]	$2n_{u_a} + 2n_{m_a} + 1$	3	5	1	1	7(+2)	⊠	⊠	No
[60]	$2n_{u_a} + 2n_{m_a} + 1$	7	5	1	1	7(+2)	√	√	No
[64]	$2n_{u_a} + 2n_{m_a} + 3$	7	5	1	1	7	√	√	No
Ours	$2n_{u_a} + 2n_{m_a} + 1$	7	3	2	2	8	√	√	Yes

Legend: $''G_{Mult}$": multiplication in group G, $''G''_{Exp_a}$: exponentiation in group G, $''G''_{T_{Exp_a}}$: exponentiation in group G_T, $G_{T_{Inv}}$: inverse computation in group G_T, $''n_m, n''_u$: length of identity in bits string, $''Pr$": pairing operations in the form $x(+y)$ as in [23], $''ET''$: equality test, $''\boxtimes''$: not supportive, $''\surd''$: supportive .

Table 2: Running Times with Symbols

Symbols	Description	Times
G_{Exp_a}	G exponentiation operation	6.3937
$G_{T_{Exp_a}}$	G_T exponentiation operation	1.9518
T_{p_a}	Operation with pairing	11.4173
T_{h_a}	Hash functions	000853
G_{Mult}	Multiplication operation in G	0.047
$G_{T_{Mult_a}}$	Multiplication operation in G_T	0.0119

The Pairing-Based Cryptography (PBC) Library [65] is used to quantify the time consumption of signcryption, unsigncryption and test operations. We use the code of a program in VC++ 6.0 and executed on a computer (Windows 10 Pro, operating system), Capacity of Intel(R) Core (TM) i5-4460 CPU with 3.20GHZ and 4Gb RAM. The code was executed several times and average time of execution extracted (see Table 2). With respect to the scheme in [66], and other pairing-based constructions with a security level of 1024-bit RSA, a supersingular curve $z^2 = x^3 + x$ with an embedded degree of 2 is adopted. Also, $q = 2^{159} + 2^{17} + 1$ noted as a 160 bit Solinas prime noted as a $512 - bit$ prime. With regards to ECC-based schemes, an equivalent security level of Koblitz elliptic curve of $y = x^3 + ax^2 + b$ defined on a $F_{2^{163}}$ is used to provide the same security level in the ECC group. The computational units are in millisecond (ms) and bytes respectively. The execution times of each respective algorithm were calculated and Matlab program was used to generate Table 1 and computational results in Table 2.

Therefore, it is clear that our scheme support equality test as compared with other schemes without equality test. The computation of trapdoor and trapdoor delegation to the tester is equally achieved in our scheme. In terms of computational cost, our scheme has a lower computation cost as compared to existing schemes. Also, the support for IND-ID-CCA2 is featured in our sccheme as compared to other existing scheme with IND-CCA support. However, the inverse compuatation to recover the message achieves a ligh computational cost as compared to others in Table 1. This is pardonable due to the additional trapdoor computations featured in our scheme.

8 Conclusion

The construction $MWAR - ID - SET$ achieves a security improvement in mobile money service security in Ghana via the adoption signcryption cryptosystem. Data forgery and re-play attack is curtailed in our construction and the simultaneous benefit of PKE and digital signature is achieved in our scheme using the random oracle model. A desirable security property of EUF-CMA is achieved in our construction. In spite of other applications and extentions of identity based cryptosystem [67]–[70], our scheme achieves a remarkable achievement in indentity based cryptosystem.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment We will like to use this opportunity to thank the anonymous reviewers for their kind support and consideration.

References

[1] S. Alornyo, K. K. Mireku, A. Tonny-Hagan, X. Hu, "Mobile Money Wallet Security against Insider Attack Using ID-Based Cryptographic Primitive with Equality Test," in 2019 International Conference on Cyber Security and Internet of Things (ICSIoT), 82–87, IEEE, 2019, doi:10.1109/ICSIoT47925.2019.00021.

[2] F. Li, H. Xiong, Y. Liao, "A generic construction of identity-based signcryption," in 2009 International Conference on Communications, Circuits and Systems, 291–295, IEEE, 2009, doi:10.1109/ICCCAS.2009.5250509.

[3] K. Bentahar, P. Farshim, J. Malone-Lee, N. P. Smart, "Generic constructions of identity-based and certificateless KEMs," Journal of Cryptology, **21**(2), 178–199, 2008, doi:10.1007/s00145-007-9000-z.

[4] L. Chen, Z. Cheng, J. Malone-Lee, N. P. Smart, "An Efficient ID-KEM Based On The Sakai-Kasahara Key Construction." IACR Cryptol. ePrint Arch., **2005**, 224, 2005.

[5] E. Kiltz, D. Galindo, "Direct chosen-ciphertext secure identity-based key encapsulation without random oracles," in Australasian Conference on Information Security and Privacy, 336–347, Springer, 2006, doi:https://doi.org/10.1007/11780656_28.

[6] A. Shamir, "Identity-based cryptosystems and signature schemes," in Workshop on the theory and application of cryptographic techniques, 47–53, Springer, 1984, doi:https://doi.org/10.1007/3-540-39568-7_5.

[7] Y. Zheng, "Digital signcryption or how to achieve cost (signature or encryption) cost (signature)+ cost (encryption)," in Annual international cryptology conference, 165–179, Springer, 1997, doi:https://doi.org/10.1007/BFb0052234.

[8] A. Fiat, A. Shamir, "How to prove yourself: Practical solutions to identification and signature problems," in Conference on the theory and application of cryptographic techniques, 186–194, Springer, 1986, doi:https://doi.org/10.1007/3-540-47721-7_12.

[9] L. C. Guillou, J.-J. Quisquater, "A "paradoxical" indentity-based signature scheme resulting from zero-knowledge," in Conference on the Theory and Application of Cryptography, 216–231, Springer, 1988, doi:https://doi.org/10.1007/0-387-34799-2_16.

[10] J. Malone-Lee, "Identity-Based Signcryption." IACR Cryptol. ePrint Arch., **2002**, 98, 2002.

[11] P. S. Barreto, B. Libert, N. McCullagh, J.-J. Quisquater, "Efficient and provably-secure identity-based signatures and signcryption from bilinear maps," in International conference on the theory and application of cryptology and information security, 515–532, Springer, 2005, doi:https://doi.org/10.1007/11593447_28.

[12] X. Boyen, "Multipurpose identity-based signcryption," in Annual International Cryptology Conference, 383–399, Springer, 2003, doi:https://doi.org/10.1007/978-3-540-45146-4_23.

[13] S. S. Chow, S.-M. Yiu, L. C. Hui, K. Chow, "Efficient forward and provably secure ID-based signcryption scheme with public verifiability and public ciphertext authenticity," in International Conference on Information Security and Cryptology, 352–369, Springer, 2003, doi:https://doi.org/10.1007/978-3-540-24691-6_26.

[14] B. Libert, J.-J. Quisquater, "A new identity based signcryption scheme from pairings," in Proceedings 2003 IEEE Information Theory Workshop (Cat. No. 03EX674), 155–158, IEEE, 2003, doi:10.1109/ITW.2003.1216718.

[15] T. H. Yuen, V. K. Wei, "Constant-Size Hierarchical Identity-Based Signature/Signcryption without Random Oracles." IACR Cryptol. ePrint Arch., **2005**, 412, 2005.

[16] Y. Zheng, H. Imai, "How to construct efficient signcryption schemes on elliptic curves," Information processing letters, **68**(5), 227–233, 1998, doi:https://doi.org/10.1016/S0020-0190(98)00167-7.

[17] F. Bao, R. H. Deng, "A signcryption scheme with signature directly verifiable by public key," in International Workshop on Public Key Cryptography, 55–59, Springer, 1998, doi:https://doi.org/10.1007/BFb0054014.

[18] J.-B. Shin, K. Lee, K. Shim, "New DSA-verifiable signcryption schemes," in International Conference on Information Security and Cryptology, 35–47, Springer, 2002, doi:https://doi.org/10.1007/3-540-36552-4_3.

[19] D. H. Yum, P. J. Lee, "New signcryption schemes based on KCDSA," in International Conference on Information Security and Cryptology, 305–317, Springer, 2001, doi:https://doi.org/10.1007/3-540-45861-1_23.

[20] F. Li, M. K. Khan, "A survey of identity-based signcryption," IETE Technical Review, **28**(3), 265–272, 2011, doi:https://doi.org/10.4103/0256-4602.81236.

[21] Y. Yu, B. Yang, Y. Sun, S.-l. Zhu, "Identity based signcryption scheme without random oracles," Computer Standards & Interfaces, **31**(1), 56–62, 2009, doi:https://doi.org/10.1016/j.csi.2007.10.014.

[22] K. G. Paterson, J. C. Schuldt, "Efficient identity-based signatures secure in the standard model," in Australasian Conference on Information Security and Privacy, 207–222, Springer, 2006, doi:https://doi.org/10.1007/11780656_18.

[23] Z. Jin, Q. Wen, H. Du, "An improved semantically-secure identity-based signcryption scheme in the standard model," Computers & Electrical Engineering, **36**(3), 545–552, 2010, doi:https://doi.org/10.1016/j.compeleceng.2009.12.009.

[24] F. Li, Y. Liao, Z. Qin, "Analysis of an identity-based signcryption scheme in the standard model," IEICE transactions on fundamentals of electronics, communications and computer sciences, **94**(1), 268–269, 2011, doi:https://doi.org/10.1587/transfun.E94.A.268.

[25] F. Li, J. Gao, Y. Hu, "ID-based threshold unsigncryption scheme from pairings," in International Conference on Information Security and Cryptology, 242–253, Springer, 2005, doi:https://doi.org/10.1007/11599548_21.

[26] S. Duan, Z. Cao, R. Lu, "Robust ID-based threshold signcryption scheme from pairings," in Proceedings of the 3rd international conference on Information security, 33–37, 2004, doi:https://doi.org/10.1145/1046290.1046298.

[27] C. Peng, X. Li, "An identity-based threshold signcryption scheme with semantic security," in International Conference on Computational and Information Science, 173–179, Springer, 2005, doi:https://doi.org/10.1007/11596981_26.

[28] F. Li, Y. Yu, "An efficient and provably secure ID-based threshold signcryption scheme," in 2008 International Conference on Communications, Circuits and Systems, 488–492, IEEE, 2008, doi:10.1109/ICCCAS.2008.4657820.

[29] S. S. D. Selvi, S. S. Vivek, C. P. Rangan, N. Jain, "Cryptanalysis of Li et al.'s identity-based threshold signcryption scheme," in 2008 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, **2**, 127–132, IEEE, 2008, doi:10.1109/EUC.2008.187.

[30] F. Li, X. Xin, Y. Hu, "ID-based threshold proxy signcryption scheme from bilinear pairings," International Journal of Security and Networks, **3**(3), 206–215, 2008, doi:https://doi.org/10.1504/IJSN.2008.020095.

[31] X. Li, K. Chen, "Identity based proxy-signcryption scheme from pairings," in IEEE International Conference onServices Computing, 2004.(SCC 2004). Proceedings. 2004, 494–497, IEEE, 2004, doi:10.1109/SCC.2004.1358050.

[32] M. Wang, H. Li, Z. Liu, "Efficient identity based proxy-signcryption schemes with forward security and public verifiability," in International Conference on Networking and Mobile Computing, 982–991, Springer, 2005, doi:https: //doi.org/10.1007/11534310 103.

[33] Q. Wang, Z. Cao, "Two proxy signcryption schemes from bilinear pairings," in International Conference on Cryptology and Network Security, 161–171, Springer, 2005, doi:https://doi.org/10.1007/11599371 14.

[34] S. Duan, Z. Cao, Y. Zhou, "Secure delegation-by-warrant ID-based proxy signcryption scheme," in International Conference on Computational and Information Science, 445–450, Springer, 2005, doi:https://doi.org/10.1007/11596981-65.

[35] M. Wang, Z. Liu, "Identity based threshold proxy signcryption scheme," in The Fifth International Conference on Computer and Information Technology (CIT'05), 695–699, IEEE, 2005, doi:10.1109/CIT.2005.129.

[36] J. Zhang, J. Mao, "A novel identity-based multi-signcryption scheme," Computer Communications, **32**(1), 14–18, 2009, doi:https://doi.org/10.1016/j.comcom.2008.07.004.

[37] S. S. D. Selvi, S. S. Vivek, C. P. Rangan, "Breaking and fixing of an identity based multi-signcryption scheme," in International Conference on Provable Security, 61–75, Springer, 2009, doi:https://doi.org/10.1007/978-3-642-04642-1 7.

[38] M. Zhang, B. Yang, S. Zhu, W. Zhang, "Efficient secret authenticatable anonymous signcryption scheme with identity privacy," in International Conference on Intelligence and Security Informatics, 126–137, Springer, 2008, doi:https://doi.org/10.1007/978-3-540-69304-8 14.

[39] F.-g. Li, S. Masaaki, T. Tsuyoshi, "Analysis and improvement of authenticatable ring signcryption scheme," Journal of Shanghai Jiaotong University (Science), **13**(6), 679–683, 2008, doi:https://doi.org/10.1007/s12204-008-0679-2.

[40] Z. Zhu, Y. Zhang, F. Wang, "An efficient and provable secure identity-based ring signcryption scheme," Computer standards & interfaces, **31**(6), 1092–1097, 2009, doi:https://doi.org/10.1016/j.csi.2008.09.023.

[41] Y. Yu, F. Li, C. Xu, Y. Sun, "An efficient identity-based anonymous signcryption scheme," Wuhan University Journal of Natural Sciences, **13**(6), 670–674, 2008, doi:https://doi.org/10.1007/s11859-008-0607-1.

[42] L. Zhun, F. Zhang, "Efficient ID–based ring signature and ring signcryption schemes," in 2008 International Conference on Computational Intelligence and Security, volume 2, 303–307, IEEE, 2008, doi:10.1109/CIS.2008.51.

[43] H. Xiong, Z. Qin, F. Li, "Cryptanalysis of an Identity Based Signcryption without Random Oracles," Fundamenta Informaticae, **107**(1), 105–109, 2011, doi:10.3233/FI-2011-395.

[44] F. Li, B. Liu, J. Hong, "An efficient signcryption for data access control in cloud computing," Computing, **99**(5), 465–479, 2017, doi:https://doi.org/10.1007/s00607-017-0548-7.

[45] V. Kirtane, C. P. Rangan, "RSA-TBOS signcryption with proxy re-encryption," in Proceedings of the 8th ACM workshop on Digital rights management, 59–66, 2008, doi:https://doi.org/10.1145/1456520.1456531.

[46] J. Malone-Lee, W. Mao, "Two birds one stone: signcryption using RSA," in Cryptographers' Track at the RSA Conference, 211–226, Springer, 2003, doi:https://doi.org/10.1007/3-540-36563-X_14.

[47] C. Wang, X. Cao, "An improved signcryption with proxy re-encryption and its application," in 2011 Seventh International Conference on Computational Intelligence and Security, 886–890, IEEE, 2011, doi:10.1109/CIS.2011.200.

[48] W. Huige, W. Caifen, C. Hao, "ID-based proxy re-signcryption scheme," in 2011 IEEE International Conference on Computer Science and Automation Engineering, volume 2, 317–321, IEEE, 2011, doi:10.1109/CSAE.2011.5952478.

[49] X. Tian, X. Wang, A. Zhou, "DSP RE-Encryption: A flexible mechanism for access control enforcement management in DaaS," in 2009 IEEE International Conference on Cloud Computing, 25–32, IEEE, 2009, doi:10.1109/CLOUD.2009.65.

[50] Q. Liu, C. C. Tan, J. Wu, G. Wang, "Reliable re-encryption in unreliable clouds," in 2011 IEEE Global Telecommunications Conference-GLOBECOM 2011, 1–5, IEEE, 2011, doi:10.1109/GLOCOM.2011.6133609.

[51] S. Yu, C. Wang, K. Ren, W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," in 2010 Proceedings IEEE INFOCOM, 1–9, Ieee, 2010, doi:10.1109/INFCOM.2010.5462174.

[52] M. Li, S. Yu, Y. Zheng, K. Ren, W. Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption," IEEE transactions on parallel and distributed systems, **24**(1), 131–143, 2012, doi:10.1109/TPDS.2012.97.

[53] M. Nabeel, N. Shang, E. Bertino, "Privacy preserving policy-based content sharing in public clouds," IEEE Transactions on Knowledge and Data Engineering, **25**(11), 2602–2614, 2012, doi:10.1109/TKDE.2012.180.

[54] K. Yang, X. Jia, "Expressive, efficient, and revocable data access control for multi-authority cloud storage," IEEE transactions on parallel and distributed systems, **25**(7), 1735–1744, 2013, doi:10.1109/TPDS.2013.253.

[55] J. Hur, "Improving security and efficiency in attribute-based data sharing," IEEE transactions on knowledge and data engineering, **25**(10), 2271–2282, 2011, doi:10.1109/TKDE.2011.78.

[56] F. Li, H. Zhang, T. Takagi, "Efficient signcryption for heterogeneous systems," IEEE Systems Journal, **7**(3), 420–429, 2013, doi:10.1109/JSYST.2012.2221897.

[57] Y. Sun, H. Li, "Efficient signcryption between TPKC and IDPKC and its multi-receiver construction," Science China Information Sciences, **53**(3), 557–566, 2010, doi:https://doi.org/10.1007/s11432-010-0061-5.

[58] Q. Huang, D. S. Wong, G. Yang, "Heterogeneous signcryption with key privacy," The Computer Journal, **54**(4), 525–536, 2011, doi:10.1093/comjnl/bxq095.

[59] J. H. An, Y. Dodis, T. Rabin, "On the security of joint signature and encryption," in International Conference on the Theory and Applications of Cryptographic Techniques, 83–107, Springer, 2002, doi:https://doi.org/10.1007/3-540-46035-7_6.

[60] F. Li, T. Takagi, "Secure identity-based signcryption in the standard model," Mathematical and Computer Modelling, **57**(11-12), 2685–2694, 2013, doi:https://doi.org/10.1016/j.mcm.2011.06.043.

[61] L. Chen, J. Malone-Lee, "Improved identity-based signcryption," in International Workshop on Public Key Cryptography, 362–379, Springer, 2005, doi:https://doi.org/10.1007/978-3-540-30580-4_25.

[62] B. Waters, "Efficient identity-based encryption without random oracles," in Annual International Conference on the Theory and Applications of Cryptographic Techniques, 114–127, Springer, 2005, doi:https://doi.org/10.1007/11426639_7.

[63] Y. Ming, Y. Wang, "Cryptanalysis of an Identity Based Signcryption Scheme in the Standard Model." IJ Network Security, **18**(1), 165–171, 2016, doi:10.6633/IJNS.

[64] G. Zhu, H. Xiong, Z. Qin, "Fully secure identity based key-insulated signcryption in the standard model," Wireless personal communications, **79**(2), 1401–1416, 2014, doi:https://doi.org/10.1007/s11277-014-1936-3.

[65] B. Lynn, et al., "The stanford pairing based crypto library," Privacy preservation scheme for multicast communications in smart buildings of the smart grid, **324**, 2013.

[66] H. Xiong, Q. Mei, Y. Zhao, "Efficient and provably secure certificateless parallel key-insulated signature without pairing for IIoT environments," IEEE Systems Journal, **14**(1), 310–320, 2019, doi:10.1109/JSYST.2018.2890126.

[67] S. Alornyo, M. Asante, X. Hu, K. K. Mireku, "Encrypted Traffic Analytic using Identity Based Encryption with Equality Test for Cloud Computing," in 2018 IEEE 7th International Conference on Adaptive Science & Technology (ICAST), 1–4, IEEE, 2018, doi:10.1109/ICASTECH.2018.8507063.

[68] S. Alornyo, E. Aidoo, K. K. Mireku, B. Kwofie, X. Hu, M. Asante, "ID-Based Outsourced Plaintext Checkable Encryption in Healthcare Database," in 2019 International Conference on Cyber Security and Internet of Things (ICSIoT), 48–53, IEEE, 2019, doi:10.1109/ICSIoT47925.2019.00016.

[69] S. Alornyo, Y. Zhao, G. Zhu, H. Xiong, "Identity Based Key-Insulated Encryption with Outsourced Equality Test." IJ Network Security, **22**(2), 257–264, 2020, doi:10.6633/IJNS.860.

[70] S. Alornyo, A. E. Mensah, A. O. Abbam, "Identity-based Public Key Cryptographic Primitive with Delegated Equality Test Against Insider Attack in Cloud Computing," International Journal of Network Security, **22**(5), 743–751, 2020, doi:10.6633/IJNS.202009_22(5).04.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 995-1000 (2020)

www.astesj.com

Special Issue on Innovative Research in Applied Science, Engineering and Technology

ASTESJ

ISSN: 2415-6698

Advanced Design of Current-mode Pass-band Filter using Ant Colony Optimization Technique

Kritele Loubna[1,*], Benhala Bachir[2], Zorkani Izeddine[1]

[1]*Faculty of Sciences Dhar el Mahraz, University of Sidi Mohamed Ben Abdllah, Fez, 30050, Morroco*

[2]*Faculty of Sciences, University of Moulay Ismail, Meknes, 50000, Morroco*

ARTICLE INFO

Article history:
Received: 02 September, 2020
Accepted: 02 November, 2020
Online: 08 December, 2020

Keywords:
Ant colony optimisation
Analog circuit design
Positive second-generation current conveyor

ABSTRACT

Ant Colony Optimization (ACO) algorithm, a well-known robust technique to solve easily both a simple and multiple objective optimization problems. This article presents an application of the ACO in order to achieve the optimal sizing of analog circuit. The proposed technique is employed to optimizing the sizing of a positive second-generation current conveyor (CCII+). Results show better objective functions than previously achieved by other optimization procedure. PSPICE simulations are used to confirm the validity of the reached optimal results and the accuracy of the proposed procedure.

1. Introduction

Current Conveyor (CC) represents the first construction block developed for current signal processing [1]. Two years later, in 1970, another variant of a CC, namely, Second Generation Current Conveyor (CCII) was published [2].

Recently, a Current Conveyor is seen as a good alternative for the voltage-mode. CC can be widely employed in different applications. The CCII is the most popular configuration [3].

Despite the number of positive proprieties ensured by the CCS [4], it seems that designing integrated CC circuits with high performance is in effect still open particularly in CMOS technology [5].

Currently, an important interest has been devoted to achieve the best sizing of the analog component efficiently and with high accuracy. Some (meta) heuristics were used in the scientific publications such as Genetic Algorithms (GA) [6, 7], Ant Colony Optimization (ACO) [8, 9], Particle Swarm Optimization (PSO) [10], Differential Evolution algorithm (DE) [11], Simulated Annealing (SA) [12] and Artificial Bee Colony Algorithm (ABC) [13, 14].

ACO algorithm is proposed in this work, with view to get an optimal design of positive second-generation current conveyor (CCII+). To this end, we applied three of the most important variants of ACO, specifically, the Ant System (AS), the Min-Max Ant System (MMAS) and the Ant Colony System (ACS). In fact, the ACO has already efficiently been used to treat real optimization problems, particularly in the analog circuit area.

A high performance CCII+ depends on the high-end cut-off frequency for the current waveform (F_{ci}) and the input impedance at port X. Hence, two objectives functions are taken into consideration: achieve a low input resistance R_X and a very high cut-off frequency F_{ci}. However, the suggested technique is a mono-objective one, to this end; the optimization process is treated using two distinct manners; firstly, each objective function is solved separately. In the second way, a mono-objective problem is considered by using a weighting technique.

This paper is organized in the following manner: The second section is about a detail description of the ACO technique. The third section proposes an example for the optimal sizing of CCII+. Then, a comparison with some similar works and simulation results are represented in the fourth section. Through the fifth section, the optimized CCII+ is exploited in order to form a current mode filter. Finally, the sixth section is the conclusion.

2. The ACO Technique Presentation

Ant Colony Optimization (ACO) is a nature-inspired algorithm, used to resolve various hard problems. The use of ACO has proved a capacity to deal with complex optimization problems successfully, for instance we can cite the traveling salesman

*Corresponding Author: Kritele Loubna, loubnakritele@gmail.com

problem (TSP) [15], vehicle routing problem [16], the quadratic assignment problem (QAP) [17].

The ACO takes inspiration from the real ant colonies behavior. The fundamental idea of this metaheuristic consists in simulating the social intelligence of ants in searching the shortest road from the nest and a source of food. In fact, real ants communicate indirectly through trails of chemical pheromone. Concretely, the agents of the colony deposit pheromone on the road to determine certain favorable route to be pursued by other ants [18].

2.1. Ant System

« Ant System » (AS) is the basic variant of the ACO. The modelling approach to resolve a given optimization problem is established as follows:

Ants choose the vertex to travel to by applying the state transition rule [19, 20]. Ant k in node a will move to vertex b according to the probability offered by (1):

$$P_{ab}^{k}=\begin{cases}\dfrac{(\tau_{ab})^{\alpha}\cdot(\eta_{ab})^{\beta}}{\sum_{l\in J_a^k}(\tau_{ab})^{\alpha}\cdot(\eta_{ab})^{\beta}} & \text{if } a\in J_a^k\\ 0 & \text{if } a\notin J_a^k\end{cases} \quad (1)$$

while J_a^k corresponds to the list of possible moves, which are possible for ants starting form vertex a, τ_{ab} represents the intensity of the trace associated to brink (a, b). α and β identify the relative impact of the trace intensity and the visibility value, η_{ab}.

where

$$\eta_{ab}=\frac{1}{d_{ab}} \quad (2)$$

The d_{ab} represents the distance between nodes a and b

The pheromone is updated by means of formula (3), when all agents of the colony construct their path:

$$\tau_{ab}=(1-\rho)\cdot\tau_{ab}+\sum_{k=1}^{m}\Delta\tau_{ab}^{k} \quad (3)$$

While ρ corresponds to the pheromone decay parameter, m represents the number of ants, and $\Delta\tau_{ab}^{k}$ is the amount of pheromone deposited by ant k on brink (a, b):

$$\Delta\tau_{ab}^{k}=\begin{cases}\dfrac{Q}{L_k} & \text{if } (a,b) \text{ belongs to tour done by k}\\ 0 & \text{otherwise}\end{cases} \quad (4)$$

Q is a constant; L_k represents the tour length built by ant k.

2.2. Min-Man Ant System

The Min-Max Ant System is an adjustment of the AS [21].

Min-Max Ant System (MMAS) is designed to attain the best result. To this end, only the optimal result is authorized to intensify the pheromone quantity. In fact, MMAS uses an instrument to border the trace intensity, and so that avoid the premature stagnation.

MMAS presents some new features and they are as follows:

- Only the ants that produced the best solution are allowed to update trails;
- To avoid stagnation, MMAS introduces upper and lower bounds so that the pheromone intensities lie inside the interval $[\tau_{min}, \tau_{max}]$;
- In each brink, the trail strength is initialized to τ_{max}.

2.3. Ant Colony System

The Ant Colony System (ACS) algorithm incorporates three main differences with respect to the AS algorithm:

- ACS algorithm uses a parameter q_0 ($0\le q_0\le 1$), which controls the tuning between exploitation and exploration. thus, the probability that ant moves to node b is given by this rule:

$$b=\begin{cases}\arg\max_{u\in B_a^k}\left[\left(\tau_{au}(t)*(\eta_{aB})^{\beta}\right)\right] & \text{if } q\le q_0\\ (1) & \text{if } q\succ q_0\end{cases} \quad (5)$$

where q is a random variable evenly distributed in [0, 1]

- The global pheromone update is applied only to edges that form part of the best solution, as in the following:

$$\tau_{ab}=(1-\rho)\cdot\tau_{ab}+\rho\cdot\Delta\tau_{ab} \quad (6)$$

where

$$\tau_{ab}=\begin{cases}1/L & \text{if } (a,b)\in best-global-tour\\ 0 & \text{otherwise}\end{cases} \quad (7)$$

- The updating of local pheromone is applied when each ant produce a solution:

$$\tau_{ab}=(1-\rho)\cdot\tau_{ab}+\rho\cdot\tau_{int} \quad (8)$$

The procedure of the ACO technique can be illustrated by means of the flowchart bellow:

3. Application of the ACO to the optimal sizing of CCII+

CCII is the most popular configuration of CCs; it replaces the conventional OPAMP in different applications as for example active filters, oscillators, and analog signal processing.

The CCII is three terminal devices; the current conveyor's response is illustrated by expression (9) [22]:

$$\begin{pmatrix}I_Y\\V_X\\I_Z\end{pmatrix}=\begin{pmatrix}0&0&0\\1&0&0\\0&\pm\beta&0\end{pmatrix}\begin{pmatrix}V_Y\\I_X\\V_Z\end{pmatrix} \quad (9)$$

The voltage at node X is duplicated from the voltage applied at node Y. Similarly, the current at Z port is copied from the current flowing through node X. A positive Z output current conveyor is considered when $\beta=+1$ and it is denoted CCII+. For optimal

circuit performance, it is desirable to have very high impedance on Y and Z nodes and small Rx, i.e. the resistance on X [22].

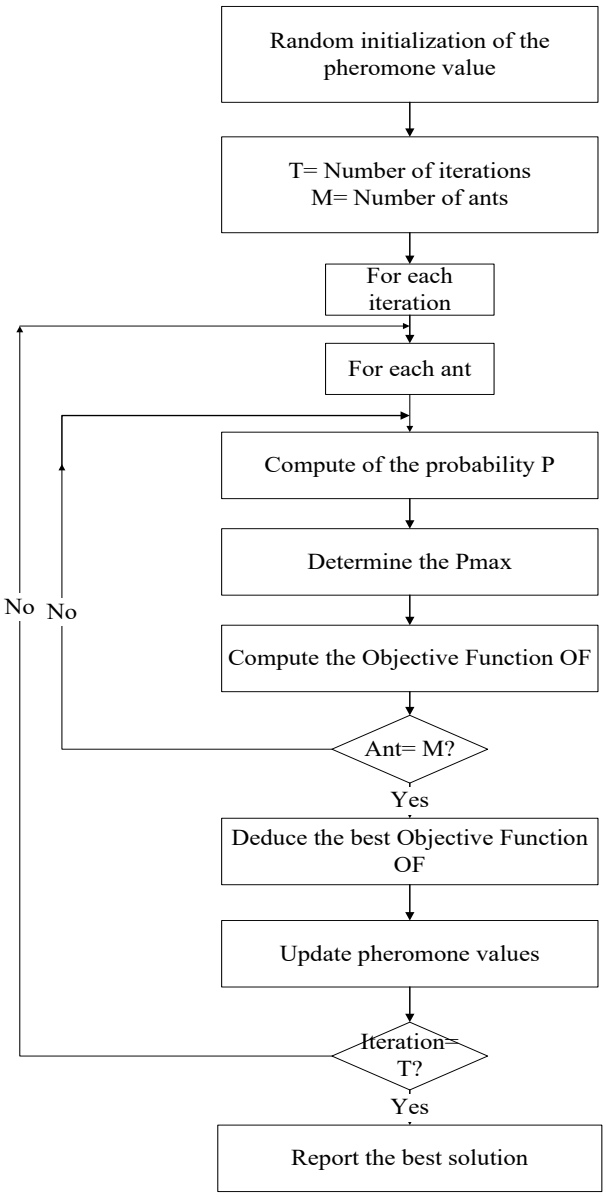


Figure 1: Flowchart of ACO

In this work, the ACO is used to optimize the performances of a positive second-generation CMOS current conveyor (CCII+) (figure2). Optimizing the most affecting parameters of the current conveyor is highlighted:

This optimization problem is resolved using two methods; in the first one, two mono objective function are considered. The objective functions are:

- Minimize the Rx, which represents the X-port input parasitic resistance

$$R_x \approx \frac{1}{gm_2 + gm_4} \tag{10}$$

- Maximize the Fci, which represents the current high cut off frequency

$$F_{ci} = \frac{1}{2\Pi}\sqrt{\frac{gm_5(gm_4 + go_4)}{Cgs_4(Cgs_5 + Cgs_6 + Cgd_4)}} \tag{11}$$

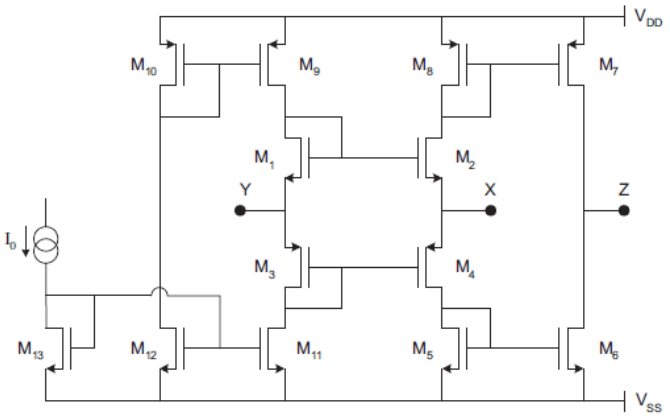


Figure 2: A positive second generation current conveyor

In the second method, the proposed objective function allows simultaneously minimizing Rx and maximizing Fci, and it is configured as:

$$f_{obj} = \frac{1}{F_{ci}} + R_X \tag{12}$$

The MOS transistors size represents the geometric variables that will be considered: the gates widths (WN and WP) and channels lengths (LN, and LP). The binding functioning conditions require that all the transistors of CCII+ must function in the saturation mode. So that, the main constrain of the optimization problem are presented by expressions (13) and (14):

$$\frac{V_{DD}}{2} + V_{TP} + \sqrt{\frac{2I_0}{\mu_N C_{ox}\frac{W_2}{L_2}}} \geq \sqrt{\frac{2I_0}{\mu_P C_{ox}\frac{W_8}{L_8}}} \tag{13}$$

$$\frac{V_{DD}}{2} - V_{TN} - \sqrt{\frac{2I_0}{\mu_N C_{ox}\frac{W_5}{L_5}}} \geq \sqrt{\frac{2I_0}{\mu_P C_{ox}\frac{W_4}{L_4}}} \tag{14}$$

where C_{ox}, μ_N, and μ_P are MOS technology parameters. V_{TN} and V_{TP} denote threshold voltages of NMOS and PMOS transistors respectively. The employed supply voltages are V_{SS}/V_{DD}= -2.5V/2.5 V and a bias current, Ibias=100 µA.

4. Results and Discussion

The proposed variants of the ACO technique were applied to deal with the optimal sizing of CCII+. The optimization technique works on C++ codes.

4.1. Optimization problem result using tow separated objective functions

The optimal values for the parameters of CCII+ achieved by the ACO algorithms and the circuit's performances linked with these values are indicated in tables 1 and 2.

Table 3 presents the performances provided by the AS, the MMAS and the ACS techniques versus to the ones ensured by Particle Swarm Optimization (PSO) and the Genetic Algorithm (GA) algorithms.

Table 1: the obtained results for RX

	LN (µm)	WN (µm)	LP (µm)	WP (µm)	$R_{X\text{-}min}$ (Ω)	
					Opt.	Sim.
AS	0.59	36.60	0.35	58.64	333.01	315.89
MMAS	0.60	36.34	0.35	59.00	335.63	318.69
ACS	0.59	36.68	0.35	59.00	**332.65**	**315.19**

Table 2: the obtained results for Fci

	LN (µm)	WN (µm)	LP (µm)	WP (µm)	Fci-max (GHz)	
					Opt.	Sim.
AS	0.55	5.48	0.35	9.52	1.795	1.739
MMAS	0.55	5.60	0.35	9.60	1.775	1.731
ACS	0.55	5.44	0.35	9.50	**1.801**	**1.754**

The results of the simulation under SPICE for Rx and Fci employing the optimal results acquired by the suggested techniques for CCII+ are exposed in figures 3 and 4. We can observe that simulation results are in good correspondence with theoretical values achieved by the ACO technique.

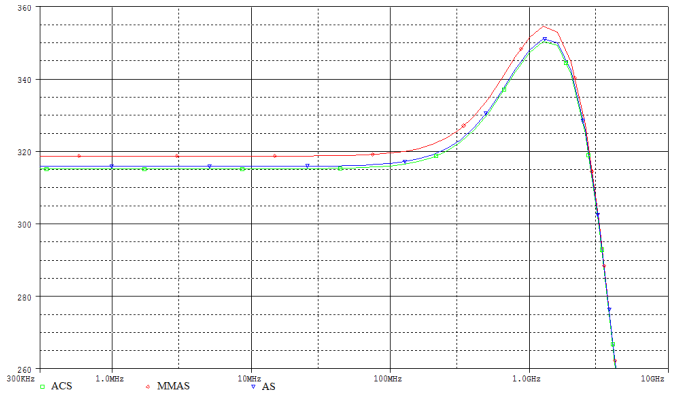


Figure 3: Variation of resistance (ohm) versus frequency (Hz) employing AS,MMAS and ACS

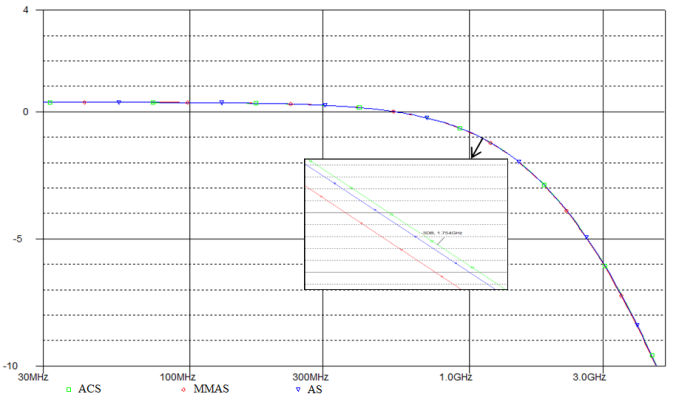


Figure 4: Current gain (dB) versus frequency (Hz) employing AS,MMAS and ACS

Table 3: Comparison results

	ACS	PSO[23]	GA[23]
R_{X}-min (Ω)	**332.65**	464	510
F_{ci}-max (GHZ)	**1.801**	1.751	1.541

It is shown that the three variants of the ACO algorithm offer similar results. The comparison with two techniques belonging to metaheuristics indicates that the ACS accomplished the greatest results with regard to objectives compared to Particle Swarm Optimization (PSO) and the Genetic Algorithm (GA) algorithms.

4.2. Optimization problem result using weighting approach:

The optimal physical parameters values of MOS transistors achieved employing the three variants of ACO algorithm are given in table 4. The comparison of the acquired results with those obtained by another metaheuristic namely, particle swarm optimization (PSO) algorithm, shows that ACS could reach a greatest F_{ci} value. However, PSO could achieve a smaller R_X value.

Table 4: Optimal parameter values and optimization results for R_X and Fci

	LN (µm)	WN (µm)	LP (µm)	WP (µm)	$F_{ci\text{-}max}$ (GHz)	$R_{X\text{-}min}$ (Ω)
AS	0.58	7.95	0.35	13.00	1.446	711.42
MMAS	0.59	8.08	0.35	13,00	1.420	708.75
ACS	0.55	7.53	0.35	12.97	**1.531**	721.08
PSO-2S [24]	0.52	14.30	0.35	23.62	1.358	**444**

The following table presents the simulation results:

Table 5: The optimal results for Rx and Fci

	F_{ci_max} (GHz)		R_{X_min} (Ω)	
	Opt.	Sim.	Opt.	Sim.
AS	1.446	1.495	711.417	825.896
MMAS	1.420	1.479	708.75	825.896
ACS	1.531	1.552	721.085	827.704

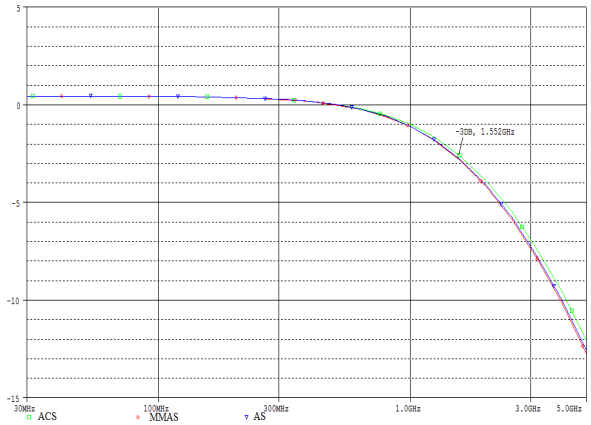


Figure 5: Current gain (dB) as a function of frequency (Hz) employing AS,MMAS and ACS

The figure 5 gives the SPICE simulation for current gain (in dB) as a function of frequency, and the figure 6 presents the SPICE simulation results for R_X (in ohm) as a function of frequency and this by employing the acquired optimal values by the three proposed variants of ACO algorithm for CCII+.

5. A Current Mode Filter based on the CCs

A second order filter can be implemented based on the use of simulated inductance [25]. In fact, Current conveyors present the basis of simulated inductance topology. Figure 7 shows a current mode active filter constituted of 4 current conveyors and two capacitors [26].

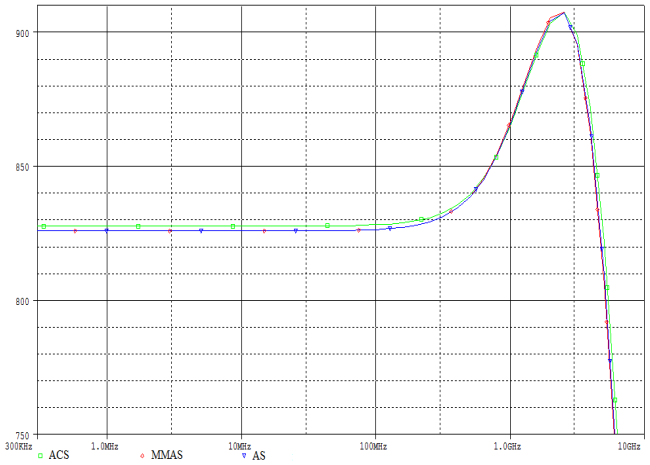


Figure 6: Variation of Rx (ohm) as a function of frequency (Hz) employing AS,MMAS and ACS

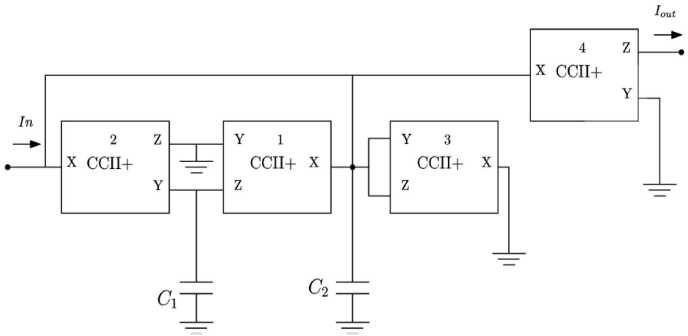


Figure 7: Current- mode pass-band filter

The quality factor and the resonance frequency of this filter are respectively presented by:

$$Q = \frac{R_G}{\sqrt{R_{X1}R_{X2}}}\sqrt{\frac{C_1}{C_2}} \tag{15}$$

$$f_0 = \frac{1}{2\pi}\frac{1}{\sqrt{R_{X1}R_{X2}C_1C_2}} \tag{16}$$

where:

$$\frac{1}{R_G} = \frac{1}{R_{X1}} + \frac{1}{R_{X2}} + \frac{1}{R_{X3}} + \frac{1}{R_{X4}} \tag{17}$$

R_{X1}, R_{X2}, R_{X3} and R_{X4} represent the parasitic resistances on port X of the considered CCII+.

The considered filter is simulated using SPICE, and it is designed based on the optimized CCII+ with optimal sizes determined by ACS for this cases:

a. Firstly, it is designed by employing the obtained CCII+ offering the maximum F_{ci}.
b. Secondly, it is designed by employing the obtained CCII+ offering the minimum R_X.
c. Finally, it is designed by employing the obtained CCII+ offering simultaneously maximum F_{ci} and minimum R_X.

The optimal parameters values of MOS transistors for the three cases are summarized in the table 6.

Table 6: Optimal physical parameters values of MOS transistors

	Minimum R_X	Maximum F_{ci}	Minimum R_X and Maximum F_{ci}
WN/ LN (µm)	36.68/0.59	05.44/0.55	7.53/0.55
WP/ LP (µm)	59.00/0.35	09.50/0.35	12.97/0.35

Figure 8 shows the SPICE simulation of the filter gain, where C1=C2=0.1pF.

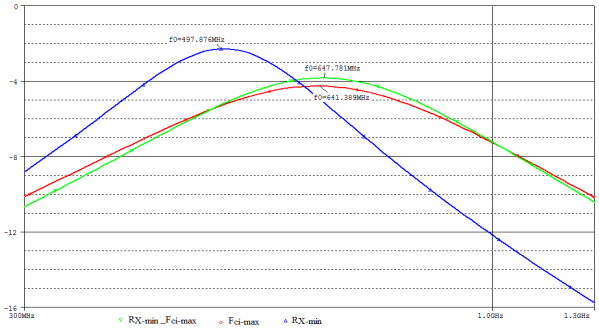


Figure 8: Resonance frequency of the second order current mode band-pass filter

The resonance frequencies reached by the ACS are presented in the following table; we can observe that the best resonance frequency of the current mode second order band pass filter is recorded in the third case.

Table 7: The obtained resonance frequencies

	$R_{X\text{-}min}$	$F_{ci\text{-}max}$	$R_{X\text{-}min}$ and $F_{ci\text{-}max}$
f_o (MHz)	497.896	641.389	647.781

6. Conclusion

In this work, the Ant System (AS), the Min-Max Ant System (MMAS) and the Ant Colony System (ACS), which are the most important variants of Ant Colony Optimization technique, are applied in order to optimize the sizing of conventional CMOS CCII circuit. The performances of a CCII+, especially the current high cut off frequency and the X-port parasitic resistance are optimized. The acquired results prove that the ACO algorithm offers a competitive optimizing capability for the design of CCII+. Then, the improved CCII+ structure was utilized as an elementary

block to build current mode band -pass filter. SPICE simulations and comparison with already published works were highlighted.

References

[1] K. C. Smith, A. Sedra. "The Current Conveyor-A new circuit building block" Proceedings of the IEEE, **56**(8), 1368-1369, 1968. DOI: 10.1109/PROC.1968.6591

[2] A. Sedra, K. C. Smith, "A Second-Generation Current Conveyor and Its Applications" IEEE Transactions on Circuits Theory, **17**(1), 132-134, 1970. DOI: 10.1109/TCT.1970.1083067

[3] D. Moro Frias, "Design and Applications of CMOS Current Conveyors," Master in electronics science, National Institute for Astrophysics, Optics and Electronics, 2008.

[4] P. Brandstetter, L. Klein, "novel wien bridge oscillator design using functional block structure with current conveyors", Advances in electrical and electronic engineering, **10**(1), 13-16, 2012. DOI: 10.15598/aeee.v10i1.556

[5] A. Chatterjee, M. Fakhfakh, P. Siarry, "Design of second-generation current conveyors employing bacterial foraging optimization", Microelectronics Journal, **41**(10), 616-626, 2010. doi:10.1016/j.mejo.2010.06.013

[6] J. B. Grimbleby, "Automatic analogue circuit synthesis using genetic algorithms", IEE Proceedings-Circuits, Devices and Systems, **147**(6), 319-323, 2000. DOI: 10.1049/ip-cds:20000770

[7] A. El Beqal, B. Benhala, and I. Zorkani, A Genetic algorithm for the optimal design of a multistage amplifier, Int. J. of Electrical and Computer Engineering (IJECE), **10**(1), 129-138, 2020. DOI: 10.11591/ijece.v10i1

[8] B. Benhala. Sizing of an inverted current conveyors by an enhanced ant colony optimization technique, Conference on Design of Circuits and Integrated Systems (DCIS), Granada, 1–5, 2016.

[9] L. Kritele, B. Benhala, and I. Zorkani. Ant Colony Optimization for Optimal Analog Filter Sizing, Chapter 10, Book: Focus on swarm intelligence research and applications, Eds., B. Benhala, P. Pereira and A. Sallem, NOVA Science Publishers, 193–220, 2017.

[10] F. T. S. Chan, M. K. Tiwari, Swarm Intelligence: focus on ant and particle swarm optimization, I-Tech Education and Publishing, 2007

[11] I. El hajjami, B. Benhala, and H. Bouyghf. Optimal Design of RF Integrated Inductors via Differential Evolution Algorithm, 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, 1-6, 2020.

[12] M. Fakhfakh, M. Boughariou, A Sallem and M Loulou, Design of Low Noise Amplifiers through Flow-Graphs and their Optimization by the Simulated Annealing Technique. Book: Advances in Monolithic Microwave Integrated Circuits for Wireless Systems: Modeling and Design Technologies, IGI global, 2012.

[13] B. Benhala, P. Pereira, A. Sallem, (Editors), Focus on swarm intelligence research and applications, NOVA Science Publishers, 2017.

[14] B. Benhala. Artificial bee colony technique for optimal design of folded cascode OTA, 2018 International Conference on Applied Mathematics and Computer Science, ICAMCS 2018, Paris, 59-64.

[15] Y. Jinhui, S. Xiaohu , M. Maurizio and L.Yanchun, "An ant colony optimization method for generalized TSP problem", Progress in Natural Science, **18**(11), 1417-1422, 2008. doi: 10.1016/j.pnsc.2008.03.028.

[16] Qi Chengming, "Vehicle Routing Optimization in Logistics Distribution Using Hybrid Ant Colony Algorithm", TELKOMNIKA Indonesian Journal of Electrical Engineering, **11**(9), 5308-5315, 2013. doi: 10.11591/telkomnika.v11i9.3284.

[17] L. M Gambardella, E. D Taillard, M. Dorigo. Ant colonies for the quadratic assignment problem. The Journal of the Operational Research Society, **50**(2), 167–176, 1999. doi:10.1057/palgrave.jors.2600676.

[18] M. Dorigo, M. Birattari, and T. Stützle, "Ant colony optimization: artificial ants as a computational intelligence technique", IEEE computational intelligence magazine, **1**(4), 28-39, 2006. doi:10.1109/CI-M.2006.248054.

[19] M. Dorigo and S. Krzysztof, An Introduction to Ant Colony Optimization, IRIDIA, 2006. doi: 10.1162/106454699568728.

[20] M. Dorigo, G. DiCaro and L. M. Gambardella, "Ant algorithms for discrete optimization", Artificial Life Journal, **5**(2), 137-172, 1999.

[21] T. Stützle, H. Hoos, "MAX-MIN Ant System", Future Generation Computer System, **16**(2000), 889-914.

[22] V.C. Gaudet, "Design of a CMOS current conveyor-based field-programmable analog array", Master of applied science, Department of electrical and cornputer engineering, University of Toronto, 1997.

[23] B. Benhala, "Metaheuristic Techniques for the Analog Circuits Performances Optimization-A comparison Issue", WSEAS Transactions on Circuits and Systems, **14**(20), 173-181, June 2015,

[24] A.El Dor, M. Fakhfakh, P. Siarry, "Performance optimization of CMOS second generation current conveyors using a multi-swarm algorithm", International Journal of Electronics and Communications (AEÜ), **8**(68), 496-503, 2014. DOI: 10.1016/j.aeue.2013.12.007

[25] A. Fabre, O. Saiid, F. Wiest, and C. Bouchern, "High frequency high-Q bICMOS current-mode bandpass filter and mobile communication application", IEEE Transactions on Circuits and Systems-I, **33**(4), 839–842, 1998. DOI: 10.1109/4.663567

[26] S.B. Salem, M.Fakhfakh, D.S.Masmoudi, M.Loulou, P.Loumeau and N.Masmoudi, "A High Performances CMOS CCII and High Frequency Applications", Journal of Analog Integrated Circuits and Signal Processing, Springer US, **49**(1), 2006. doi.org/10.1007/s10470-006-8694-4

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 1001-1006 (2020)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

Feature Gate Computational Top-Down Model for Target Detection

Aarthi Ramachandran *, Amudha Joseph, Shunmuga Velayutham

Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore & Bengaluru, Amrita Vishwa Vidyapeetham, India

ARTICLE INFO

Article history:
Received: 22 August, 2020
Accepted: 10 November, 2020
Online: 08 December, 2020

Keywords:
Target search
Top-Down model
Human detection
Moving object detection

ABSTRACT

Computer vision is a technique used for processing images and videos which are increasingly becoming ubiquitous day by day. Technologies developed are revolving around human needs and demands high computational power as volume of data increases. The extraction of the necessary information for processing, that is independent of various scene complexity is a challenging task. Computation visual attention methods channelize a way to select the information using psychological studies on the human visual system. This work aims to develop a computational visual attention method to select the target in a scene. Feature Gate Top-Down model is proposed to filter the significant region of a target in the scene. The proposed model is extended as a choice-based system to detect target or salient movement in surveillance videos. Experimental analysis is performed on various scenes for detecting human as a target followed by the analysis on surveillance videos is evaluated. The metrics such as Kullback-Leibler divergence (KL div), Normalized Scanpath Saliency (NSS), Correlation Coefficient (CC) and similarity reveals that the proposed model is more adaptive in identifying the target region by suppressing other dominant objects.

1 Introduction

In this era of technology, many human activities largely demand on various kind of information processing by machines. The simplification and accessibility of devices make the accumulation of data that is increasing day-by-day. Consequently, modern computing methods focus on processing such voluminous data especially images and videos. The demand is to manipulate and maintain the data in an intelligent way and to extract the information whenever required. The computer vision methodologies provide a solution to the problem in terms of image labelling, classification and compression etc. Many applications focus on human-centric technology especially detecting the target in the image or video for further manipulation. Target detection is a challenging problem, due to variations in natural factors like background, illumination and posture of object. So the thrive for the robust target detection algorithm is continuing as an active area of research. In this work, the effectiveness of adapting visual attention technologies in target detection is proposed. Visual attention methods are derived by analyzing the behaviour of the human eye while processing the information that is in form of image or video. The computational models devised from the hypothesis of the visual system have often helped to improve and simplify the task of many vision applications. Visual attention model has been developed and studied through two approaches viz. Bottom-Up (BU) and Top-Down (TD) approach. BU approaches detects the salient region in the image and TD approaches detects the regions based on intention. As target search as intention in the scene, object may or may not be salient in the scene. Hence the proposed work is on TD computational model that identifies the significant regions of target in an image. The organization of the paper is as follows: Section 2 presents the preliminaries of the visual attention mechanism and human detection system. Section 3 proposes computational Top-Down model. Section 4 discusses experimental design and simulation results. Conclusion is given in section 5.

2 Related Work

The proposed method is demonstrated to detect the human as a target in the scene. In [1] surveyed many models that detect all moving objects and classified humans using features like shape, texture etc. The state-of-the-art human detection methods were analyzed in [2] and the advantages of each method with a guide to the choice for applications were briefly discussed. The proposed visual attention techniques detect both target in the scene as well as moving object. In psychological attention theories, the BU is handled by early vision regions where the component is fast and instantaneous. There are models developed to extract BU component by extending theories and image-based techniques. Saliency detection based on

*Corresponding Author: Aarthi Ramachandran, Email: r_aarthi@cb.amrita.edu

regional contrast that can simultaneously evaluate global contrast differences and spatial coherence was discussed in [3]. Histogram-based Contrast(HC) method [4] was used to define saliency values for image pixels using colour statistics of the input image. The method extended to Colour Co-occurrence Histogram (CCH) that captures saliency based on location information and statistics of image pixels were composed into a visually perceivable image. It was inspired by the behaviour of the human visual system and the fact that visual attention was driven by the low-level stimulus. The images were analyzed in various resolutions, then the maps were computed on multiple low-level features like intensity, colour, orientation, and texture. The top-down component was highly dependent on the task, the human visual system, attention could be focused volitionally to cues determined by the current task (e.g., looking for something). In [5], the author derived the TD (Top Down) map generated using the decision tree classifier on the test date. A fuzzy rule-based system was used in [6], [7] and [8], where the feature is used for the specific image property (Intentional features) introduced in fuzzy inference system on learned features. A Visual Attention System for Object Detection and Goal-Directed Search (VOCUS) [9] model derived the TD map by tuning the features of the target region in the map. The weights were calculated from the ratio of mean value between the Region of Interest (RoI) and the surrounding region. The contextual information of the object was used in many works as TD component. The method in [10] method used spatial information on global context information to detect the target. The gist of the scene was determined to categorize as indoor and outdoor and to perform the task based on that. In [11] learned the behaviour of human eye fixations while playing a video game mapped to scene as TD model. Natural Language Processing (NLP) and then interfaced with vision using Language Perceptional Translator (LPT) for parsing the sentence and to extract the corresponding properties of an object like location, colour, size and shape for the object. Few features that influence the visual search task were context, features and background [12]. Context was unimportant for the detection of human due to its dynamic behaviour in surveillance scenes. Features that are highly related to the target was elicited other than the distractor feature. Brain system had functional areas as a pre-attentive stage where the parallel processing of basic visual cues were performed and subsequently followed by the region where complex-operation such as recognition was performed [13]. In Top-Down model the location that satisfied the target features from inhibiting the other distracting region was favoured. Boolean map model [14] divided the region as selected and not selected by sequentially producing Boolean map considering one feature at a time. It combined the result of one feature that becomes the input for searching the other feature. The main contribution of this work is the development of computational visual attention method to select the target in a scene. We have proposed a model namely feature gate model using biological theories, that makes use of the influence of target features and integrates the distinctive features for filtering the target. Then the model is extended to video by integrating with moving object detection algorithm. The method is analysed with map generated without intention and prominence of a target concerning surrounding environment complexity. The proposed method is also compared with other existing models to evaluates its performance.

3 Feature Gate Top Down Model

Feature gate model [15], terms derived from visual attention system as gate control the flow of information from each level of the hierarchy to the next. The gating of each location depends on the features present. It is a fast and efficient method for inhibiting distractor locations in the search task. The specific aim of the work is:
1. Implement the Feature Gate model concept for effective feature combination of target detection.
2. Evaluating the influence of pre-learned feature extending to video processing. The results are analysed for visual target search application, where the choice for selecting the target region in frames is also included.

The system architecture is presented in Figure 1. The sequence of 'k' frames(fr) from fr_{i-k}, fr_{i-k+1},... fr_i are considered to be entitled as processed in the system. The ith frame (fr_i) in the video sequence is considered for finding static TD map. The low-level features like colour, orientation and intensity are tuned to pops out the target or salient region as static map. The proposed Feature Gate Top-Down model (FGTD) is discussed in subsection 3.1 and pop out the probable location of the target(human) as TD map. The 'k' frames are used for finding the potential region of moving target in the scene by entropy-based moving object detection discussed in section 3.2. The saliency map is calculated from low-level feature by procedure discussed in [16]. The model has the flexibility to select a specific target in the video or it can detect the potential salient moving object. Hence, the weights W_m and W_s is help to increase or decrease the proposition of selecting the target. The final map generated as discussed in subsection 3.3.

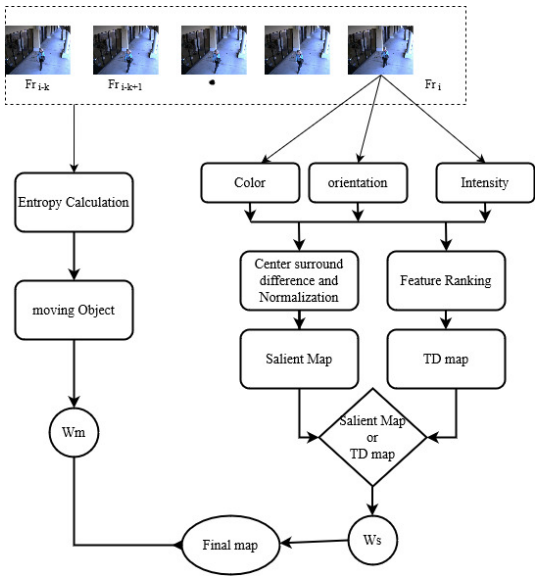


Figure 1: Overview of Proposed model

3.1 TD Map Generation

The human brain fires the neurons based on the task that need to be coordinated. In case of developing a TD model, it requires to enable the relevant low-level feature information. The feature related and necessary for the target passes through the open gate and other features are inhibited by the closed gate. The features that

are passed through the open gate will be kept in the output and others are restricted from appearing in the output. Some of the pre-defined set of local features are learned in the model. Color, intensity, and orientation play a significant role in determining the salient regions in images. The architecture for FGTD model is given in Figure 2. The details of the target and environment to be searched are given as input to the system. The visual feature and salient features are considered for evaluation of TD Information. These features are well suited for target search concerning various background. Features are expressed as a feature map and then combined to form conspicuity maps. Feature map is mapped to form a sub-conspicuity map. The features are learned from low-level features in different scales. The input image is broken down into intensity, colour and orientation maps. Four broadly-tuned color channels Red(R), Green(G), Blue(B) and Yellow(Y) are created. Gaussian pyramids R(s), G(s), B(s), and Y(s) are created from these four colour channels. Centre-surround differences (Θ) between a "centre" fine-scale c and a "surround" coarser scale s yield the feature maps as represented in equation 1. Scales are obtained from pyramid with center c= {2,3,4} and surround scale S= {3,4}. Feature maps across color, orientation and intensity are computed using Itti model [16]. Conspicuity feature map of intensity, color opponent as Red-Green (RG) in (1) and orientation in (2) of target region are also calculated. Three-level orientation feature for θ={0°,45°,90°,135°} of target region are calculated. Gabor filter is used to obtain different orientation features. The mean of features is extracted from the selected target region of the image to create training samples.

$$RG(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))| \quad (1)$$

$$O(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)| \quad (2)$$

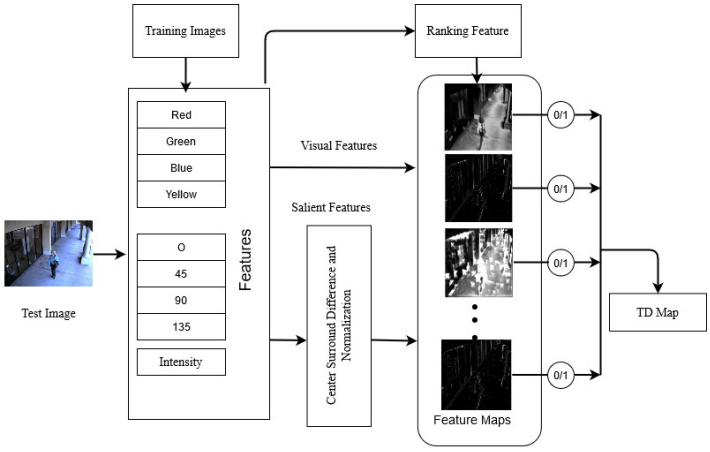


Figure 2: Architecture of FGTD map generation

In the learning phase, each feature is observed across different environments. The proposed FGTD Model combines the few relevant feature of the target as selecting weight as '1' (open) and other as weight '0'(close). Thus the features that are relevant to object are given uniform weightage and irrelevant features are removed. The target-specific local features are learned and used as cues while searching. The details of the target as features are given as training data to the system. The target features are learned and stored based on the dominancy as TD Information. Relief algorithm[17] is a successful attribute ranking method, which has been well studied and adopted for classification problems. The experimental analysis is performed by having the target as human in various scene from other background region. The mean value of the features are used to classify target from non-target. The algorithm used in this method is proposed in [17] as ReliefF to rank the features. Rather than single hit and miss while detection, the approach relies on parameter k that specifies the use of k-nearest hits and misses in the scoring update for each target instance. The features rank and weights are calculated based on scores. TD map has been generated based on combining the first 'n' ranked features. There are mathematical operations for combining the feature map. The summation is used to combine the feature from TD candidate map. Further the TD map on static images is combined with movie object map to extend to videos.

3.2 Moving Object Map

Visual saliency helps the human brain to select the most salient region in real-time. In human vision system apart from the static salient region, the moving object attracts attention in a sequence of images. To extend the method for video analysis, motion is an added feature required to process by analyzing the set of sequence of images, which gives moving object map. The map is obtained by the method in [18] that uses entropy calculation to detect the moving object map from every instance. It also eliminates the background noise like the movement of the tree, flag, elevator, fan, etc based on processing location information. The same procedure adopted to detect the moving object map(M) is given below,

Input:'k' successive frames(fr) in the video (fr_{i-k}, fr_{i-k+1} .....fr_i)
Output:Detection of moving object in ith frame as 'M'.

Step 1: Convert fr_i to gray scale and quantize into 'L' levels.
Step 2: Normalize the frame to the range [0, L-1].
Step 3: Calculate entropy by designing probability mass function for 'k' frames.
Step 3.1: Obtain the histogram integer values.
Step 3.2: Let R(x, y) is a local region in coordinate (x, y). The information content M(x,y) is calculated as (3)

$$M(x, y) = \sum_{\forall k} p(R(x, y), k) log_2 p(R(x, y), k) \quad (3)$$

Step 3.3: Eliminate the static elements by analyzing M(x, y).
Step 4: Continue the step 1 to 3.3 for the entire video sequence.

3.3 Final Map Generation

Regions related to TD Knowledge have higher saliency value than other regions. The contribution of features in the field is not uniform and its weights are varied between the targets. Assigning the appropriate weight to the feature channel pops out the target region. The weights help to determine the amount of feature that

Table 1: Metrics used for analysis

Evaluation method	Equation	Significance of value
Similarity: closeness between maps ground truth $S_h(i)$ and predicted values $S_s(i)$	$\sum_{i=1}^{n} min(S_s(i) - S_h(i)))$	'1' determines the results are same; 0 depicts it is different
Normalized Scan path Saliency (NSS): Average response value of human eye position in a model's saliency map	$NSS = \frac{1}{n}\sum_{i=1}^{n} \frac{S(x_h^i, y_h^i) - \mu_s}{\sigma_s}$	Score>1 saliency map shows significantly higher value than the human fixated location map
Kullback–Leibler(KL) Div.: KL divergence predicts the distance Between two distribution S(x) and h(x).	$D(s,h) = \sum_x s(x) log \frac{s(x)}{h(x)}$	It is non negative value, the score is 0 if distribution s=h
Correlation Coefficient (CC): Measures the strength of linear relationship between the model and human saliency map	$D(s,h) = \sum_x s(x) log \frac{s(x)}{h(x)}$	CC varies between -1 to 1. Positive values shows the better correlation in prediction than -ve values

Table 2: The Ranking of features by ReliefF

Features	**R**	**G**	**B**	**Y**	**I**	**RG**	**GR**	**BY**	**YB**	**0°**	**45°**	**90°**	**135°**
Ranking	10	8	6	12	13	11	3	1	9	4	5	2	7

helps to separate the target region from other regions. In this work, weights for moving object map (Wm) and static map (Ws) is used to fine-tune the result. Each intermediate feature map is multiplied by its corresponding weight to form the final map as in (4). The weight stands for the amount of information that is decided based on the required property on output. The weight can be within the range of [0-1], and according to the probability theory the sum of weight is '1'. To summarize, the FGTD model that uses biological inference to select the salient location of target pixels is given by,

$$Finalmap = Wm * (M) + Ws * (BUorTDMap) \quad (4)$$

4 Results and Discussion

In this section the experiments done on proposed FGTD model on static images followed by combining it with movie object map is discussed. The dataset is taken from various sources on internet having the target with different illumination and scale complexities. Experiments are done with the aim to find the influence of TD knowledge compared with map generated by without intention. The application of computer vision domain namely human detection having target as human are evaluated using metrics like similarity, NSS, KLdiv and CC of visual attention models. In Table 1, $S(x_h^i, y_h^i)$ saliency value of human fixation at ith position, mean μ_s and variance σ_s of saliency region, Cov(s,h) covariance of saliency(s) and human map(h). In Figure 3 gives the comparison of FGTD with BU Map. First Column is input image, followed by result obtained by BU model and proposed FGTD model and region of human target in the scene.

The FGTD model in section 3.1 is evaluated to detect the human target with dataset consisting of more than 200 images that were collected under varying illumination, scale and background. The features discussed in section 3.1 are learned, and relief method is used to rank the features. The ranking of features for human as target is given in Table 2. The TD map is generated for test image by combining first '5' features according to ranking starting from Blue-Yellow(BY), orientation 90° e.t.c. The sample result obtained for 6 images are presented in Figure 3. The first column shows test image and second column gives the result map without intention [16]. The result obtained by proposed FGTD is highlighted with red color square box and finally the expected target region is highlighted with red color oval shape.

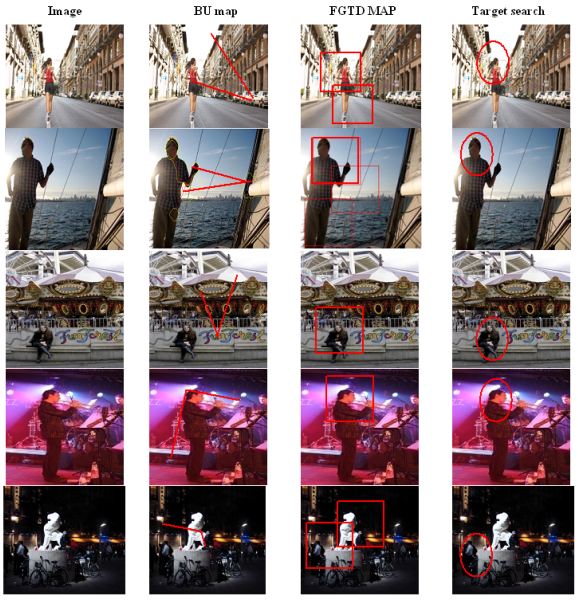


Figure 3: Comparison of FGTD with BU Map

A comparison between Map [16] and FGTD Map with target has been carried out to measure how the model helps to pop out the relevant region as it decreases in prominence in image. Hence the performance is compared with human perception without intention measures like similarity, Normalized Scan path Saliency (NSS), KL div and Correlation Coefficient (CC) in Table 1 used for analysis. Hundreds of image samples were collected under various environments, illumination and scaling conditions. In Figure 4, KL Div score is minimized in the variation of 0 ~ 10. NSS, CC and similarity score is maximized compared with map[16] Model. The scores improved especially in case of target in cluttered background. Observation shows that FGTD model helps to move closer to target region even if it uses minimum feature of the target object to filter out the irrelevant information from image. But, it shows similar performance of saliency map when the entire test image has the same feature as target. In Figure 4 the performance Analysis of FGTD model with map[16] metrics are compared with NSS, KL div, Similarity and CC, where the X-Axis in the graph is the image samples arranged according to the background clutter.

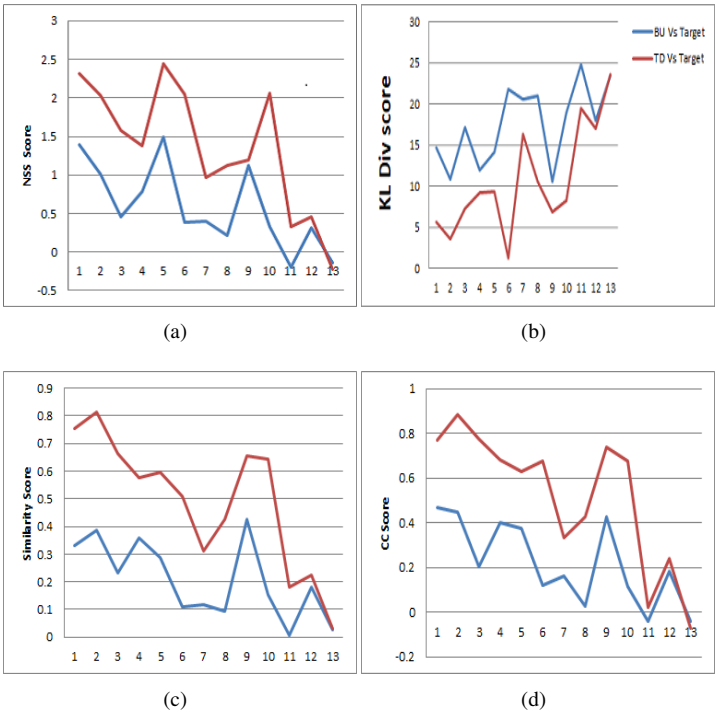


Figure 4: Performance Analysis of FGTD model with map[16](a)NSS, (b)KL div,(c)Similarity and (d) CC , where X-Axis in the graph images arranged according to the background clutter

The static map is further combined with moving object map to extend in video. The dataset consists more than 100 video clips collected from surveillance video of varying outdoor and indoor scene. The model can select Most Salient Region (MSR) as human or it can directly consider as salient region by BU model. The impact of selection between TD Map or salient map are analyzed by varying the weight parameter Wm and Ws are represented in Figure 5. The result obtained for indoor and outdoor scene with and without human presences are evaluated. The sample results obtained for 6 video frames are presented in Figure 5.

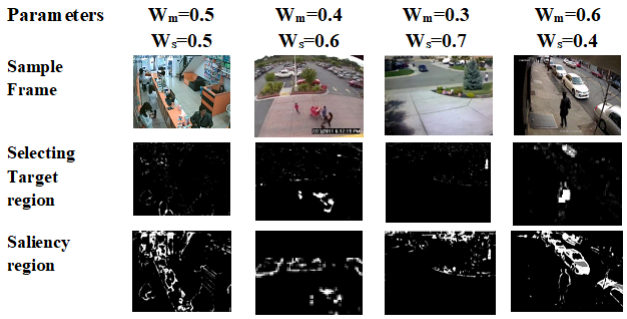


Figure 5: Comparison of BU model with TD map where First row weight parameters are given,second row has TD map with moving target and third row has BU map combined with Moving target

Hence the background distraction is minimum in indoor scene, the outdoor scene sample are highlighted. The first column represents an indoor scene, where other represent outdoor scene. In order to analyze the result, the result by selecting as human region (TD map) and saliency region (without intention) are presented. A comparison with Map [16] and FGTD Map with target Evaluated by factors prescribed in Table 1 are presented in Figure 6. Performance Analysis of proposed model with map[16] of NSS, KL div, Similarity and CC is represented. The X-Axis in the graph has samples arranged based on increasing complexity of scene.

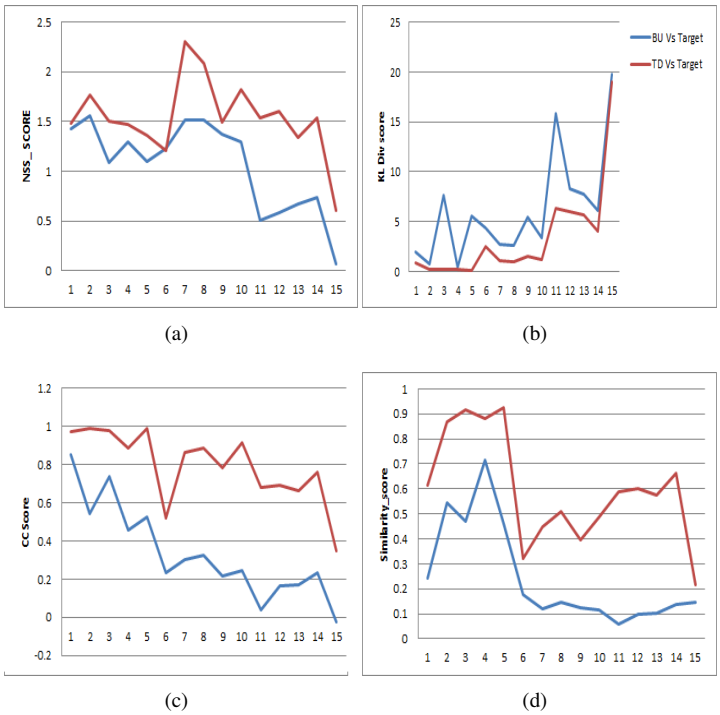


Figure 6: Performance Analysis of proposed model with map[16] (a)NSS, (b)KL div,(c)Similarity and (d) CC , where the X-Axis in the graph has samples arranged on increasing complexity of scene

In Figure 6, KL Div Score is minimized in the variation of 0.2 ~ 5. NSS, Similarity and CC Score is maximized in the variation of 0.02 ~1. The score shows reliable performance even in case of moving object that is not salient. The model is also compared

with moving object detection algorithms like optical flow [19] and dynamic background subtraction [20] methods. The existing algorithm detects all the movie objects in the scene. So, the evaluation is performed on the scene where the pedestrians or vehicle is present in the video. The analysis done more than 30 video sequence with frame rate of 30 to 45 frames per second. The average pedestrians speed varied from 3 km/hour to 5 km/hour and vehicle speed in the video varied from 40 km/hour to 80 km/hour. The performance metrics like True Positive Rate(TPR), False Positive Rate(FPR), True Negative Rate(TNR) and False Negative Rate(FNR) are listed in Table 3. It gives the average detection rates on all video clips. The proposed model shows better performance in compared with other methods.

Table 3: Comparison of FGTD method with existing techniques

Model	TPR	FPR	TNR	FNR
Optical flow method	80.56 %	15.11 %	82.89 %	17.44 %
Background Subtraction method	69.19 %	15.75 %	82.25 %	32.81 %
Proposed Model	85.75 %	73.19 %	26.81 %	24.25 %

5 Conclusion

The main contribution of this work is a TD model that binds the low-level features and pops out the target region. The FGTD model combines the features-based rank order derived by relief method. Further the model is combined with moving object map for analysing the performance in video processing. The weight based final map generation provides the significance of weight modulation in region detection. The method is evaluated with the computer vision problem to detect a human in a scene. The procedure has been extended to detect the target in videos and compared with existing methods. As compared with the scene perceived without intention, the proposed approach shows the capability of suppressing the dominant region of the image to pop out the target. The methods can be further extended to detect different targets in a real-world scene. It can be further extended to simultaneously detect multiple targets in the scene. The features like depth and texture can be included for better robotic navigation in the scenarios.

Conflict of Interest The authors declare no conflict of interest.

References

[1] M. Paul, S. M. Haque, S. Chakraborty, "Human detection in surveillance videos and its applications - a review," 2013, doi:10.1186/1687-6180-2013-176.

[2] D. T. Nguyen, W. Li, P. O. Ogunbona, "Human detection from images and videos: A survey," Pattern Recognition, **51**, 148–175, 2016, doi:10.1016/j.patcog.2015.08.027.

[3] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, S.-M. Hu, "Global contrast based salient region detection," IEEE transactions on pattern analysis and machine intelligence, **37**(3), 569–582, 2014, doi:10.1109/TPAMI.2014.2345401.

[4] S. Lu, J. H. Lim, "Saliency modeling from image histograms," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012, doi:10.1007/978-3-642-33786-4_24.

[5] J. Amudha, K. Soman, P. S. Reddy, "A Knowledge Driven Computational Visual Attention Model," International Journal of Computer Science Issues (IJCSI), **8**(3), 134, 2011.

[6] E. M. De Almeida Neves, J. E. Borelli, A. Gonzaga, "Target search by bottom-up and top-down fuzzy information," in Brazilian Symposium of Computer Graphic and Image Processing, 2000, doi:10.1109/SIBGRA.2000.883895.

[7] J. Amudha, K. V. Divya, R. Aarthi, "A fuzzy based system for target search using top-down visual attention," in Journal of Intelligent and Fuzzy Systems, 2020, doi:10.3233/JIFS-179712.

[8] W. S. Lin, Y. W. Huang, "Intention-oriented computational visual attention model for learning and seeking image content," in 2009 4th IEEE Conference on Industrial Electronics and Applications, ICIEA 2009, 2009, doi:10.1109/ICIEA.2009.5138402.

[9] S. Frintrop, "VOCUS: A visual attention system for object detection and goal-directed search," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2006.

[10] N. Bergboer, E. Postma, J. Van Den Herik, "A context-based model of attention," in Frontiers in Artificial Intelligence and Applications, 2004.

[11] R. J. Peters, L. Itti, "Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007, doi:10.1109/CVPR.2007.383337.

[12] J. Amudha, R. K. Chadalawada, V. Subashini, B. Barath Kumar, "Optimised computational visual attention model for robotic cognition," in Advances in Intelligent Systems and Computing, 2013, doi:10.1007/978-3-642-32063-7_27.

[13] A. M. Treisman, G. Gelade, "A feature-integration theory of attention," Cognitive Psychology, 1980, doi:10.1016/0010-0285(80)90005-5.

[14] J. Zhang, S. Sclaroff, "Saliency detection: A boolean map approach," in Proceedings of the IEEE International Conference on Computer Vision, 2013, doi:10.1109/ICCV.2013.26.

[15] K. R. Cave, M. S. Kim, N. P. Bichot, K. V. Sobel, "The feature gate model of visual selection," in Neurobiology of Attention, 2005, doi:10.1016/B978-012375731-9/50094-X.

[16] L. Itti, C. Koch, E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, doi:10.1109/34.730558.

[17] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, J. H. Moore, "Relief-based feature selection: Introduction and review," 2018, doi:10.1016/j.jbi.2018.07.014.

[18] R. Aarthi, J. Amudha, K. Boomika, A. Varrier, "Detection of Moving Objects in Surveillance Video by Integrating Bottom-up Approach with Knowledge Base," in Physics Procedia, 2016, doi:10.1016/j.procs.2016.02.026.

[19] H. S. P. DGT, K. J. Udesang, "A survey on object detection and tracking methods," International Journal of Innovative Research in Computer and Communication Engineering, **2**(2), 2014.

[20] R. D. Sharma, S. L. Agrwal, S. K. Gupta, A. Prajapati, "Optimized dynamic background subtraction technique for moving object detection and tracking," in 2nd International Conference on Telecommunication and Networks, TEL-NET 2017, 2018, doi:10.1109/TEL-NET.2017.8343526.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1007-1012 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Design of a Remote Real-time Groundwater Level and Water Quality Monitoring System for the Philippine Groundwater Management Plan Project

Carlos M. Oppus*,[1], Maria Aileen Leah G. Guzman[2], Maria Leonora C. Guico[1], Jose Claro N. Monje[1], Mark Glenn F. Retirado[1], John Chris T. Kwong[1], Genevieve C. Ngo[1], Annael J. Domingo[1]

[1]*Ateneo de Manila University, Electronics, Computer, and Communications Engineering Department, Quezon City, 1108, Philippines*

[2]*Ateneo de Manila University, Environmental Science Department, Quezon City, 1108, Philippines*

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 21 November, 2020
Online: 08 December, 2020

Keywords:
Groundwater Level Monitoring
Groundwater Quality Monitoring
Sensor Systems
Remote Real-time Systems
Water Quality Sensors

ABSTRACT

Recent technological advances allow us to utilize remote monitoring systems or real-time access of data. While the use of remote monitoring systems is not new, there are still numerous applications that can be explored and improved on, one such is groundwater level and quality monitoring. In the Philippines, the extraction of groundwater for both domestic use and industrial use are manually monitored by the government's concerned agency and is done at least once per year. With this current setup, the real and significant state of the groundwater is not reflected in a way that is most valuable to the government and to the community. This project aims to design and develop a remote real-time groundwater level and quality monitoring system. It is intended to provide quantitative data for policy makers in addressing recurrent water shortages in the Philippines. This paper discusses the designed system composed of three modules: power module, sensors and control, and data visualization. These three modules provide real-time data from far-flung locations while being energy-sustainable. Dry runs of the system in a controlled environment yielded excellent results — average data accuracy of 96.63% for all six (6) groundwater quantity and quality parameters namely: pH, temperature, electrical conductivity, total dissolved solids, salinity, and static water level (SWL), and 90.63% data transmission reliability. Initial deployment of the system on one of the groundwater monitoring well in Metro Manila, Philippines returned a 91.16% data transmission reliability. The system is currently installed in 20 groundwater monitoring sites all-over the Philippines and is scheduled for more installations.

1. Introduction

In the Philippines, the main freshwater sources are inland surface water and groundwater. These water sources are used by the agricultural, industrial, and domestic sectors [1]. For the domestic sector, the ideal setup utilizes the surface water as the main source and the groundwater as a reserve during a water crisis [2]. But in the current recurring water shortage crisis, it is inevitable that these groundwater reservoirs are tapped as additional water resources.

The monitoring of water supply and water quality using observation wells is paramount. There are calls for deployment of sensor systems for monitoring wells for dense communities with high water usage and industrial areas producing wastes and contaminants [2]. If the rate of potential groundwater recharge is lower than the rate of extraction, the extraction well could dry up. Additionally, water contamination will harm the consumers, including agricultural farms and livestock located downstream [3].

Regular monitoring of the water is crucial. There is a need for automated water quality monitoring systems that regularly gather data from observation wells at designated and crucial times of the day. This lessens manual collection of samples, particularly in hard-to-reach and dangerous areas [4]. The use of wireless sensor networks (WSNs), deployed on-site to transmit data back at specified intervals, has been found to be an efficient way of regularly monitoring and ensuring that water quality remains consistent and compliant with standards [5]-[7]. This setup provides needed information to relevant agencies that allow them to make well-informed decisions and guidelines for water conservation.

In full, this paper focuses on the design of the sensor system

*Carlos M. Oppus, Quezon City, NCR, Philippines, coppus@ateneo.edu

composed of three modules, namely: the power supply, sensor and control, and data visualization.

2. Overall System Overview

Figure 1 shows the overall framework of the groundwater monitoring system which is composed of: (A) Power Supply, (B) Sensor and Control, (C) Data Visualization, and (D) Monitoring Well. While the block diagram of the whole system and components for each sub-system are shown in Figure 2.

The power supply sub-system provides power to the sensor and control sub-system through a rechargeable battery. The battery and solar panel are both managed by the solar charge control to prevent unwanted voltage or current discharges during low sunlight or nighttime.

The sensor and control sub-system uses a microcontroller to fetch data from the sensor units. The sensor data collected are pH level (pH), temperature, electrical conductivity (EC), total dissolved solids (TDS), salinity, and groundwater level (static water level). These water parameters are listed under the water quality monitoring guidelines for groundwater from the Philippines' Department of Environment and Natural Resources Administrative Order No. DAO-2016-08 [8]. This set of sensor data is consolidated and sent to the cloud server using a GSM/GPRS module [9], [10]. This sub- system uses an on-site micro-SD card for data back-up in case of remote communication loss.

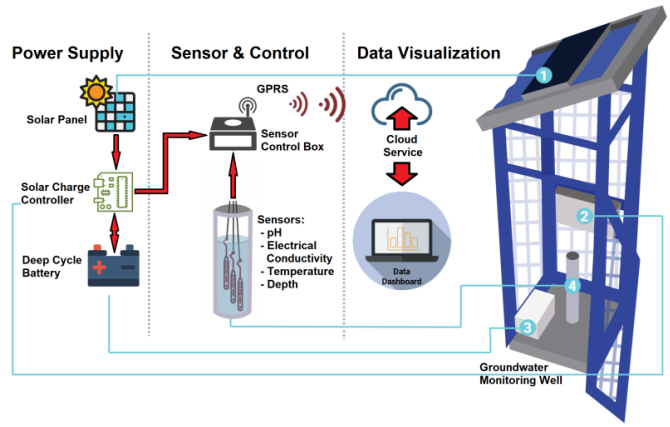


Figure 1: GMP System Framework

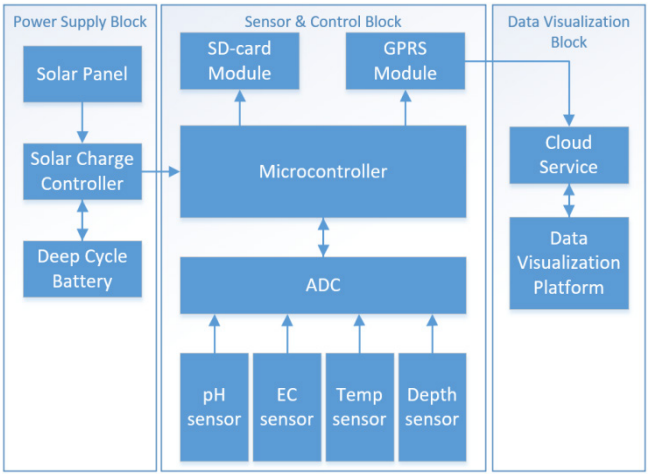


Figure 2: GMP System Block Diagram

Once the data is received by the cloud server, they are processed by the data visualization sub-system. The processed visualized data can be accessed through a dedicated website. The website includes a presentation of the project with its goals and partner institutions; a map showing the monitoring well sites with sensor systems; a view of available data; and a page on helpful information regarding water level and water quality parameters.

3. Design Implementation

3.1. Power Supply

It is essential that the system has a reliable power supply to operate and transmit data at required intervals. The system's power supply is composed of the solar battery, the solar panel, and the charge controller. Actual components used are shown in Figure 3.

The solar battery acts as the main source of power for the sensor system. It provides power while being rechargeable by harnessing solar energy.

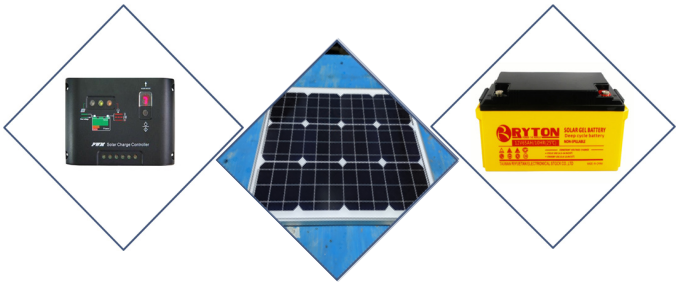


Figure 3: Solar charge controller (left), Solar panel (center), Deep cycle battery (right)

The solar panel charges the battery to keep it above the cutoff voltage. The solar panel harvests energy from the sun and converts it to electrical energy. To replenish the energy consumption of the system, the solar panel is located and properly positioned to be directly exposed to sunlight for at least 3 hours a day. Unable to meet the minimum exposure time might not guarantee full recharge of the solar battery. In this case, the sensor system will only work until the energy of the solar battery is depleted. The specification of the solar panel is shown in Table 1.

Table 1: Solar Panel Specifications

Type	Mono Crystalline
Power	50W
Open-Circuit Voltage	22.5V
Current at Max Power	2.75A
Material	Weatherproof Tempered Glass Aluminum Frame
Dimensions	70x54 cm^2

The solar charge controller maintains the proper recharging and disconnection of the solar panel and the battery. The solar charge controller disconnects the solar panel when the solar panel output voltage does not meet the minimum requirement for charging the battery. Also known as a solar charge regulator, it protects a solar battery's life by limiting the electric current allowed to the connected solar battery when charging, and from the solar battery while used to power a device.

The solar charge controller in the setup has 3 pairs of positive

and negative terminals where three types of devices can be attached: a solar panel, a battery, and a load. The load in this case is the sensor control box. A cutoff voltage is set to prevent the battery from fully draining itself. This allows the system to work for roughly two (2) weeks without sunlight exposure.

3.2. Sensor and Control

A Bluno Mega microcontroller is the central unit of the sensor system. It features enough serial ports to accommodate the various sensors and can be managed via Bluetooth communications for the calibration method. It collects sensor data before sending it using a GSM/GPRS module.

Figure 4: pH sensor probe (left), Submersible pressure sensor (center), and EC sensor probe (top right)

The sensor system measures groundwater quality by collecting the parameters pH, temperature, electrical conductivity, total dissolved solids, and salinity. A sixth parameter, static water level (SWL), is included to measure groundwater quantity relative to the ground level. The sensor system uses off-the-shelf environmental sensors from Atlas Scientific and a pressure transducer from Sendo Sensors. The sensor components are shown in Figure 4. The sensors used with the corresponding units are summarized in Table 2.

Table 2: Water Level and Water Quality Sensors

Sensor	Parameter	Unit
Atlas Scientific Industrial pH Sensor	pH, Temperature	pH °Celsius
Atlas Scientific Industrial Conductivity K 1.0 Sensor	Conductivity, Total Dissolved Solids, Salinity	microSiemens/centimeter ppm, ppt
Sendo Sensors Pressure Transducer	Static Water Level	meter

The Atlas Scientific sensor probes are interfaced with Atlas Scientific carrier boards for better electrical isolation. Each EZO™ circuit carrier board from Atlas Scientific contains calibration data and receives the analog sensor information to be sent to the Bluno Mega's serial ports as digital data. The system contains three EZO™ circuits: pH, temperature, and electrical conductivity. The pH sensor probe is also being used for temperature detection, hence it is connected to both the pH and temperature EZO™ circuit carrier boards.

The Sendo Sensors pressure transducer sends signal directly as current in milli-Ampere values. However, the Bluno Mega's Analog-to-Digital Converter can only read voltage signals. The sensor probe's 4-20mA current loop is connected in parallel to a resistor to convert the electrical signal as a corresponding voltage value. Since the maximum voltage the Bluno can read is 5-volts, we need a 250-ohm parallel resistor. The available single-resistor resistance value closest is 220 ohms. Therefore, our highest analog voltage reading peaks at 4.4 volts, and the pressure data is calibrated accordingly.

3.3. Data Visualization

For data telemetry, a minimum-standard 2.5G network is used for data transmission to cloud servers since there are areas in the Philippines that do not have access to even a 3G network and above. With this limitation, the Elecrow SIM800C GSM/GPRS module is chosen as the data telemetry module.

Since the GPRS module is designed for Arduino, minimal modification is required to interface it with the Bluno Mega. However, the pin assignments for software serial ports are different. The GSM module's default software serial ports are at pins 7 and 8. Since these pins are not usable as software serial pins for Bluno Mega, pins 7 and 8 of the GSM module are rewired to pins 10 and 11 of the Bluno Mega, respectively.

4. Results

4.1. The Sensors and Control Box

Figure 5 shows the setup of the system in a controlled environment, as well as the setup installed on some of the groundwater monitoring wells in Metro Manila, Philippines. The sensors and control box are tested for accuracy after calibration by comparing the sensor readings to that of the known values of several specified buffer solutions.

Note that prior to actual calibration, initial trend tests should show the sensor reacting to its specific parameter, e.g., pH level. Once the trend test readings show a correct trend, such as pH level sensor value increasing when dipped in higher pH level, the sensor is then accepted for calibration.

For calibration, the pH sensor is tested using pH buffer solutions 4, 7, and 10, being acidic, neutral, and basic solutions respectively. The sensor's electrical responses for each of the buffer solution are recorded. Then by using linear regression method on the three calibration points, any pH values outside the buffer solution range can be obtained. As for the electrical conductivity sensor, three-point calibration method is also used. The first point being the dry-state or 0µS (zero), second and third points uses standard electrical conductivity solutions 1,440µS and 12,880µS respectively from Atlas Scientific, with the higher value being the more conductive or saline solution. The same linear regression method is used to determine electrical conductivity values outside the calibration points. As for the pressure sensor, it is calibrated based on water depth and pressure relationship. The sensor is dipped in a container filled with water and with known depth. Two-point calibration is used for determining water depth with respect to pressure read by the sensor. To test the accuracy of the pH and EC sensors after calibration, the sensors are re-tested in parallel with a pH level digital testing stick, a litmus paper, and an EC testing stick.

Shown in Figure 6 are the results of the pH level test after calibration. The different instruments used showed almost consistent values with the pH solution, thus validating the performance of the system's pH sensor. This results in a computed general accuracy for the pH sensor of 95%.

Figure 5: Test setup in a controlled environment (top-left), the setup installed on some of the groundwater monitoring wells (top-right, bottom-left), and the team during one of the system installations (bottom-right)

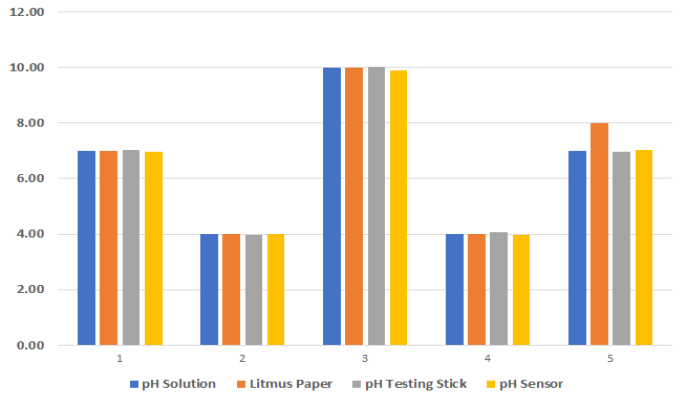


Figure 6: pH level test using litmus paper, digital tester, and the pH sensor

4.2. Test Setup and Initial Deployment

To further test the different sensors, the setup equipped with data telemetry capability was left in a controlled environment with distilled water for 4 days, as can be seen in Figure 5 (top-left). Table 3 indicates the summary of data gathered for 4 days: showing the minimum, maximum, and average values transmitted for each parameter; the accuracies of the transmitted values compared to the expected values; and the overall average accuracy. The accuracy for parameters pH, temperature, EC, TDS, salinity, and SWL are 92.14%, 94.69%, 97.26%, 99.06%, 100%, and 100% respectively, which then translate to an overall average accuracy of 96.63%. These accuracy values are very well comparable with the sensors' technical specifications which are listed at 95%-98% accuracy.

Table 3: Summary of Test Data from Controlled Environment

	pH	Temp	EC	TDS	Sal	SWL
Min	7.09	23.33	640	346	0.31	0.17
Max	7.95	29.82	763	412	0.37	0.19
Average	7.55	27.38	719.17	388.63	0.35	0.18
Expected	7	26	700	385	0.35	0.18
Accuracy (%)	92.14	94.69	97.26	99.06	100	100
Average Accuracy (%)	**96.63**					

After yielding promising results from the controlled test setup, the system was then deployed initially on one of the groundwater monitoring well in Metro Manila, Philippines, as shown in Figure 5 (top-right). The setup was left on the monitoring well for 31 days equipped with data telemetry and solar panel for power sustainability. Table 4 shows the summary of data gathered from the deployed system. Different with the data from controlled setup, data from the deployed setup were not compared with any expected values due to the nature of the setup being a monitoring system. Accuracy checks for sensor readings were done prior to installing the system on the groundwater monitoring well by comparing them with the values obtained using pH level and EC digital testing kits, as well as the manual measurement of the actual SWL.

Table 4: Summary of Data from initially deployed Monitoring System

	pH	Temp	EC	TDS	Sal	SWL
Min	9.06	36.05	1105	597	0.55	17.05
Max	11.82	36.37	1391	751	0.70	17.97
Average	9.83	36.23	1237.02	668.27	0.62	17.44

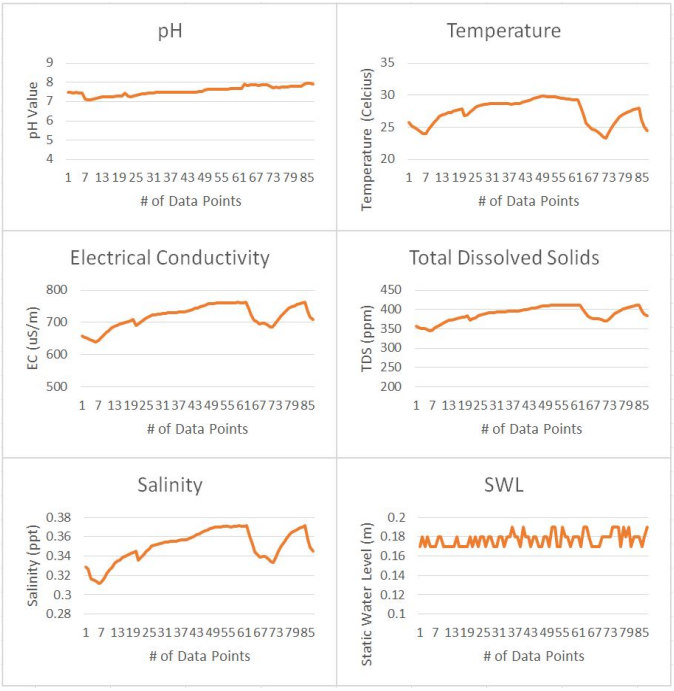


Figure 7: Sensor Test Results at 100 data points (x-axis)

4.3. The Telemetry

In terms of data transmission reliability, we have the system send data around every hour. In this paper, we define data

transmission reliability as the capacity of the system to transmit data to the remote server via GPRS communication [11]. Table 5 shows the summary for the data telemetry reliability. For the controlled-environment setup, the remote server was able to receive 87 data sets out of the expected 96 from the 4-day test setup, thus yielding 90.63% of data telemetry reliability. While for the field-deployed setup, the remote server was able to receive 650 out of the 713 data sets, for 91.16% transmission reliability. These reliability values also translate to about 21 data sets per day. The complete 87 data sets from the controlled-setup can be seen represented by the graphs shown in Figure 7.

Table 5: Summary for Data Telemetry Reliability

# OF DATA SET			
Controlled-Setup		**Deployed-Setup**	
Actual	87	**Actual**	650
Expected	96	**Expected**	713
Reliability (%)	**90.63**	**Reliability (%)**	**91.16**

4.4. The Power and Sustainability

A simple voltage divider circuit is installed in the control system for the Bluno to read the voltage level of the battery power source. This allows for remotely monitoring the power level of the system for its consumption and recharging cycles. In Figure 8, the battery voltage trend both for controlled- and deployed-setups having 87 data sets can be seen. The battery voltage for the setup under controlled-environment have a decreasing trend with the setup being in a laboratory setting where the battery is not connected to a solar panel or any charging source. While the voltage trend for the deployed-setup shows a cyclic trend for charging and discharging phases that signifies that the solar panel is decently exposed to sunlight.

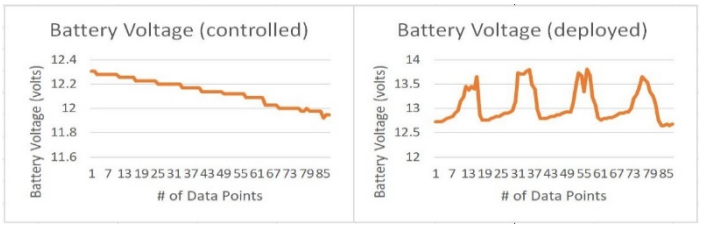


Figure 8: Battery voltage trend, controlled-environment (left), deployed-setup (right)

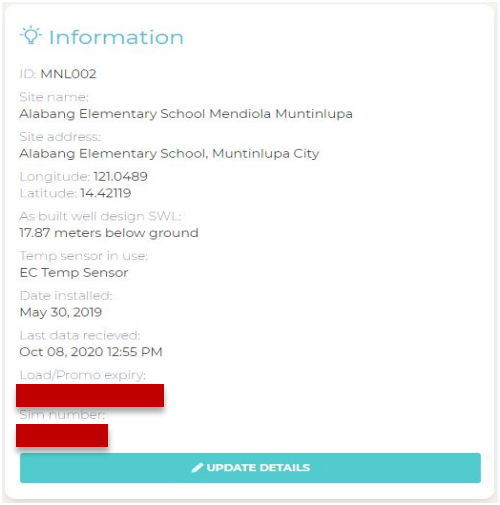


Figure 9: Screenshot of the Data Visualization Platform showing the information regarding the Monitoring Well

4.5. Data Visualization

After gathering these water quantity and quality parameters, the control box sends the data to a cloud server for visualization purposes in a web-based platform. These can be viewed publicly on the website of the project. The data are sorted according to location, data type (pH, temperature, EC, TDS, Sal, SWL), and in chronological order. A screenshot of the data visualization in Figures 9 shows the information about one of the monitoring wells in Metro Manila, Philippines, while Figure 10 shows the monitoring well's actual SWL readings for about a month.

Figure 10: Screenshot of the Data Visualization Platform showing actual SWL data

5. Conclusion

The research group is able to develop a fully-functional energy-sufficient remote monitoring system that can gather and record accurate groundwater quality and quantity data. The data can be transmitted reliably to a separate web-server. Further testing for robustness and accuracy during extended field deployments is recommended. In conclusion, the designed and developed system can operate with an overall average accuracy of 96.63% for 6 water quantity and quality parameters, and 91% data transmission reliability. Considering the data transmission requirement of at least 1 data set per day by the funding agency, it is therefore recommended for installation for the remaining groundwater monitoring wells all over the Philippines.

One of the challenges encountered during the designing and development of the system is the calibration of the sensors as there are occasions where the manufacturer's calibration kits are off. Worth mentioning is the pH sensor since its EZO^{TM} circuit is very sensitive to electrical static discharge and mishandling. It is also problematic to manually control its calibration offsets. The analog pH sensor is used instead. By mathematically computing for the calibration equations and offsets, more consistent readings of pH values are obtained.

Another challenge is the power consumption of the system. The sensors consume more power than the charging capability of the power sub-system due to unforeseen lower solar exposure. This is remedied by having the sensors undergo sleep mode majority of the time and varying the remote sending frequency of the system.

Acknowledgment

This research is a project implemented by the Ateneo de Manila University (ADMU), funded by the Department of Science and

Technology (DOST) - Philippine Council for Industry, Energy and Emerging Technology Research and Development (PCIEERD), National Water Resources Board (NWRB), and partially funded by the University Research Council (URC) of the Ateneo and the Ateneo Research Institute of Science and Engineering (ARISE).

References

[1] The World Bank Group, Philippines: Environment Monitor 2006 (English), 1, Washington, D.C., 2006.

[2] A.C. Rola, J.M. Pulhin, A.H. Rosalie, Water supply and demand and the drivers of change in Water Policy in the Philippines: Issues, Initiative, and Prospects, Springer International Publishing: 15–40, 2018.

[3] National Water Resources Board, Water Resources Assessment for Prioritized Critical Areas (Metro Manila), Metro Manila, 2004.

[4] National Water Resources Board, 2018 Annual Report, National Water Resources Board, Metro Manila, 2018.

[5] K. Sumathi, E. Christina Dally, G.V. Ashwini, S. Saigokul, "Water Pollution Indicator using Arduino Uno," International Journal of Innovations in Engineering and Technology, **9**(2), 28–39, 2017, doi:10.21172/ijiet.92.05.

[6] S. Pasika, S.T. Gandla, "Smart water quality monitoring system with cost-effective using IoT," Heliyon, **6**(7), e04096, 2020, doi:10.1016/j.heliyon.2020.e04096.

[7] L.B. Imran, R.M.A. Latif, M. Farhan, H. Aldabbas, "Smart City Based Autonomous Water Quality Monitoring System Using WSN," Wireless Personal Communications, **115**(2), 1805–1820, 2020, doi:https://doi.org/10.1007/s11277-020-07655-x.

[8] Department of Environment and Natural Resources, Water Quality Guidelines and General Effluent Standards of 2016 (DAO 2016-08), 2016.

[9] Elecrow, Elecrow GPRS/GSM Shield v1.0 Guide, 2019.

[10] J. Du, J. Guo, D. Xu, Q. Huang, "A remote monitoring system of temperature and humidity based on OneNet cloud service platform," in 2017 IEEE Electrical Design of Advanced Packaging and Systems Symposium (EDAPS), IEEE, Shaanxi, China: 1–3, 2017, doi:10.1109/EDAPS.2017.8277059.

[11] S. Gokulanathan, P. Manivasagam, N. Prabu, T. Venkatesh, "GSM Based Water Quality Monitoring System Using Arduino," Shanlax International Journal of Arts, Science and Humanities, **6**(4), 22–26, 2019, doi:10.34293/sijash.v6i4.341.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1013-1020 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Determinants of Technological and Innovation Performance of the Nepalese Cellular Telecommunications Industry from the Customers' Perspective

Rewan Kumar Dahal*, Ganesh Bhattarai, Dipendra Karki

Tribhuvan University, Faculty of Management, Nepal Commerce Campus, New Baneshwor, Kathmandu, 44600, Nepal

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 21 November, 2020
Online: 08 December, 2020

Keywords:
Innovation
Measures
Organization
Performance
Technology
JEL Classification
L25
L32
L86

ABSTRACT

The study aimed to understand and analyze the technological and innovation performance measures of the Nepalese cellular telecommunications industry from the customers' perspective. The measures like network service quality, signal strength and coverage, voice quality, and calls drop are used for technological performance, whereas measures like product/service innovation, process innovation, customization, competitive innovation, and marketing innovation are used for innovation performance. It followed a descriptive research design employing a structured questionnaire survey instrument. Two market leader companies of the Nepalese telecommunications industry (Nepal Telecom and Ncell) and their customers were considered as objects of the study. For this study, 391 respondents were selected by using the judgmental sampling technique. Targeted respondents of the study were postgraduate understudies, service holders, business persons, and freelancers. The survey instrument was set in three segments with 19 questions to collect and analyze information with the help of a statistical package for the social sciences and analysis of moment structures software. Two constructs with nine test variables found as the determinants of the technological and innovation performance of the Nepalese cellular telecommunications industry. The results of the study would provide helpful guidelines in understanding the key drivers of the technological and innovation performance in the Nepalese scenario.

1. Introduction

One of the most important inquires in business has been why some organizations succeed while others fizzle. Organizational performance is one of the most articulated concerns for all organizations, either profit or not-for-profit. It has been immensely important for stakeholders, especially for managers, to identify the components that influence an organization's performance so that they can initiate appropriate actions. However, describing, conceptualizing, and measuring organizational performance has never been an easy undertaking [1]. Various researchers have made an attempt to define and assess organizational performance, yet it remains an issue of debate among analysts [2]. Basically, organizational performance refers to the outcomes of daily administrative processes [3] that the organization intends to accomplish in the long-term [4]. The primary concern of the study of organizational performance is the suitability of various approaches to deal with the idea usage and its measurement [5].

Organizational performance is a multifaceted concept, and the use of a financial measures-based performance system alone is not sufficient to measure performance of organizations operating in highly competitive, innovative, and advanced technological environments [6]. As such, the contemporary business environment demands a multidimensional performance measurement system that can provide more information to stakeholders, more specifically managers. Out of several dimensions of the performance measurement system, the telecommunication industry's two inherent components are *technology* and *innovation* due to the fact that adoption of technological advancements through innovative actions distinguishes products/services or companies in the competitive market.

Although the determinants of technological and innovation performance are still unclear, accomplishing and sustaining superior organizational performance, especially in the cellular telecommunications setting has become a top priority today. The study therefore tries to find the measures to assess and quantify the technological and innovation performance of the Nepalese

*Corresponding Author: Rewan Kumar Dahal, Tribhuvan University, Nepal
Email: dahalrewan@gmail.com

cellular telecommunications industry. Alongside that, the study's objective is to examine the intensity of technological and innovation performance measures in the Nepalese scenario and initiate some theoretical and practical discussions for better organizational performance.

This study is extremely significant in the scenario of Nepalese telecommunications industry as the cellular telecommunications has been growing by leaps and bounds and gradually covering the broad spectrum of Nepali populace. As per [7], the number of Global System for Mobile (GSM) network subscribers has surpassed 3.5 million as of June 2020. The industry's main focus as it can be observed is on mobile phones and their primary functions, such as calling and hearing. Telecommunication companies are under persistent pressure to deliver creative items more rapidly, at lower cost, and with enhanced quality [8]. This shows that cellular telecommunications companies have a special relationship with the customer, and necessitates the assessment of organizational performance in order to strength that relationship by incorporating new features at lower cost with better services in the future.

2. Literature Review

The literature contends that technological and innovative performance measures are endogenous and are impacted by more profound organizational growth determinants. In today's knowledge-based economy, corporate growth is driven primarily by creative capacity, invigorated by the pertinent information and technological externalities, instead of gathering capital [8]. So, performance measures concerning technology and innovation stand at the point of convergence of organizational performance measurement system.

2.1. Technological performance measures

The world of technology is changing at a rapid pace, and telecommunications technology is not an exception. Over time, technological changes have a significant impact on the creation of marketing opportunities [9]. The preferences of customers change rapidly and pose numerous difficulties for any business [10]. Technological advances, especially in information and communication technology (ICT), make product/service innovation conceivable [11]. New technologies bring new challenges to business players and seek opportunities to offer customers innovative products/services to pursue long-term relationships [12] and distinguish them from other competitors [13]. Previous studies like [14]-[16], have analyzed the crucial role of technology in the performance and competitiveness of organizations. This study thus examined and evaluated some technological performance measures among the various customers.

Network and service quality has a significant role in organizational performance for telecom service providers [17]. In [18], [19], the authors acknowledged that network quality is one of the crucial factors that determine the organization's performance in evaluating its service quality.

Signal strength and coverage have always been essential criteria for selecting telecom services. Multiple studies demonstrate that signal strength and range positively affect consumers' satisfaction and the organization's image [20], leading to overall organizational performance.

Voice quality in telecommunications networks has been a fundamental part of service qualification. Because of the inherent characteristics of a converged voice and voice-over internet protocol (VoIP), administrators face specific problems such as echo, jitter, delay, side-tone, background noise, etc. Quick and efficient mitigation of VoIP issues to maximize call quality leads to better service provider performance [21].

Calls drop is another key performance indicator to assess the performance of operators' networks. It is expected to have a direct impact on customer satisfaction and retention [22].

Electronic customer relationship management (e-CRM) allows users to frequently communicate with one another and maintain their databases as clearly as possible while improving customer relationships [23]. Implementation and uses of e-CRM enhance organizational effectiveness and significantly connect with the overall performance [24].

2.2. Innovation Performance Measures

New methods of products, processes, and other components are being applied to organizations by innovation [25]. It transforms into a product/service as it is produced and bought by the customer and provides the vendors with financial benefits. Innovation performance measures introduce new strategic tactics, working environment and guidelines, decision-making techniques, and better external relations approaches [26]. Studies like [27], [28], [29], have been conducted to examine the crucial role of innovation on a firm's performance and competitiveness. Despite several dimensions of the innovation performance measures, the study secured just five general business criteria and assessed them from the customers' perspective.

Product/service innovation satisfies customers by turning the invention or idea into a product/service and enhancing company value at acceptable risk [30]. Innovation analyzes the needs of the customers and provides the products/services accordingly. Therefore, product/service innovation helps escalate customer benefits [29] tantamount to better organizational performance.

Process innovation refers to doing business in a new and innovative way [31]. Conversion of new technology into processes, especially in ICT based companies, has a significant impact on the performance of the firm. Process innovation therefore represents a type of service innovation that provides an immense advantage in creating new streams of value for better customer experiences, thereby enhancing a firm's competitiveness and overall performance of a company [32].

Customization is the process of making and changing something to the buyer's or user's needs. The telecommunication sector's service competitiveness, including variations of services and facilities and the degree of customization, influences technological and innovative performance. Additionally, customization capability encourages service systems to suit a specific customer within a particular use-circumstance and to make a unique value for that customer [33].

Competitive innovation is a strategy that organizations use to make a competitive advantage by developing products that

nobody else can, showing improvements over every other person, or presenting prevalent, less expensive, and quicker services [34]. It is a type of social design that focuses on creating and elaborating the best and practical thoughts originating from the best innovators for superior organizational performance.

Marketing innovation is implementing a new advertising technique, including significant changes in product design, placement, packaging, distribution, communication, promotion, and pricing strategy [35]. According to [36], improvisation, modification, augmentation, or transformation of existing trade channels using technology can reduce transaction costs that lead to better organizational performance. [37] also argued that companies developing innovative practices were more able to position themselves on the market against their competitors and achieve emphatic retention of customers.

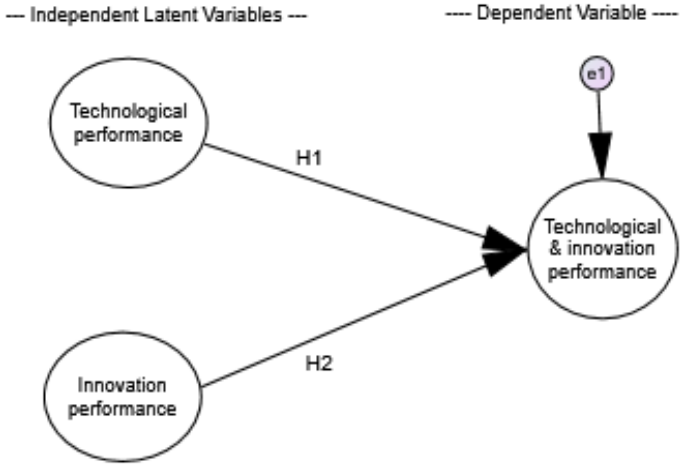


Figure 1: Conceptual framework of the study

2.3 Technological and Innovation Performance

Various metrics can be used to assess a firm's performance [38]. Technological and innovation initiatives are considered to be primary indicators of the cellular telecommunication industry. They allow the firm to improve capability to meet the demands of its evolving domestic and international market [39], thus giving the firm a lucrative outcome.

The overall technological aspect is multidimensional and can be decomposed appropriately from different perspectives [40]. Studies showed that technological advancement is a primary determinant of superior long-term success [41] and leads to better performance than rivals [42].

The overall innovation aspect allows better delivery of the services offered and a more significant impact of any campaign and promotes organizational changes standardizing certain activities [43]. In [44], the authors evidenced that in firms, where there is a willingness to innovate its products/services, there is a significant and positive impact on overall organizational performance.

3. Conceptual Framework and Hypotheses

Organizational performance shows the development and progress of an organization [45]. The analysis of organizational success is one of the most investigated variables. It is a criterion for management and tends to depend on the fitness between corporate systems and situational factors. Therefore, the study is driven by the contingency theory, which provides a practical perspective for examining and measuring different contingent variables.

Within the framework of prior literature and conceptual sketching, the study has the following two hypotheses:

H1: Technological performance positively and significantly affects technological and innovation performance.

H2: Innovation performance positively and significantly affects technological and innovation performance.

4. Methodology

The quantitative research strategy was employed to extract information for the study, and the required information was acquired through a structured questionnaire survey from the targeted respondents. The study used a statistical package for social sciences (SPSS) version 23 and analysis of moment structures (AMOS) version 21 software to dissect and decipher the information. The study's targeted populace comprised all the Global System for Mobile (GSM) customers of Nepal Telecom and Ncell. According to [7], the sample companies' customers possess more than 93 % of the Nepalese cellular telecommunications market share. The study employed a judgmental sampling technique because there was no proof of active and the dead Subscriber Identification Module (SIM) card issued with the sample organizations. Targeted respondents of the study were postgraduate understudies, service holders, business persons, and freelancers. A sampling plan was used as proposed by [46], [47] and ensured at least 385 respondents' responses. Therefore, the study employed 391 respondents' perceptions as sample.

The designed questionnaire for the survey comprised 19 questions and structured them into three segments. In the first segment, seven items were requested, relating to demographic and general information. In the second segment, 10 items mentioned about technological and innovation performance measures in various dimensions. In the last segment, two items inquired about the overall technological and innovation performance. The last two segments centered on the study variables with a series of close-ended questions to get the appropriate information. With a 6-point Likert-type scale of 1 = strongly disagree to 6 = strongly agree, all the survey queries of the study variables were estimated. The 6-point scale was picked because it excludes the undecided middle rating.

The researchers distributed the questionnaires through an online and a field survey. Based on the data provided by different colleges/universities, an online survey conducted among university graduates from the recent five years and who are working in Nepal's various parts. A total of 500 respondents approached during the 30-days of January/February 2020. Out of this, 143 responses were received from the online survey and recorded in SPSS for further analysis. On the other hand, the researchers themselves disseminated 500 questionnaires among Tribhuvan, Kathmandu, Pokhara, and Purbanchal Universities'

graduate students and their faculties to conduct the field survey. Out of them, 276 responses were collected during the period of 30 days in January/February 2020. From the collected responses through the field survey, 28 were rejected due to incomplete information. The remaining 391 responses were considered valid and were used in this study.

5. Reliability and Validity of the Survey Instrument

The study assessed the reliability and validity to confirm the constructs' suitability before evaluating the hypothesized model. The survey instrument's overall reliability statistics showed an excellent Cronbach's alpha ($\alpha = 0.819$) and the average inter-item correlation coefficient ($r = 0.312$) from the 10 test items. Despite having excellent values of the statistics, construct-wise Cronbach's alpha was also analyzed. The value of Cronbach's alpha of the technological performance construct having five test items (VAR_8, 9, 10, 11, and 12) was 0.775. The analysis found that the VAR_12 'the company alters and provides me all information through SMS' required to delete because of the higher value of alpha 0.807 if the item deleted. The variable VAR_12 also had a low correlation (0.317) to total correlation, while the rest were in the area of 0.576 to 0.680. Therefore, VAR_12 was also not relevant in the study and deleted from the analysis. None of the items were found to delete at the innovation performance construct. Thus, reliability statistics promoted nine variables within two constructs for further study. Table 1 presented the analytical result for the reliability statistics of each construct.

Table 1: Reliability Statistics

Construct	Cronbach's Alpha	The average inter-item correlation coefficient	Name of the items	No of items
Technological performance	0.807	0.512	VAR_ 8, 9, 10, 11	4
Innovation performance	0.765	0.394	VAR_13, 14, 15, 16,17	5
			Total no of items	9

The construct-wise statistics showed that both constructs had good alpha values, i.e., 0.7 or higher, as suggested by [48]. The innovation performance construct had good average inter-item correlation ($r = 0.394$) whereas the technological performance construct had ($r = 0.512$) slightly higher than the recommended ideal range of 0.15 to 0.50 by [49]. A higher value of the average inter-item correlation indicated that the measuring variables might be capturing a small bandwidth and tended to be very similar to each other.

The study had an adequate sample size concerning the validity issue since the Kaiser-Meyer-Olkin (KMO) value of 0.816 from nine test items was more significant than the threshold value of 0.5 as recommended by [50]. The significant value of Bartlett's test of sphericity (1035.011, $df = 36$, $p = 0.000$) indicated that the variables had some correlations to each other. The constructs' validity was assessed by using convergent and discriminant validity. Table 2 presented the validity statistics of the technological and innovation performance model.

Table 2: Validity Statistics

Constructs	Composite Reliability (CR)	Average variance extracted (AVE)	Max./Average shared variance (MSV/ASV)	Constructs	
				Technological Performance	Innovation Performance
Technological performance	0.79	0.49	0.212	**0.70**	
Innovation performance	0.77	0.40	0.212	0.47	**0.63**

Validity statistics showed that both constructs had a good value of composite reliability, i.e., greater than 0.7, as suggested by [51]. They stated that if the average variance extracted is less than 0.5, but composite reliability is higher than 0.70, the latent measures' convergent validity is adequate. In addition, in [52], the authors proclaimed that AVE values > 0.4 are considered acceptable convergent validity measures. In the case of discriminant validity, there was no multicollinearity problem since inter latent variables correlation ($r = 0.47$) was less than the threshold value of 0.7, as recommended by [53]. Similarly, as per [51] criterion, AVEs of the constructs were greater than MSV/ASV. The square root of the constructs' AVEs (presented in bold in Table 2) was higher than the correlation between the constructs for satisfactory discriminant validity. Therefore, the constructs were distinct and measuring the theoretically-driven facts. Besides, absolute values of skewness (- 0.515 to - 0.152) and kurtosis (- 0.949 to - 0.142) of the observed variables proved normality of the data as they were reported within the recommended ± 2 range [54]. The absolute value of standardized residual covariance of the variables (- 1.862 to + 1.589) within the recommended range of ± 2.58 [55] also indicated that the latent variables had a significant effect on the model.

6. Results and Analysis

By administrating the organized questionnaire to 391 cell phone users of Nepal Telecom and Ncell, the required information was accounted for, dissected, and deciphered. According to the sample encircled, responses were taken from understudies (56.7 %), service holders (38.9 %), business persons (2.3 %), and freelancers (2.0 %). Female (51.7 %) respondents were somewhat higher than the male (48.3 %) respondents. Even though the field study was conducted in Kathmandu valley, it covered all provinces of Nepal since the respondents regularly used to come to Kathmandu valley at one time or another. Demographically, the respondents were from: Province 1 (19.4 %), Province 2 (9.5 %), Bagmati (36.8 %), Gandaki (11.5 %), Lumbini (12.5 %), Karnali (4.6 %) and Sudur Pashchim (5.6 %). The study ensured almost equal participation from Nepal Telecom (50.1 %) and Ncell (49.9 %) and showed that 79.0 % of users carrying the pre-paid subscriber identity module (SIM) card exclusively. All the respondents had at least experience of five years or more in carrying the cell phone services in Nepal.

Descriptive statistics observed that all nine test variables except VAR_9 (signal strength and coverage) averaged more than 3.5 on a six-point Likert scale and revealed that all the test items within the constructs had the right intensity. The signal strength and coverage (VAR_9) was perceived to the lowest mean value of 3.21 (below the average mean of 3.5). Such a result indicated that cellular companies need to improve their signal strength and coverage. The respondents' response dispersions from the mean of all indicator items were quite similar.

The study applied structural equation modeling (SEM) to evaluate the significance of the hypothesized paths and the model's explanatory power by computing multiple correlation coefficients for each endogenous variable. Technological and innovation performance was measured by nine test variables within two latent variables, as presented in Figure 2, and the key parameter estimates of the model shown in Table 3.

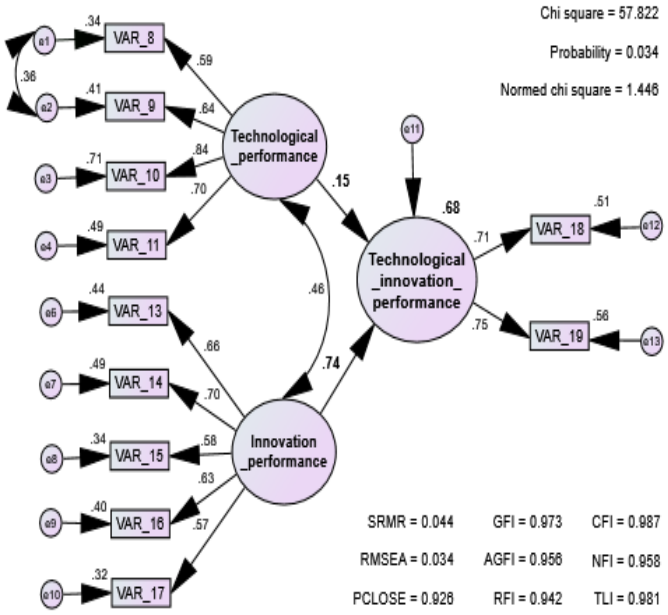


Figure 2: Technological and innovation performance model

Table 3: Parameter Estimates of the Technological and Innovation Performance Model

INDICATORS	URW	SRW	SE	CR	p-value
Technological performance					
---> Tech & innovation performance	0.142	0.152	0.061	2.326	0.020
Innovation performance					
---> Tech & innovation performance	0.941	0.743	0.119	7.920	***
Network and service quality (VAR_8)					
<--- Technological performance	0.857	0.586	0.086	9.950	***
Signal strength and coverage (VAR_9)					
<--- Technological performance	0.885	0.643	0.082	10.841	***
Voice quality (VAR_10)					
<--- Technological performance	1.179	0.844	0.097	12.198	***
Calls drop (VAR_11)					
<--- Technological performance	1.000	0.701			
Product/service innovation (VAR_13)					
<--- Innovation performance	1.220	0.663	0.131	9.329	***
Process innovation (VAR_14)					
<--- Innovation performance	1.274	0.702	0.132	9.637	***
Customization (VAR_15)					
<--- Innovation performance	1.115	0.584	0.130	8.604	***
Competitive innovation (VAR_16)					
<--- Innovation performance	1.185	0.629	0.131	9.032	***
Marketing innovation (VAR_17)					
<--- Innovation performance	1.000	0.565			
Overall technological aspect (VAR_18)					
<--- Tech & innovation performance	0.964	0.711	0.087	11.037	***
Overall innovation aspect (VAR_19)					
<--- Tech & Innovation performance	1.000	0.749			
Technological performance					
<--> Innovation performance	0.225	0.460	0.040	5.647	***
e1 <--> e2	0.300	0.364	0.055	5.416	***

*Note: URW = Unstandardized regression weights; SRW = Standardized regression weights; SE = Standard error; CR = Critical ratio; *** = Significant at 0.01 level*

As presented in Table 3, the technological and innovation performance model produced statistically significant critical ratios at $p \leq 0.05$ for all measuring test variables, covariance relationships between latent variables, and error terms. The testing model results produced standardized regression weights that were all significantly different from zero and above the 0.5 threshold level for a good fit, as recommended by [56]. The standardized regression weights indicated that innovation performance was a better predictor of technological and innovation performance ($\beta = 0.743$, $p < 0.01$) than technological performance ($\beta = 0.152$, $p < 0.01$). Various indices also observed to test the constructs' model fit, as presented in Table 4.

Table 4: Model Fit Indices of the Technological and Innovation Performance Model

Model fit measures	Cut-off value	Recommended by	Model of the study
Chi square (χ^2)	Smaller the better	Wan, 2002	57.822
Probability (p)	> 0.05	Wan, 2002	0.034
Normed chi square [χ^2/df]	≤ 3.00	Kline, 1998	1.446
Standardized Root Mean-square Residual [SRMR]	≤ 0.08	Hu & Bentler, 1999	0.044
Root Mean Square Error of Approximation [RMSEA]	≤ 0.08	Hu & Bentler, 1999	0.034
RMSEA Associated p Value [PCLOSE]	≥ 0.05	Garson, 2009	0.926
Goodness of Fit Index [GFI]	≥ 0.90	Bagozzi & Yi, 1988	0.973
Adjusted Goodness of Fit Index [AGFI]	≥ 0.90	Bagozzi & Yi, 1988	0.956
Relative Fit Index (RFI)	≥ 0.80	Hair et al., 2006	0.942
Comparative Fit Index [CFI]	≥ 0.90	Hu & Bentler, 1999	0.987
Normed Fit Index [NFI]	≥ 0.80	Bentler & Bonnet, 1980	0.958
Tucker Lewis Index [TLI]	≥ 0.90	Hu & Bentler, 1999	0.981

All the model fit indices were fallen in the range of recommended cutoff values and evidenced an excellent fit with the technological and innovation performance model. Exemplary values of the chi-square, normed chi-square, and probability as recommended by [57], [58], revealed the absolute and parsimonious fit of the model. Values of root-mean-square residuals, the root-mean-square error of approximation, associated p-value, and goodness of fit index as recommended by [59]-[61], further proved the absolute fit of the model. Values of adjusted goodness of fit index, comparative fit index, normed fit

index, and Tucker Lewis index as recommended by [59], [61], [62], supported the incremental fit of the model.

7. Discussion and Conclusion

Traditional financial measures-based backward-looking accounting systems' dissatisfaction led to a performance measurement revolution in the late 1970s [63]. Each organization has its reasons to measure performance. So, multidimensional performance measurement systems have been acknowledged since the 1990s. Scholars from different schools of thought have used the concept of subjective information to scale organizational performance. As the guidance of contingency theory, this study explores the subjective measures of technological and innovation performance in the Nepalese scenario. A literature review of the study identified 10 observed variables within two constructs to assess technological and innovation performance. The study did not accept the test variable e-CRM (VAR_12) as the determinant of technological performance in the Nepalese scenario though it was significant in the earlier studies like [23], [24].

The latent variable 'technological performance' was assessed from four observed variables: network and service quality ($\beta = 0.586$, $p < 0.01$); signal strength and coverage ($\beta = 0.643$, $p < 0.01$); voice quality ($\beta = 0.844$, $p < 0.01$); and calls drop ($\beta = 0.701$, $p < 0.01$) that directly influence the technological and innovation performance. Technology and information facilitate innovation in competitive markets [64]. As consistent with the past studies like [17]-[22], [65], the study confirms that network and service quality, signal strength and coverage, voice quality, and calls drops to have significant influence to choose the cell phone operators as the service provider. That means cell phone operators must ensure good and strong technological performance measures to satisfy their subscribers. Such evidences lead to an inference that technological performance at the Nepalese scenario was significantly associated with technological and innovation performance ($\beta = 0.152$, $p < 0.01$).

The latent variable 'innovation performance' was assessed from five observed variables: product/service innovation ($\beta = 0.663$, $p < 0.01$); process innovation ($\beta = 0.702$, $p < 0.01$); customization ($\beta = 0.584$, $p < 0.01$); competitive innovation ($\beta = 0.629$, $p < 0.01$); and marketing innovation ($\beta = 0.565$, $p < 0.01$). Service innovation influences customer satisfaction and retention [66] that eventually leads to better organizational performance. Process innovation applies a new method for doing something that helps an organization remain competitive and meets customer demands. Customization can govern the effectiveness of other capabilities and create higher competitive advantages [67]. Competitive innovation differentiates and adds competitive value to their products/services by adopting innovations and new technologies. Marketing innovation aims to give value to the customers and to improve competitive advantage. As consistent with previous research like [68]-[70], the study showed that the observed variables have a positive and significant effect on innovation performance. Besides, it supports the study conducted in [71] that innovation performance positively and significantly impacts technological and innovation performance. Hence, innovation culture has been pronounced as a pre-condition for improving organizational, marketing, and managerial entrepreneurship in a competitive environment [72]. In other words, the innovative capabilities have a direct and positive impact on organizational performance [69]. All the evidence lured that innovation performance in the Nepalese scenario was significantly associated with technological and innovation performance ($\beta = 0.743$, $p < 0.01$). As per the conceptual framework of the study, the overall model summary is presented in Figure 3.

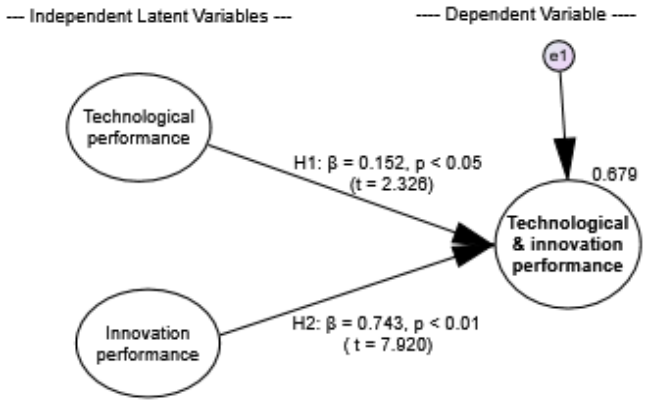


Figure 3: The model summary

The study outcomes inferred that each of the subjective performance measures provides a partial explanation of the synergetic impacts on organizational performance. The study observed nine contingent variables that quantified two response variables to measure technological and innovation performance. The technological and innovation performance depends on a fit or math between various contingent factors. Consequently, the study concludes that the model's validation and fit were vital because they provide empirical evidence for measuring non-financial organizational performance. More precisely, it offers the power of the Nepalese cellular telecommunications industry's technological and innovation performance measures. It is believed that the level of technological advancement with innovativeness is an essential determinant of organizational performance.

8. Limitations and Future Research

Studies of performance measurement systems cover several disciplines, and their methodological approaches differ considerably. It made a comprehensive examination extremely challenging. As a result, only two aspects of the performance measures, technological performance measures, and innovation performance measures have been considered in this study. The study variables were developed the most straightforward ways since they could be easily understandable and the respondents could mark the structured questionnaire's option precisely. The technological performance measures have been associated with the characteristics of mobile phone networks. In contrast, innovation performance measures have been associated with the business in general because a common respondent cannot grasp the technological essence of innovation performance measures in Nepal.

Although the analysis made for the study is rigorous and the findings significantly relatable, this study is a representative instead of a comprehensive effort. The key drawback was that the analysis used only the quantitative survey to collect the necessary

information with a structured questionnaire, using the terms accessible to ordinary users. The structured questionnaire denies the opportunity to discuss many of the relevant issues of the responses. Nonetheless, all the tests have been carried out to assess the authenticity and reliability of the information obtained. The reliability could depend on the number of participants that could have been further improved by a larger sample further emphasizing on the validity and generality of the findings.

The study examined new contingent variables in the cellular industry's organizational performance measurement system. The findings of the study will serve as a foundation for future studies to refine the theory. Researchers may consider more variables such as additional packages provided by operators, latency, security, etc. not covered in this analysis. Similarly, the ownership of firm can be integrated into the research model to identify whether the findings vary with various corporate holdings. On the other hand, future studies can concentrate on other sectors or other countries to address and compare related topics with this report. These results thus unravel the avenue for future research in various contingent variables to generalize the theory. Finally, this study hopes that the analysis findings will be useful for practitioners, researchers, or policy makers and will lead to related studies and future research as a guide.

9. Implications

Studies have shown that various performance assessment frameworks have been maintained across the world. In the case of Nepal, no corporation tends to make specific use of technological and innovative performance measures explicitly in the organizational performance process. Therefore, this study sought to analyze the effects of performance measures that could further improve the Nepalese cellular telecommunications industry's organizational performance. The consequences of this study have been divided into two parts: practical and theoretical perspectives.

9.1. Practical Implications

The study results would help telecommunications industry managers recognize the customers' perception of organizational success as a path to technical acceptance and innovation in the Nepalese scenario. The analysis provides more evidence for the various management decisions. Managers can focus on variables that obtained low weights from respondents. This study would also provide valuable guidance for identifying the critical driving variables of technical and innovative change and offer insightful knowledge on the system's strategic areas of organizational performance evaluation.

9.2. Theoretical Implications

The study is driven by the contingency theory, which provides a reliable lens to consider different variables and new ways to measure them. Every measure of a firm's performance has advantages and disadvantages, and provides a prospect of success for researchers. It endeavored to fill the literature gap and address the Nepalese scenario's issues not considered by previous researchers. On the other hand, the government can help companies innovate by leading and supervising their innovation efforts and by providing them with an amiable institutional environment.

References

[1] A. Elrahman, A. El-Borsaly, S. Hassan, "The impact of service quality on organizational performance in the mobile telecommunications sector in Egypt," Proceedings on Engineering Sciences, **2**(1), 93-104, 2020, doi: 10.24874/PES02.01.010.

[2] J. Barney, "Firm resources and sustained competitive advantage," Journal of Management, **1**(17), 99-120, 1991.

[3] N. Hussein, A. Mohamad, F. Noordin, N.A. Ishak, "Learning organization and its effect on organizational performance and organizational innovativeness: A proposed framework for Malaysian Public Institutions of Higher Education," Procedia-Social and Behavioral Sciences, **130**, 299-304, 2014.

[4] N.A. Morgan, D. W. Vorhies, C.H. Mason, "Market orientation, marketing capabilities, and firm performance," Strategic Management Journal, **30**(8), 909-920, 2009. doi:10.1111/1467-88963B.00097.

[5] N. Venkatraman, V. Ramanujan, "Measurement of business performance in strategy research: A comparison of approaches," Academy of Management Review, **11**(4), 801-14, 1986.

[6] T.Z.T. Mat, M. Smith, "The impact of changes in environment and AMT on management accounting practices and organizational strategy, structure and performance," Journal of Applied Management Accounting Research, **12**(1), 55-82, 2014.

[7] Nepal Telecommunication Authority, "Management information system report," **140**(188), 1-10, July 2020.

[8] M.U. Din, I.U. Mangla, M. Jamil, "Public policy, innovation and economic growth: An economic and technological perspective on Pakistan's telecom industry," The Lahore Journal of Economics, **21**, 369-391, 2016.

[9] J.A. Varela, M. del Río, "Market orientation behavior: An empirical investigation using MARKOR," Marketing intelligence & Planning, **21**(1), 6-15, 2003. https://doi.org/10.1108/02634500310458108

[10] J.S. Chen, T.H. Tsou, A.Y.H. Huang, "Service delivery innovation: Antecedents and impact on firm performance," Journal of Service Research, **12**(1), 36-55, 2009.

[11] K.H. Huarng, "A comparative study to classify ICT developments by economies," Journal of Business Research, **64**(11), 1174-1177, 2011. doi:10.1111/1467-98X.00097.

[12] S.P. Yeh, H-W. Fu, "Effects of service innovation on customer satisfaction in recreation industry," International Journal of Applied Mathematics and Statistics, **38**(8), 107-116, 2013. http://ir.lib.isu.edu.tw/handle/987654321/16547

[13] M.R. Saco, P.A. Goncalves, "Service design: An appraisal," Design Management Review, **19**(1) 10-19, 2008.

[14] S.K. McEvily, K.M. Eisenhardt, J.E. Prescott, "The global acquisition, leverage, and protection of technological competencies," Strategic Management Journal, **25**(8), 713–722, 2004. https://doi.org/10.1002/smj.425

[15] W.J. Orlikowski, "Using technology and constituting structures: A practice lens for studying technology in organizations," Organization Science, **11**(4), 404–428, 2000. https://doi.org/10.1287/orsc.11.4.404.14600

[16] Y. Xu, D.C. Yen, B. Lin, D.C. Chou, "Adopting customer relationship management technology," Industrial Management & Data Systems, **102**(8), 442-452, 2002. doi: https://doi.org/10.1108/02635570210445871

[17] N.K. Saha, M.R. Islam, A.U. Hoque, "Factors affecting customers' satisfaction of mobile phone subscribers: An empirical study on mobile telecommunication industry in Bangladesh. International Journal of Business and Management, **11**(6), 252-261, 2016.

[18] R. Negi, "Determining customer satisfaction through perceived service quality: A study of Ethiopian mobile users, International Journal of Mobile Marketing, **4**(1), 31-38, 2009.

[19] S. Rahman, A. Haque, M. Ahmad, "Choice criteria for mobile telecom operator: Empirical investigation among Malaysian customers," International Management Review, **7**(1), 50-57, 2011.

[20] K.S. Woo, H.K.Y. Fock, "Customer satisfaction in the Hong Kong mobile phone industry," The Service Industries Journal, **19**(3), 162–174, 1999.

[21] E.T. Aire, B.T. Maharaj, L.P. Linde, "Implementation considerations in a SIP based secure voice over IP network," in Proceedings of the 7th AFRICON, 167-172, Botswana, Africa, 2004.

[22] B. Eljaam, Customer satisfaction with cellular network performance: Issues and analysis, Ph. D. Dissertation, Iowa State University, United States, 2005.

[23] C. O'Leary, S. Rao, C. Perry, "Improving customer relationship management through database/internet marketing," European Journal of Marketing, **38**(3/4), 338-354, 2004.

[24] N. Kim-soon, M.F. Zulkifli, "The impact of electronic customer relationship management (e-CRM) on the business performance of small company," Journal of Engineering and Technology, **3**(12), 139-153, 2012.

[25] M. Dachyar, L. Hananto, "Innovation and quality service factors to customer loyalty in Indonesia telecommunication company by using structural equation modeling method," Management and Technology in

Knowledge, Service, Tourism & Hospitality, **1**, 41-44, 2014, https://doi.org/10.1201/b16700-10.
[26] M. Polder, G.V. Leeuwen, P. Mohnen, W. Raymond, "Product, process and organizational innovation: Drivers, complementarity and productivity effects. CIRANO–Scientific Publications, 2010s-28, 2010, http://dx.doi.org/10.2139/ssrn.1626805
[27] D.M. DeCarolis, "Competencies and imitability in the pharmaceutical industry: An analysis of their relationship with firm performance," Journal of Management, **29**(1), 27–50, 2003.
[28] S. Thornhill, "Knowledge, innovation and firm performance in high-and low-technology regimes," Journal of Business Venturing, **21**(5), 687–703, 2006.
[29] G. Zaefarian, S. Forkmann, M. Mitręga, S.C. Henneberg, "A capability perspective on relationship ending and its impact on product innovation success and firm performance," Long Range Planning, **50**(2), 184–199, 2017.
[30] T. Dotzel, V. Shankar, L.L. Berry, "Service innovativeness and firm value," Journal of Marketing Research, **50**(2): 259–276, 2013, http://dx.doi.org/10.1509/jmr.10.0426.
[31] A.F. Akgün, H. Ince, S.Z. Imamoglu, H. Keskin, I. Kocoglu, "The mediator role of learning capability and business innovativeness between total quality management and financial performance," International Journal of Production Research, **52**(3), 888–901, 2014.
[32] T. Anning-Dorson, "Organizational culture and leadership as mediators of service innovation and firm competitiveness: A study of an emerging economy," International Journal of Innovation Management, **20**(7), 1-29, 2016b.
[33] T. Anning-Dorson, "Interactivity innovations, competitive intensity, customer demand and performance," International Journal of Quality and Service Sciences, **8**(4), 536–554, 2016. doi: https://doi.org/10.1108/02635570210445871
[34] N.N.A. Aziz, S. Samad, "Innovation and competitive advantage: Moderating effects of firm age in foods manufacturing SMEs in Malaysia," Procedia Economics and Finance, **35**, 256-266, 2016.
[35] M.U. Hassan, S. Shaukat, M.S. Nawaz, S. Naz, "Effects of innovation types on firm performance: An empirical study on Pakistan's manufacturing sector," Pakistan Journal of Commerce and Social Sciences, **7**(2), 243-262, 2013.
[36] A. Rodríguez-Pose, R. Crescenzi, "Mountains in a flat world: Why proximity still matters for the location of economic activity," Cambridge Journal of Regions, Economy and Society, **1**(3), 371–388, 2008.
[37] H.V. Auken, A. Madrid-Guijarro, D. Garcia-Perez-de-Lema, "Innovation and performance in Spanish manufacturing SMEs," International Journal of Entrepreneurship and Innovation Management, **8**(1), 36-56, 2008.
[38] K. B. Clark, T. Fujimoto, Product development performance: Strategy, organization, and management in the world auto industry, Boston: Harvard Business School Press, 1991.
[39] S. A. Zahra, G. George, "Absorptive capacity: A review, reconceptualisation and extension. Academy of Management Review, **27**(2), 185-203, 2002, https://doi.org/10.5465/AMR.2002.6587995.
[40] D. Leonard-Barton, "Core capabilities and core rigidities: A paradox in managing new product development", Strategic Management Journal, **13**(S1), 111–125, 1992.
[41] R. Blundell, R. Griffith, J. Van Reenen, "Market share, market value and innovation in a panel of British manufacturing firms", Review of Economic Studies, **66**(3), 529-554, 1999, DOI:10.1111/1467-937X.00097.
[42] A. De Massis, F. Frattini, E. Pizzurno, L. Cassia, "Product innovation in family versus nonfamily firms: An exploratory analysis", Journal of Small Business Management, **53**(1), 1-36, 2013, DOI:10.1111/jsbm.12068.
[43] C. Carmen, G. M. Jose, "The role of technological and organizational innovation in the relation between market orientation and performance in cultural organizations, European Journal of Innovation Management, **11**(3), 413-434, 2008, DOI: 10.1108/14601060810889035.
[44] B. Menguc, S. Auh, "Creating a firm-level dynamic capability through capitalizing on market orientation and innovativeness", Academy of Marketing Science Journal, **34**(1), 63-73, 2006.
[45] A. Koohang, J. Paliszkiewicz, J. Goluchowski, "The impact of leadership on trust, knowledge management, and organizational performance: A research model, Industrial Management & Data Systems, **117**(3), 521-537, 2017.
[46] R. Krejcie, D. Morgan, "Determining sample size for research activities," Educational & Psychological Measurement, **30**, 607-610, 1970.
[47] B.L. Bowerman, R.T. O'Connell, J.B. Orris, Essential business statistic*s* (3rd ed), McGrawHill, 2004.
[48] J.C. Nunnally, Psychometric theory (3rd ed.), McGraw-Hill, 1993.
[49] L.A. Clark, D. Watson, "Constructing validity: Basic issues in objective scale development," Psychological Assessment, **7**(3), 309-319, 1995.
[50] H.F. Kaiser, "An index of factorial simplicity," Psychometrika, **39**, 31–36, 1974.
[51] C. Fornell, D.F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," Journal of Marketing Research, **18**(1) 39-50, 1981.
[52] R.P. Bagozzi, H. Baumgartner, The evaluation of structural equation models and hypothesis testing, In R. P. Bagozzi (ed.), Principles of marketing research 386–422, Cambridge: Blackwell, 1994.
[53] L.S. Meyers, G. Gamst, A.J. Guarino, Applied multivariate research: Design and interpretation, Sage Publications, Inc., 2006.
[54] D. George, M. Mallery, SPSS for windows step by step: A simple guide and reference, 17.0 update (10a ed.), Pearson, 2010.
[55] B.M. Byrne, Structural equation modeling with AMOS: Basic concept, applications, and programming (2nd ed.), Taylor and Francis Group, 2010.
[56] J.F.Hair, W.C. Black, B.J. Babin, R.E. Anderson, R.L. Tatham, Multivariate data analysis, Prentice Hall, Pearson Education, Inc., 2006
[57] T.T.H. Wan, Evidenced-based health care management: Multivariate modeling approaches, Kluwer Academic Publishers, 2002.
[58] R.B. Kline, Methodology in the social sciences: Principles and practice of structural equation modeling, Guilford Press, 1998.
[59] L. Hu, P. Bentler, "Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives," Structural equation modeling - A multidisciplinary journal, **6**(1), 1-55, 1999.
[60] G.D. Garson, "Structural equation modeling", *2009,* http://faculty.chass.ncsu.edu/garson/PA765/structur.htm
[61] R. Bagozzi, Y. Yi, "On the evaluation of structural equation models, Journal of the Academy of Marketing Science, **16**(1), 74-94, 1988.
[62] P.M. Bentler, D.G. Bonnet, "Significance tests and goodness of fit in the analysis of covariance structures," Psychological Bulletin, **88**, 588-606, 1980.
[63] S. S. Nudurupati, U.S. Bititci, V. Kumar, "State of the art literature review on performance measurement," Computers & Industrial Engineering, **60**(2), 279-290, 2011.
[64] A. Sood, G. J. Tellis, "Do innovations really pay off? Total stock market returns to innovation, Marketing Science, **28**(3), 442–456, 2009.
[65] S. Khan, S. Afsheen, "Determinants of customer satisfaction in telecom industry–A study of telecom industry Peshawar KPK Pakistan," Journal of Basic and Applied Scientific Research, **2**(12), 12833-12840, 2012.
[66] S. Salunke, J. Weerawardena, J.R. McColl-Kennedy, "Competing through service innovation: The role of bricolage and entrepreneurship in project-oriented firms," Journal of Business Research, **66**(8), 1085-1097, 2013.
[67] C. K. Prahalad, V. Ramaswamy, "Co-creation experiences: The next practice in value creation," Journal of Interactive Marketing, **18**(3), 5–14, 2004. doi: https://doi.org/10.1108/02635570210445871
[68] S. Afriyie, J. Du, K. Appiah, A-A. I. Musah, "The nexus between types of innovation and marketing performance of SME in an emerging economy," International Review of Management and Marketing, **8**(6), 78–92, 2018, https://doi.org/10.32479/irmm.7042
[69] L.F. Hsieh, L.H. Lin, "A performance evaluation model for international tourist hotels in Taiwan - An application of the relational network DEA," International Journal of Hospitality Management, **29**, 14–24, 2010.
[70] S. Yıldız, F. Baştürk, I. T. Boz, "The effect of leadership and innovativeness on business performance," Procedia-Social and Behavioral Sciences, **150**, 785–793, 2014.
[71] V.H. Nam, "The roles of human and social capital in the development of manufacturing SMEs in Vietnam, Journal of Economic Development, **16**(1), 5-22, 2014, https://doi.org/10.33301/2014.16.01.01.
[72] H. Aksoy, "How do innovation culture, marketing innovation and product innovation affect the market performance of small and medium-sized enterprises (SMEs)?" Technology in Society, **51**, 133-141, 2017, https://doi.org/10.1016/j.techsoc.2017.08.005.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1021-1026 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

An Analysts' Skills: Bespoke Software vs Packaged Software at Small Software Vendors

Issam Jebreen*

Department of Software Engineering, Zarqa University, Zarqa, 13110, Jordan

ARTICLE INFO

Article history:
Received: 10 October, 2020
Accepted: 21 November, 2020
Online: 08 December, 2020

Keywords:
Small Software Vendors
Packaged Software
SMEs
Analysts Skills

ABSTRACT

SMEs usually succeed in today market since they achieve unique things in the business market. Therefore, SMEs look to keep their business benefit by avoiding any regulation conducted of software. Executed software packaged at SMEs consequently provides obstacle concerning to misfits between software services process and SMEs business process. Hence software analysts' skills are a critical factor for successfully overcome obstacle relating to misfits. This study conducted a questionnaire by quantitative data to understand in-depth analysts' skills need during the process of software mismatch adjustment. Based on our analysis of the data gathered, we represented a list of skills that analysts applied to detect misfits between software service provided and SME's business process. This study, therefore, explains skills that analysts used to overcome the obstacle of software implementation at SMEs.

1. Introduction

In current years Small Software Vendors (SSVs) for Packaged Software (PS) is increased they are competitive by selling their products to large numbers of SMEs [1]. SSVs have, therefore activated to mark SMEs less complicated products that have developed [2]. SMEs are an essential business target many SSVs as mentioned by [3] SMEs "with less than 500 employees provided 51 per cent of all employment in the USA as of March 2004 and 64 per cent of all Canadian private sector employment in 2005. In the European Union, firms with 250 employees or less provided 67 per cent of employment" [3]. A study by [4] stated that SMEs subsidize to economic growing to a great degree and incessantly generating employ chances through the business. SMEs are an essential economies part, SMEs surface particular challenges once they are implementing a PS for their business [5, 6].

SMEs seem to remain different comparing to large companies in several points [7]. Some individual features of SMEs contain types of ownership, business structure, and target market [7]. Several studies have stated that SMEs have limited IT knowledge and resource regarding IT adaptation/adoption [7]. These obstacles lead to the implementation of PS presence a challenge for SMEs. Researchers of SSVs have found that they cannot conduct the process of implementing PS at a large business to SMEs [7]. Regardless of the importance of Packaged Software (PS) conducting it is recognized by these SSVs researchers, there is a little investigation to exploring these obstacles. In particular, considerations about SMEs hardly investigate in the SSVs literature and the question of how SSVs have overcome these obstacles and what are the analysts' skills to successfully conducted a PS at SMEs [2].

In [8], the author conducted research relating to the soft skills of software development staff across four significant parts of the world, which included North America, Asia, Europe, and Australia. researchers conducted on the belief that soft skills are complementary to technical skills, and thus essential attributes for IT staff to have [8] claimed that during different project stages, incredibly soft skills required. Therefore, various team members should have a variety of soft skills. Those skills were communication skills, interpersonal skills, analytical and problem-solving skills, organizational skills, innovative skills, adaptability to change, and the ability to work in a team and to work independently. Other study by [9] regarding analysts' communication skills a key factor for requirement gathering, they found that knowledge about business process of users' companies and domain knowledge different business are the most critical factor for requirements gathering.

It can claim that none of the previous studies related to analysts' skills focuses on any detail on how analysts manage software specificity and generality when implementing packaged software compare to bespoke software [10]. Hence, they do not provide any overview of what domain knowledge and estimation skills required from analysts implementing packaged software. Nor do they focus on the fact that there are pronounced differences

*Corresponding Author: Issam Jebreen, ijebreen@zu.edu.jo

between bespoke software and packaged software. Therefore, there is a need for an investigation of which domain knowledge and estimation skills are required when analysts manage software specificity and generality.

This study provides an overview of analysts' skills at small software vendors (SSVs) regarding PS conducted at SMEs. It explores this area from the perspective of small software vendors. In focusing on an understanding of analysts' skills at SSVs, we construct a questionnaire by quantitative data to represent high demand for skills used during the conduct of PS at SMEs. This study organized as follows: review literature; research method; results and discussed, conclusion and considers future work.

2. Study Background

The purpose of this section is to provide a perception of the theoretical concept of software packaged; analysts' skills and the processes involved, and packaged software implementation. In this section, we assumed that PS conducted at SMEs depend on a set of techniques such as software integrations, customization, adoption/adaption, and the identification of misfits between PS services and SMEs business process. Hence these processes require from PS analysts to interact with SMEs employees that require different skills. Additionally, in this section, during the review of previous studies relating to packaged software, we present and define several PS elements. There are several elements involved in the process of PS conduction such as PS customization, adaption, and adoption [11]. If the efficiency of these elements enhanced, this might lead to the increased success of PS conduction at SMEs.

The market of PS is the fastest rising business for small software vendors (SSVs), [12] study shown that PS market for SSVs is growing to $64.88 billion in 2009. Therefore, such procedures of software may be very targeted to viable SMEs business [12].

Small software vendors shape their business to developing small packaged software (SPS) instead of bespoke software. However, those vendors have faced new management development challenges [13], [14] recognized the lack of in research investigate dedicated to reviewing the packaged software area, he recognizing main differences between the packaged software development process and bespoke software. The study acknowledged variances at four levels that contain "industry forces, approaches to software development, work culture, and development team efforts". His argument of the conflicts increased by recognizing and discoursing the five different participant groups participate in both bespoke and PS development. In [14], the author stated that "custom IS are those made by either an organization's internal staff or by direct subcontract to a software house". The study goes on to classify ERP product as the wildest rising example of a packaged software product. PS requires wide modifying for their application, which regularly involves the help of variances parties such as training, consultants, and support staff [15-17].

When associating noticeable differences between bespoke and PS developing process, Sawyer states that time pressures are a main critical factor for PS development rather than software cost compare to bespoke software. Meanwhile, the successful implementation criteria for PS software is different compare to bespoke software, PS software assesses by software profit, several target markets, and famous the product. However, successful implementation criteria for bespoke software, that is measured by specific company satisfaction about the services provided by the software. Moreover, the PS development is relying on developers' experience of business domain knowledge, but bespoke software is depending on the needs of users.

In [14-17], and other research by [13] and [18], the authors stated that software companies treat PS as a product. Therefore, the effort is on carrying a product that can sell to many, and the product vision relying on releases planned. Meanwhile, bespoke software, the effort is on carrying on software development process in order to have a user's stratification. The primary assumption of these researchers is that the process of software development is different from PS and bespoke software, the most distinct is at software company structure level, the process of development regarding users' involvement, the culture of work, development effort by software teams.

In recognition of these concerns, systematic studies have been undertaken in the current year to increase the success of the method of packaged software creation [11]. For example, in order to understand the process of requirement engineering for PS, [14] researched the process and challenges of the requirement engineering process for PS in Swedish software companies. Many of the problems they exposed were exclusive to PS and not suitable for personalized applications. They therefore noted that specifications for engineering approaches carried out by a tailor-made software team might not be especially helpful in supporting the development process for PS. In another study by [10], it was reported that the personalized process of software requirement engineering and the methods used have inadequate utility for PS.

For analysts, these present a primary challenge in creating the preliminary software for future marketing. However, as possible consumers are accepted, they have become the primary source for the selection of requirements. Nevertheless, the original software / product specifications are typically generated by the creators of PS, who have used their thoughts on business goals, understanding of the business domain or a product vision [14]. This method makes it rational to assume that not all packaged software services and features are suitable for the business of SMEs [15]. In other words, there can be misfits between the business processes of PS services and SMEs. In order to minimize any unintended effects, SSVs and recognized SMEs should reduce the possible gaps (a skills require from analysts) in the PS to be effective.

Therefore, several studies conducted to investigate the critical success factors (CSFs) for PS implementation. Instead of that, [19] surveyed 86 organizations by an essential characteristic of success PS implementation to generate the most important critical success factors for PS implementation. They propose that the CSFs list may include assistance leaders of PS projects to better apply inadequate resources by using the CSFs that are existences at SMEs and have a primary effect on the PS success implementation. The study offers a brief clarification of critical success factors, along with an interpretation of the value of each aspect. (1) Management support, (2) product champion, (3) consumer

training, (4) expectation management, and (5) supplier / customer relationships are the leading five variables recognized.

The purpose of [17] is not only to test the accuracy of the top 10 list of [16], but to build on the list by theorizing some of the causal relationships between the individual CSFs. The research question was: (1)" Can the list of [16] helps to gain a deeper understanding of the root causes of success and failure in the implementation of ERP?" [17] agreed to evaluate the top 10 CFSs that have a mix of 'hard' and 'soft' success elements, and they noted that the most significant factors are: top management support, project team experience, interdepartmental cooperation, consistent expectations and goals, and project management. Any of the stuff on the top 10 list were originally ranked lower by [16]. Still, after feedback received from 52 company managers approached to provide input by [17], he moved the ranking up. In other IT literature related to implementation, many of the items on the top 10 list also seem to appear frequently.

In [4], the author reviewed in the words of SMEs, the most recent literature on the adoption and application of ERP systems in SMEs. Noting that ERP systems have now been almost universally implemented by large organizations, [4] said that ERP vendors have now begun to turn their attention to small-medium-sized organizations (SMEs). Although ERP systems can help small and medium-sized enterprises, "the risks for small and medium-sized enterprises of implementing an ERP system are different because small and medium-sized enterprises are likely to have limited resources and business features that are different from those of large organizations." In [4], the author provided insight into the areas missing from the current ERP adoption analysis in SMEs and provided expertise to assist 'professionals, suppliers and SMEs' while embarking on ERP projects.' In fact, 'Small and medium-sized enterprises (SMEs) recognized as fundamentally different circumstances relative to large enterprises'[18] have not published any relevant literature on the implementation process in SMEs [4].

The literature mentioned by [4] shows that academic research has shown a significant increase in the use of ERP in small and medium-sized enterprises, and that case studies and surveys are the most common methods used in research papers on this topic. They discovered that the implementation method was the most discussed in the literature on ERP use in SMEs. This result contributes to the crucial topics of literature debate within larger organizations on ERP systems. However, in the literature on the use of ERP in SMEs, the adoption decision, the acquisition process, and the use and maintenance step are also given acceptable levels of emphasis. The processes for which literature has been rather scarce or non-existent are ERP evolution and ERP system retirement [4]. Further suggested that only two published papers found 'in-house built systems' to be a viable option for small and medium-sized enterprises. At the same time, "normal ERP solutions could require straight lines and a lack of versatility for particular SMEs." It is therefore safe to say that, from the point of view of SSVs, the existing literature has paid little attention to the implementation of PS activities by SSVs.

In [5], the author make some more remarks about the existing literature after giving specific critical of the reviewed literature and suggest more avenues for research. First of all, they say that, even though 77 papers identified and examined, this was still a minimal number of documents to be conducted within ten years on the subject, considering the increasing significance of ERP technology in response to SMEs. They conclude that "in contrast with ERP in LEs, SMEs have not earned sufficient attention." "that include a lack of" ex-ante cost estimate, economic viability and investment appraisal studies of ERP ventures "research, a lack of contrast between" SME-specific ERP and general ERP systems " or "industry-specific ERP vs general ERP packages" studies.

In [4], the author found that only a few studies were performed on the growth of ERP systems in small and medium-sized enterprises, and no study on small and medium-sized enterprises discussed the implementation phase of an ERP system. Finally, while they found 77 articles related to ERP systems in small and medium-sized enterprises, [4] recorded that the majority of small and medium-sized enterprises were engaged in traditional manufacturing, and that it could be important to achieve results compared to different industries, or it could be beneficial if recent studies on the use of the ERP system in small and medium-sized enterprises were carried out. They also found that almost all research studies have found that businesses are concentrated in America, Australia, Europe, and Asia. There has been a shortage of studies that examine SMEs in Africa or the Middle East. Therefore, a one-sided viewpoint (in data collection) has been adopted by current literature, e.g. on the client-side, whereas other views might strengthen the perception of such phenomena. Any studies that examine instances of failed ERP implementations within SMEs are also absent.

After reviewing the literature on this subject, package software often created in many sequential releases but that there is intense competition among various formats within the packaged software industry. This is only one of the components unique to packaged software, which explains part of why the features of packaged software vary significantly from the elements of customized software. Alternatively, there are potential clients, an imagined group of people who may fit the profile of the product's intended customer. The elicitation of specifications from this group of users and customers is one of the tasks that separates bundled applications from bespoke. The elicitation of such requirements is primarily handled by ads, technical assistance, user groups and trade publication testers. Therefore, it seems reasonable to say that not all the applications functionalities will be appropriate for the customers' business requirements. In other words, when packaged software is implemented by tech firms, there would be inconsistencies between the features provided by packaged software and customers ' business requirements. Therefore, different skills are required from PS analysts in order to successfully implement a PS.

Packaged program development can be a massive undertaking that needs a substantial amount of time, effort and adjustments in the implementation organization. The implementation of the PS is often the single largest project ever launched by an organization.

3. Research Design

A survey questionnaire on the skills adopted was used to collect quantitative data based on [8-10] and shows the skills required in PS. The skills mentioned in the table below in terms of how analysts used them.

In cases where expectations of one's competence affect the way he or she interacts, the Self-report instruments are suitable [20]. Besides, claimed that people could have more contextual information about themselves than anyone else does, offering some encouragement for the use of such a system.

As show in table 2, our sample of software companies involved in this research covers the area of software development, integration of systems, and localization of software. A total of 60 participants presented, including analysts and developers. Seven of the participants made up of team leaders.

Table 1: Survey Questionnaire Sample

Skills	
Skills & Knowledge	Assessment Level
Communication skills	Low, Moderate, and High
Interpersonal skills	Low, Moderate, and High
Organizational skills	Low, Moderate, and High
Team Player	Low, Moderate, and High
Ability to Work Independently	Low, Moderate, and High

Table 2: Sample descriptions

Area	%Sample	Participants	Experiences
development	85%	Team Leader	12%
Integration	70%	Analysts & Developers	75%
localization	40%		
Experiences	12%	75%	

System analysts and developers were most of the participants at the same time. Most participants had a cumulative experience of 3-10 years in the industry, while a few had the experience of more than ten years. Most of the participants had experience working concurrently as researchers, designers, and developers. As analysts only, some participants had the expertise and some as developers only. Most participants had experience with software for company applications and software for database systems.

The data has analyzed as the percentage of participants answers regarding skills required for bespoke software and packaged software.

Table 3: Responses regarding software type.

Software type	#Responeses
Bespoke	27
Packaged	33

As show in table 3, it was 27 responses from bespoke software analysts and developers and 33 responses from packaged software. The top of questionnaire, the participants were chosen the type of software they would like to fulfill up the questionnaire for.

4. Result & Discussion

The table 4 below follows a list of skills adopted by [8-10] and shows the skills required in PS vs Bespoke software. The skills mentioned in the table in terms of how analysts used them.

It can be seen from the table 4 that the skills generally required of analysts practicing Packaged software implementation (PSI) are much the same as those required of analysts engaged in bespoke software, but that some of the skills are required to be practiced at a higher level or to use more often. There are some differences involved in terms of how skills are needed and practiced about skills and knowledge, development skills, software knowledge, business skills, problem-solving, and hardware knowledge.

Table 4: Skills Assessment Level

Skills		
Skills & Knowledge	**Bespoke software**	**PS**
Communication skills	High	High
Interpersonal skills	Moderate	***High***
Organizational skills	Low	Low
Team Player	Moderate	Moderate
Ability to Work Independently	Moderate	***High***
Development skills		
Programming	High	High
General knowledge of development	High	High
Implementation	High	High
Operations/maintenance	High	High
Design	Moderate	Moderate
Analysis	High	High
Documentation	Moderate	Moderate
Development methodologies	Moderate	Moderate
Integration	Low	***High***
Knowledge of technological trends	Low	***High***
Quality assurance	Low	***High***
Software		
Programming language	High	High
Database	High	High
Operating systems / platforms	Moderate	***High***
Packages	Moderate	***High***
General knowledge of s/w	Moderate	Moderate
Business skills		
General knowledge of business	High	High
Function specific	Moderate	***High***
Industry specific	Low	***High***
Enterprise-wide	Low	***High***
Problem Solving		
Technical expertise	Low	***High***
General problem solving	Low	***High***
Adaptive / flexible	Low	Low
Analytical / critical / logical	Low	***High***
Customer-oriented	Low	***High***
Hardware		
Server	Low	***High***
General knowledge of h/w	Low	***Moderate***
Desktop/PC	Low	***Moderate***
Devices/printers/storage	Low	Low

Note that the bolded italics in the PS column show the differences

Bespoke software requires the analyst to show a certain level of skills and knowledge related to communication skills, interpersonal skills, organizational skills, ability to work in a team, and ability to work independently. PSI requires the same kinds of skills, but some of these skills demanded at a different level as shown in figure 1.

For example, the analyst engaged in PSI will need to meet higher demands in terms of their development of interpersonal skills, and their ability to work independently.

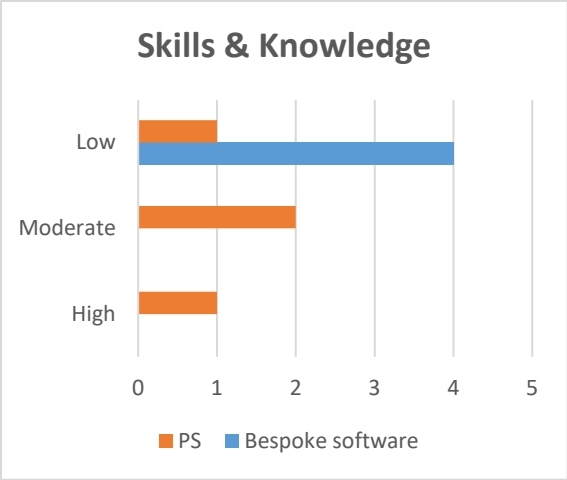


Figure 1: Skills & Knowledge

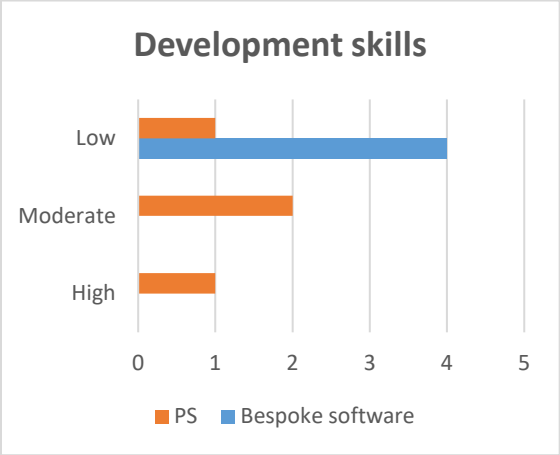


Figure 2: Development Skills

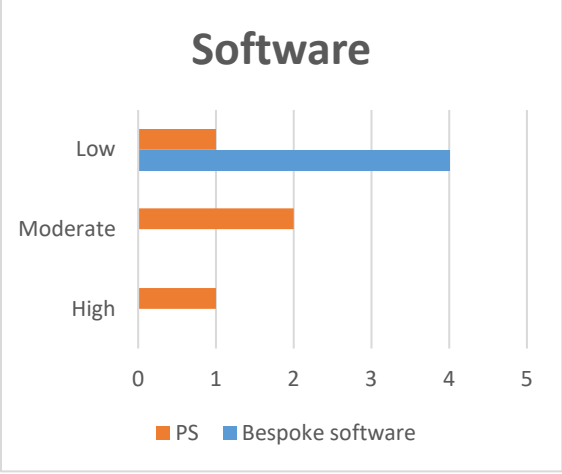


Figure 3: Software

When it comes to development skills as shown in figure 2, the analyst engaged in PSI will need to have a higher level of skill when it comes to dealing with software integration, quality assurance, and knowledge of technological trends. In bespoke software, these skills only required at a 'low' level. In PSI, the need for these skills categorized as 'high'.

When it comes to knowledge of software and the ability to use the software as shown figure 3, the skills required of bespoke software analysts and PSI analysts also differ. In Bespoke software, the analyst's knowledge about operating systems/platforms and packages only needed to be at a 'moderate' level. In PSI, these skills required to be at a 'high' level.

Bespoke software and PSI also require a different use of business skills as shown in figure 4. The analyst engaged in bespoke software needs a high level of knowledge about the business, as does the analyst employed in PSI.

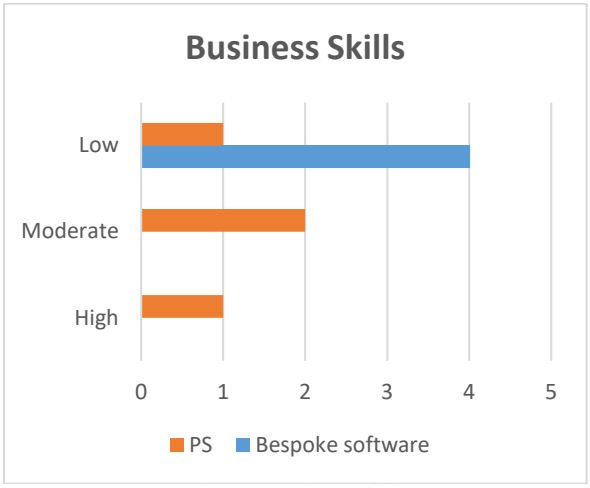


Figure 4: Business Skills

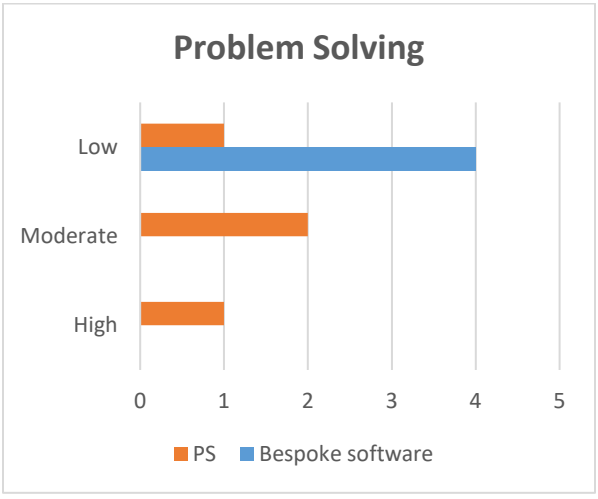


Figure 5: Problem Solving Skills

However, in bespoke software, the analyst needs only a 'moderate' understanding of specific functions. In PSI, this knowledge of particular procedures must be high. In bespoke software, the analyst's industry-specific and enterprise-wide skills are 'low', whereas, in PSI, they practiced at a 'high' level.

The skills related to 'problem-solving' also differ between bespoke software and PSI as shown in figure 5.

The ability to be adaptable or flexible required at the 'low' level in both forms of RE. Several other skills needed to differ in terms of use, however. For example, skills related to general problem-solving, technical expertise, the ability to be analytical, critical, or logical, and to engage in customer-oriented problem solving

practiced at a 'low' level in bespoke software. These same skills practiced at a 'high' level in PSI.

Lastly, analysts' skills related to hardware also differ as shown figure 6. The analyst doing bespoke software requires only a 'low' level of skills concerning general knowledge about hardware, knowledge about servers, desktop PCs, and knowledge about devices and storage. The requirement of these skills is slightly different for the PSI analyst. In PSI, the analyst has 'moderate' skills concerning general knowledge about hardware, desktop computers, and PCs. As in bespoke software, they need only 'low' knowledge about devices, storage, and printers. However, they will need to have a 'high' level of knowledge about servers.

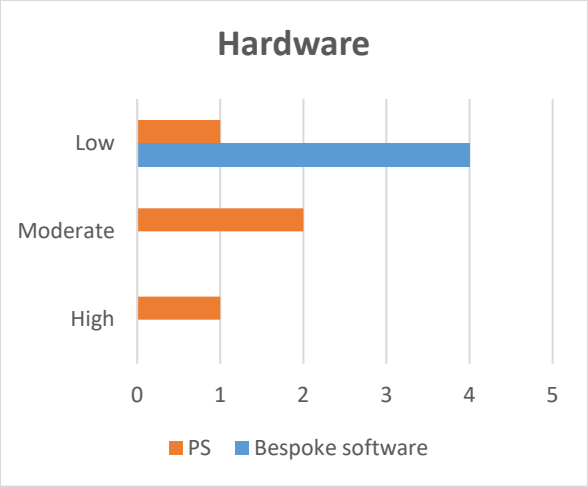


Figure 6: Hardware

5. Conclusion

This study provides an account of the distinction between the skills of analysts for packaged software and bespoke software with an emphasis on the activities of analysts. Through this report, we offer a detailed discussion of the skills of analysts required to be effectively implemented by PS and the challenges faced when implementing packaged software at SMEs. The numerous factors involved in seeking buyers, eliciting requirements and detecting misalignments, designing a packaged software product, and updating or adapting existing packaged software have been addressed.

The expertise and abilities of analysts to communicate with consumers should be considered as crucial factors when attempting to enhance the actions and results of the application of the PS. By enhancing their development skills, in particular integration, awareness of technical developments, and quality assurance skills, analysts will gain confidence, strengthen their critical thinking capabilities and problem-solving skills, and enhance their ability to connect, engage, and cooperate with the implementation of PS.

Business skills are the most important for understanding the business process of SMEs based on the responses of our 60 participants. In view of this, it is important to improve well-developed function-specific skills to minimize misunderstandings and increase the efficacy of such communication. In addition, the involvement of analysts with inadequate skills to work efficiently during the implementation of PS, showing weak or poorly performed interaction skills, may also lead to fragmented teams and disagreements between team members, resulting in an unsatisfactory experience for both users and analysts. This analysis of skills analysts in PS can help to shed light on areas of practice that can be focused on by an analyst to reduce the risk of problematic implementation of PS. Future studies will seek to establish a theoretical framework that describes the interaction techniques that analysts apply during the production and implementation of the PS, as this research is ongoing.

References

[1] V. Morabito, S. Pace, and P. Previtali, "ERP marketing and Italian SMEs," European Management Journal, **23**, 590-598, 2005. doi.org/10.1016/j.emj.2005.09.014.

[2] O. Zach and B. E. Munkvold, "Identifying reasons for ERP system customization in SMEs: a multiple case study," Journal of Enterprise Information Management, **25**(5), 462-478, 2012, https://doi.org/10.1108/17410391211265142

[3] F. Kramer, T. Rehn, M. Schneider, and K. Turowski, "ERP-adoption within SME—challenging the existing body of knowledge with a recent case," in Multidimensional Views on Enterprise Information Systems, ed: Springer,**12**, 41-54, 2016, doi.org/10.1007/978-3-319-27043-2_4

[4] S. B. Gunjati and C. Adake, "Innovation in Indian SMEs and their current viability: A review," Materials Today: Proceedings, **28**(4), 2325-2330, 2020, doi.org/10.1016/j.matpr.2020.04.604.

[5] M. Haddara and O. Zach, "ERP systems in SMEs: A literature review," in 2011 44th Hawaii International Conference on System Sciences,1-10, 2011, doi: 10.1109/HICSS.2011.191

[6] B. Snider, G. J. da Silveira, and J. Balakrishnan, "ERP implementation at SMEs: analysis of five Canadian cases," International Journal of Operations & Production Management, **29**(1), 4-29. 2009, https://doi.org/10.1108/01443570910925343.

[7] S. Laukkanen, S. Sarpola, and P. Hallikainen, "Enterprise size matters: objectives and constraints of ERP adoption," Journal of enterprise information management, **20**(3), 319-334, 2007, https://doi.org/10.1108/17410390710740763.

[8] F. Ahmed, L. F. Capretz, S. Bouktif, and P. Campbell, "Soft skills requirements in software development jobs: A cross-cultural empirical study," Journal of systems and information technology, **14**(1), 58-81, 2012, https://doi.org/10.1108/13287261211221137.

[9] I. Jebreen and A. Alqerem, "Critical proficiencies for requirements analysts: reflect a real-world needs," Int. Arab J. Inf. Technol., **15**(3A), 626-632, 2018,

[10] I. Jebreen and A. Al-Qerem, "Empirical Study of Analysts' Practices in Packaged Software Implementation at Small Software Enterprises," International Arab Journal of Information Technology (IAJIT), 14, 2017.

[11] M. Tamimi and I. Jebreen, "A Systematic Snapshot of Small Packaged Software Vendors' Enterprises," International Journal of Enterprise Information Systems (IJEIS), **14**(2), 21-42, 2018, doi: 10.4018/IJEIS.2018040102

[12] L. Staehr, G. Shanks, and P. B. Seddon, "An explanatory framework for achieving business benefits from ERP systems," Journal of the Association for Information Systems, **13**(6), 2, 2012, doi: 10.17705/1jais.00299

[13] T. Gorschek, A. Gomes, A. Pettersson, and R. Torkar, "Introduction of a process maturity model for market-driven product management and requirements engineering," Journal of software: Evolution and Process, **24**, 83-113, 2012, https://doi.org/10.1002/smr.535

[14] P. Sawyer, "Packaged software: challenges for RE," in Sixth International Workshop on Requirements Engineering: Foundations of Software Quality (REFSQ 2000) Stockholm, 2000.

[15] C. Moon and V. Torossian, "System and method for classification of documents," ed: Google Patents, 2007.

[16] D. L. Olson and J. Staley, "Case study of open-source enterprise resource planning implementation in a small business," Enterprise Information Systems, **6**, 79-94, 2012, doi.org/10.1080/17517575.2011.566697

[17] S. Sawyer, "Effects of intra-group conflict on packaged software development team performance," Information Systems Journal, **11**(2), 155-178, 2001, doi.org/10.1046/j.1365-2575.2001.00100.x

[18] L. Karlsson, Å. G. Dahlstedt, B. Regnell, J. N. och Dag, and A. Persson, "Requirements engineering challenges in market-driven software development–An interview study with practitioners," Information and Software technology, **49**(6), 588-604, 2007, doi.org/10.1016/j.infsof.2007.02.008.

[19] T. M. Somers and K. Nelson, "The impact of critical success factors across the stages of enterprise resource planning implementations," in Proceedings of the 34th Annual Hawaii International Conference on System Sciences, 2001, doi**:** 10.1109/HICSS.2001.927129

[20] I. Jebreen, "Requirements determination as a social practice: Perceptions and Preferences of novice analysts," Computer and Information Science, **8**(3), 134, 2015, doi:10.5539/cis.v8n3p134

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1027-1032 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Ozone, Nitrogen Dioxide, and $PM_{2.5}$ Measurement at Three Urban Parks in Manila, Philippines using Portable Sensors

Maria Cecilia Galvez[*,1,2], Daniel Paulo Tipan[1], Angelo Ashtin Valera[1], Edgar Vallar[1,2], Alma Nacua[3]

[1]*Environment And RemoTe sensing researcH (EARTH) Laboratory, Physics Department, De La Salle University, 2401 Taft Avenue, Malate, Manila, 0922, Philippines*

[2]*Applied Research for Community, Health and Environment Resilience and Sustainability (ARCHERS), Center Natural Sciences and Environmental Research (CENSER), De La Salle University, 2401 Taft Avenue, Malate, Manila, 0922, Philippines*

[3]*Universidad de Manila, Cecilia Muñoz St., Ermita, Manila, 0922, Philippines*

ARTICLE INFO

Article history:
Received: 29 July, 2020
Accepted: 18 November, 2020
Online: 14 December, 2020

Keywords:
Air Quality
Portable Sensors
Urban Parks

ABSTRACT

The presence of urban parks and green spaces in the city can play a role in maintaining urban residents' quality of life. Parks in Manila are located near main thoroughfares. Since people usually go to parks for recreation and relaxation it is important to have an idea of the concentration of criteria pollutants at these parks as they have great effects on people's health. Using portable sensors, one-hour average ambient concentration of O_3, NO_2, and $PM_{2.5}$ was measured once a month between 8 am to 11 am local time from May 2018 to April 2019 at three popular urban parks in Manila, namely, Arroceros Forest Park, Rizal or Luneta Park, and Manila Zoo. Mean concentrations of O_3 and NO_2 are highest in Luneta Park at 0.071 ppm and 0.032 ppm, respectively. In the case of $PM_{2.5}$, the mean concentrations at Luneta and Arroceros Park are the same at 0.070 mg/m^3. Manila Zoo had the lowest mean concentrations of the three criteria pollutants among the three urban parks. In terms of proximity to major thoroughfares, Luneta is closer and surrounded by four busy thoroughfares compared to Manila Zoo, and Arroceros Forest Park. It was also observed that measured concentrations of the three criteria pollutants were relatively higher during the cool dry months from October to February, with January 2019 recording the highest concentrations. Compared to a background site in Halang, Batangas, results of the measurement showed that the mean concentrations of O_3, NO_2, and $PM_{2.5}$ at the three urban parks are higher demonstrating the effect of proximity of the parks to vehicular traffic on the air quality inside the parks.

1. Introduction

According to the State of Global Air Report 2019, air pollution is the fifth leading risk factor for mortality worldwide with the Philippines ranking number 10 in terms of the highest mortality burden attributable to air pollution [1]. Exposure to short-term and long-term air pollution is associated with respiratory and cardiovascular diseases [2]. In highly urbanized places like the National Capital Region (NCR) in the Philippines, popularly known as Metro Manila, air pollution aside from urban heating is one of the region's major challenges [3]. In the latest National Emissions Inventory by source conducted in 2015, 88% of air pollution in Metro Manila comes from mobile sources, 10% from stationary sources, and a mere 2% from area sources. The increase in the number of vehicles is attributable to the increase in population to cater to the transport needs of the people [4]. In Metro Manila alone, there was an average increase of 7.77% in vehicular registration the year 2016-2018 (NCA, 2018). One of the 16 highly urbanized cities (HUCs) in Metro Manila is the City of Manila, the capital of the Philippines. It was the most densely populated among the HUCs with 71,263 persons per square kilometer in 2015 [5]. Criteria pollutants such as Nitrogen Dioxide

[*]Corresponding Author: Maria Cecilia D. Galvez, EARTH Lab, Physics Dept. De La Salle University, 2401 Taft Avenue, Malate, Manila, Philippines. Contact No.: +639163526517 Email: maria.cecilia.galvez@dlsu.edu.ph

Figure 1: The location of the sampling sites relative to each other. Yellow and white lines indicate roads for vehicular traffic

Table 1: Location, coordinates, and addresses of the sampling sites

Location	Coordinates	Address
Halang (Background site)	13°57'26.9"N 121°04'55.8"E	Halang, Lipa City, Batangas
Arroceros Forest Park	13°57'26.9"N 121°04'55.8"E	Antonio Villegas St, 659 A Ermita, Manila, 1000 Metro Manila
Japanese Garden (in Luneta Park)	14°35'3" N 120°58'44" E	Maria Orosa Street corner Padre Burgos Street, Manila, Metro Manila 1004, Philippines
Manila Zoological and Botanical Garden	14°33'50" N 120°59'18" E	Adriatico St, Malate, Manila, 1004 Metro Manila

Table 2: Summary of the sampling dates for each site

Site	May 2018	Jun 2018	July 2018	Aug 2018	Sept 2018	Oct 2018	Nov 2018	Dec 2018	Jan 2019	Feb 2019	Mar 2019	Apr 2019
Halang	13	17	15	19	23	28	30	16	27	25	17	14
Arroceros	17	27	19	24	27	22	21	14	31	27	20	29
Luneta	18	25	30	29	26	24	23	13	30	28	21	11
Mla Zoo	25	16	31	30	25	23	22	12	24	26	27	30

(NO_2), Ozone (O_3), and fine particulate matter ($PM_{2.5}$) mostly come from mobile sources [4].

The increasing population and vehicles make Manila an example of unsustainable urbanization [6]. Hence, the presence of urban parks and green spaces in the city can play a role in maintaining urban residents' quality of life. Urban green space, a category of land cover that includes public parks and other (public or private) vegetated areas in densely populated places plays a great role in improving not only human health and but also the air quality [7–11]. To date, there were no available data on the level of air pollution in urban parks in Manila. People come to public parks to exercise, sometimes have family picnics, and just for relaxation. Some urban poor and street-dwellers make parks as their temporary home. Measurement of air pollution level is important especially if the parks are located or surrounded by main thoroughfares where traffic congestions are always present. These can increase the potential of park users to be exposed to pollutants associated with vehicular emissions. Lam et al made a study in Hongkong urban parks, and they have shown that the air quality in urban parks is better compared to the roadside but not significantly different from that in the ambient conditions [12]. Three of the popular parks in Metro Manila are located within the City of Manila. These are Manila Zoo, Luneta or Rizal Park, and

Arroceros Forest Park. These three parks are located near the main thoroughfares, hence the proximity to vehicular traffic, which is the main source of air pollution in the city. There have been no air quality measurements in these parks as most static reference standard instruments are usually positioned near the roadside. The use of low-cost and portable air quality monitoring systems has become very popular as there is a need to provide a more accurate assessment of human air pollution exposure assessments [13]. In this study, the ambient concentrations of criteria pollutants such as nitrogen dioxide (NO_2), fine particulate ($PM_{2.5}$), and ozone (O_3) at these three public parks were measured once a month for one year using portable sensors. The measured ambient concentrations at the three parks are then compared to a background site in Halang, Batangas which is far from any vehicular traffic. The comparison between the urban parks and the background site will demonstrate the effect of vehicular traffic on the ambient concentrations of $PM_{2.5}$, NO_2, and O_3. Furthermore, this study will explore the usefulness of portable sensors to provide real-time determination and assessment of air quality in public areas so that people will know the quality of the air they are breathing and, once the air quality drops to unhealthy levels, people can be requested to vacate the area. This will also provide information to authorities if there is a need to place static reference standard air quality monitoring system in such public areas.

2. Materials and Methods

2.1. Sampling Sites

The study was conducted at the three urban parks located in the city of Manila. These are Arroceros Forest Park, Luneta or Rizal Park, and Manila Zoo. For comparison, a background sampling site was included, and this was located in Halang, Batangas which is about 70 km south of Manila. The location and addresses of the sampling sites are given in Table 1 and Figure1 shows the relative location of each sampling site. Yellow and white lines indicate roads for vehicular traffic. Permits were secured with the park administrators for us to conduct the measurements, which should be done only during office hours. Sampling Site 1 is the Arroceros Forest Park (AFP), also called as the "Last Lung of Manila". It is a 2.2-hectare (5.4-acre) manmade urban forest along Pasig River, at the foot of Quezon Bridge. The second sampling site is Rizal Park also known as Luneta Park or simply Luneta is a historical urban park in the Philippines and one of the major tourist attractions of Manila. It is approximately 16.24 hectares (40.01) acre and located along Roxas Boulevard, Manila, and also adjacent to Taft Avenue, Manila. The last urban park is the Manila Zoological and Botanical Garden or Manila Zoo. It is a 5.5-hectare (14-acre) zoo located in Malate, Manila, Philippines.

2.2. Measurement

The criteria pollutants measured in this study are $PM_{2.5}$, NO_2, and O_3. A DustTrak™DRX aerosol monitor (Model 8533, TSI incorporated) that used a light scattering technique to infer the mass concentration of particles was used for $PM_{2.5}$ measurement [14]. It was set on a 1-s time resolution at 3L/minutes. Before the start of each measurement, zero calibration was performed using a TSI 800663 zero filter. NO_2 and O_3 concentrations were measured by a real-time portable battery-operated gas sensor monitor AEROQUAL Series 500 with NO_2 and O_3 sensor heads. The sensor heads used gas-sensitive electrochemical (GSE) technology where it measures the concentration of a target gas by using oxidation or reduction reactions to generate a positive or negative current flow through an external circuit and the magnitude of this current is proportional to the gas concentration. The operating parameters of the sensor heads are as follows; O_3 (detection range 0–10 ppm; resolution and minimum detection limits 0.01 ppm; response time 60 secs; temperature 0 – 40°C; relative humidity 15 – 90%), NO_2 (detection range 0–1 ppm; resolution 0.001 ppm; minimum detection limits 0.005 ppm; response time 30 secs; temperature 0 – 40°C; relative humidity 15 – 90%). A 1-min time resolution was used during the measurement. Although the DustTrak and Aeroqual are not regulatory-grade monitors they are widely used in prior air quality research studies [13,15–24]. Measurements done for this paper were performed during the warranty period of the instruments. Being under warranty, instrument calibration and performance was within factory specifications. During the measurements, researchers observed the performance of the instruments, and temperature and humidity on the site were within their specified operating conditions. Simultaneous measurement on all sampling sites was not possible because only one set of instruments was available. Background measurement in Halang, Batangas was done on a Sunday, and the measurements at the three parks in Manila were conducted on the following weekdays if weather permits. The sampling dates are summarized in Table 2. To observe the same atmospheric and vehicular traffic conditions in the site, measurements were done on successive days in the morning between 8 am to 11 am local time for one hour, once a month for each sampling site for one year. Measurement was not conducted during bad weather conditions and sampling time was chosen to also avoid the effect of sunlight on O_3 concentration.

3. Results

One Way Analysis of Variance (ANOVA) was the statistical method used for identifying if there are any significant differences in the measured concentration of a particular criteria pollutant at the different sampling sites. If there's a significant difference, a Tukey honestly significant difference (HSD) test is then carried out to identify what sampling locations are significantly different, and then based on the mean concentration, the locations are ranked from highest to lowest concentration. RealStat (http://www.real-statistics.com/) Microsoft Excel Add-In was used to carry out the statistical calculation. An alpha value of 0.01 was used. To illustrate the variation of air pollution concentrations at each sampling site, Box and Whisker plots were used in terms of the lower quartile, upper quartile, median, mean, minimum, and maximum in each of the four study locations.

3.1. Ozone (O_3) Measurement

Shown in Figure 2 is the box plot of the O_3 concentrations from the four different locations from May 2018 to April 2019. Table 3a shows the result of the statistical analysis for O_3. In all the sampling months, there was a significant difference in O_3 concentration between each sampling site. The O_3 concentrations at the background site always have the lowest concentration on all sampling months. The DENR EMB National Ambient Air Quality Guideline Values (NAAQGV) for 1-Hour Averaging is 0.07 ppm [4], [25]. As shown in Figure 3, in some months, the measured O_3 concentration was 0 ppm at the background site

Table 3: One way Anova P-value between each sampling sites per month for (a) Ozone, (b) Nitrogen Dioxide, (c) $PM_{2.5}$.

(a) Ozone (in ppm)

Month	P value	Location	Mean
May 2018	7.52E-05	Arroceros	0.067
		Batangas	0.045
		Luneta	0.041
June 2018	7.35E-37	Arroceros	0.063
		Batangas	0.041
		Luneta	0.082
		Manila Zoo	0.056
July 2018	4.61E-31	Arroceros	0.061
		Batangas	0.040
		Luneta	0.049
		Manila Zoo	0.049
August 2018	5.06E-05	Arroceros	0.060
		Batangas	0.046
		Manila Zoo	0.054
September 2018	4.18E-33	Arroceros	0.078
		Batangas	0.047
		Luneta	0.064
		Manila Zoo	0.063
October 2018	2.45E-38	Arroceros	0.081
		Batangas	0.057
		Luneta	0.085
		Manila Zoo	0.069
November 2018	4.06E-79	Arroceros	0.060
		Batangas	0.052
		Luneta	0.089
		Manila Zoo	0.056
December 2018	4.34E-15	Arroceros	0.061
		Batangas	0.053
		Luneta	0.067
		Manila Zoo	0.054
January 2019	6.59E-37	Arroceros	0.078
		Batangas	0.062
		Luneta	0.082
		Manila Zoo	0.074
February 2019	1.86E-24	Arroceros	0.066
		Batangas	0.055
		Luneta	0.071
		Manila Zoo	0.067
March 2019	4.58E-25	Arroceros	0.064
		Batangas	0.048
		Luneta	0.058
		Manila Zoo	0.072
April 2019	7.47E-62	Arroceros	0.066
		Batangas	0.047
		Luneta	0.076
		Manila Zoo	0.072

(b) Nitrogen Dioxide (ppm)

Month	P value	Location	Mean
June 2018	1.51E-13	Arroceros	0.024
		Batangas	0.018
		Luneta	0.040
		Manila Zoo	0.012
July 2018	8.28E-05	Arroceros	0.011
		Batangas	0.007
		Luneta	0.012
		Manila Zoo	0.003
August 2018	0.053493	Arroceros	0.015
		Batangas	0.013
		Manila Zoo	0.020
September 2018	7.4E-11	Arroceros	0.021
		Batangas	0.008
		Luneta	0.021
		Manila Zoo	0.005
October 2018	6.7E-33	Arroceros	0.028
		Batangas	0.016
		Luneta	0.040
		Manila Zoo	0.013
November 2018	2.84E-51	Arroceros	0.022
		Batangas	0.008
		Luneta	0.047
		Manila Zoo	0.004
December 2018	4.24E-14	Arroceros	0.010
		Batangas	0.016
		Luneta	0.028
		Manila Zoo	0.034
January 2019	2.32E-29	Arroceros	0.039
		Batangas	0.026
		Luneta	0.061
		Manila Zoo	0.034
February 2019	5.42E-28	Arroceros	0.042
		Batangas	0.013
		Luneta	0.040
		Manila Zoo	0.042
March 2019	6.1E-13	Arroceros	0.035
		Batangas	0.016
		Luneta	0.031
		Manila Zoo	0.034
April 2019	2.52E-19	Arroceros	0.023
		Batangas	0.037
		Luneta	0.047
		Manila Zoo	0.029

(c) $PM_{2.5}$ (mg/m^3)

Month	P value	Location	Mean
May 2018	1.79E-79	Batangas	0.017
		Arroceros	0.074
		Luneta	0.033
		Manila Zoo	0.086
June 2018	2.09E-152	Batangas	0.023
		Arroceros	0.044
		Luneta	0.117
		Manila Zoo	0.055
July 2018	4.26E-95	Batangas	0.002
		Arroceros	0.104
		Luneta	0.028
		Manila Zoo	0.033
August 2018	4.08E-138	Batangas	0.039
		Arroceros	0.113
		Luneta	0.044
		Manila Zoo	0.054
September 2018	3.16E-214	Batangas	0.023
		Arroceros	0.088
		Luneta	0.036
		Manila Zoo	0.077
October 2018	1.62E-43	Batangas	0.043
		Arroceros	0.071
		Luneta	0.077
		Manila Zoo	0.048
November 2018	3.47E-112	Batangas	0.018
		Arroceros	0.040
		Luneta	0.077
		Manila Zoo	0.059
December 2018	2.61E-157	Batangas	0.014
		Arroceros	0.040
		Luneta	0.077
		Manila Zoo	0.024
January 2019	2.23E-135	Batangas	0.029
		Arroceros	0.112
		Luneta	0.107
		Manila Zoo	0.079
February 2019	5.06E-118	Batangas	0.045
		Arroceros	0.026
		Luneta	0.086
		Manila Zoo	0.071
March 2019	5.87E-67	Batangas	0.026
		Arroceros	0.046
		Luneta	0.032
		Manila Zoo	0.077
April 2019	7.47E-133	Batangas	0.027
		Arroceros	0.073
		Luneta	0.079
		Manila Zoo	0.047

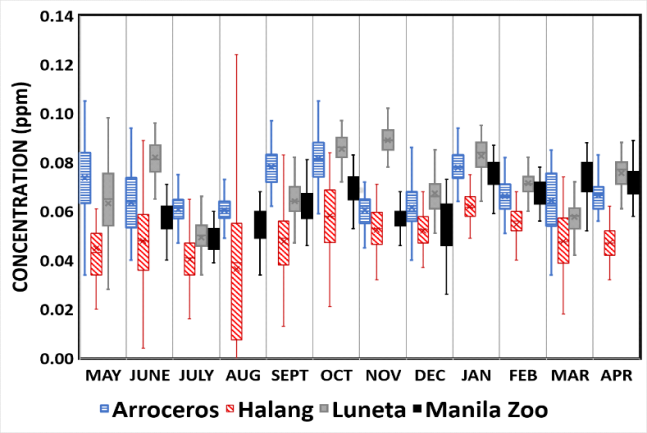


Figure 2: Box plot of monthly concentration of O_3 at the four sampling sites from May 2018-April 2019

For the whole year, as shown in Figure 3, Luneta had the highest concentration of 0.071 ppm and Arroceros was a close second at 0.067 ppm, followed by Manila Zoo at 0.063 ppm, and the background site at 0.050 ppm. A high concentration of O_3 exceeding the NAAQGV was observed for all sites except the background site in January 2019 which is one of the coldest months in the Philippines.

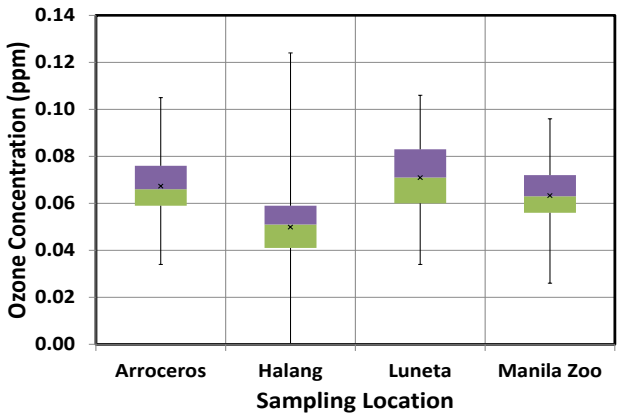


Figure 3: Box plot of the one year mean O3 concentration at the four sampling sites from May 2018-April 2019.

3.2. Nitrogen Dioxide (NO_2) Measurement

Shown in Figure 4 is the box plot of the NO_2 concentration from the four different locations from May 2018 to April 2019. Table 3b shows the result of the statistical analysis. In all the sampling months, there was a significant difference in NO_2 concentration between each sampling site. The NO_2 concentrations at the background site have the lowest concentration on all sampling months. On some months, the NO_2 concentration was 0 ppm indicating that the NO_2 concentration was below the detection limit of the NO_2 sensor. This was true most especially in the case of the background site where the lower quartile value was equal to the minimum value measured by the NO_2 sensor. For the whole year, as shown in Figure 5, there is a significant difference in all sampling locations. Luneta having the highest concentration of 0.032 ppm and Arroceros is a close second at 0.025 ppm, followed by Manila Zoo at 0.022 ppm and the background site at 0.017 ppm.. The National guideline value for NO_2 is 150 µg/Ncm (0.08 ppm) maximum for a 24-hour exposure [25]. As in the case of O_3, the coldest month of January and February 2019 registered higher NO_2 concentrations for all the urban parks except the background site.

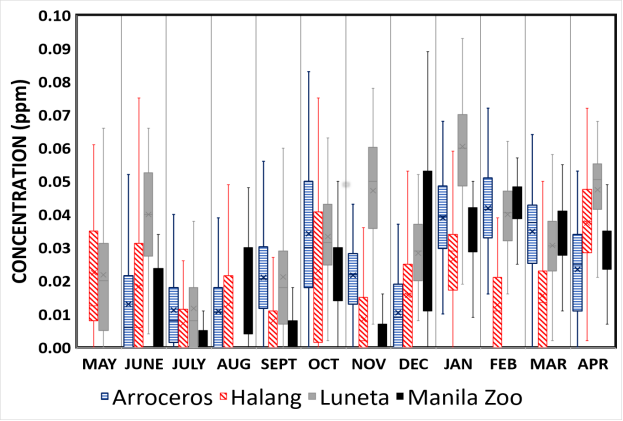


Figure 4: Box plot of monthly concentration of NO2 at the four sampling sites from May 2018-April 2019.

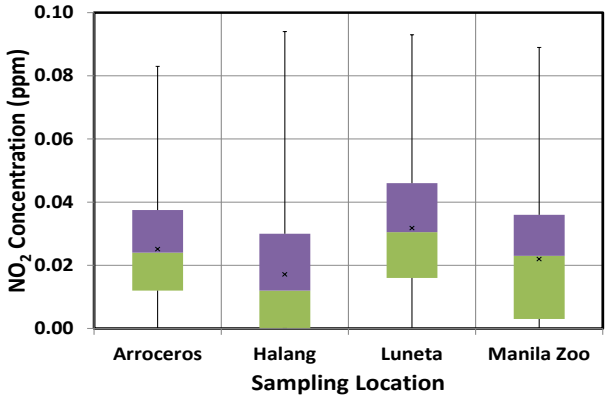


Figure 5: Box plot of the one year mean NO_2 concentration at the four sampling sites from May 2018-April 2019.

For Particulate Matter measurement, TSI DUSTTRAK DRX Aerosol Monitor 8533 was used and this can measure PM_{10}, $PM_{2.5}$, and $PM_{1.0}$ but only $PM_{2.5}$ will be presented. For statistical analysis, only $PM_{2.5}$ measurement will be used since all the other PM sizes follow the same trend. Furthermore, in terms of Air Quality Guidelines, $PM_{2.5}$ is the one that is reported. No air quality guidelines yet for $PM_{1.0}$. Shown in Figure 6 are the box plots of the $PM_{2.5}$ concentration measured by the DUSTTRAK. The lowest concentration of PM is in Halang, Batangas while Arroceros always gives the highest concentration. The result of One-way ANOVA indicates a significant difference in the PM concentration for all months between each sampling location. Arroceros always ranks the highest while Halang is the lowest. The WHO Air Quality Guideline values for $PM_{2.5}$ is 10 µg/m^3 annual mean and 0.025 mg/m^3 24-hour mean [4]. NAAQGV are 0.025 mg/m^3 annual mean and 0.050 mg/m^3 24-hour mean [4]. It can be seen from Figure 7 that the $PM_{2.5}$ concentration measured from TSI DUSTTRAK at the sampling locations in Manila exceeds the annual mean guideline values. However, please note that the instrument is not the reference instrument. Like O_3 and NO_2, it was also observed that in January 2019, high values of $PM_{2.5}$ concentration were measured at the urban parks exceeding the 24-hour mean NAAQGV.

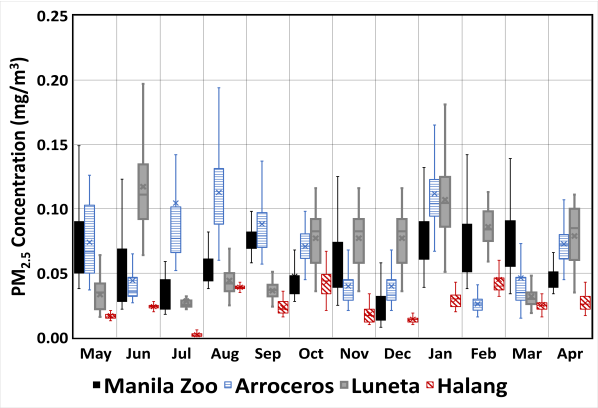


Figure 6: Box plot of monthly concentration of PM2.5 at the four sampling sites from May 2018-April 2019

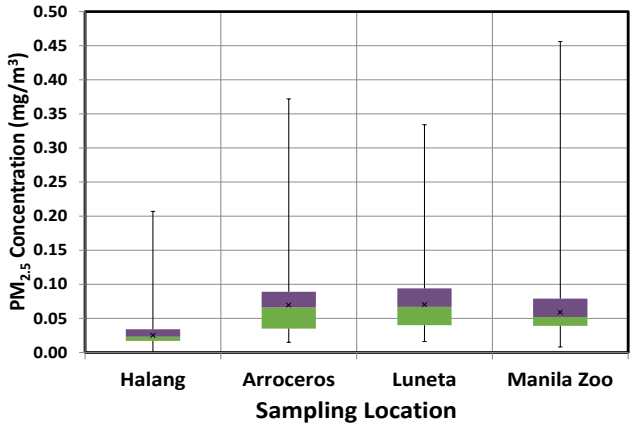


Figure 7: Box plot of the one year mean $PM_{2.5}$ concentration at the four the four sampling sites from May 2018-April 2019

With the unsustainable urbanization in Manila, the presence of urban parks can provide residence with quality of life. However, the location of these parks affects the air quality within the park. The measured values of O_3, NO_2, and $PM_{2.5}$ at the three well known urban parks in Manila are higher compared with a background site in Halang, Batangas. Using portable sensors, the

ambient concentration of O_3, NO_2, and $PM_{2.5}$ was measured once a month between 8 am to 11 am local time from May 2018 to April 2019 at three known urban parks in Manila, namely, Arroceros Forest Park, Rizal or Luneta Park, and Manila Zoo. Mean concentrations of O_3 and NO_2 are highest in Luneta Park at 0.071 ppm and 0.032 ppm, respectively. Among the three parks, Luneta is the one surrounded by major roads. For mean concentration of $PM_{2.5}$, Luneta, and Arroceros Park are the same at 0.070 mg/m^3. One of the main reasons is the proximity of the three parks to vehicular traffic. It was also observed that measured concentrations of the three criteria pollutants were relatively higher during the cool dry months from October to February. Although the instruments used were not the reference instruments, the result provides information on the necessity of monitoring the air quality within the parks. Many residents most especially people who cannot afford, and don't have time to go out of the city need these urban parks for relaxation and recreational activities.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to acknowledge the support from Commission on Higher Education (CHED) project entitled UDM CHED DARETO "Using wireless environmental monitoring sensors in assessing the impact of megacity environmental pollution and local climate on butterfly diversity in Manila, Philippines" and CHED –Grant-In-Aid Project "Evaluation of Human Skin Analogues" and we also acknowledge ARCHERS, CENSER through EARTH lab and De La Salle University Research Coordination Office (URCO) 47 F U 2TAY18-2TAY19 for partially supporting this project.

References

[1] M.E.I. Health Effects Institute. 2019. State of Global Air 2019. Special Report. Boston, "A SPECIAL REPORT ON GLOBAL EXPOSURE TO AIR POLLUTION AND ITS DISEASE BURDEN What is the State of Global Air? Who is it for? How can I explore the data?," Boston, MA:Health Effects Institute., 24, 2019.

[2] B. Chen, H. Kan, "Air pollution and population health: A global challenge," Environmental Health and Preventive Medicine, **13**(2), 94–101, 2008, doi:10.1007/s12199-007-0018-5.

[3] J. Oliveros, E. Vallar, M. Galvez, "Investigating the Effect of Urbanization on Weather Using the Weather Research and Forecasting (WRF) Model: A Case of Metro Manila, Philippines," Environments, **6**(2), 10, 2019, doi:10.3390/environments6020010.

[4] Department of Environment and Natural Resources Philippines - Environmental Management Bureau, "Air Quality in the Philippines (2008-2015)," 2016.

[5] Philippine Population Density (Based on the 2015 Census of Population) | Philippine Statistics Authority, Nov. 2019.

[6] S. Kecorius, L. Madueño, E. Vallar, H. Alas, G. Betito, W. Birmili, M.O. Cambaliza, G. Catipay, M. Gonzaga-Cayetano, M.C. Galvez, G. Lorenzo, T. Müller, J.B. Simpas, E.G. Tamayo, A. Wiedensohler, "Aerosol particle mixing state, refractory particle number size distributions and emission factors in a polluted urban environment: Case study of Metro Manila, Philippines," Atmospheric Environment, **170**, 169–183, 2017, doi:10.1016/j.atmosenv.2017.09.037.

[7] C.R. Chang, M.H. Li, "Effects of urban parks on the local urban thermal environment," Urban Forestry and Urban Greening, 2014, doi:10.1016/j.ufug.2014.08.001.

[8] A. Chiesura, "The role of urban parks for the sustainable city," Landscape and Urban Planning, 2004, doi:10.1016/j.landurbplan.2003.08.003.

[9] B. Givoni, "Impact of planted areas on urban environmental quality: A review," Atmospheric Environment. Part B, Urban Atmosphere, 1991, doi:10.1016/0957-1272(91)90001-U.

[10] L.R. Larson, V. Jennings, S.A. Cloutier, "Public parks and wellbeing in urban areas of the United States," PLoS ONE, 2016, doi:10.1371/journal.pone.0153211.

[11] J. Vieira, P. Matos, T. Mexia, P. Silva, N. Lopes, C. Freitas, O. Correia, M. Santos-Reis, C. Branquinho, P. Pinho, "Green spaces are not all the same for the provision of air purification and climate regulation services: The case of urban parks," Environmental Research, **160**(December 2016), 306–313, 2018, doi:10.1016/j.envres.2017.10.006.

[12] K.C. Lam, S.L. Ng, W.C. Hui, P.K. Chan, "Environmental quality of urban parks and open spaces in Hong Kong," Environmental Monitoring and Assessment, 2005, doi:10.1016/j.cccn.2005.02.014.

[13] G.R. McKercher, J.A. Salmond, J.K. Vanos, "Characteristics and applications of small, portable gaseous air pollution monitors," Environmental Pollution, **223**, 102–110, 2017, doi:10.1016/j.envpol.2016.12.045.

[14] TSI, "DustTrak DRX Aerosol Monitors Models 8533, 8533EP And 8534," 2014, 4, 2014.

[15] K.N. Jallad, C. Espada-Jallad, "Characterization of Road Traffic Emissions in a Densely Populated Residential Area of Kuwait," Environment and Natural Resources Research, **2**(2), 2–15, 2012, doi:10.5539/enrr.v2n2p2.

[16] I. Rivas, M. Mazaheri, M. Viana, T. Moreno, S. Clifford, C. He, O.F. Bischof, V. Martins, C. Reche, A. Alastuey, M. Alvarez-Pedrerol, J. Sunyer, L. Morawska, X. Querol, "Identification of technical problems affecting performance of DustTrak DRX aerosol monitors," Science of the Total Environment, **584**–**585**, 849–855, 2017, doi:10.1016/j.scitotenv.2017.01.129.

[17] M. Masiol, S. Squizzato, D. Chalupa, D.Q. Rich, P.K. Hopke, "Spatial-temporal variations of summertime ozone concentrations across a metropolitan area using a network of low-cost monitors to develop 24 hourly land-use regression models," Science of the Total Environment, **654**, 1167–1178, 2019, doi:10.1016/j.scitotenv.2018.11.111.

[18] E. Ezani, N. Masey, J. Gillespie, T.K. Beattie, Z.K. Shipton, I.J. Beverland, "Measurement of diesel combustion-related air pollution downwind of an experimental unconventional natural gas operations site," Atmospheric Environment, **189**(November 2017), 30–40, 2018, doi:10.1016/j.atmosenv.2018.06.032.

[19] C. Lin, J. Gillespie, M.D. Schuder, W. Duberstein, I.J. Beverland, M.R. Heal, "Evaluation and calibration of Aeroqual series 500 portable gas sensors for accurate measurement of ambient ozone and nitrogen dioxide," Atmospheric Environment, **100**(2), 111–116, 2015, doi:10.1016/j.atmosenv.2014.11.002.

[20] J.Y. Kim, S.R. Magari, R.F. Herrick, T.J. Smith, D.C. Christiani, "Comparison of fine particle measurements from a direct-reading instrument and a gravimetric sampling method," Journal of Occupational and Environmental Hygiene, **1**(11), 707–715, 2004, doi:10.1080/15459620490515833.

[21] D. Liu, Q. Zhang, J. Jiang, D.R. Chen, "Performance calibration of low-cost and portable particular matter (PM) sensors," Journal of Aerosol Science, 2017, doi:10.1016/j.jaerosci.2017.05.011.

[22] J.D. Yanosky, P.L. Williams, D.L. MacIntosh, "A comparison of two direct-reading aerosol monitors with the federal reference method for PM2.5 in indoor air," Atmospheric Environment, **36**(1), 107–113, 2002, doi:10.1016/S1352-2310(01)00422-8.

[23] S.E. Bartington, I. Bakolis, D. Devakumar, O.P. Kurmi, J. Gulliver, G. Chaube, D.S. Manandhar, N.M. Saville, A. Costello, D. Osrin, A.L. Hansell, J.G. Ayres, "Patterns of domestic exposure to carbon monoxide and particulate matter in households using biomass fuel in Janakpur, Nepal," Environmental Pollution, **220**, 38–45, 2017, doi:10.1016/j.envpol.2016.08.074.

[24] Z. Ali, K. Shahzadi, S. Sidra, Z. Zona, I. Zainab, K. Aziz, M. Ahmad, S.T. Raza, Z.A. Nasir, I. Colbeck, "Seasonal variation of particulate matter in the ambient conditions of Khanspur, Pakistan," Journal of Animal and Plant Sciences, **25**(3), 700–705, 2015.

[25] Department of Environment and Natural Resources Philippines - Environmental Management Bureau, "2002 National Air Quality Status Report," 2002.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1033-1041 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Evaluation of the Living Condition in Fishermen Settlement in the Coastal Area of Marisa City, Gorontalo Province

Irwan Wunarlan*,1,2, Sugiono Soetomo[1], Iwan Rudiarto[3]

[1]School of Architecture and Urban Science, Diponegoro University, Semarang, 50275, Indonesia

[2]Department of Architecture, Faculty of Engineering, State University of Gorontalo, Gorontalo, 96128, Indonesia

[3]Department of Urban and Regional Planning, Diponegoro University, Semarang, 50275, Indonesia

ARTICLE INFO

Article history:
Received: 25 August, 2020
Accepted: 03 November, 2020
Online: 14 December, 2020

Keywords:
Fishermen's Settlement
Village
Living Conditions
Slum
Coastal Area

ABSTRACT

The population living in the residential area in two villages, including Pohuwato village and Pohuwato Timur village, are mostly fishermen, who live under the poverty line, meaning that the people there have low income. The objective of the present study was to explore the living condition in terms of the physical, social, and economic aspects of the fishermen communities in Pohuwato Village and Pohuwato Timur Village, Marisa City, Gorontalo; it took into account the minimum service standard of a settlement area. The primary data was obtained by conducting field observations, interviews with fishermen, and Focus Group Discussion (FGD) forums; secondary data was obtained from the related literature, including books, research results and documentation as well as various articles. The quantitative analysis was implemented by scoring, in which the indicators and variables of slum will be assessed with the standards of the Ministry of Public of Works and Housing, Directorate General of Human Settlement. Living conditions in the slums spread in the studied villages were in the moderate category. Findings showed that, issues, such as the provision of lavatory, drainage system for wastewater, clean water supply, waste management, and fire protection system, were most visible. Developing a disaster management system for the settlement area is also crucial. Therefore, the fulfillment of basic needs was deemed essential as the solution to the sustainability of the fishermen community.

1. Introduction

The term "urbanization" is defined as the population shift from rural areas to urban areas, aiming at creating a better quality of life. This phenomenon can be found in many parts of the world, both in metropolitan cities and small cities, with a feasible livelihood option that people can find compared to the one in their previous settlement. In [1], the authors argue that urbanization has encouraged and attracted rural communities to move to urban areas. Marisa city, one of the small cities, has experienced a rise in its population numbers due to urbanization. The rapid advancement of industrial cities, in addition to other contributing factors, e.g., ease of access to infrastructures, education, health, and job opportunity, is cited as the cause of urbanization. However, the population growth is not accompanied by the initiative of the city government to provide basic infrastructure, include adequate or appropriate settlements, for the people; thus leading to socio-economic inequality. In [2], the authors claim that the inability to provide the infrastructures contributes to the rise in the number of slums with insufficient necessary infrastructures, and this problem is found in the fishermen's settlement in Marisa city.

The population living in the residential area in two villages, including, Pohuwato Village and Pohuwato Timur Village, are mostly fishermen. In those villages, the houses are adjusted to fit the needs of the fishermen so that the pattern of the settlement results in a linear shape, with wooden moorings and wild plants along the beach. Many people there live under the poverty line, meaning that they have low income. In [3], the authors propose that the settlement in those villages is mostly inappropriate houses with poor environmental conditions due to dense populations. In this research, the term "slum" refers to unsuitable building and/or high-density housing, which is inappropriately arranged, with

*Corresponding Author: Irwan Wunarlan, irwan.wunarlan001@gmail.com

below standard facilities and infrastructures. Another defining characteristic of a slum is that this area does not meet the spatial planning or housing standard. A slum is notoriously known for its settlement that has poor services; it is an overcrowded and unhealthy area with high criminality rates. In general, settlements located in a slum represents an area with poor quality in terms of the environmental aspect, overpopulated, and high-density housing, with below-standard public infrastructures.

Conditions in an abandoned small city could trigger some problems, such as the rise in the number of slums, poor development of the area, visually undesirable city, and a decline in the health of the population due to the absence of the health standard, as well as socio-economic problems. The present work discusses the life of the fishermen community residing in areas nearby a small city of Marisa, including Pohuwato Village and Pohuwato Timur Village; these areas are notable for their urban-rural characteristics and far from the activity center.

Based on the research background, the main question of the study is as follows: "how is the living condition of fishermen's settlement in the coastal area of Marisa City, Gorontalo province according to the minimum standard of services? Therefore, the objective of the present study was to explore the living condition of the fishermen's settlement. It focused on examining the physical, social, and economic aspects of the fishermen communities in Pohuwato Village and Pohuwato Timur Village, Marisa City, Gorontalo; it took into account the minimum service standard of a settlement area.

It is further expected that the study will contribute to: (1) broadening the insight regarding the living condition, facilities, and infrastructures of the fishermen's settlement in the coastal area of Marisa City, Gorontalo; (2) providing references and inputs for the government of Pohuwato regency in terms of the housing and settlement in the area, by which it helps their development program and the management of the coastal area in Marisa City, including the provision of housing facilities and infrastructures that fit the physical, social, and economic condition of the people.

2. Literature Review

2.1. Settlement

A settlement refers to a process of functional establishment underpinned by the pattern of human activities and the influence of the aspect of the physical and non-physical setting (which affects the activities and establishment process). In [4], the author infers that the setting is impactful on the physical environment that is shaped by the location and the community with its socio-cultural aspects. A settlement can be established through particular processes, with or without a plan. The processes are long-term and continuous. On the other hand, in [5], the author opines that the formation of a settlement pattern is the outcome of the distribution of a settlement area based on a particular geographical factor, e.g., a settlement in a coastal area or nearby a river and linear settlement.

According to the Law of the Republic of Indonesia, Number 1 of 2011 Concerning Housing and Areas for Settlement, the term settlement refers to a part of the environment, outside the protecte d areas, including both urban and rural areas serving as a residential area and a place that supports livelihood. The term settlement area, in this context, is defined as the residence of a person or a family, consisting of a house and yards. In [6], the author states that this notion later confirms that housing is among the components of a settlement area. Furthermore, In [7], the author mentions five major elements in a settlement area, namely (1) residential area, (2) workplace, (3) movement area, (4) recreational area, and (5) facilities and infrastructures.

In [9], (as cited in R.W. Diah [8] I.N Nasution, A. Syahreza), the author argues that issues revolving around settlement areas are notoriously complicated due to many underlapping factors that contribute to the concerns. As a living place, the settlement areas take into account several aspects other than physical and technical; those aspects include the socioeconomy and culture of t he residents. The problem concerning the provision of housing mostly occurs in urban areas. It blames the population migration from the rural areas to cities due to cities' appealing aspects and internal urges from the villages. Such a condition results in continuous increases in the urban population by which it is considerably impactful on the provision of land for housing. In [10], the authors point out that the growth rate in cities in many developing countries is accompanied by the growth in the number of population living in a slum. This is signified by the condition of people living with undesirable standards.

In [11], the author defines the condition of a building is seen from its age, density, and arrangement. Building's age is one of the factors causing damage and fires. In [12], the author describes an area becomes a slum due to the density of the building and the poor standard of the area of the construction. On top of that, the types of construction can determine the level of the building's durability. As mentioned in [13], the author states that most constructions in a slum are made of combustible materials, e.g., wood, zinc, fabric, iron, and bamboo.

2.2. Urban Settlement

A settlement develops along with the rise in the population number of an area. The constant condition of the land and the increase in the population number is the leading cause of the increase in the needs for settlement areas. For this reason, finding another land to create more suitable residential is a necessity.

Settlements in urban areas are generally categorized into three groups, namely: (i) well-arranged settlement with quality facilities, infrastructures, and utilities; (ii) unplanned settlement, i.e., areas where the housing is not in compliance with the regulation, and it has no facilities, infrastructures, and utilities that meet the standard, and; (iii) informal settlement, where the construction of the houses are not well-planned. In [14], the author explains that in the last group of the settlement, poorly arranged houses can be found amid well-constructed houses. The inability of spatial planning in urban areas to make ends meet results in improper management of the settlement areas; which also causes rapid growth of slums. In [15], the author states that these situations will, in turn, lead to a distortion in the environment in terms of the physical and social aspects.

A metropolitan city has people with different elements; this condition is the outcome of the high attractiveness value of the city. Some of the people have inhabited the city for a long time, and the others are newcomers who expect to earn a better living. One

inexorable trend of a metropolitan city near a coastal area or ship port is the presence of the fishermen community in the structure of the urban population. According to [16], the author reckons that fishers are among the components of society with the lowest financial well-being compared to other social groups. In the same tune, in [17], the authors state that the consequence of low-quality human resources in the fisheries sector is a declined productivity level, leading to the low earning and impoverishmen t among the fishermen.

2.3. Slum

The term slum refers to a settlement area with undesirable environment qualities caused by the gap between the rapid population growth and slow economic growth. A guideline published by the Directorate General of Human Settlements has established the criteria to identify a slum area. Indicators or criteria of a slum area based on the facilities and infrastructures include the condition of the road drainage system, clean water, wastewater, and waste management.

The slum area is an unplanned settlement area with inadequat e basic services, such as clean water and sanitation, and poor ownership security. Further, this area is infamous for its inapprop riate and dense housing, and unsafe construction of the houses. In [18], it considers that there are many terms associated with the slum area at the national and regional levels, such as informal settlement and squatter. In other words, the slum area represents a squalid and overcrowded area in many cities. The slum area is represented by many defining characteristics, such as inhabited by unemployment, high crime rate, and demoralization rate, inha bited by unemotionally stable residences with low purchasing power and income level, and squalid and unhealthy environment. Most of the population in this area are migrants from villages, and they are prone to serious health issues, such as malnutrition [19], since authors have limited access to the public facilities. In [20]-[22], the authors express the majority of the people here work in informal sectors as labor, and they live in informal houses.

According to [23], the author states that common problems revolving around a slum area include (1) overcrowded housing with poor health and social standard, (2) dense building, which is prone to fires, (3) lack of clean water, (4) insufficient electrical energy supplies, (5) poor drainage system, (6) bad roads, and (7) lack of lavatories. Preventions of slum conditions are possible if the development of the settlement area adheres to the regulations; one of the practices is by establishing a healthy housing standard. A slum area represents a residential area with low quality of housing and poor infrastructures with a constant decline in the environmental aspect. Despite these conditions, the residences of the area have guaranteed land ownership; this regulation encompasses legal protection for the owner of the land, a residence, or the landlord. In [24], the author states that slum refers to an area located in urban areas, in habited by marginalized communities that experience a drop in the quality of life.

2.4. Typology of Fishermen's Settlement Areas

The fishermen's settlement area in Pohuwato and Pohuwato Timur villages is positioned to adjust with the area, facing the coastline, i.e., a linear settlement. This adjustment is based on the fishermen's activities. In general, this typical settlement is categorized into two: (1) settlement area with houses facing towards the beach, and (2) settlement area with houses facing towards the street with access to the beach.

The settlement is regarded as a high-density area, thus leaving the impression that the settlement resembles a slum area with bad sanitation. Most of the houses are made of wood and bamboo, although there are some concrete houses with gable roofs. Furthermore, the materials of the roof of the houses in the area are zinc and thatch. Each house has a mooring. Currently, the two villages have no fishing ports that function as a port or harbor for landing and distributing fish.

Each fisher who owns a boat is accompanied by 25-30 subordinates who are helpful during fishing. They can catch 1.5 to 3 tons of fish per day.

As many as 1.5 to 2 tons of the catch are distributed to the tra ditional market in Marisa and Paguat for fulfilling the food supply of the population in the cities and nearby areas, while 3 tons of the catch is sold to the markets in the city center in Bitung, Gorontalo, Palu, and even to other islands, such as Kalimantan to earn more income for the community in Marisa city. Should the fishermen want to land their catches from their boat to the coastal area, they have no choice but to rent a smaller boat because the port in the area does not meet the standard.

Figure 1: Fish Drying (Source: Authors, 2018)

Figure 2: Fishermen's Activities in Pohuwato Village and Pohuwato Timur Village (Source: Authors, 2018)

Figure 3: Fishermen's Activities in Drying the Fish (Source: Authors, 2018)

Figure 4: Fishermen's Activities in Scaling the Fish (Source: Authors, 2018)

Figure 5: Permanent house of a Fisherman in Pohuwato and Pohuwato Timur (Source: Authors, 2018)

3. Research Methodology

3.1. Research Sites

This study was conducted in the fishermen's settlement in Po huwato and Pohuwato Timur village, Marisa city, Pohuwato regency; the area of the settlement is 392.91 Hectares. (see Figure 6). In these densely packed villages, the majority of the population works as fishermen living in slums. Such a situation leaves the impression of an inelegant city since the fishermen settlement is located at the center of Marisa City.

3.2. Data Collection

The data involved primary and secondary data. The primary data functioned to gather information regarding the overview of the condition in the settlement area of the fishermen community in Pohuwato and Pohuwato Timur village, and information regarding the preferences and perceptions of related stakeholders in terms of the improvement of the quality of these two villages. Methods to obtain the primary data involved field observation, interviewing 148 fishermen (or 25% of the total population selected purposively), and Focus Group Discussion (FGD). Each closed-ended question (with responses using a Likert scale system) is intended to retrieve information regarding the condition of the fishermen settlement. The secondary data were collected from reviewing the literature, books, research report, documents, and articles that correlated with the present study.

3.3. Data Analysis

The data were analyzed quantitatively using a scoring system, where the indicator and the variables of slum area were examined based on the standard of Ministry of Public of Works and Housin g, Directorate General of Human Settlement; this was aimed at determining the living conditions in the fishermen's settlements in the studied villages.

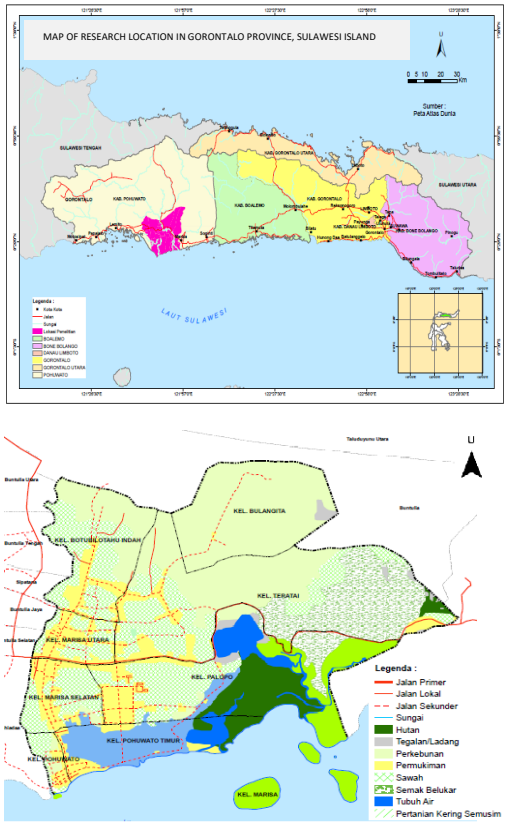


Figure 6: Research Site Map

4. Results and Discussion

4.1. Evaluating the living conditions in the fishermen's settlements

The population living in the residential area in two villages, including, Pohuwato village and Pohuwato Timur village, are mostly fishermen. The total population is 4333. The two villages share land borders with several areas. Palopo and Marisa Selatan villages are located in the north border of Pohuwato and Pohuwato Timur. On the east border and the southern border, the two villages share land borders with Maleo village, Paguat district, and Tomini bay, respectively. Bulili village, Duhiadaa district, is located on the east border of Pohuwato and Pohuwato Timur.

Table 1 shows the living conditions of the settlement area of the fishermen community in Pohuwato and Pohuwato Timur village.

Table 1: Living conditions in fishermen's settlements in Pohuwato and Pohuwato Timur village units

Number	Aspect	Conditions of slums in Pohuwato Village and Pohuwato Timur Village
1.	**Information of Population**	
	Total population	4333
	Number of the head of the household	867
2.	**Conditions of house**	
	Total houses	864 units
	Number of unorganized houses	529 units

	Number of houses that do not meet the standard	335 units
3.	**Environmental Condition**	
	Road width > 1.5 m.	6634 meter.
	Road width < 1.5 m.	1957 meter.
	Width of bad roads	1796 meter.
	Width of the road without canal	5519 meter.
4.	**Condition of Drainage System**	
	Width of the drainage canal	5034 meter
	Width of the bad drainage canal	3045 meter
5.	**Clean Water Services**	
	Number of households who has no access to clean water	651
	Number of households who has no access to clean water	651
6.	**Condition of Waste Water Management**	
	Number of houses without access to public lavatories	304
	Number of houses without standard lavatories	475
7.	**Condition of Waste Management**	Management of domestic waste in Pohuwato Village and Pohuwato Timur Village are yet categorized good since the officers do not take the waste to the garbage dump nearby.
8.	**Condition of Fire Protection System**	The villages have no fire protection systems; some locations are hard-to-reach areas, thus hindering the fire trucks from reaching the areas (the road is < 1.5m).

Source: Survey reports, 2018

According to the findings, the general information depicting the condition of the building in Pohuwato and Pohuwato Timur village based on eight main indicators of a slum area are as follows:

1. As many as 61% of the total houses in the two villages are categorized as informal housing, and the majority of them are not facing directly to the main road with a minimum width of 1.5m.
2. The percentage of houses that do not meet the standard is 38.64%. All of those buildings need to be repaired.
3. The width of some of the roads is less than 1.5m.
4. Other roads (83.19%) with a width of more than 1.5m have no drainage system.
5. As many as 33.89% of the drainage system is in bad condition, hampering the function of the system to flow the water.
6. Although the access to clean water for the community in the research sites has been provided by Indonesian Regional Water Utility Company (hereinafter referred to as PDAM) of Marisa city, 75.34% of the population in the two villages still have a problem regarding the clean water supply. As a result, they rely on public wells.
7. Regarding the wastewater management, the majority of the people in Pohuwato and Pohuwato Timur village have standard private lavatories, while others still use the public lavatories or defecate in the seawater.
8. The disaster management system in Pohuwato and Pohuwato Timur village is yet available. On top of that, some areas are hard to reach by a fire truck as the roads are too narrow (< 1.5m). The two villages also do not have sufficient disaster management systems (such as mangrove forest as shoreline protection) to keep the settlement area from abrasion and rob flood.

Figure 7: Simple fishermens' settlement in Pohuwato Village and Pohuwato Timur Village (Source: Authors, 2018)

Figure 8: Houses with mooring (Source: Authors, 2018)

Figure 9: Condition of roads (Source: Authors, 2018)

Figure 10: Daily activities of the fishermen (Source: Authors, 2018)

Two scoring methods were employed to determine the living conditions in the fishermen's settlement in the studied villages. Table 2 demonstrates the criteria and indicators of the slum area, along with its scoring, prepared by the Directorate General of Human Settlements

Table 2: Criteria score of living conditions in fishermen's settlements in Pohuwato and Pohuwato Timur village units

No.	Aspect	Criterion	Indicator and Parameter	Score
1.	Condition of building	Building irregularity	Building located in an irregular location	3
		Building density	High-density building that does not meet the standard	3
		The discrepancy with building	Building in a location that is not in compliance	5

		technical requirements	with the technical requirements	
2.	Condition of roads	Road coverage	Areas not covered by the roads	1
		Quality of the surface of the roads	Areas with poor-quality roads	5
3.	Condition of clean water supply	Availability of the access to clean water supply	Total population with no access to clean water	1
		Unfulfilled demand for clean water supply	Total population, whose demand for clean water supply, is not met	3
4.	Condition of the drainage system	Inability to manage water flow	Areas with >30cm of puddle, in > 2 hours, and > 2 times a year	3
		Unavailability of the drainage system	Areas with no drainage system	3
		A local drainage system that is not connected to the urban drainage system	A drainage system that is not connected with other systems	5
		Poorly-managed drainage system	Dirty and unpleasant smell	5
		Quality of the construction of the drainage system	Areas with a poorly-constructed drainage system	5
5.	Condition of the wastewater management system	The wastewater management system is not in compliance with the technical standard	Areas with a wastewater management system that is not in compliance with the technical standard	3
		Infrastructures in the wastewater management system are not in compliance with the technical standard	Areas with poor infrastructures in its wastewater management system	3
6.	Condition of waste management	Infrastructures in the waste management systems are not in compliance with the technical standard	Areas with poor infrastructures in its waste management system	5
		The waste management system that does not meet the technical standard	Areas with poor waste management systems	5
		Poorly-managed infrastructures	Areas with poorly-managed infrastructures	5
7.	Condition of the fire protection system	The availability of fire protection system	Areas with no fire protection system	5
		Unavailability of the fire protection system	Areas with no fire protection system infrastructures	5
			Slum criteria score	73
	Category	Moderate		

Source: Analysis Results, 2018

Description: Score 1 = 25% – 50%, Score 3 = 51% – 75%, Score 5 = 76 %– 100%, Low slum level = 19 – 44, Moderate slum level = 45 – 75, Severe slum level = 76 – 95.

The condition of an area with a disproportional number betwe en the total population, the availability of facilities, and infrastructures would evince the quality of a slum area. As a result, the provision of the facilities and infrastructures is considered as something that is likely over compelled.

Facilities and infrastructures found at the fishermen's settlement are discussed below.

- Lavatories

The needs of lavatories are something unavoidable for the fishermen community in Pohuwato and Pohuwato Timur village. According to the data, 31 units of the house (3.57%) are regarded unlivable, although those houses have lavatories. This condition urges a need for a public lavatory for the fishermen community, considering that 96.43% of the houses in the area have no private lavatories. Consequently, the majority of the people have no choice but to defecate in seawater or a river. Table 3 provides information regarding the distribution of public toilets in Pohuwato and Pohuwato Timur village.

Table 3: The Number of Public Lavatories in Marisa City, 2019

No.	Village	Total	Percentage
1	Pohuwato	21	67,74
2	Pohuwato Timur	10	32,26
Sum Total		31	100

Source: Village Profile, 2020

Table 3 shows that Pohuwato village has more public rest rooms totaling 21 units or 67.74% compared to Pohuwato Timur Village. Pohuwato Timur Village has the least lavatories (10 units, 32.36%) for its people. There are 31 public lavatories spread in the residential area and the settlement of the fishermen community at Marisa City. However, people's awareness to manage the cleanliness of the public facility is low. To worsen, the public lavatories in the two villages are far from satisfying the minimum standard. Such a condition demands the involvement of the people to start taking care of the cleanliness of this facility.

- Roads

Roads are crucial for the fishermen's settlement at Marisa city since these infrastructures connect the activities between the city and the village. The roads in the research site are mostly collector roads, residential streets, and pathways. In this context, the collector road and the residential street refers to a road that connects a group of houses with others, while the pathway connects one house with another.

Table 4: Roads in the fishermen's settlement, Marisa City, 2019

No.	Road classification	Height (km)	Width (m)	Existing condition	SPM (Minimum Service Standard)	Description
1	Collector	22.65	10	0.00522	0.0006 m/individual	Meet criteria
2	Residential street	9.56	8	1.59703	40-60/Ha	Does not meet criteria
3	Pathway	7.1	0.8-2	1.18608	50-110/Ha	Does not meet criteria

Source: Field Survey, 2020

Table 4 reveals that only the collector road that satisfies the minimum public service standard compared to other road types in Marisa city. The other two roads, however, urge the local government to improve, such as expanding new road networks and enhancing the quality of the roads to fulfill the standard.

- Wastewater system

According to [25], the author states that an urban drainage system is designed as flood control in a settlement. The drainage system in the residential area and fishermen's settlement in Marisa city is adjusted according to the model of the road. However, some of the current drainage canals cannot be used properly. The drainage systems that are properly functioned serve to manage the rainwater and domestic wastewater thus flow the water into a river and sea. Problems in some canals blame the rubbish heap covering the drainage surface. The rubbish clogs the rainwater, thus resulting in floods. For this reason, the maintenance of the drainage system is crucial in preventing the rise in the water level surface and flood during the rainy season. The types of drainage systems in the residential area and fishermen's settlement in Marisa city consist of arterial, primary, and secondary drainage.

Table 5: The drainage system in Marisa City, 2019

Number	Classification of Drainage System	Height (m)	Width (m)	Description
1	Primary	2108	1	Good
2	Secondary	514	0.9	Poor
3	Tertiary	326	0.7	Poor
Sum Total				

Source: Survey Reports, 2020

According to the interview with the people living in the residential area and the fishermen's settlement area in Marisa city, flood never hit their areas since the drainage canals can collect the rainwater, thus preventing puddles for two hours. On that ground, it could be argued that the drainage system in the residential area and the fishermen's settlement within Marisa city has met the minimum standard. Expansion of canals in the areas without the drainage systems allows the rainwater to flow into the point of final disposal.

- Clean water resources

In general, PDAM has provided affordable access to clean water for people in the residential area and fishermen's community in Marisa city. The following table provides brief information regarding the use of clean water of the people in the research sites.

Table 6: Clean water resources in Marisa City, 2019

No.	Village	Water source			Percentage		
		Dug well	Drilled well	Indonesian Regional Water Utility Company (PDAM)	Dug well	Drilled well	Indonesian Regional Water Utility Company (PDAM)
1	Pohuwato	15	0	0	4,76	0,00	0,00
2	Pohuwato Timur	300	0	10	95,24	0,00	100,00
Sum Total		315	0	10		100,00	0,00

Source: Village Profile, 2020

According to the results of Table 6, the clean water supply of the fishermen community in Marisa City includes 315 dug wells. PDAM contributes to ten units of water resource facilities. Pohuwato Timur Village has the most access to clean water resources (dug wells), totaling to 300 units or 95.24%, while Pohuwato village has limited resources with only 15 units or 4.76%. Resources provided by PDAM in Pohuwato Timur Village comprise ten units (100%). Both Pohuwato Village and Pohuwator Timur village, however, have no injection wells due to funding problems and the fact that the water from such wells contains salt, and thus it is undrinkable. Clean water resources are one of the primary needs for the fishermen community residing in the residential area and settlement area found within Pohuwato Village and Pohuwato Timur Village. From the above discussion, it can be said that 37.84% of the urban population has access to clean water, while the remaining 62.16% is the opposite.

Thereby, the local government is urged to provide facilities of clean water resources for 1000 houses in the fishermen's settlement; it is aimed at ensuring a healthy life of the fishermen community.

- Waste management

The residential area and fishermen's community in Marisa city has been perplexed by the issue of waste. Such a condition blames the lack of waste management and temporary waste dump sites (hereby referred to as TPS) in all residential areas and fishermen's settlement area in the city. As a result, the majority of the people have no choice but to litter and to burn their waste.

A survey conducted in the residential areas and fishermen's settlement in Marisa city reveals that some areas, such as temporary drainage systems, have been turned into a garbage disposal, thus resulting in the spread of diseases.

Provision of facilities, such as TPS, in a specific area that serves 100 units of the house per radius is, thereby, deemed essential. According to Table 7, the waste management system in the research sites is below the minimum standard, urging the stakeholders to design waste management planning.

Table 7: SPM of waste management

No.	Representing Parameter	Existing condition	SPM (Minimum Service Standard)	Description
1	TPS	0	Per 50-100 m	Below standard
2	Public waste container	0	Per 150m	Below standard
3	Domestic trash bin	26%	60-100%	Below standard
4	Frequency of waste collection	2 times	Every day	Below standard
5	Population served	20%	80%	Below standard

Source: Survey Reports, 2019

- Electricity

Electricity is the source of energy and lighting, which is crucial for all human activities in every sector, e.g., industry, commercial, and household. It also functions as the resource of energy for

electric supplies and industrial machines. There are no significant issues regarding the electrical distribution system in the residential area and fishermen's settlement in Marisa city since the system has been distributed within these areas. The only notable concern for the State Electricity Company (PLN) is the management of cable management for better power transmission to all houses.

- Fire Protection System

There were fires in the residential area and fishermen's settlement in Marisa city in 2008, 2016, and 2017, damaging the houses and public facilities, e.g., schools and regent's official residence, within the city. However, the issues had been addressed properly due to the preparedness of fire brigades in Marisa city, with four units of fire trucks.

The capability of the fire trucks to reach the area correlates with the road and the availability of water supply in every house within the residential areas and settlements in Marisa city. However, there are still some issues regarding the availability of hydrant and road conditions (the trucks are unable to reach the site due to narrow roads, such as in Pohuwato and Pohuwato Timur village). Another issue worth-mentioned is the density of building in the areas, in which it could hinder the officers from extinguishing the fire.

5. Conclusion

It can be concluded that the living conditions in the studied settlements are in the moderate category. Several issues, such as the provision of lavatory, drainage system for wastewater, clean water supply, waste management, and fire protection system, are most visible. Formulating coping strategies by the local government is necessary to address the problem in the areas. According to the living conditions in the studied slums, the fulfilment of basic needs is deemed essential for the sustainability of the fishermen community. This approach thus addresses the scarcity of lavatories and clean water supply, limited drainage system for wastewater, and poor management of waste. Developing a disaster management system for the settlement area is also crucial.

Oftentimes people residing in Pohuwato Village and Pohuwato Timur Village tend to build their houses in a vacant land without any legal permission. The houses can be found around the riverbanks, beaches, or nearby the workplace of the people and in the city center. The needs for housing and settlement areas for the fishermen in Marisa City are considered as their primary needs. Meanwhile, the provision of facilities and infrastructure of a settlement area is supposed to fulfill the spatial planning criteria, involving education facilities, health facilities, praying facilities, public services (office) and open spaces, roads, clean water system, drainage system, waste management (such as garbage dump), electrical energy, telephone network, and fire protection systems. For this reason, the government of Marisa City must strictly monitor the issuing of a building permit, since carrying out building work without such a permit is considered illegal and violate the regulation of urban planning.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] B. Marx, T. Stoker, T. Suri, "The economics of slums in the developing world" Journal of Economic Perspectives, **27**(4), 187–210, 2013, https://dspace.mit.edu/openaccessdisseminate/1721.1/88128/J.S. https://doi.org/ 10.1257/jep.27.4.187

[2] M. Ishtiyag, S. Kumar, "Typology of informal settlements and distribution of slums in the NCT, Delhi" Journal of Contemporary India Studies: Space and Society, Hiroshima University, **1**, 37-46, 2011, http://doi.org/10.15027/31427

[3] B. Ferguson, N. Jesus, "A financial framework for reducing slums: lessons from experience in Latin America," Environment and Urbanization, **15**(2), 201-216, 2003, https://doi.org/10.1177/095624780301500210

[4] A. Rapoport, House Form and Culture, the University of Michigan, Prentice-Hall, 1969.

[5] J.T. Jayadinata, Tata Guna Tanah dalam Perencanaan Pedesaan, Perkotaan dan Wilayah [Land Use in Rural, Urban and Regional Planning], ITB, 1999.

[6] Sumartono, B. Deliyantono, "Pengembangan Kawasan Permukiman dan Keberlanjutan Kota [The Development of Settlement Area and Urban Sustainability]," in National Seminar FMIPA on Universitas Terbuka, 2018.

[7] H.S. Yunus, Struktur Tata Ruang [Spatial Structure], Pustaka Pelajar, 2008.

[8] R.W. Diah, "Penataan permukiman di kawasan segiempat Tunjungan Kota Surabaya [Settlement Arrangement in Tunjungan square area, Surabaya City]," Jurnal Tata Kota Dan Daerah, **2**(2), 2010, https://doi.org/10.26905/mintakat.v18i2.1692

[9] I.N. Nasution, A. Syahreza, "Riverside Sustainability-Overcome Problems and Reliable Approaches" in MATEC Web of Conferences 138, 08001, 2017, https://doi.org/10.1051/matecconf/201713808001

[10] F.J. Chowdhury, A.T.M. Nurul Amin, "Environmental assessment in slum improvement programs: some evidence from a study on infrastructure projects in two Dhaka slums," Environmental Impact Assessment Review, **26**(6), 530-552, 2006, https://doi.org/10.1016/j.eiar.2005.11.004

[11] I. Akmal, Membetulkan Dan Merawat Rumah [Home Repair and Maintenance], Gramedia Pustaka Utama, 2012.

[12] Komarudin, Menelusuri Pembangunan Perumahan Dan Permukiman [Tracing the Development of Housing and Settlements], Yayasan Realestatet Indonesia, 1997.

[13] E.E. Sutriani, Faktor-Faktor yang Mempengaruhi Terciptanya Kawasan Permukiman Kumuh di Kawasan Pusat Kota [Factors Affecting the Rise in the Number of Slums at a City Center], Universitas Diponegoro, 2006.

[14] S. Yudohusodo, Rumah Untuk Seluruh Rakyat [Houses for All People], Direktorat Jendral Cipta Karya, 1991.

[15] E. Suharini, "Menemukenali agihan permukiman kumuh di perkotaan melalui interpretasi citra penginderaan jauh [Identifying the characteristics of slum areas through the interpretation of remote sensing imagery]," Jurnal Geografi Universitas Negeri Semarang, **4**(2), 77-85, 2007, https://doi.org/10.15294/jg.v4i2.100

[16] Kusnadi, Konflik Sosial Nelayan: Kemiskinan dan Perebutan Sumber Daya Perikanan [Fishermen Social Conflict: Poverty and Conflicts over Fisheries Resources], in LKIS, 190, 2002.

[17] H.M. Putri, Nurlaili, "Tata kelola pemukiman nelayan di wilayah perkotaan pesisir utara Jakarta [Fisherman settlement governance in the north coastal area of Jakarta]," Buletin Ilmiah "MARINA" Sosial Ekonomi Kelautan dan Perikanan, **4**(1), 7-13, 2018, http://dx.doi.org/10.15578/marina.v4i1.2048

[18] P. Jones, "Formalizing the informal: Understanding the position of informal settlements and slums in sustainable urbanization policies and strategies in Bandung, Indonesia" Sustainability, **9**(8), 14-36, 2017, https://doi.org/10.3390/su9081436

[19] A. Khan, H. Akter, S. Zakia, J.K. Tania, "Displacement and deplorable living conditions of slum dwellers: With special reference to Sylhet City," International Letters of Social and Humanistic Sciences, **46**, 51-64, 2015, https://doi.org/10.18052/www.scipress.com/ILSHS.46.51

[20] U. Nasir, "Assessing urban sustainability of slum settlements in Bangladesh: Evidence from Chittagong city," Journal of Urban Management, **7**(1), 32-42, https://www.sciencedirect.com/science/article/pii/S2226585617301279?via%3Dihub, 2018, https://doi.org/10.1016/j.jum.2018.03.002

[21] W. Malau, "Dampak urbanisasi terhadap permukiman kumuh (slum area) di daerah perkotaan [The impact of urbanization on urban slums]," Jurnal Pendidikan Ilmu-Ilmu Sosial, **5**(2), 2013, https://doi.org/10.24114/jupiis.v5i2.1113

[22] M. Basir, "Hubungan sosial dan akses sosial masyarakat pada lingkungan pemukiman kumuh di Kota Makasar [Social relations and social access of communities in slums in Makassar City]," Jurnal Perkotaan, **4**(1), 2012.

[23] Kadir, "Penanganan kawasan kumuh Kota Bau-Bau [Management of slums in Bau-Bau City]," Metropilar, **8**(2), 251-259, 2010,

[24] Indriani, "Formasi spasial permukiman kumuh kota studi kasus: perubahan pola ruang bermukim pada lahan di Jalan Sersan Sani Palembang [A case study of spatial formation of urban slums: changes in the pattern of living space on land in Jalan Sersan Sani, Palembang]," Jurnal Arsir, **1**(1), 30-37, 2017, https://doi.org/10.32502/arsir.v1i1.855

[25] Sarbidi, "Kriteria desain drainase kawasan permukiman kota berwawasan lingkungan [Criteria of the design of environmental-friendly urban drainage system]," Jurnal Permukiman, **9**(1), 1-16, 2014, doi: 10.31815/jp.2014.9.1-16

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1042-1046 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Design of a Mobile Application for the School Enrollment Process in Order to Prevent Covid-19

Alexi Delgado[*, 1], Enrique Lee Huamaní[2], Alfredo Chiara-Sotomayor[3], Florencio Roman-Casahuamán[3]

[1]*Mining Engineering Section, Pontificia Universidad Católica del Perú, 15088, Perú*

[2]*Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, 15307, Perú*

[3]*Systems Enginerring Program, Universidad de Ciencias y Humanidades, 15307, Perú*

ARTICLE INFO

Article history:
Received: 19 August, 2020
Accepted: 09 November, 2020
Online: 14 December, 2020

Keywords:
Android
Registration process
Mobile Application
Agile Methodology
Health

ABSTRACT

In 2020, a great change occurred around the world due to the covid-19 pandemic, daily activities were restructured and thanks to technological advance it has been an ideal means to carry out remote work. With regard to education in Peru, it was paralyzed for a long time and only obtained in later months the option of video offerings software for the interaction of teachers and students, given these new activities carried out. In schools, we are proposing a mobile application prototype that is adaptable to any school in Perú in order to facilitate the registration process of parents for their children due to the global crisis that we are going through. Parents will have a follow-up and ideal control for your children. The methodology that will help us to carry out this application will be the agile methodology with the scrum framework, since it adapts to the needs of the client, and to the constant changes in the development process of the mobile application. Since high student demand requires fast delivery of the application with cost and time efficiency. The mobile application to which students and tutors will be automatically transferred will have easy access to use it in the registration processes and in their academic controls. Our task is created in three points of view, student, tutor and teacher. The student was validated, with the registration or registration and login for the students already registered. Where the student can validate their personal data already registered and see their school monitoring, such as grades, attendance, courses, schedules and payments. The result obtained is the proposal of a productive approach, optimizing the license plate control processes, through a mobile application, where each agent can perform this function, and thus reduce the risk of contagion in these moments of pandemic. And so, in this way, interconnect all the enrollments, the benefactors in a system, saving their information and the data of the students. With these results, it will be of great benefit for the school to minimize the time and thus increase the competitiveness between the schools, also for the parents and for the students who despite the difficulties continue to progress in life.

1. Introduction

In December 2019, a large and growing outbreak of a new coronavirus was reported in Wuhan, Hubei Province, China [1]. The global COVID-19 pandemic spreads easily to people nearby, especially in crowds with mobile people (for example, city centers). A widely accepted strategy to mitigate its spread is social distancing, avoiding crowds [2].

One of the first broad-spectrum measures has been the closure of educational centers at all levels, which has affected 91.3% of the total number of students enrolled in the world: more than 1.5 billion people have been left without power attend their face-to-face classes, according to UNESCO [3].

Several methodologies were found for the construction of the mobile application, but in this case the Scrum methodology was used, which is a framework for agile software development that has expanded to other industries.It is a process in which a set of

[*]Corresponding Author: Alexi Delgado, kdelgadov@pucp.edu.pe

good practices are applied on a regular basis to work collaboratively, in teams and to obtain the best possible project results. The evolution of software development methods has taken place since 1970. The Agile Methodology also offers services to companies to create the right product that makes them aggressive in the market [4], [5]. The Agile methodology is considered one of the best approaches for the development of mobile applications [6]. Many unavoidable developmental changes could also be identified [7]. Finally, Balsamiq was chosen for the development of the design, as it is useful for communicating ideas [8].

This article uses the design of a mobile application to analyze the enrollment and school monitoring process in the Comas District [9], in order to allow parents to access enrollment without the need to go to schools. It will also include a search and database of school monitoring information per student.

The objective of this research work is to help parents maintain control of their children at school, using a tool that will be easy to use for the different processes that the institution handles. In addition, with the study of information that we carry out it will allow us to avoid spreading the coronavirus.

The present work is structured in the following way, in the next section number II the methodology used for the construction of the application will be described in detail. In section III the results obtained will be evidenced and finally in section IV the own results will be discussed and the conclusions will be presented.

2. Methodology

In this section we will detail what steps will be followed for the development of the mobile application, so we will use the Agile Scrum Methodology as shown in Figure 1 which explains in a general way the behavior of Scrum, using this methodology is advantageous due to the flexibility in the requirements and ease in adapting to changes.

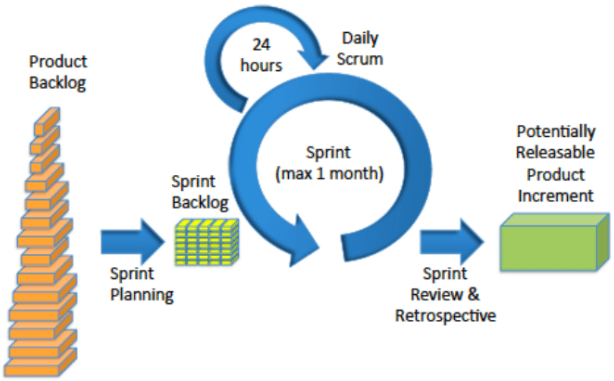


Figure 1: Diagram of the Scrum methodology

2.1. Analysis

Agile Scrum Methodology: As Troy Dimes wrote, scrum is like a lifeline for those companies that face difficulties in not even having a methodology to develop software. Scrum is a framework for creating complex software and delivering it on time in a much simpler way. Scrum is not a methodology; it is a reference framework within the agile software development methodology, which will enable it to create excellent software, through the application of a set of guidelines to be followed by work teams and the use of specific roles [10].

Scrum defines three roles: The Scrum master, the product owner, and the development team. The scrum master's role is to ensure that the team is adopting the methodology, its practices and values and norms; is the team leader, but does not manage development [11]. For its part, the development team is responsible for converting what the customer wants, the Product Backlog, into functional iterations of the product; the development team does not have hierarchies, and all its members have the same level and position: developer. The optimal team size is between three and nine people [12].

Sprint planning

This stage takes place at the Sprint Planning meeting where the activities to be carried out are defined. This activity takes eight hours for a month of Sprint. If Sprint is shorter, the time is allocated proportionally [12].

Development stage

At this stage, the sprint managers must guarantee that last minute changes that affect the objectives are not generated. In addition, they must ensure that they meet the established deadlines.

Sprint review

The Sprint Review occurs at the end of the Sprint and lasts for four hours for a one-month project (or a proportion of that time if the duration is shorter). At this stage: the project owner reviews what was done, identifies what was not done, and discusses the Product Backlog; the development team counts the problems it found and the way they were solved, and shows the product and how it works.

Feedback

Empirical process control is retrospective and not predictive, which would imply that decisions are made based on evidence and knowledge is nurtured in experience. The lessons to be learned in this stage will allow us to learn how sprints can be much more effective and agile.

2.2. Design and tools

Programming platform (Android Studio)

It is the official integrated development environment for the Android platform, is one of the most used IDES for the development of Android applications due to its intuitive interface for the arrangement of the elements used in the architecture of the project [13].

Database manager (Microsoft SQL Server)

The storage and management of data is mainly done through the database of the business management system, so it is very important to study the design and implementation of the database in the system [14].

Design Tool (Balsamiq Wireframes)

In this research, the Balsamiq Mockups tool is used as a tool that will facilitate the development of the project's prototypes [15].

2.3. Procedure

Product blacklog

The Product Backlog is a list, ordered by value, risk, priority and need of the requirements that the product owner defines, updates and orders. The list has as a particular characteristic that it is neveFinished, since it evolved during the development of the project [12].

Sprint backlog

It is the task list that contains selected tasks from the Product Backlog, in which it subdivides the user stories that describe the functionalities that make up the project. This listing is defined and estimated at the Sprint Planning meeting at the start of the iteration [16].

Sprint planning meeting

It is the Sprint planning meeting based on the Product Backlog and they participate: Product Owner (responsible for the subject) who prioritizes the tasks to include in the Sprint Backlog, the Scrum Master and the Development Team [16].

3. Case study

This section will develop what is the planning of the sprints of each of the modules and the estimates of the times in which it is being carried out in the mobile application development stage.

3.1. Sprint planning

- As a user you can enter the application through an account generated by the school.
- As a user you can view the courses with their grades, schedules and pensions.
- As a user, you may be able to register me once you have paid the registration fee.
- As a proxy you can know the names of the teachers with their respective courses that my son will take.
- As a proxy you can know the respective qualifications of my son with their respective states.
- As administrator, you can keep track of student pensions.
- As administrator you can generate detailed reports of each student.

3.2. Development Stage

Time estimation

In this phase we will see the duration of each of the sprints as shown in Table 1.

Product scope

This is the estimate of the time it takes for the equipment to be able to have the user's history points, for which the estimates of the scope of the equipment will be noted in Figure 2.

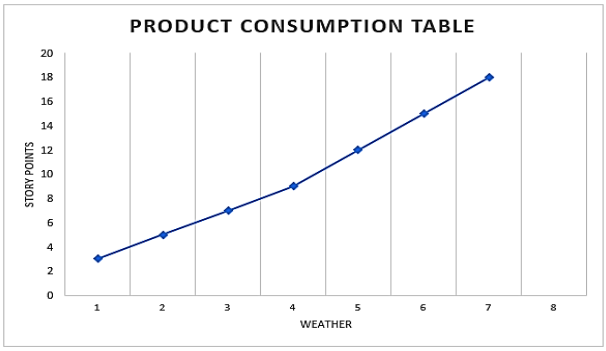


Figure 2: Product scope

Table 1: Duration of Sprints

Name	Duration	Start	Final
Mobile Application for the Enrollment Process	3 months	01/06/2020	01/10/2020
1. User Module	2 weeks	01/06/2020	15/06/2020
2. Enrolment Module	2 weeks	16/06/2020	30/06/2020
3. Schedule Module	2 weeks	31/06/2020	13/07/2020
4. Courses Module	2 weeks	14/07/2020	27/07/2020
5. Ratings Module	2 weeks	28/07/2020	10/07/2020
6. Pension Module	4 weeks	10/07/2020	10/08/2020
7. Reporting Module	5 weeks	11/08/2020	31/09/2020

3.3. Prototypes of the mobile application

In this stage we will see the creation of the mobile application design that we will carry out and we will detail the functions that each of the modules will have as we see in Figure 3 (a), Figure 3 (b), Figure 3 (c) and Figure 3 (d).

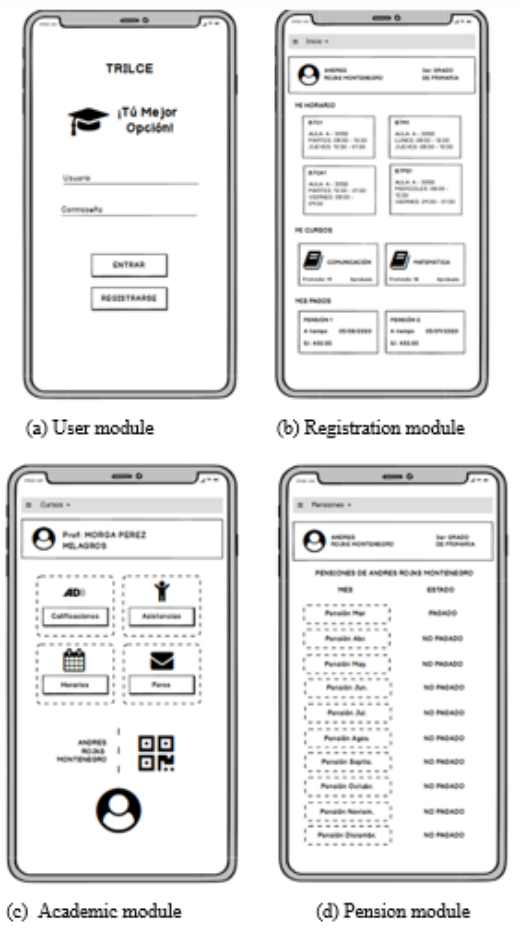


Figure 3: Prototype phase of the mobile application

Increment 1: Main Start

This increase shows the start of the application which, if we have a user already generated by the institution, we could enter the account that maintains personal data of the students Figure 3 (a).

Increment 2: Enrollment module

This increase shows the process that must be carried out to enroll each student, and as a requirement to proceed, they would have to pay the enrollment for the school year Figure 3 (b).

Increment 3: Detailed courses module

In this increase it will show the detailed course with its qualifications, assists, schedules and forums, in addition to having a QR code to download the report Figure 3 (c).

Increment 4: Pension module

In this increase it shows the pensions of each student, detailed by month and the state if it is already canceled or even just canceled Figure 3 (d).

4. Results and discussion

4.1. About the case study

Compared to other research works, we have observed the design of the work of an Android mobile application for an online Travel Agency, where the use of the NinjaMock tool is specified, where we can see that, on the contrary, its interface does not provides a professional visual finish, increasingly difficult to use [17].

The google API services will be used to obtain the latitude and longitude of the reported location [18]. The importance of the interface is a means for the user to interact with the software and hardware. The interface will help the user to interact with the system much easier than with the old method of giving a command at the command prompt. The following image is the Trusted Halal interface. Malaysia Cosmetic Products System Application [19] [20].

Due to the covid-19 pandemic, enrollment in Peru was 7,834,543, unlike last year, which was 8,024,672. Figure 4 extracted from the Ministry of Education [21] shows the number of enrollments in 2020 and a large decrease in enrollment is observed, therefore the proposal of this research will help students to enroll successfully using mobile applications.

URBANO		RURAL	TOTAL
6,637,116	BÁSICO REGULAR*	1,197,427	7,834,543
1,336,742	INICIAL*	292,411	1,629,153
3,021,779	PRIMARIA*	614,344	3,636,123
2,278,595	SECUNDARIA*	290,672	2,569,267
13,161	BÁSICA ESPECIAL*	63	13,224
214,004	BÁSICA ALTERNATIVA**	1,172	215,176
248,402	TÉCNICO PRODUCTIVO**	3,647	252,049
521,353	SUPERIOR NO UNIVERSITARIA**	4,975	526,328
	SUPERIOR UNIVERSITARIA*		1,492,470

*Avance de matrícula al 17.07.2020. En el 2019, la matrícula en EBR fue de 8,024,672 y en EBE fue de 20,722.
**Al 2019.

Figure 4: Statistical data of enrolled in 2020 under the covid-19 pandemic

4.2. About the methodology

The tool used in the present work, which is Balsamiq Mockups, is easier and more adaptable to any user, and that shows us a good effectiveness before its prototypes made. In Balsamiq Mockups, mobile applications are designed in a smarter and easier way [22]. We have observed as an important factor in a project and the first decision of the application is the development prototype therefore the prototyping software Balsamiq Mockups served as a benefit to be able to create our Wireframes easily and quickly [23].

Riadh proposed designing the notification system using GCM (Google Cloud Messaging) of the Android app and connecting to the website and providing information that helps instructor to select who will send notification [24].

In this project, the methodology to be used is Agile Scrum Methodology. Scrum is a lightweight project management framework with the ability to iteratively and incrementally control and manage all types of projects. Through use. The "Product

Owner" Scrum methodology will work closely with the team to detect and prioritize system functionality [25].

Differences in Methodologies: This part will detail the differences of using the agile methodology, as opposed to using other methodologies that exist. Table 2 will detail the differences that each methodology will have.

Table 2: Traditional Methodology vs Agile Methodology

Traditional methodologies	Agile methodologies
Predictive	Adaptive
Process-oriented	People-oriented
Rigid process	Flexible process
It is conceived as a project	A project is subdivided into several smaller projects.
Little communication with the client	Constant communication with the client.
Software delivery at the end of development.	Constant software deliveries
Extensive documentation	Little documentation

5. Conclusions

In this research we designed the proposal of a mobile application, helping to achieve a better solution at the time of registration of enrollment and monitoring of students.

This proposal does not try to meet one level but to seek a comprehensive solution in the short term and can be adapted to any educational institution as has been determined in Table 2 with respect to the duration of the Spring. This work can be improved and will be carried out when they go through the final users in which the satisfaction of the users will be determined with statistical data. We can conclude that the design proposal of the mobile application will be of great benefit to parents when registering their children and monitoring and controlling their daily activities, a proposal that is being promoted in schools in Lima, Peru.

References

[1] C. Huang et al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," The Lancet, **395**(10223), 497–506, 2020, doi:10.1016/S0140-6736(20)30183-5.

[2] Centers for Disease Control and Prevention, Coronavirus (COVID-19), 2019.

[3] UNESCO, Education: From disruption to recovery, 2019.

[4] B. Mathur, S.M. Satapathy, "An analytical comparison of mobile application development using agile methodologies," Proceedings of the International Conference on Trends in Electronics and Informatics, **4**, 1147–1152, 2019, doi:10.1109/icoei.2019.8862532.

[5] P. Abrahamsson, A. Hanhineva, H. Hulkko, T. Ihme, J. Jäälinoja, M. Korkala, J. Koskela, P. Kyllönen, O. Salo, "Mobile-D: An agile approach for mobile application development," Proceedings of the Conference on Object-Oriented Programming Systems, Languages, and Applications, 174–175, 2004, doi:10.1145/1028664.1028736.

[6] S. Holla, M.M. Katti, "ANDROID BASED MOBILE APPLICATION DEVELOPMENT and its SECURITY," International Journal of Computer Trends and Technology, **3**(3), 486–490, 2012, doi:10.1080/10304312.2012.706462.

[7] A. Cockburn, L. Williams, Agile Software Development: It's about Feedback and Change, IEEE Computer Society: 39–43, 2003.

[8] I. Bouchrika, L. Ait-Oubelli, A. Rabir, N. Harrathi, "Mockup-based navigational diagram for the development of interactive web applications," ACM International Conference Proceeding Series, (7), 27–32, 2013, doi:10.1145/2503859.2503864.

[9] R.C. Jisha, M.P. Mathews, S.P. Kini, V. Kumar, U. V. Harisankar, M. Shilpa, "An Android Application for School Bus Tracking and Student Monitoring System," International Conference on Computational Intelligence and Computing Research, 1–4, 2018, doi:10.1109/ICCIC.2018.8782320.

[10] A. Mundra, S. Misra, C.A. Dhawale, "Practical scrum-scrum team: Way to produce successful and quality software," International Conference on Computational Science and Its Applications, 119–123, 2013, doi:10.1109/ICCSA.2013.25.

[11] K. Schwaber, J. Sutherland, The Scrum Guide: The Definitive The Rules of the Game, 2017.

[12] M. Wadhwa, N. Sharma, "Review of Agile Software Development Methodologies," International Journal of Advanced Research in Computer Science and Software Engineering, **3**(2), 370–374, 2015.

[13] R. V Golhar, P.A. Vyawahare, P.H. Borghare, A. Manusmare, "Design And Implementation Of Android Base Mobile App For An Institute," International Conference on Electrical, Electronics, and Optimization Techniques, 3660–3663, 2016.

[14] L.J. Zhu, "Research and application of SQL Server in the trade management system," International Conference on Computer Research and Development, **3**, 209–213, 2011, doi:10.1109/ICCRD.2011.5764282.

[15] A. Delgado, J. Sosa, "Mobile application design of geolocation to collect solid waste: A case study in Lima, Peru," International Conference on Electronics, Electrical Engineering and Computing, 1–4, 2019, doi:10.1109/INTERCON.2019.8853594.

[16] P.L. Alfonzo, "Implementación de SCRUM en el diseño del proyecto del Trabajo Final de Aplicación," Scientia Et Technica, **19**(4), 413–418, 2014, doi:10.22517/23447214.9021.

[17] A. Gómez, Aplicación Android para la empresa Travelling-Service, 2014.

[18] Z. Jiantao, P. Xuan, "Design and implementation of field staff management system on android mobile," International Conference on Instrumentation and Measurement, Computer, Communication and Control, 791–794, 2013, doi:10.1109/IMCCC.2013.176.

[19] V.N. Vithana, D. Asirvatham, M.G.M. Johar, "Investigating the Issues of Using Agile Methods in Offshore Software Development in Sri Lanka," Studies in Computational Intelligence, **710**(March), 515–523, 2017, doi:10.1007/978-3-319-56660-3_44.

[20] A. Delgado, "Social conflict analysis on a mining project using shannon entropy," International Congress on Electronics, Electrical Engineering and Computing, 3–6, 2017, doi:10.1109/INTERCON.2017.8079661.

[21] MINEDU, Propuesta de metas educativas e Indicadores al 2021 Lima, 2010.

[22] K. Kalimullah, D. Sushmitha, "Influence of Design Elements in Mobile Applications on User Experience of Elderly People," Procedia Computer Science, **113**, 352–359, 2017, doi:10.1016/j.procs.2017.08.344.

[23] Balsamiq Studios, Quick and Easy Wireframing Tool, 2020.

[24] M. H., "Notification System to Students using an Android Application," International Journal of Computer Applications, **140**(1), 22–27, 2016, doi:10.5120/ijca2016909189.

[25] A. Srivastava, S. Bhardwaj, S. Saraswat, "SCRUM model for agile methodology," International Conference on Computing, Communication and Automation, **6**, 864–869, 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1047-1053 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Estimating LAI of Rice Using NDVI Derived from MODIS Surface Reflectance

Rushikesh Kulkarni[1,*] Kiyoshi Honda[2]

[1]*Symbiosis Institute of Technology, Symbiosis International (Deemed University), Lavale, Pune, Maharashtra, 412115 India*

[2]*Department of Astronautics and Aeronautics, Chubu University, 1200 Matsumoto-cho, Kasugai, Aichi 487-8501, Japan*

ARTICLE INFO

Article history:
Received: 29 October, 2020
Accepted: 20 November, 2020
Online: 14 December, 2020

Keywords:
Rice LAI
In-situ LAI
MODIS Surface reflectance

ABSTRACT

The measurement of Leaf Area Index is of fundamental importance in agriculture and ecological research because it reflects plant growth. This study aims to estimate LAI from, NDVI derived from standard eight days composite MODIS surface reflectance, acquired during the rice growing period.

A two-stage model was developed to relate the LAI of rice to NDVI derived from MODIS surface reflectance. In the first stage, a correlation has been developed between 'MODIS NDVI' and 'in situ NDVI' derived from spectral reflectance measured with spectrophotometer, in rice field of Suphanburi Province, Thailand. In the second stage, correlation has been developed between in situ NDVI and, in situ LAI measured with handheld LAI-2000 instrument. Eventually, a model has been derived using these two relationships, which estimate LAI from MODIS Surface Reflectance, in an irrigated rice field, central Thailand.

1. Introduction

In agriculture and ecology, historically, the Leaf Area Index (LAI) concept is used for the estimation and evaluation of crop production and crop yield and the scheduling of irrigation and amendments during crop cultivation [1]. In order to assess growth and vegetation intensity across the globe, monitoring the distribution and change of LAI is significant. It is directly related to the interception and absorption of light by the canopy. It also influences heat balance and evaporation from the cropping field and thus, accurate assessment of LAI is necessary [2]. However, many factors are important if the reflectivity of the crop canopy is to be assessed. These include the complexity of canopy design, improvements to the internal arrangement of leaves and background effects of the soil [1].

Direct measurement of canopy LAI is accurate but extremely labor intensive and destructive [3]. Remote sensing techniques have been used to measure LAI on a landscape scale or even global scale. The benefit of the remote sensing technique is that it can collect plant canopy information on a global scale without interfering with normal plant growth [4, 5]. The introduction of simple NDVI-LAI relationships has yielded reliable results over a 12-year period in the estimate of green vegetation abundance [6]. There are two common approaches to estimate LAI using remote sensing [7]. The first one is associated with spectral vegetation index (SVI) and the second one, is by modelling approach [7]. In the first approach, an empirical relationship between a Vegetation Index (VI) and LAI has to be established. One major drawback of this method is that equation have to be established for each vegetation type. Thus, it requires substantial LAI measurements and corresponding remote sensing data [8]. This method is unsuitable for pixels of low-resolution remote sensing data enclosing a heterogeneous field. Another approach to estimate LAI is by using Bi-Directional Distribution Function (BRDF) models such as Spectral Absorption of Inclined Leaves (SAIL). These models usually consist of a set of equations that relate surface physical properties to observed signals as a function of wavelength. Physical properties may include soil reflectance, canopy architecture, optical properties and geometric configuration of sensing systems as well as illumination sources. . The input parameters needed for BRDF models are single-scattering albedo of individual leaves, leaf inclination distribution and anisotropic properties of both canopy and soil substrates. BRDF models can then be theoretically inverted to obtain the model parameters, and some vegetation parameters can be estimated [7].

The relationship between growth traits and Normalized Differential Vegetation Index (NDVI) to assess their potential use in estimating the growth of rice crops has been examined [9].

*Corresponding Author: Rushikesh Kulkarni, rushikeshk@sitpune.edu.in

Experiments were conducted at the Taiwan Agriculture Research Institute (TARI) experimental farm on loam soil during the first and second cropping seasons (1996-1997). Plants were sampled at two weeks intervals after transplanting till harvest and the LAI was computed at each new sampling date. A LI-1800 high spectral resolution portable Spectroradiometer (LI-COR Inc., Nebraska, USA) was used to measure the reflectance to calculate in-situ NDVI. Yang and Su then developed a correlation between the computed LAI and in-situ NDVI. A linear regression gave results with R^2=0.982 for SPOT HRV and R^2= 0.985 for Landsat MSS. As these spectral transformations are highly correlated, estimation of growth traits based on NDVI can be accurate for both in-situ NDVI and satellite-derived NDVI. Thus reflectance band of HRV and MSS sensors could be used as predictors of canopy characteristics [9].

In-Situ LAI and spectral reflection data were assessed for the 2004-2005 period in part of Suphan Buri, Thailand [10]. This was done in order to develop a relationship between LAI and NDVI at ground level. In 2004, the researchers developed a model to estimate ground based LAI from ground-based NDVI. This is a specific case, since the remote sensing data corresponding to the field data collection period was not available. It was assumed that NDVI from Landsat MSS and from SPOT VGT to be equal and thus, used the model developed by Yang and Su (1998) for MSS to estimate ground based NDVI from SPOT VGT [10]. Yang and Su's LAI - NDVI correlation estimated LAI in the range of 0.082 to 1.848 for SPOT VGT data which seemed too low compared to the LAI field data collected by Srinuandee in 2004-05, which were measured in the range of 0.0 to 7.15. Srinuandee's field based correlation estimated LAI in the range of 2.03 to 5.3002 for a pixel in SPOT VGT.

Srinuandee's model showed little improvement compared to Yang and Su's model however, impressive variation in result have not been observed. Improvements made by Srinuandee on Yang and Su's model are not proving conclusive. Thus, the main goal of this research was to estimate LAI from Moderate Resolution Imaging Spectroradiometer (MODIS) based NDVI with major changes on Srinuandee's efforts. Indeed, in-situ LAI data recorded by Srinuandee seems to be overestimated and BRDF effect has not been considered during data collection. BRDF effect is an important factor to be considered in case of low-resolution remote sensing data where satellite have wide swath since the measurements vary according to different viewing angles.

2. Materials and methods

2.1. Description of Study Area

The case study was conducted in rice field of Song Phi-Nong district, Suphan Buri Province, central part of Thailand during January' 2006 to April' 2006 which is the growing period of first crop in Suphan Buri province. The geographic location of study site as shown in Figure 1. Experiment was carried out in the homogeneous rice field which was detected using Landsat 7 ETM +, land use map and aerial photographs.

The topography of study area was mostly flat lowland terrain. Thus the rice fields in this area remains completely flooded during rainy season. The climate of the Province characterized mainly by three seasons: (i) rainy season from May to October (ii) cool and dry season from November to February (iii) hot and dry season from March to May. The average annual rainfall in the central plain is between 900-1300 mm, average solar radiation about 17 MJ $m^{-2}d^{-1}$ and temperature ranges from 33-34^{0}C to 23-24 ^{0}C.

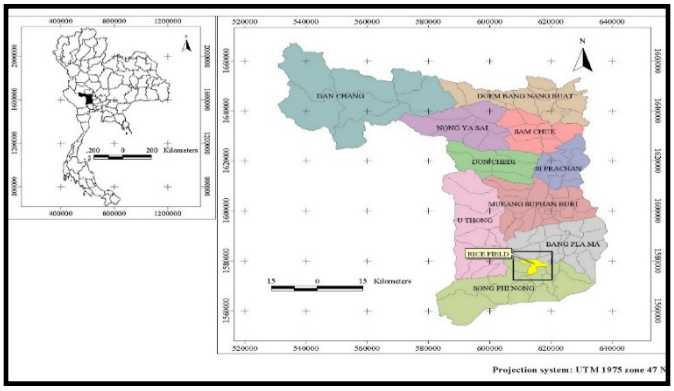


Figure 1: Location of study area in Suphan Buri Province, Central Thailand

3. Field Data Collection

The spectral reflectance and LAI were measured using a Spectroradiometer and a LAI-2000 Plant Canopy Analyser (PCA) respectively from the first stage of rice growth till the harvesting period. LAI was measured at 9:00 am to 11:00 am because LAI canopy analyser doesn't need direct solar radiation to measure LAI and LAI measurements should be done with low sun elevation angle which is in between morning time or in the evening time. Five different locations were sampled at each time of ground trothing.

In-situ spectral reflectance readings used to calculate NDVI have been recorded at five different angles in East and West directions separately in order to create similar circumstances as satellite data; since, satellite moves in North-South direction with acquiring reflectance in East-West direction. Each measurement was recorded with 15^0 intervals, starting from 0^0 to 60^0, marked on a simple handmade clinometer in order to examine BRDF effect in the field refer Figure 2. Five readings of spectral reflectance at each five different angles in East and West directions and five LAI readings, from five sampled locations were averaged for each day of ground truthing. Total eight readings of spectral reflectance at each five different angle and eight readings of LAI, were measured in field during the period of 12th February to 2nd April. The fibre optics wire of the spectrometer was clamped at each angle and spectral reflectance readings were recorded in both directions, so total ten readings of spectral reflectance has been measured each day of ground truthing. One end of fibre optic wire was at a height of 1.5 m from ground level and another end was connected to spectrometer which was further connected to computer through USB port. The spectrometer measures spectral reflectance in the range of 304 nm to 1135 nm with an increment of 1 nm. In order to compare in situ NDVI with MODIS data, average reflectance for red band was calculated from wavelength of 620 nm to 670 nm and that for near infrared band was calculated from wavelength of 841 nm to 876 nm.

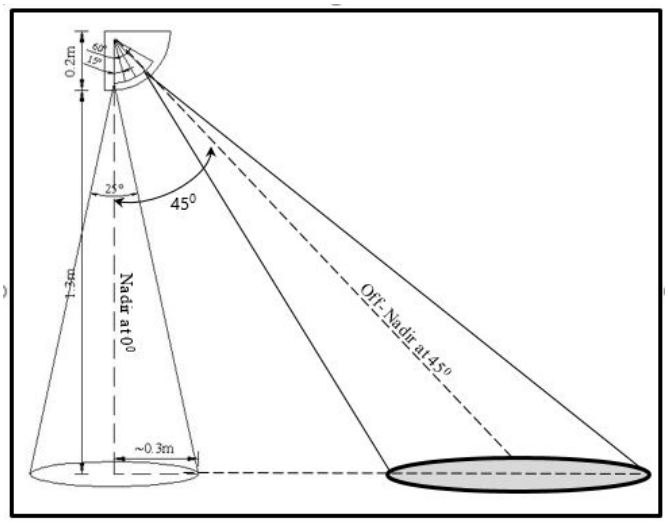


Figure 2: Schematic diagram of clinometer setup used to measure surface reflectance in the field with various angles.

In-situ NDVI estimated from field spectral reflectance measured in each week of data collection has been compared with NDVI estimated from 8 days composite MODIS surface reflectance acquired on corresponding week of ground thruthing, refer figure 5. Two field data collection dates, 26th Feb and 3rd March, are members of 8 days composite of 26th February to 5th March (Table 1).

The general flow chart of the methodology is shown in figure 3. The atmospheric corrected MODIS surface reflectance is obtained and compared to the field NDVI derived using a fiber-optic spectroradiometer.

The relationship between MODIS derived NDVI and LAI is established as shown in flow chart of methodology figure 3.

4. Results and discussion

4.1. Correlation analysis of in situ NDVI and NDVI derived from MODIS

Thus, eight values of in situ NDVI were compared with seven values of MODIS derived NDVI. NDVI derived from MODIS surface reflectance and in situ NDVI did not show the same pattern in early crop development.

Table 1: Comparison of field data collection dates and MODIS 8-days composite dates

Field Data collection dates	MODIS 8-days composite dates
12th Feb	10th Feb - 17th Feb
19th Feb	18th Feb - 25th Feb
26th Feb	26th Feb - 05th Mar
03rd Mar	26th Feb - 05th Mar
12th Mar	06th Mar - 13th Mar
19th Mar	14th Mar - 21st Mar
26th Mar	22nd Mar - 29th Mar
02nd Apr	30th Mar - 06th Mar

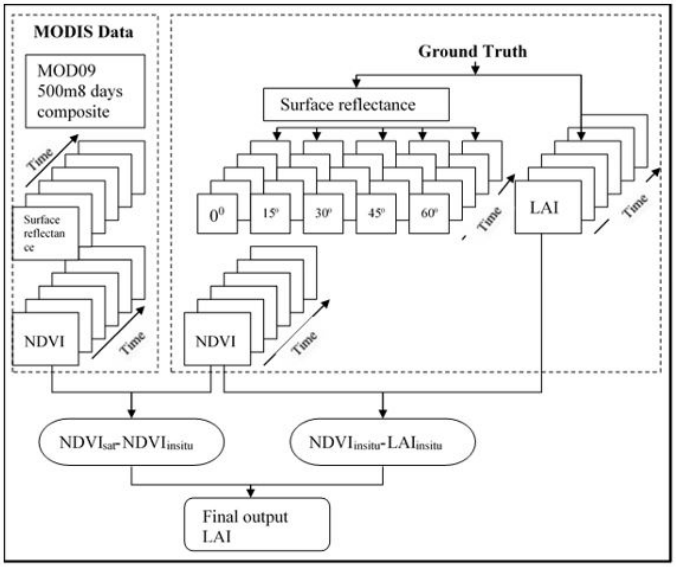


Figure 3: Overall methodology of model development

However, they agree in the full vegetative stage. In the field, early growing stage paddy has a dispersed pattern of newly germinated rice plants along with water, refer figure 4. In this case, muddy soil was easily visible at this stage, which induces canopy background contamination error. While in case of observation from MODIS, angular effect make the field looks different from the way it actually looks with naked eyes. Thus, angular effect was responsible for higher NDVI in case of MODIS observations. Difference between "true" NDVI, as would be measured at the surface, and that actually determined from the space are sensitive to attenuation by atmosphere and by aerosols [11]. In this case MODIS BRDF nadir adjusted 16 days composite data was compared with in situ NDVI figure 6. Similar growing pattern was obtained with in situ ground truth measurement measured at nadir. However, it was rather difficult to compare 16 days composite data with a data acquired on single day.

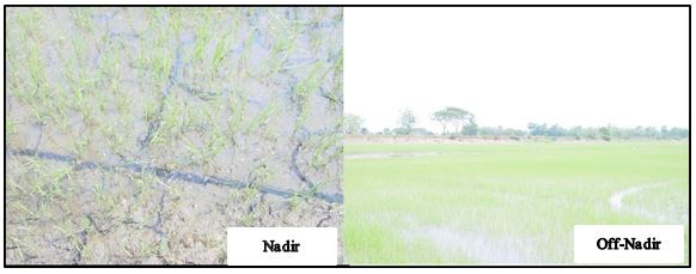


Figure 4: Photograph of paddy field taken at nadir and at off-nadir.

MODIS NDVI was compared with average in situ NDVI estimated from NDVI from all five locations in the field, in both directions i.e. East and West, for all five observation angles. Ground truth surface reflectance was collected in morning time from 9:00 am to 11:00 am.

Linear correlation coefficient (R^2) was calculated between in situ NDVI estimated in five angles and NDVI derived from MODIS surface reflectance,(Table 2). The results obtained were more or less seems to be equal in all cases figure 7[a-e]. The in situ NDVI was linearly interpolated to corresponding view zenith angle of MODIS.

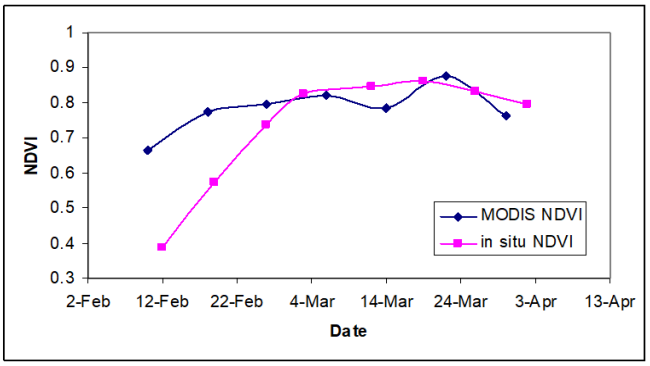


Figure 5: Plot between estimated NDVI in field at 00 viewing angle and NDVI from MODIS surface reflectance.

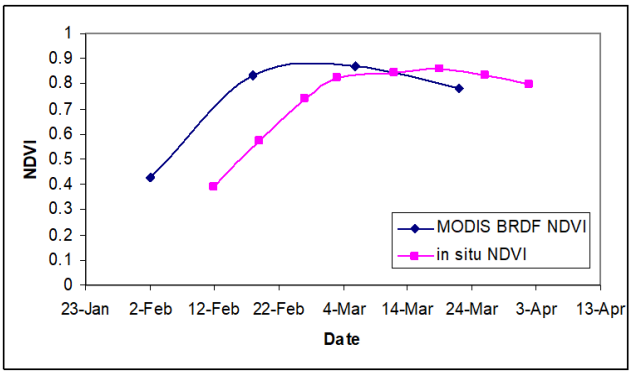


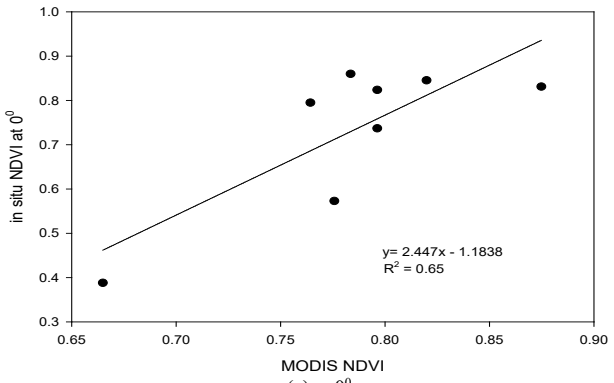
Figure 6: Plot of NDVI computed from MODIS Nadir adjusted data, NDVI estimated from field-collected data

Table 2: Linear relationships between in situ NDVI at different angle and MODIS NDVI

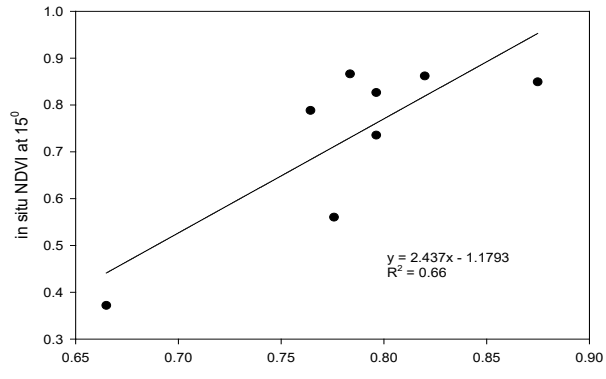
Observation angle	Relationship	R^2
0^0	y= 2.447x – 1.1838	0.65
15^0	y= 2.437x – 1.1793	0.66
30^0	y= 2.447x – 1.1838	0.65
45^0	y= 2.532x – 1.2469	0.67
60^0	y= 2.488x – 1.2068	0.68

Table 3: Characteristics of MODIS on respective dates.

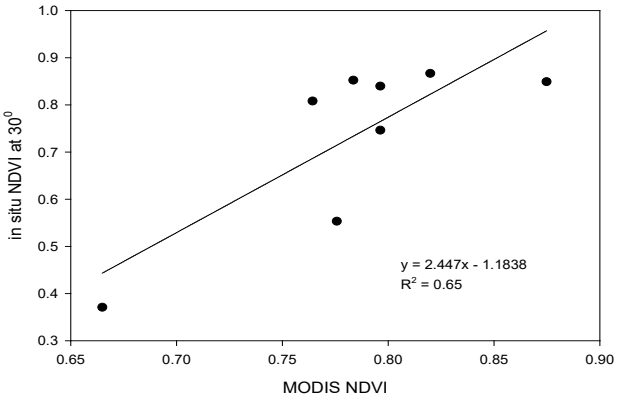
MODIS data available day	DOY	View zenith angle	solar zenith angle	NDVI
02 Feb	33	6.77	39.20	0.5987
10 Feb	41	7.93	36.59	0.6649
18 Feb	49	6.43	30.69	0.7757
26 Feb	57	8.39	32.14	0.7962
06 Mar	65	22.26	28.93	0.8199
14 Mar	73	20.07	29.22	0.7835
22 Mar	81	22.45	24.01	0.8748
30 Mar	89	19.39	25.14	0.7643



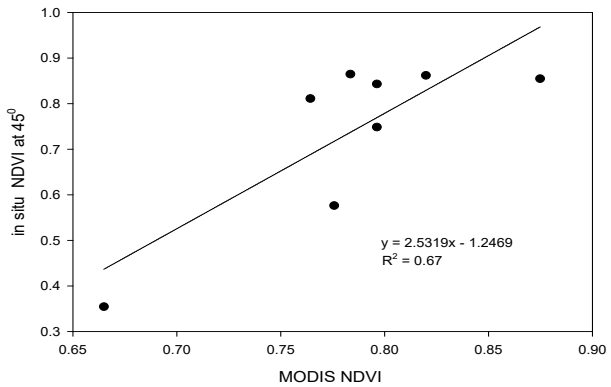
(a) 0^0



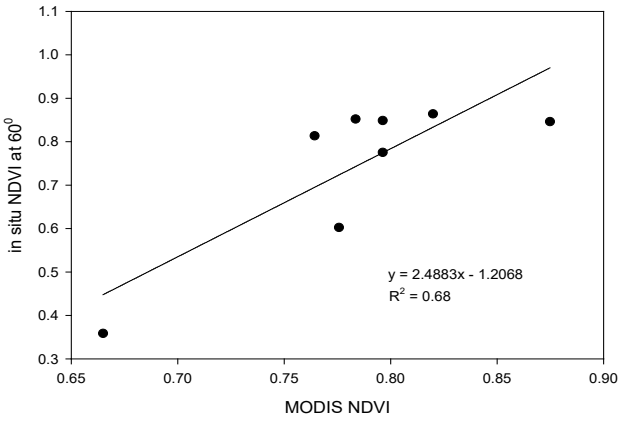
(b) 15^0



(c) 30^0



(d) 45^0



(e) 60^0

Figure 7: Relationship between in situ NDVI estimated at all five different angles and MODIS NDVI

$$\text{In-situ NDVI}_{interpolated} = 2.3959 * \text{NDVI}_{MODIS} - 1.1448 \quad (1)$$

$$\text{Model derived in-situ NDVI}_{at\,0^0} = 0.9586 * \text{NDVI}_{interpolated} + 0.024 \quad (2)$$

$$\text{Model derived in-situ NDVI}_{at\,0^0} = 2.2967 * \text{NDVI}_{MODIS} - 1.0727 \quad (3)$$

MODIS metadata file represents the specifications of the MODIS surface reflectance product. (Table 3). The in-situ NDVIS measured at various angles is interpolated to the corresponding zenith angle of the MODIS. The results are shown in Table 4.

Table 4: Interpolated in situ NDVI corresponding to view zenith of MODIS

Field data collection Day	Interpolated in situ NDVI
12 Feb	0.3794
19 Feb	0.5675
26 Feb	0.7362
03 Mar	0.8252
12 Mar	0.8642
19 Mar	0.8616
26 Mar	0.8492
02 Apr	0.7941

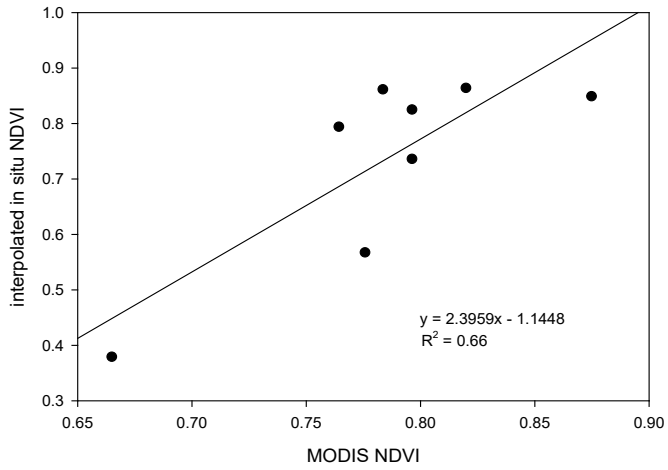


Figure 8: Relationship between MODIS NDVI and interpolated.

The interpolated in situ NDVI was adjusted to standardized form i.e. at nadir observation. Thus, the in situ NDVI was derived from MODIS NDVI in multistage approach.

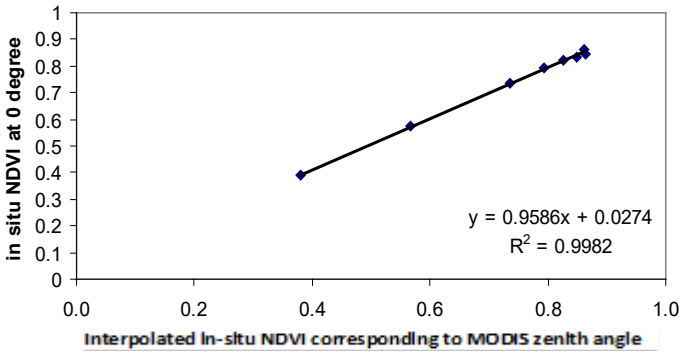
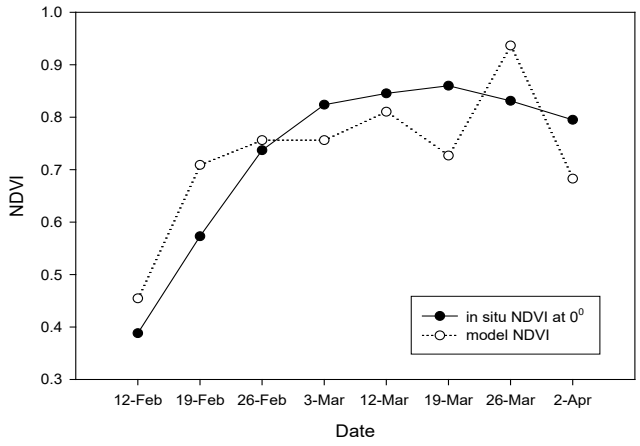
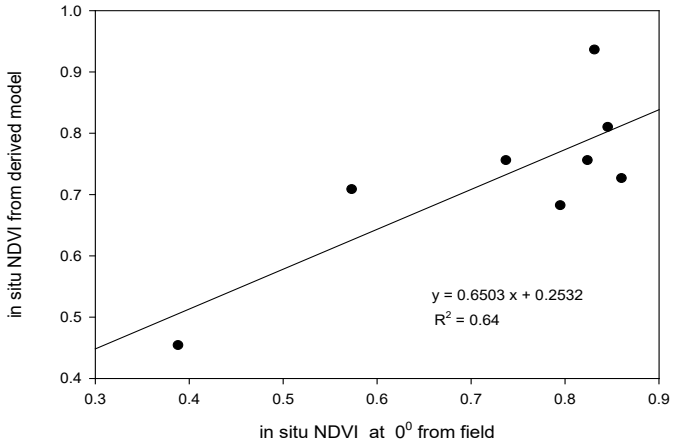


Figure 9: Plot of in situ NDVI estimated at 00 and interpolated in situ NDVI.



(a)



(b)

Figure 10: Comparison of in situ NDVI estimated at 00 in field and in situ NDVI obtained by derived model

The presence of clouds has been shown to have a major impact on the field data set. The remote sensing data collected on the day 19th March and 22nd March have lots of errors which induced errors in calculation of NDVI and LAI.

The field data collection in this study started with data collection from seven locations but due to proper care and maintenance of rice at two locations in the field, data collection from only five sites could be considered till the end of study.

5. Validation of Model

The derived NDVI from model showed nearly same pattern with an in situ NDVI estimated at 0^0 in the field, figure 10 [a-b]. In the beginning of rice growth, modelled in situ NDVI seems to be approaching to get better fitness with in situ NDVI. However, in the last stage of rice growth, few contaminations induced in remote sensing data were intending to the modelled in situ NDVI to oscillate from expected output.

6. Correlation analysis of in situ NDVI and in-situ LAI

In-situ NDVI and in-situ LAI were not linear together. The sensitivity of NDVI to LAI becomes increasingly weak with increasing LAI beyond a threshold value, which was typically between 2 and 3 [11]. NDVI increases almost linearly with increasing LAI and then asymptotic region was observed in which NDVI increases very slowly with increasing LAI. Ground surface was almost completely covered by leaves in asymptotic region.

A linear correlation coefficient (R^2) was calculated between in situ NDVI estimated at 0^0 and in situ LAI. However, most of the LAI measurements were in the range of 4 to 5. A non-linear correlation analysis between in situ NDVI and in situ LAI was used in this study.

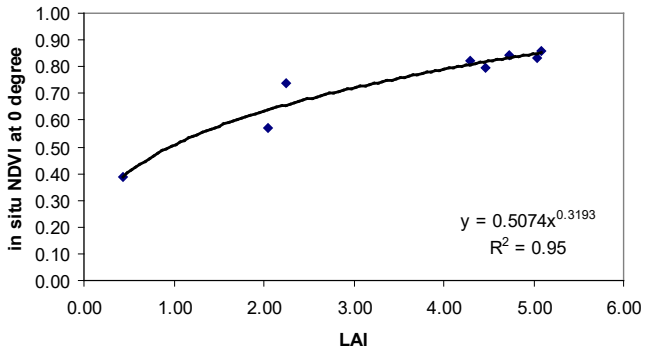


Figure 11: In situ NDVI computed at Nadir and in-situ LAI collected from filed.

Eventually the result obtained by non-liner correlation analysis between in situ NDVI estimated at 0^0 and in situ LAI is:

$$\text{In situ NDVI at } 0^0 = 0.5074\ \text{LAI}^{0.3193} \quad (4)$$

7. Final Model development

The final model involves two parts, the first one is a NDVI transformed from satellite to in situ and the second one from in situ NDVI to LAI. In the first part, similar variables (NDVI) were compared. In the second part, the spectral reflectance and an in-situ LAI were measured at the same location and at the same time in the rice field, which makes this multi stage modeling approach more realistic than single stage model. However direct relationship between MODIS NDVI and in situ LAI was not used in this study.

In situ NDVI was calculated with RMSE of 0.603 and that of for LAI was estimated at 2.120. The high RMSE could be induced because of the contamination error in the remote sensing data. The remote sensing data acquired on 22nd of March estimated higher value of NDVI which caused an error in LAI calculation.

$$\text{In situ NDVI at } 0^0 = 2.2967\ \text{NDVI}_{\text{MODIS}} - 1.0727 \quad (5)$$

$$\text{LAI} = (\text{NDVI}_{\text{in situ}} / 0.5074)^{3.1318} \quad (6)$$

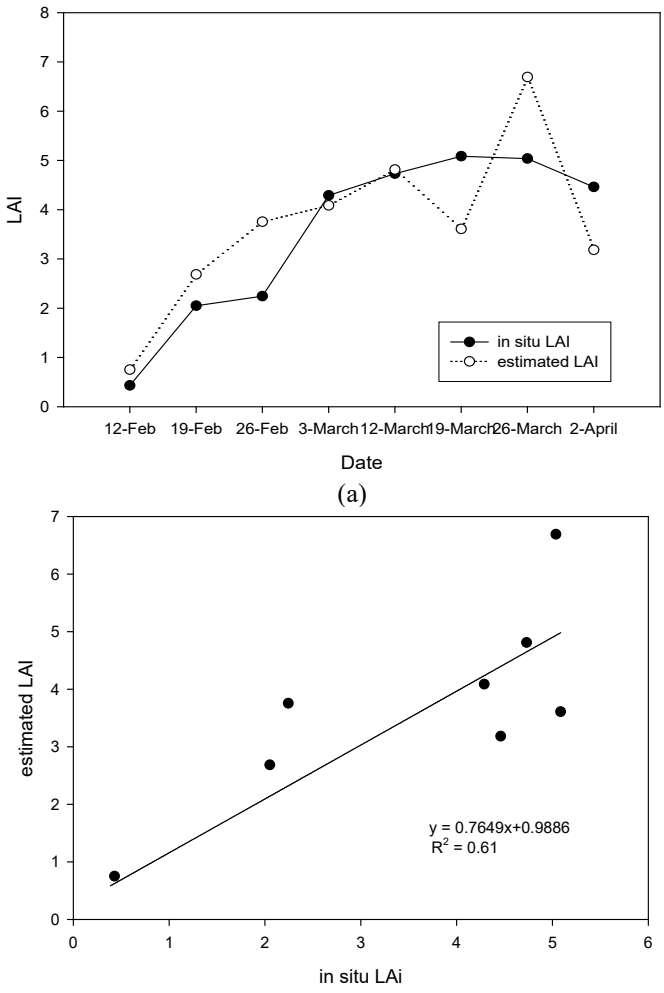


Figure 12: Validation of model LAI with in situ LAI

The LAI estimated with developed model grow similar as the in situ LAI measured. Coefficient of correlation (R^2) was found to be 0.61 which could be increased with increasing accuracy of remote sensing data.

8. Conclusions

In this study, field-based rice LAI is estimated from standard MODIS surface reflectance product. We proposed multi stage approach to derive LAI from MODIS surface reflectance. In the first stage, in situ NDVI is derived from MODIS NDVI and in the second stage, field-based LAI is derived from in situ NDVI. BRDF corrected relationship is considered in the first stage.

The angular effect in MODIS NDVI causes little higher values of modeled in situ NDVI than in situ NDVI estimated at 0^0. In the case of in situ measurements of surface reflectance small

area is sampled by fiber optics wire as compared to remote sensing data. If the number of locations in the site is increased more realistic result can be obtained. The data information of rice growth in a year definitely improves the result. The accuracy of the model could be increased with increasing in the field data collection for entire period of rice cropping cycle in a year.

The model could be used for regional scale with MODIS 250m and accuracy can be checked with MODIS 500m. The model can also be further developed for other crops like sugarcane, maize etc. The model could be improved for mixed pixel condition. The LAI could be estimated by using some other vegetation indices such as Soil Adjusted Vegetation Index (SAVI) and accuracy check can be done with the derived model in this study.

In order to enhance the applicability of LAI-NDVI relationship, more crops and more satellite remote sensing data sets should be assessed and checked. The use of multi-angular remote sensing for LAI estimation may be further tested and investigated using radiative transfer models.

Acknowledgements

We thank Ms. Kamal Thapa for her assistance and support during in-situ data measurement and also thankful to Thai research students from the RS-GIS Field of Study for helping as interpreter, Thai- English-Thai during field data measurement.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] C. Tan, P. Zhang, X. Zhou, Z. Wang, X. Zi-qiang, "Quantitative monitoring of leaf area index in wheat of different plant types by integrating NDVI and Beer-Lambert law," Scientific Reports, Nature Research, 1–10, 2020, doi:10.1038/s41598-020-57750-z.

[2] R.R. Jensen, M.W. Binford, "Measurement and comparison of Leaf Area Index estimators derived from satellite remote sensing techniques," International Journal of Remote Sensing, **25**(20), 4251–4265, 2004, doi:10.1080/01431160410001680400.

[3] Y. Guangjian, H. Ronghai, L. Jinghui, W. Marie, J. Hailan, M. Xihan, "Agricultural and Forest Meteorology Review of indirect optical measurements of leaf area index : Recent advances , challenges , and perspectives," Agriculture and Forest Meteorology, **265**, 390–411, 2019, doi:10.1016/j.agrformet.2018.11.033.

[4] Q. Jiao, B. Zhang, J. Liu, L. Liu, "A novel two-step method for winter wheat-leaf chlorophyll content estimation using a hyperspectral vegetation index," International Journal of Remote Sensing, **35**(21), 7363–7375, 2014, doi:10.1080/2150704X.2014.968681.

[5] C. Tan, D. Wang, J. Zhou, Y. Du, M. Luo, Y. Zhang, "Remotely Assessing Fraction of Photosynthetically Active Radiation (FPAR) for Wheat Canopies Based on Hyperspectral Vegetation Indexes," Frontiers in Plant Science, **9:776**, 1–9, 2018, doi:10.3389/fpls.2018.00776.

[6] P. Taylor, D.C. Garson, B. Lacaze, "Monitoring Leaf Area Index of Mediterranean oak woodlands : Comparison of remotely-sensed estimates with simulations from an ecological process-based model," International Journal of Remote Sensing, **24**(17), 3441–3456, 2003, doi:10.1080/0143116021000024267.

[7] J. Qi, Y.H. Kerr, M.S. Moran, M. Weltz, A.R. Huete, S. Sorooshian, "Leaf Area Index Estimates Using Remotely Sensed Data and BRDF Models in a Semiarid Region," Remote Sensing of Environment, (73), 18–30, 2000, doi:10.1016/S0034-4257(99)00113-3.

[8] M.W. Kappas, P.A. Propastin, "Review of Available Products of Leaf Area Index and Their Suitability over the Formerly Soviet Central Asia," Journal of Sensors, 2012, doi:10.1155/2012/582159.

[9] Yang. C.M. and Su M.R., "Correlation of Spectral Reflectance to Growth in Rice Vegetation," in Asian Conference on Remote Sensing (ACRS), Proceeding of the 19th Asian Conference on Remote Sensing ACRS: 16-20 November 1998, Manila., 1998.

[10] I.A. Srinuandee Panithan, Honda Kiyoshi, Chemin Yann, "SWAP Crop Model Parameter Identification using SPOT Vegetation in Suphanburi, Thailand.," in Proceeding of the 25th Asian Conference on Remote Sensing, ACRS 2004, November 22-16, 2004, Chiang Mai, Thailand, 2004.

[11] T.N. Carlson, D.A. Riziley, J. Ji, "On the Relation between NDVI , Fractional Vegetation Cover , and Leaf Area Index," Remote Sensing of Environment, **62**, 241–252, 1997.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1054-1063 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Developing a Modular Material-Based Independent Training Model for Primary School Teacher Training

Asep Herry Hernawan[1,*], Mustari Bosra[2]

[1]*Department of Curriculum and Educational Technology, Fakultas Ilmu Pendidikan, Universitas Pendidikan Indonesia, 50604, Indonesia*

[2]*Department of History of Education, Fakultas Ilmu Sosial, Universitas Negeri Makassar, South Sulawesi, 50004, Indonesia*

ARTICLE INFO

Article history:
Received: 13 June, 2020
Accepted: 08 July, 2020
Online: 14 December, 2020

Keywords:
Technology for teaching
Primary education
Modular material
Teachers

ABSTRACT

This study is based on the challenges related to the low quality of education in Indonesia. One of the reasons for these challenges is the low ability of teachers to conduct professional tasks. Future demands for teachers' professionalization will be higher, considering the rapid development of society, science and technology, and current global competitiveness. In aspiring for improved quality education in the future, efforts are needed to optimize teachers' professional development, particularly for primary school teachers, through continuous and comprehensive education and training with broader access. Basing on this, the present paper analyzes a model and material of education and training programs which is believed to improve teachers' competence on a broader scope and ensures quality training outcomes. The research method study used was research and development. It has been used depending on the objectives to be achieved, which included developing a training model aimed at improving the mastery of cognitive abilities in the pedagogical competence of elementary school teachers. The preliminary study indicates that the implementation of the existing teacher education and training program is sufficient. However, further improvements are needed to reach a broader scope. Teachers' perceptions concerning their mastery of competencies required to perform their professional tasks indicated weaknesses regarding several aspects of the pedagogic and professional competencies. This exists because the of education and training teachers undergone which varies; yet, some teachers still have not had the opportunity to attain sufficient education and training. Identifying the needs of primary school teachers training and education program highlights the importance of developing a training and education model which optimizes modular material-based independent learning.

1. Introduction

The Indonesian Law No. 14 of 2005 [1] on Teachers and Lecturers requires teachers to have suitable academic qualifications, academic competence, teaching certificate, physical and mental health, and capability to realize the objectives of the national education system. The competencies a teacher must have, according to the law, include pedagogic competence, personal competence, social competence, and professional competence.

Possession of those four competencies is expected to facilitate a condition that allows teachers to perform excellently during work, serve their functions, and play their roles as accountable and professional educators. To ensure that teacher's services, tasks, and responsibilities as professional educators are implemented under the existing law. A Teacher's Performance Assessment System (or PK Guru) was developed and applied to every teacher in every formal education unit (schools) by the central government, regional government, or society. The result of Teacher's Performance Assessment (PK Guru) becomes the basis for determining the amount of credit obtained by the teacher for teacher's career development, as set in the Decree of the Minister of Administrative and Bureaucratic Reform No. 16 of 2009 [2] concerning Teachers' Functional Positions and Their Credit Values.

Then, the result of PK Guru is utilized to develop a profile of teacher's performance as an input for the obligatory Sustainable

*Corresponding Author: Asep Herry Hernawan, Universitas Pendidikan Indonesia (Indonesia University of Education), Street of Dr. Setiabudi, 229 Bandung, 40154, Phone (022)2013163, Email: asepery@yahoo.com / asepherry@upi.edu

Profession Development (PKB) program held for teachers in every academic year. This program intends to ensure that teachers are always able to satisfy the standard of competence and improve their competencies to anticipate future demands. This will make all teachers gain the appropriate image, respect, honor, and welfare because their competencies will promote the improvement and development of high-quality education services. The Sustainable Profession Development program is obligatory for teachers from the III/a level of position, in which they have to undergo self-development; teachers in the III/b level are required to publish academic and or innovative works. To move from IV/c level to the IV/d rank, teachers must do an academic presentation.

The Decree of the Minister of Administrative and Bureaucratic Reform No 16 of 2009 [2] mentions three elements of teacher's activities in Sustainable Profession Development (PKB) that earn them credit value, i.e., self-development, scientific or academic publication, and innovative work. Self-development is essentially an effort to improve teachers' skills and capabilities through, among others, functional training and education that may develop teachers' competencies and or profession. With such self-development, teachers will be able to perform their main and additional tasks and responsibilities. Functional training and education are some of the in-job training, implemented so that teachers can achieve the required competencies appropriate for their type and level of functional position (Government Regulation No 101 of 2000 concerning In-Job Training and Education for Civil Servants)[3]. The Minister of Education Regulation No 35 of 2010 [4] defines functional training and education as teachers' activities participating in education or training aimed to improve the teachers' profession in question over a specified period. As a consequence of this policy, all future teachers who have to develop their professional careers must have physical evidence of participating in Functional Education and Training.

The data from the Directorate of Educators (PMPTK) in the Ministry of Education shows that the Ministry is responsible for 2,607,311 educators from all levels of education in Indonesia. With this large number of teachers, it is impossible to provide Functional Training and Education for all of them in the current conventional way. An alternative, more progressive form of training, with broader access while maintaining a high level of quality, is needed so that all teachers can participate in functional training without leaving their places of work and their primary responsibilities at schools.

One of the alternatives to conduct such training and education for the development of teachers' competencies is by utilizing the concept of independent learning using modular material. Therefore, the study and development of a modular material-based independent training model are needed to promote sustainable professional teacher development functional training.

The professional career improvement of future teachers is highly dependent on the result of teachers' performance assessments (PK Guru), which require sustainable professional development (PKB) through in-job functional training. Several problems that often hinder the implementation of teachers' in-job functional training are, among others: (a) the materials provided in training are not the ones the teachers need; (b) the trainers often do not have more qualified knowledge and are not more experienced than the trainees; (c) the training is commonly held in the effective working hours when teachers are supposed to teach; (d) teachers generally view training as merely a requirement to have an efficiency report; and (e) the training organizers are usually too project-oriented, which means that they often focus only on the delivery of the training, not the quality. Another problem concerning teacher training is that the training is usually delivered in lecture form, which tends to provide a comprehensive understanding of the competencies needed by the teachers. These problems lead to the low quality of teachers' performances and, in turn, affect the result of students' learning.

Many aspects of the main competencies are understood incomprehensively by the teachers. Besides, teachers do not have enough opportunity to develop their understanding while performing their jobs. The training that teachers have participated in intend to be short and not optimal, as well as focus directly on practical skills only, leaving the other aspects untouched.

Identifying these problems warrant a scientific and empirical study to deal with the obstacles in teachers' competencies development through various alternatives embedded in the teachers' training model. In general, this study aims to answer the main problems of "what kind of model and training material is effective in developing teachers' competencies in a wider scope, with quality, and without requiring teachers to leave their primary jobs?" The questions to be studied to answer the main research problem include: (1) How are the implementation of the existing teacher training, the type of training, and the teachers' mastery of competencies? (2) What training model and material design can be developed to improve the ability of teachers' competencies? (3) How is the effectiveness of the implementation of a teacher training model using a modular material-based learning approach?

2. Literature Review

To be a teacher is to be a professional, meaning that the duties and responsibilities therein cannot be carried out effectively without appropriate training. A profession is a job or activity carried out by a person as a source of income that requires expertise and skills that meet specific quality standards or norms and require professional education and training. In-Law Number 14 of 2005 [1] on Teachers and Lecturers, it is stated that a teacher is a professional educator whose primary task is to educate, teach, guide, direct, train, and evaluate students through formal education in the levels of early childhood education, primary education, and middle education. However, according to Supriadi [5], the teacher's position is still a growing / emerging profession and has not yet reached the level of being a profession in the real sense.

Professional teachers must have various abilities. Sukmadinata [6] groups these abilities into three general dimensions: professional, social, and personal. Supriadi [5] assesses that capability is too ideal, so it is not easy to achieve and be evaluated by measurable criteria. Various training is needed to improve the ability of teachers. The real challenges facing the world of teacher training in the present and future according to Hasan (2001) are mainly related to improving the quality of teachers in terms of (1) mastery of subject matter being taught, (2) teaching ability, (3) vision and attitude towards the profession, (4)

ability to develop a profession, (5) ability to communicate with educators, experts, and the community, (6) community appreciation of the teaching profession, and (7) high competitiveness and professionalism. According to Evans and Brueckner [7], to become a qualified and professional teacher is strongly influenced by (a) professional preparation, (b) continuing education, (c) professional involvement, and (d) commitment.

Teacher training is being carried out by various parties, both government and private; it is considered as an effort to develop a systematic attitude, knowledge, skills, and behavior patterns needed by someone to have the ability to perform the task appropriately. It is also affirmed by Bramley [8] that "Training is a systematic development of the attitude, knowledge, skills and behavior pattern required by an individual to perform a given task or job adequately."

In training teachers, adult education methods (andragogy) must be applied. According to Abdulhak [9], the scope of the technique in adult learning involves: giving encouragement, revealing the growth of learning interest, delivering learning materials, creating a conducive learning climate, stimulating creativity, encouraging self-assessment, and identification and tackling weaknesses in learning outcomes. Considering that trainee teachers are adults, a methodological study of how to conduct adult training is more effective in optimizing intended outcomes.

Several approaches in training programs that apply self-instruction have received considerable attention in adult education [10]. These approaches include programmed instruction, individualized instruction, personalized system of teaching, learner-controlled instruction, correspondence study, and self-study.

The self-instruction approach in teacher training is related to the application of instructional technology concepts. According to Miarso [11], this instructional technology uses an isomorphic approach, which combines two complex concepts, technology, and instruction, in an integrated concept. Both academics and practitioners have now developed the science of learning technology in developed countries, especially the United States. The emergence of a book called "Instructional Technology: The Definition and Domains of the Field" is one proof of the development of scientific learning technology. Instructional technology "is defined as" the theory and practice of design, development, utilization, management, and evaluation of processes and resources for learning [12]." The application of learning technology in teacher training, according to Mukminan [13], comprises four primary characteristics: applying a systems approach, using learning resources as broadly as possible, and improving the quality of human learning, and orientation towards individual learning activities.

Programmed learning is a form of curriculum implementation in view of technological education. Lapp [14] notes that there are several advantages of this programmed learning: (a) providing positive immediate feedback and reinforcement for students (immediate positive feedback and support for the student), (b) releasing the teacher from "drill" type teaching, and (c) individual student progressions. The type of programmed learning developed by Crowder is better than that established by Skinner; however, it would be even more perfect if modified with other methods, in this case, with the tutorial learning method. Kemp and Dayton [15] put it that the "tutorial method attempts to emulate a human tutor. Instructions are provided via text or graphics on a screen. At appropriate points, a question or problem is posed. If the student's response is correct, the computer moves on to the next block of instructions. If the response is incorrect, the computer may recycle to the previous instruction or move to one of several sets of remedial instructions depending upon the nature of the error."

3. Methodology

The research method in this study uses research and development methods. It is following the objectives to be achieved that include developing a training model aimed at improving the mastery of cognitive abilities in the pedagogical competence of elementary school teachers. Research and Development (R&D) is essentially a process used to develop and validate learning products or prototypes of research-based learning tools [16]. The outcome in this research is a training development model that uses an independent learning approach to improve cognitive abilities in elementary school teachers. Research and development activities in this study are operationally carried out through three stages: the preliminary investigation, the development of a training model, and the validation test of the model. The detailed research and development procedures are described in Figure 1.

The study subjects were managers of the institutions/agencies that were given the task of organizing training of elementary school teachers in West Java Province and several elementary school teachers in cities across West Java province. Data collection techniques and instruments for gathering information on research points are adjusted to the three stages of the research activities. At the field study stage, the questionnaire was used as the data collection technique; in the limited trial stage, questionnaires and interview guidelines were used for the broader trial phase in addition to using questionnaires, and interview guidelines were used as comprehension test instruments. Finally, at the model validation test, the comprehension test instrument was used.

Data analysis techniques used were adapted to the stages of research and development carried out, the pre-survey, testing the application of the model, and testing the validation of the training model.

4. Results/Findings

4.1. Preliminary Study Results

The pre-survey results from the preliminary study were taken into consideration in developing the draft model of elementary school teacher training to address the problem in the focus of this study. The field pre-survey was conducted through interviews with the leadership of teacher training institutions. Information gathering was by questionnaire to elementary school teachers in the study sample in several districts/cities in West Java. Table 1 below sums up the results of the survey i56t7n teacher training institutions concerning the organization of teacher training.

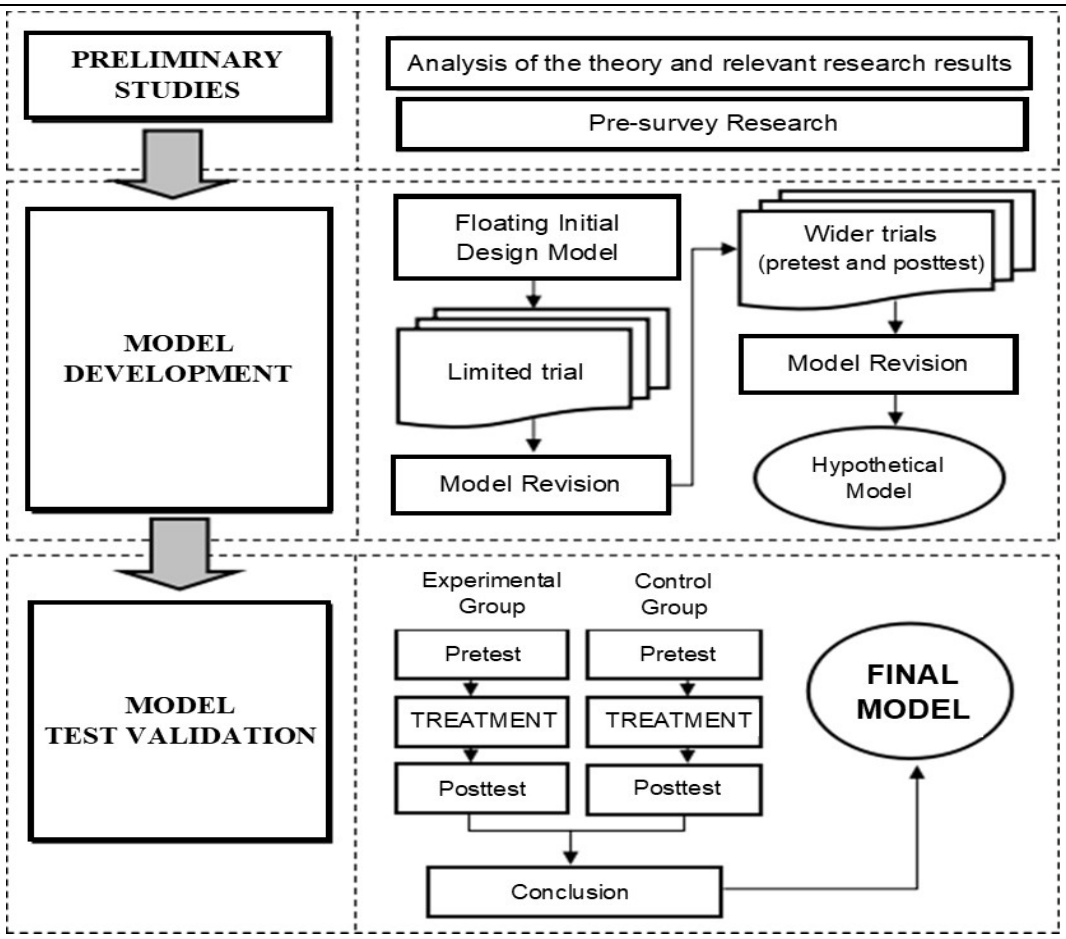


Figure 1: Research and Development Procedure

The following description is a synthesis of data obtained from the results of a survey of teachers as the basis for developing a modular model of teaching materials in elementary school teacher training. The study results on mastery of subject matter and pedagogical competencies can be seen in tables 2 and 3.

In table 2, almost all teachers stated that they had mastered teaching materials in the five primary subjects that were their responsibility. Only a small percentage of teachers still said they did not learn the teaching materials. There are no teachers who claim to be very ignorant of teaching material. However, there are only a few teachers who claim to be proficient in teaching material. With this description, according to perception, teachers, in general, have mastered the material in five elementary subjects.

If observed in table 3, it turns out that in general, teachers say that they have mastered the pedagogical competencies needed in elementary schools. A small proportion of teachers said that they had learned them. There are some interesting notes from the data above, where for the mastery of teachers relating to how to communicate with students, it seems that as many as 70% of teachers have reached the level of proficiency to master it.

However, in three cases it was revealed that there were still teachers who did not learn the necessary educational / teacher knowledge, namely relating to (1) how to develop curriculum, (2) how to utilize information and communication technology, and (3) how to carry out thematic learning.

In table 2, almost all teachers stated that they had mastered teaching materials in the five primary subjects that were their responsibility. Only a small percentage of teachers still said they did not learn the teaching materials. There are no teachers who claim to be very ignorant of teaching material. However, there are only a few teachers who claim to be proficient in teaching material. With this description, according to perception, teachers, in general, have mastered the material in five elementary subjects.

Table 1: Results of a survey on the implementation of elementary school teacher training

Teacher training aspects	Conclusion of survey results
1. Teacher training needs assessment	The needs assessment results are rarely utilized in determining and sending prospective trainees following the training institution's requirements.
2. Teacher training programs and materials	Teacher training programs and materials are developed at the level of the training institution based on field needs analysis results. The focus of the

	training program is on the development of subjects taught in primary schools.
3. Learning methods in teacher training	The training method applies the adult learning approach (andragogy) by optimizing the use of lecture and question and answer methods, simulation methods, problem-solving, practical activities, and field visits or practice.
4. Teacher training instructor qualifications and competencies	Teacher training instructor qualifications are determined based on competency or mastery of subjects, experience, and educational background following the items. The appointment of instructors is determined by the directorate and or the teacher training institution.
5. Recruitment of teacher training participants	Teacher training participants' recruitment is determined based on (a) direct appointment from the ordering party (directorate) and (b) based on training need assessment, although sometimes the local government with its authority determines itself without clear criteria.
6. Assessment of the impact of teacher training	Impact assessments are carried out by monitoring the progress of the participants, both in the form of visits, correspondence, questionnaires, and direct observation of schools.
7. Teacher training quality assurance program	Teacher training institutions do not explicitly develop quality assurance programs. It is usually done based on a review of various activities that are programmed.

Table 2: Levels of Teacher Mastery of Subject Materials

Subject	Teacher Mastery Level									
	STM		TM		CM		M		SM	
	f	%	f	%	f	%	f	%	f	%
PPKN	0	0	0	0	29	41.43	41	58.57	0	0
Religious Education	1	1.43	13	18.57	40	57.14	13	18.57	3	4.29
Indonesian	0	0	3	4.29	31	44.29	36	51.43	0	0
Mathematics	0	0	2	2.86	39	55.71	29	41.43	0	0
Science	0	0	0	0	42	60	28	40	0	0
Social Studies	0	0	6	8.57	40	57.14	24	34.29	0	0
Art and Crafts	0	0	10	14.29	42	60	18	25.71	0	0
Physical Education	5	7.14	21	30	37	52.86	7	10	0	0
Local content	5	7.14	5	7.14	40	57.14	20	28.57	0	0

Note: STM (very not mastered), TM (not mastering), CM (quite mastering), M (mastering), SM (very mastering).

Table 3: Mastery Competency Level of Elementary School Teacher Competence

Teacher's Pedagogical Competencies	Teacher Mastery Level									
	STM		TM		CM		M		SM	
	f	%	f	%	f	%	f	%	f	%
Educational foundation	0	0	7	10	40	57.14	23	32.86	0	0
Preparation of teaching plans	0	0	2	2.86	32	45.71	36	51.43	0	0
How to manage classes	0	0	1	1.43	30	42.86	39	55.71	0	0
Using media/learning resources	0	0	4	5.71	32	45.71	34	48.57	0	0
Manage Teaching and Learning Interactions	0	0	3	4.29	38	54.29	29	41.43	0	0
Assess Learning achievements	0	0	3	4.29	24	34.29	43	61.43	0	0
Counseling functions and services	0	0	6	8.57	34	48.57	30	42.86	0	0
School administration	0	0	7	10	34	48.57	29	41.43	0	0
Class Research	0	0	10	14.29	29	41.43	31	44.29	0	0

Note: STM (still not mastered), TM (not mastered), CM (quite mastered), M (mastered), SM (highly mastered).

If observed in table 3, it turns out that in general, teachers say that they have mastered the pedagogical competencies needed in elementary schools. A small proportion of teachers said that they had learned them. There are some interesting notes from the data above, where for the mastery of teachers relating to how to communicate with students, it seems that as many as 70% of teachers have reached the level of proficiency to master it. However, in three cases it was revealed that there were still teachers who did not learn the necessary educational / teacher knowledge, namely relating to (1) how to develop curriculum, (2) how to utilize information and communication technology, and (3) how to carry out thematic learning.

Based on the results of capturing information from teachers, information/data obtained about the frequency of training activities that have been attended by elementary school teachers are closely related to the main task area as a teacher. There is a tendency in the past five years. The frequency of teachers attending training is generally only 1-3 times, and only a small proportion of teachers participated in more than four training activities. There are still many teachers who have not attended training in the past five years (17.52% of 304 respondents). Based on the data obtained about the teacher's assessment of the training activities that he had participated in, there were still many teachers who stated that the training was unclear, for example relating to training on process evaluation techniques and the results of instruction used.

As for the expectations of teachers in the implementation of teacher training in the future, including:

- The existence of training material packages that have the characteristics of 1) using language adapted to the features of the participants, 2) optimizing the use of supporting media, 3) equipped with pictures and illustrations that are appropriate and interesting, 4) containing statements that are packaged in the form of disclosure of teacher problems, 5) the facts that are raised are actual, and 6) some questions are directed to solve the problem.
- Training places, in the opinion of some teachers, should not need an exceptional location, so teachers do not need to come to the training venue and do not need to leave their primary tasks. The method of training is to use material packages that can be done alone. Thus, in such training activities, there just a few instructors/trainers, but with expertise in their fields.
- According to some teachers, training time could be at any time as long as the form of training uses material packages and individualized instruction approaches.

5. Model Generated

Based on the pre-survey results, it was found that there were some loopholes the learning system implemented in elementary school teacher training activities at the time. Thus, a teacher training model that is more individualized and requires training participants to learn actively with the following characteristics is essential : (1) flexible in the sense of not needing a particular place, (2) opportunity and time are not so binding on the teacher, (3) learning adapted to the ability of learner and teacher, (4) broad-reaching and able to overcome geographical obstacles, (5) relatively cheaper training costs (on a large scale), (6) teachers do not need to come to the training site and do not need to leave their main task at school, (7) able to overcome the limitations or availability of qualified instructors, (8) standardization of teacher quality; and (9) affect the independence, responsibility, discipline, tenacity, initiative, and creativity of teachers in carrying out their duties. The teacher training model's characteristics can be accommodated in training models that use programmable modular training materials.

The model design that resulted from this study was in the form of independent programmatic learning developed in elementary school teacher training, including planning design, implementation design, and assessment design, as illustrated in Figure 2.

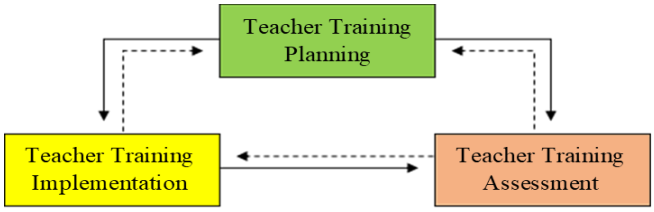


Figure 2: Teacher Training Development Model Design

5.1. Training Planning Design

The planning activity functions as a blueprint, providing an overview of what to do in teacher training activities. This model is an application of the concept of instructional technology in which the learning planning stage applies an approach that emphasizes systemic connectedness between various learning components. In this case, a systemic relationship means that all components in planning are integrated following their functions, related to one another, and form a unity. The training planning component consists of the objectives or competencies to be achieved, learning material, learning activities, and learning assessment. The training planning stage is illustrated in Figure 3.

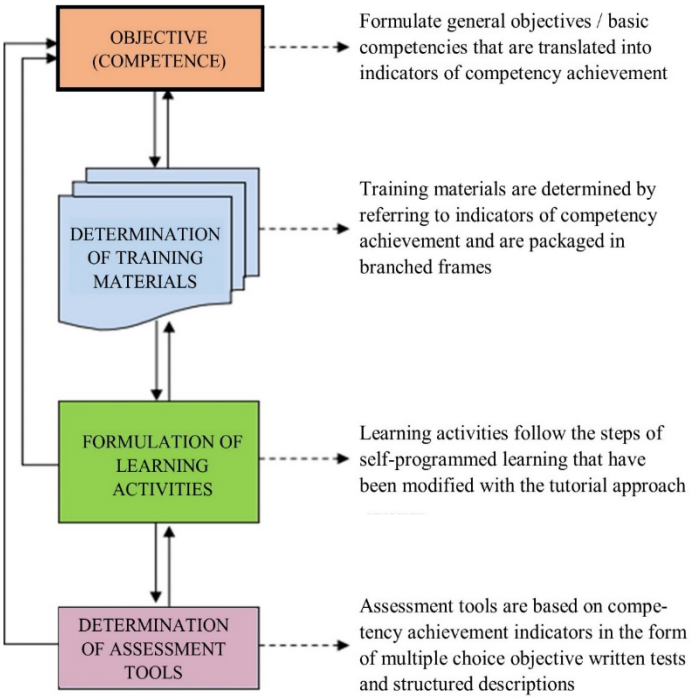


Figure 3: Teacher Training Planning Design

5.2. Training Implementation Design

The design of the implementation of learning developed in teacher training uses a self-instructional approach in which individual trainees carry out learning activities following prescribed instructions and procedures. The trainees actively interact with the training materials that have been packaged in such a way by reading instructions and presenting information, responding to questions, and seeing feedback on the alternative answers they choose. With this feedback, trainees without delay can determine whether the response given is correct or not, so mistakes can be corrected immediately. At the end of the activity, the trainees work on the tests prepared about the training objectives.

The amount of time spent by each participant in carrying out the training process may vary considerably. In this case, the trainees complete the learning at their own pace. Some participants can quickly complete, while slow participants can still finish even though not too fast. Adjustments are made in line with the principles of mastery learning.

The design of the programmed self-learning implementation developed as a modification of the model developed by Norman A. Crowder in 1962 called the Crowder's Intrinsic Program in the branching frame format. The characteristics of this type include: (a) broader and more information, (b) questions that are accompanied by several possible answers that can be chosen, and (c) checks that are accompanied by an explanation that explains why the answer/response true or false. Modifications to the learning model intended in this study emphasize improving the components or steps of the learning. The learning step applies the tutorial learning procedure developed by Allesi and Trollip [17] with the following eight stages: 1) Introduction, 2) Presentation of information, 3) Question of responses, 4) Judging responses, 5) Providing feedback about reactions, 6) Remediation, 7) Sequencing lesson segments, and 8) Closing. The procedure of implementing the self-learning program that must be taken by the trainees in this study follows the learning steps, as illustrated in Figure 4.

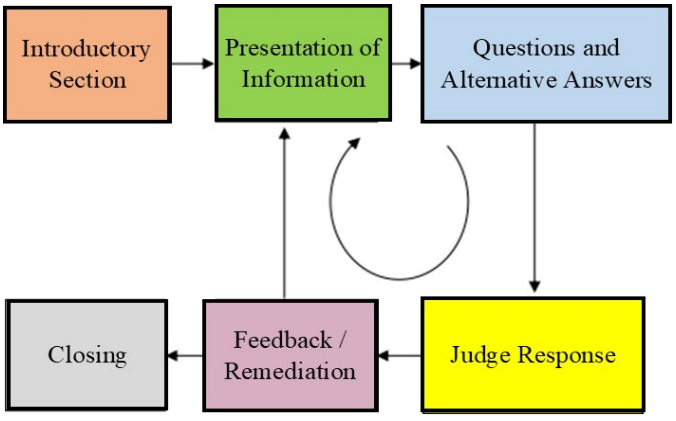


Figure 4: Design of Teacher Training Implementation

5.3. Teacher Training Assessment Design

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

The assessment design used is guided by the learning planning developed in which the trainees must achieve some objectives/competencies. The characteristics of this model emphasize the ability and speed of learning of each trainee. Therefore, the assessment design used is more individualized and directed primarily to measure the cognitive abilities of trainees. The assessment takes the pretest and post-test procedures. The pretest determines the extent of training participants' abilities before studying the training material, while the final test determines the participants' abilities after investigating the training material. If the pretest demonstrates the trainees' mastery of the content presented, then there are several possibilities. For trainees who have mastered the entire learning material, there is no need to study the material. For trainees who have mastered some or part of the learning material, these participants do not need to learn it from the beginning. In this case, because the abilities of each participant differ, it is necessary to determine carefully, which participants must start from the beginning, which participants can begin in the middle, and so on.

The pretest (TA) and the pot-test (TF) in this model can be carried out with two variations. The first variation is when the two types of tests are carried out in each piece of training material (SM). In the second variation, the pretest is done before the trainee learns the presentation of the learning material. The post-test is done after the trainee completes the entire performance of the learning material. The type and form of assessment used are appropriate in measuring competence in cognitive aspects. Thus, the type of evaluation used is with an objective test and or structured description. The design of teacher training assessments can be seen in Figure 5.

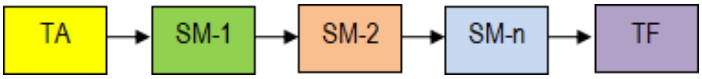


Figure 5: Teacher Training Assessment Design

5.4. The Effectiveness of the Teacher training Model

This teacher training model's effectiveness has been proven by a significant effect on improving cognitive abilities in pedagogical competence compared to conventional training models that are generally used today. To find out the description of the validity of the pretest and post-test treatment between the control group (trainees treated through the presentation with conventional procedures) and the experimental group (the trainees are given a treatment of serving with programmed modular teaching materials), the null hypothesis (Ho) proposed "there is no significant difference in ability between the control group post-test outcome variable and the experimental group post-test outcome variable." In contrast, the alternative hypothesis (Hi) proposed is that "there is a significant ability difference between the control group post-test result variable and the post-test result variable of the experimental group. "Acceptance of the null hypothesis (Ho) is done if the results of the calculation of significance are greater than 0.05 (> 0.05), and Acceptance of alternative hypotheses (Hi) is done if the results of the calculation of significance are less than 0.05 (<0.05). The calculation of the significance of the results of this post-test can be seen in table 3.

Table 3: Results of post-test calculation of control and experimental validation tests

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min.	Max.
					Lower Bound	Upper Bound		
POSTTEST CONTROL	36	7.2769	0.65877	0.10979	7.0540	7.4998	6.27	8.50
POSTTEST EXPRIMENT	36	8.1944	0.66845	0.11141	7.9683	8.4206	7.00	9.00
Total	72	7.7357	0.80474	0.09484	7.5466	7.9248	6.27	9.00

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	15.153	1	15.153	34.406	0.000
Within Groups	30.828	70	0.440		
Total	45.981	71			

The significance calculation results are equal to 0.000, and F test results are 34.406. With the results of these calculations, the null hypothesis (Ho) is rejected. The alternative hypothesis (Hi) is accepted, meaning there is a significant difference between the control group's post-test outcome and the experimental group's post-test outcome variable. In this case, it can be concluded that the trainees who were given the treatment using the training development model using a modular programmatic model have higher abilities compared to the ability of trainees treated with ordinary procedures (conventional models).

The test results obtained instructions that the developed training material package can meet the requirements as an independent learning material that allows self-instruction. In reality, the use of training development models that are self-instruction is still a new thing that is considered relatively difficult to implement. Thus, an adaptation process needs sufficient time to provide awareness to the teacher training providers that training, which is more focused on acquiring cognitive abilities, does not always have to be done through face-to-face meetings with training instructors with a tight schedule and venue for training.

6. Discussion

Several assumptions underlying the choice of the teacher training model in this study include the following:

As a whole individual, the teacher inherently possesses self-learning capability by optimally utilizing his/her potentials. With sufficient time and opportunity, every teacher has the same chance to understand and master the knowledge and skills needed to perform his professional tasks.

Teachers are adults; therefore, the training model or approach should be relevant and based on adult learning theories and principles (andragogy learning). The implementation of an independent training model is essentially an implementation of adult learning theories and principles.

The self-instructional approach utilized in professional training is a better approach to develop human resources. It allows the trainees to self-instruct according to their own time and capabilities.

Learning with modular materials is a learning method or technology based on a system approach. It is primarily characterized by the progressive development of ideas to achieve the desired behavior. This approach's other characteristics include continuous evaluation, verification, and revision, varied and flexible delivery of easy-to-learn materials, and logistically easy to administer. Modular elements are equipped with various tasks and learning models, as well as continuous feedback. Slow learners will not lose their opportunity or be left behind, even if the materials get more complex and sophisticated.

6.1. Contribution of the Study to Teacher Training

In general, this study aims to analyze and develop a model for training primary school teachers using modular material-based learning. With new policies concerning the assessment of teachers' performance (PK Guru) and the sustainable profession development (PKB) and considering the vast number of teachers in Indonesia; makes it is impossible to satisfy teachers' needs for functional training using the existing conventional method. The development of a new model of teacher training is justified.

This study's significance and urgency on primary school teachers training using modular material-based learning are as follows:

The training model is based on the individual teacher's potential and skills. Thus, it will theoretically provide benefits for the development of future primary school teachers' training models.

This study's findings are expected to provide new principles for primary school teachers' training, based on the application of instructional technology concepts. These principles are expected to strengthen the benefit of self-instructional concepts in teacher training, which will help develop primary school teachers' competencies through maximizing independent learning without face-to-face learning.

This study's findings can be used to develop a model for initial teacher training in the future, mainly to develop the training system and materials. What is more, this training model will help improve teachers' mastery of their competencies because its learning is adjusted to individual learning characteristics and pace. They can learn the materials in training based on their capabilities, opportunities, and time.

For teachers living and working in remote areas, this model will free them from the geographical limitations to participate in competencies development training. Those teachers do not need to come to the training venue and do not have to leave their jobs at the schools.

For the primary school teachers training administrators, both at the regional and national level, this training model will go a long way in addressing various issues hindering the efforts of improving teachers' competencies and qualifications.

The dissemination and dissemination of new information or knowledge to develop teachers' competencies and improve teachers' professionalism and quality can be conducted on a broader scope. This model can reach more teachers, which will accelerate the development and equalization of primary school teachers' quality.

6.2. Output

The challenges and obstacles in future education, particularly at the primary school level, are expected to be more complicated than today. Therefore, it is essential to prepare teachers who are competent and able to understand their professional function and responsibilities to realize a high-quality education process. The targets to be achieved in this study are:

The profile of current primary school teachers training implementation, the types of training the primary school teachers, have participated in, and the level of teachers' competencies mastery.

The training model to improve primary school teachers' competency mastery consisted of the model of learning design, the model of learning implementation, and the learning assessment model.

Modular materials developed and used in the training model the dual-mode system.

Data and information concerning the effectiveness of the implementation of primary school teachers training model using the modular material-based approach, both in the aspects of comprehension and application.

The administration or implementation of primary school teachers' training focuses on the following aspects: Need assessment to compile data concerning various elements needed in teachers' training, supported by the result of teachers' competencies mapping as a measure of primary school teachers' competencies after the training.

The training program and materials are based on various aspects. The development of school subjects needs to focus on integrating those subjects with other essential skills. The learning method developed in primary school teachers' training should be supported by a more robust approach, focusing on adult learning principles.

Primary school teachers training in an uncoordinated training institution needs more proper management, including the recruitment and assignment of qualified instructors. The qualification of training instructors varies based on their competence/mastery of materials, experience, and educational background.

Training activities that can facilitate the improvement of competencies for numerous teachers require a significant amount of money. Teachers training with an independent learning approach can be a solution to widen access for teachers to participate in competency development training while reducing the costs.

Cognitive competencies of primary school teachers, including the mastery of the five primary subjects and educational/teaching discourses, are not comprehensively mastered by primary school teachers. This situation indicates the importance of a training system that develops teachers' motivation to learn those competencies according to their potentials. Decentralized training hinders teachers' mastery of competencies because they limit teachers' frequency and opportunity to participate and their time to master the competencies they need.

7. Conclusion

The implementation of primary school teachers' training has generally reflected the ideal condition of the efforts to improve teachers' competency. Further development to equalize the training opportunity for teachers is still needed. The obstacles to equalize the training and development opportunity include the discrepancy between the available training facilities and frequency of training with the number of primary school teachers whose competencies need to be improved.

The development of teacher training models should represent the interrelated nature of design, implementation, and assessment aspects as a whole system to improve teachers' cognitive, pedagogic, and professional competencies mastery. This training design utilizes a systemic approach that focuses on the interrelationship among the competencies to master, the training materials, the learning activities, and the learning evaluation. The designed training requires the trainees to perform the events based on defined procedures individually. The training design is more individual-oriented and aims at measuring teachers' cognitive skills in the pedagogic and professional competencies.

Ethical Statement: There is no conflict of interest. The author followed the right academic procedures. There is no human specimen used requiring a special declaration.

References

[1] Undang-Undang Republik Indonesia Nomor 14 Tahun tentang Guru dan Dosen, 2005.

[2] Peraturan Menteri Negara Pendayaan Aparatur Negara dan Reformasi Birokrasi Nomor 16 Tahun tentang Jabatan Fungsional Guru dan Angka Kreditnya, 2009.

[3] Peraturan Pemerintah Nomor 101 Tahun tentang Pendidikan dan Pelatihan Jabatan Pegawai Negeri Sipil, 2000.

[4] Peraturan Menteri Pendidikan Nasional Nomor 35 Tahun tentang Petunjuk Teknis Pelaksanaan Jabatan Fungsional Guru dan Angka Kreditnya, 2010.
[5] D. Supriadi, "Mengangkat Citra dan Martabat Guru. Yogyakarta: Adicita Karya Nusa, 1999.
[6] N.S. Sukmadinata "Pengembangan kurikulum. Bandung: Remaja, 1997
[7] J.M. Evans, M.M. Brueckner, "Teaching and You, Committing, Preparing, and Succeding. Boston: Allyn and Bacon, 1992
[8] P. Bramley, Evaluating Taining Effectiveness: Translating Theory into Practices. New York: Mc Graw-Hill Book Compani, 1991.
[9] I. Abdulhak, "Strategi membangun motivasi dalam pembelajaran orang dewasa. Bandung: CV. Andira, 2000.
[10] Y. Miarso, Teknologi pendidikan menyongsong abad 21. Makalah pada rapat koordinasi Pustekkom Depdikbud, Bandung, 1997.
[11] R.L. Craig, Training and Development Handbook. New York: Mc.Graw Hill Book Company, 1987.
[12] B.B. Seels, R. Richey, Instructional technology : The definition and Domains of the Field, AECT, Washington DC, 1994.
[13] Mukminan, Desain Pembelajaran. Yogyakarta: Universitas Negeri Yogyakarta, 2006.
[14] D. Lapp at al. Teaching and Learning (Philosophical, Psychological, Curricular Application, New York: Macmillan Publishing Co., Inc, 1975.
[15] J.E. Kemp, D.K. Dayton, "Planning and Producing Instructional Media. New York: Happer and Row Publisher, 1985.
[16] W.R. Borg, "Educational research: an introduction. New York: Longman Inc. 1989.
[17] S.M. Alessi, S.R. Trollip, Multimedia for learning: Methods and development. Allyn & Bacon, 2001.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 1064-1070 (2020)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

Projection of Wireless Multipath Clusters Using Multi-Dimensional Visualization Techniques

Aaron Don M. Africa[*,1], Emmanuel T. Trinidad[1,2], Lawrence Materum[1,3]

[1]*Department of Electronics and Communications Engineering, De La Salle University, 2401 Taft Ave., Manila, 1004, Philippines*

[2]*Don Honorio Ventura State University, Villa de Bacolor, Pampanga, 2001, Philippines*

[3]*Tokyo City University, 1-28-1 Tamazutsumi, Setagaya, Tokyo, 158-8557, Japan*

ARTICLE INFO

Article history:
Received: 01 September, 2020
Accepted: 26 November, 2020
Online: 14 December, 2020

Keywords:
Data Visualization
Channel Modeling
Multipath Clustering
Wireless Communications

ABSTRACT

Advances in channel modeling allow wireless communication designers to accurately model and understand the channel's phenomena within different propagation scenarios. A precise channel model results in the wireless system's optimized performance while considering trade-offs due to the effects of the channel. The geometric-based stochastic channel model considers different interacting objects affecting the parameters using the indicated parameters measured, such as the delay, power profile, and angles of departure and arrival in MIMO wireless systems to multiple paths taken by the propagating signal. These different paths are known as multipath components (MPCs). Studies and measurements suggest that MPC appears in groups based on similarity or dissimilarity measures of the MPCs known as multipath clusters. The clustering of these MPCs can be done automatically and manually. The automatic approach clusters the measured or simulated data using an algorithm; on the other hand, the manual approach uses a visual representation and an expert evaluation to cluster the data. This paper aims to implement visual mapping of the multipath cluster dataset approach into parallel coordinates plot, t-SNE, principal component analysis, heatmaps, and dendrograms to find an optimal visualization of data to enhance the validation and interpretation of MPC further. The dataset is simulated and presented in a MATLAB environment.

1 Introduction

Wireless communication technology grows exponentially in connected devices and the data required in the modern information era [1]. The wireless communication system can be simplified to three main components, the transmitter, the channel, and the receiver. Both the transmitter and the receiver can be optimized based on the required use case. However, the channel is the challenging component since it varies in the environment and channel bands in which the channel cannot be engineered and manipulated easily [2]. System designers address the channel impairments through statistical modeling by acquiring parameters and effects to form the channel impulse response (CIR). Channel modeling is a broad research area that can be seen through literature, and these channel models account for different wireless propagation scenarios.

The development of the fifth-generation (5G) mobile communication standards takes into accounts the use cases with three major requirements [3, 4]. These standards require the wireless community to develop new enabling techniques and technologies to satisfy the stringent standards to provide the Quality of Service that the 5G demands. One key enabling technology is based on massive multiple-input multiple-output (MIMO) systems. Such a system has n number of antennas that are utilized at the transmitter and/or the receiver to provide trade-offs based on the use case as reliability, sensitivity, and high throughput. MIMO systems send out a multiplicity of signals going to the receiver. As these signals travel, different paths are taken by the signals known as multipaths that can create constructive or destructive combinations at the receiver. Multipath components (MPC) exhibit different parameters due to the channel state, where multiple objects that obstruct the signals can be reflected or dispersed [5]. The geometric based stochastic channel models account for these objects collectively known as scatterers. The MPC measured in different studies proves that MPC tends to cluster based on their similarity of parameters; furthermore, the clustering of these MPCs provides simple calculation while maintaining the channel model's accuracy, which can adequately measure the channel capacity and behavior.

Clustering has benefited different disciplines in analyzing mea-

[*]Corresponding Author: Aaron Don M. Africa, De La Salle University, 2401 Taft Ave., Malate, Manila, 1004, Philippines, +63(2)85244611 & aaron.africa@dlsu.edu.ph

sured data. It can also predict the future values or categories used in machine learning to analyze the data with multiple features or dimensions to be measured. Clustering measurement data also benefit the wireless channel modeling field [6]. Validating the clusters poses a challenge due to the limitation of the human eye's visual perception, especially for large datasets that crowds the common visual representation such as the 2 and 3 dimensional (2D, 3D) scatterplots. Therefore, techniques have been developed to mitigate these limitations by using dimensionality reduction techniques to embed the D- dimensional data sets to a 2D and 3D where it can be visually interpreted and analyzed.

The traditional identification of clusters was made manually through the visual analysis of the data, which provides a physical realism of the environment. However, the manual approach can be laborious due to large data sets to be clustered. This approach has changed through the development of cluster algorithms. The automated method of clustering is done through the power of computational algorithms. The measured data are fed to an algorithm and computes the clusters within the measured data, and the validity is tested using different Clustering Validity Index (CVI). However, the channel's physical realism does not account for automated clustering and relies on numerical computations [7]. Furthermore, different algorithms tend to focus on specific parameters that lead to unaccounted features of the MPC.

This study aims to project the MPC data using visual techniques to validate the cluster using visual representation, thus harvesting both the advantages of the manual and the automated approaches of clustering. The graphical representation aids the validation and interpretation of data by drawing insights and increase cognition of the wireless propagation clustered data. Proper evaluation of clusters of MPC leads to an accurate representation of the channel, thus enhancing the design and implementation of the wireless network within a given environment and specific propagation scenario.

This paper aims to map the wireless multipath cluster dataset to the state-of-the-art visualization techniques and impressively represent and analyze MPC clusters. The remaining parts of this paper are organized as follows: Section 2 presents related works for improving MPC clustering and studies in other disciplines that employ data visualization. Section 3 discusses the propagation channel reference, clustering, and data visualization principles with the methodology used in this study. Section 4 provides a comprehensive discussion of the data visualization techniques used in this work. Section 5 summarizes and analyses the results with future considerations, followed by Section 6, which concludes this work.

2 Literature Review

2.1 Multipath Clustering

Multipath clustering has been unfavorable in the past [2]; due to the algorithmic development of clustering, novel algorithms are used to determine the clusters in various propagation scenarios. This section provides a review of the literature that scopes manual and automated clustering approaches.

One of the earliest and popular indoor channel models is the Saleh-Valenzuela model [8]. It describes the time of arrival of the MPC in a 1.5 GHz indoor channel where a cluster of the measured rays is due to objects surrounding the transmitter and receiver. This model is for the Single input single output system (SISO). The author in [9], used a manual clustering approach in an indoor scenario in the MIMO 5 GHz band and identified 4 to 5 clusters in an area. In [10], the authors presented a framework for automating the clustering process that outperforms traditional visual identification clusters. The study was placed in an indoor environment and utilized the k-Means algorithm. The study also proposed the use of Multipath Component Distance (MCD), enhancing the clustering algorithm's performance. Simulation performed by Arias [11] used a cluster-based channel focusing on the angle of arrival (AoA) and the power delay profile (PDP) where the probability density function of the clusters was used to derive these parameters. The work presented in [7], the author used a middle-ground approach utilizing automatic and manual clustering to validate each of the results. The representation of data produced by automatic clustering poses crowding and overlapping data points in the scatterplot. This study was able to identify interacting objects along the wireless channel, thus real objects associated with the clusters in the channel. A proposed a framework in [12] focused on de-noising MPC data using a snapshot fusion method to address false MPC clustering. Simulation and measurement were also done, and concluded that the Fuzzy C-means algorithm was outperforming the K-means algorithm. A power-weighted Gaussian mixture model was used to cluster the measurements in a 28 GHz indoor channel reported in [13]. The clustering's focus is the elevation of arrival, delay, and the MPC's relative power. Moreover, the accuracy of K-Power Means was evaluated using the Jaccard Index [14]. The datasets used are from the COST2100 channel model, where the comparison has been visualized using angles of azimuth and elevation in contrast with the delay. One challenge of clustering is also to find the number of optimal clusters. Moreover, a graphical user interface (GUI) was developed in [15] to determine cluster counts by a factor weighing approach, and it was reported that the delay and angular data were effective weight factors in the indoor and semi-urban scenarios, respectively. The non-stationary and spherical waveform was considered parameters in [16]. They proposed considering these parameters for a broad array of antennas to improve the channel model's accuracy.

2.2 Data Visualization

In [17], the author provides a survey for visualization advances in the past decade. They also categorized the pipeline and workflow in visualizing data from data transformation, visual mapping, and the view transformation. This survey also shows that different interactions can be done by the user when visualizing data.

A GUI is presented in [18] named VISTA that provides a framework to validate the multivariate or multi-dimensionality of clusters interactively. Furthermore, in [19] the authors presented a cluster visualization web tool implemented using R. The web tool utilizes the heatmap and the Principal Component Analysis (PCA) visual tools. The ToxPi presented in [20] was developed using JAVA to explore and visualize data. The interface shows the user's good interactivity with the data resulting in a more efficient analysis and interpretation. Clustering methods with 30 CVI variables were implemented in R software presented in [21]. Another application of

visualization is finding the traffic patterns used to analyze data and improve the data-driven transport systems [22].

To summarize, studies on MPC clustering relies on the algorithms and corresponding CVI. The studies visualize datasets using scatterplots, providing the MPC parameter that is focused on the clustering procedure. Different propagation scenarios, measurement campaigns are also present in the literature, and to this point, no clustering algorithm performs best for all scenarios. Spatial parameters were also considered in [16] and resulted in increasing the measurement parameters that limit the visual representation of the 3D plot. Data visualization is also abundant in different fields, especially bioinformatics but rarely used in channel modeling. By addressing these issues as they relate to cluster identification, the authors' goal is to investigate optimal visualization techniques to draw the advantage of manual clustering as a post-process for automatic clustering.

3 Theoretical Consideration

3.1 Channel Modeling

Channel modeling is a crucial task in designing the wireless communication system. Channel models specify parameters that reduce the randomness in the channel. The parameters indicate essential features that affect the performance of wireless networks. In a MIMO system, the double-directional channel model reveals the propagation's fundamental parameters [23]. The double directional channel model parameters consist of the delay τ that corresponds to the length of the MPC going to the receiver, the azimuth of departure φ_{AOD}, the azimuth of arrivalφ_{AOA}, the elevation of departure θ_{AOD} , and the elevation of arrival θ_{AOA} , and power. These parameters can be seen graphically in Figure 1. The MPC appears in clusters based on the similarity and dissimilarity of the above parameters. This combination produces a six-dimensional vector of an MPC: $\boldsymbol{x}_\ell = [\tau_\ell \quad \theta_{\ell,\text{AOA}} \quad \varphi_{\ell,\text{AOA}} \quad \theta_{\ell,\text{AOD}} \quad \varphi_{\ell,\text{AOD}}]$. Where $\boldsymbol{x}_\ell$ represents the ℓ-th MPC. Each MPC observed or measured is stored in a matrix $\boldsymbol{X}$ which forms the set of one snapshot of the MPC. Given these parameters, the visual analysis of MPC is limited due to its multidimensional features. To improve the validation of the clusters in the channel model, the visualization addressing all the features is necessary.

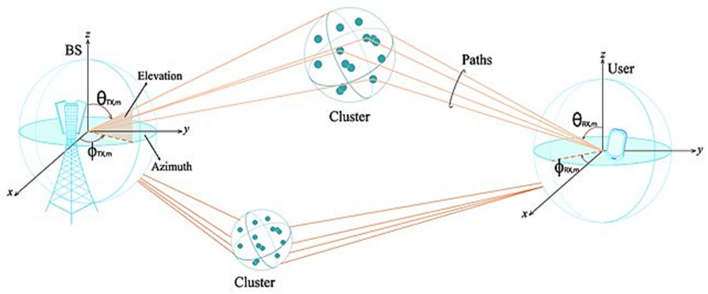


Figure 1: Parameters in the multipath wireless channel [24]

3.2 Clustering

Clustering is a process of separating and grouping the MPC into their respective similar or dissimilar parameters. Clustering allows the discovery of patterns in a given dataset. In terms of machine learning, clustering can be seen as an unsupervised learning problem [25]. When datasets are unlabeled, clustering algorithms label the data in their respective affinity based on the feature of the data specified by the algorithm. The clustering of MPC can simplify acquiring the CIR and obtain the effects of scatterer in the environment. The clustering of MPC was traditionally done using the manual visual inspection approach. Due to the manual approach's laborious and subjective drawbacks, [10] proposed an automatic framework for MPC clustering, which increases the impact on the research of clustering MPC. This work resulted in a plethora of literature that accommodates novel algorithms for specific propagation scenarios. The clustering method can be summarized into four steps: feature extraction, clustering algorithm, validation, and interpretation [26]. This work focuses on aiding the post-processing of the data, validation, and interpretation using data visualization techniques.

3.3 Data Visualization

Data visualization is a crucial method to interpret the MPC measurements. Visualization techniques unfold underlying patterns of clusters through visual analysis. Although, plots that can be drawn and perceive by the human eye for interpretation are limited to 3-Dimensional scatter plots. With the addition of colors, the designation or cluster membership of MPC can be identified. The rise of big data produced techniques to further improve the visualization of datasets to reveal patterns and clusters within the data. The methods developed reduced the features and embedded the dimension to a two-dimensional matrix, which can be visualized using a scatter plot. Dimensionality reduction techniques use a transformation to extract the essential details while preserving the features and the data [27]. The reduction techniques can be linear and non-linear. Data visualization has benefitted disciplines like bioinformatics regarding large data sets of gene expressions. Keim [28] states that to infer knowledge in a dataset effectively, it must include the human domain knowledge and interaction in the data exploration.

4 Design Consideration

This section presents a brief discussion of the main concepts of visualization techniques used in the MPC datasets in this work.

4.1 Parallel Coordinate Plot (PCP)

PCP is one axis based method for representing multidimensional datasets. The parallel plot maps all the data dimensions to parallel horizontal axes that displays the values of each data point to its respective feature. It can represent all the data dimensions and can be color-coded to show the cluster identification. The parallel plot's drawback is that large data sets crowd the graph resulting in visual clutter [29]. Another drawback is the overshadowing of values if one parameter has large magnitudes compared to others.

This disadvantage can be addressed by using data normalization before plotting. In this paper, the dataset used was normalized before visualization and mapping. The coordinate axes order can be re-arranged to uncover specific patterns and relationships among the features to interact with the data fully.

4.2 Dendrogram and Heatmaps

The dendrogram is a visualization technique that uses a tree-like structure representing clusters by nodes that interconnect similar multipaths. It uses distance metrics for computing, which among the measurements are closed to one another and projects a line graph to a node included in a more massive cluster. Heatmaps represent a colored data matrix representing the magnitude of parameters [30]. The clustergram function in MATLAB is used in this work. Hierarchical clustering is used to form the visual representation; hence, data interpretation must focus first on the dendrogram linkage of elements. This technique reorders the row elements based on the tree-like structure of the values.

4.3 Principal Component Analysis

Principal Component Analysis (PCA) is a reasonably old dimensionality reduction technique that captures multi-dimensional data variance [31]. The goal of the PCA is to construct principal components by computing the covariance matrix using the eigenvector and eigenvalues of the data. Upon applying PCAon a D-dimensional data, its principal components, each with a percentage on how it largely affects the variance. PCA projects a line in each dimension that captures the maximum variance of the data. Selecting a principal component with the highest variance score leads to an analysis of the multi-dimensional dataset. PCA projects the best fitting line is the highest scored principal component's dimension and minimizes squared distances of the points to the line in the respective dimension. The algorithm does the same for all the orthogonal dimension to the previous dimensions and readjusts the axes with respect to the projected line. Using PCA captures the global structure of the data linearly.

4.4 t-Distributed Stochastic Neighbor Embedding

t-SNE is a popular and relatively new dimension reduction technique developed in 2008 by van der Maaten [32]. It is a non-linear projection algorithm that computes the local structure data points indicated by a distance metric and is mapped to low dimensional space. t-SNE utilizes an optimized cost function to remove the crowding problem of data points and applies Student-t distribution instead of Gaussian. t-SNE requires a perplexity parameter that is a value of neighboring points the user wishes to preserve during transformation. In this work, the features of the MPC are the input variables for each of the multipaths. The t-SNE outputs matrices containing the joint probability of the L multipaths with the computed scores embedded to plot the data based on the stochastic neighbors. The plot of the t-SNE is a grouped scatter plot based on the perplexity given and groups the points with the nearest correlation. t-SNE preserves the local structure of the dataset and a powerful dimensionality reduction technique in a non-linear fashion. In MATLAB, the t-SNE function allows the user to define the distance metric to calculate the pairwise distance. In this work, the Euclidean distance metric is employed.

5 Results and Analyses

The data set used in this study is extracted from the IEEE data port presented in [33]. Datasets were first extracted from the COST2100 Channel Model [34]. The dataset was transformed from the spherical coordinates to rectangular coordinates using the directional cosine method resulting in a 7-dimension feature of each MPC. Whitening the data also converts the data into specific values to accommodate necessary adjustments. This transformation assumes that the cluster ID of the channel reference is considered the ground truth of each MPC. Two specific propagation scenarios were selected from the eight propagation scenarios in the dataset—the Indoor Line of Sight and the Semi-Urban Non-Line of Sight in a wideband channel. The dataset captures 29 and 911 multipaths for the indoor and semi-urban propagation scenarios, respectively. The visual technique's performance in drawing knowledge with the data can be observed using these two scenarios.

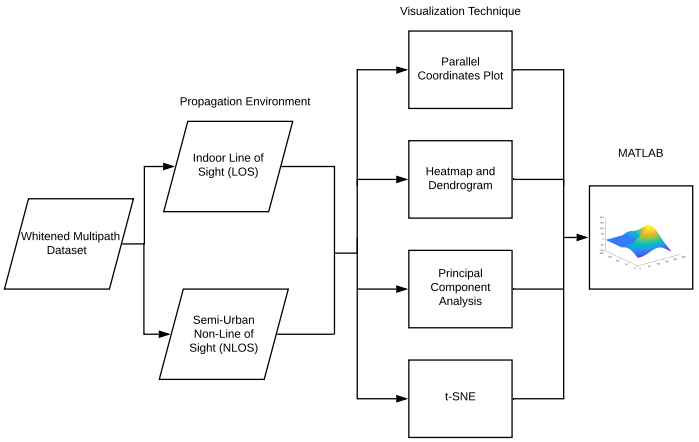


Figure 2: The methodology of the study

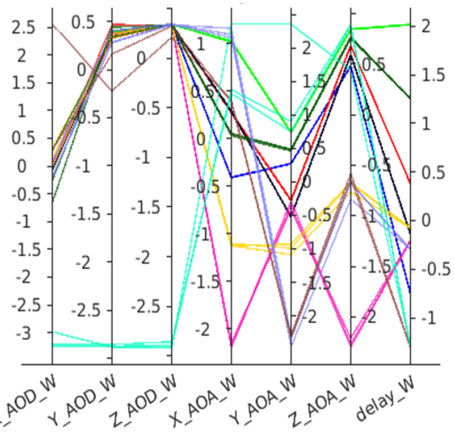


Figure 3: Parallel Coordinate Plot (Indoor)

The methodology in the implementation of this work is shown

in Figure 2. The datasets chosen are implemented in the MATLAB environment due to the availability of the previously discussed visualization techniques' tools and functions. The input data sets are stored, evaluated, and presented using visualization methods. The color represents each multipath's reference cluster in the graphs except for the heatmaps in Figure 3.

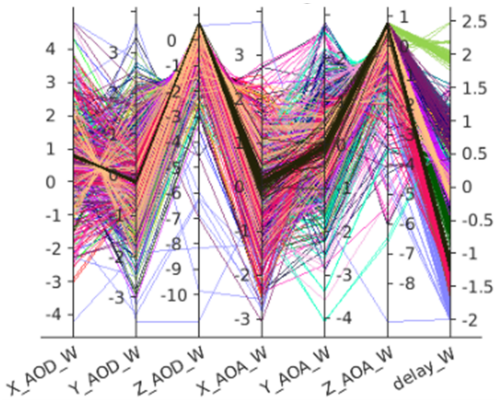


Figure 4: Parallel Coordinate Plot (Semi-urban)

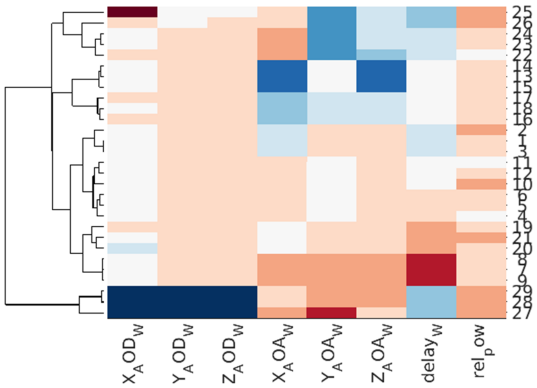


Figure 5: Heatmap and Dendrogram (Indoor)

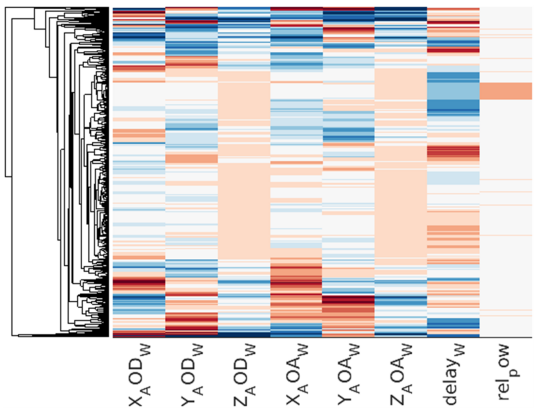


Figure 6: Heatmap and Dendrogram (Semi-urban)

The parallel coordinate plot is a useful technique in visualizing all dimensions of the data as it plots all the MPC dimensions all at once. A specific color represents the cluster ID based on the dataset. In the indoor scenario, more straightforward analysis is possible due to the small number of MPC, and the outliers can easily be observed with the light blue color cluster. The indoor scenario in Figure 3 shows the MPC clusters have the same AoD values but varies within the arrival angle and delay. This result shows the advantage of using a parallel plot for cluster analysis. As opposed to the indoor figure, the projection of the semi-urban environment in Figure 4 visually clutters the graph due to a large number of MPC data. To address the cluttering, the user can filter clusters that can be seen as validated clusters and use a density-based parallel plot as presented in [29].

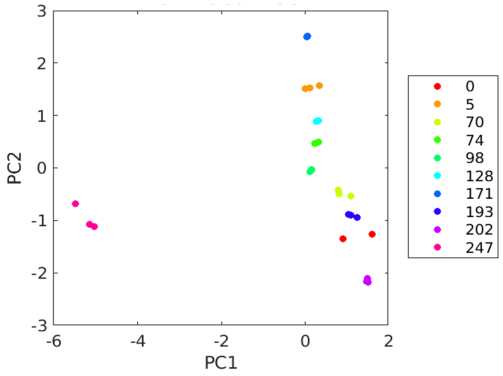


Figure 7: Principal Component Analysis (Indoor)

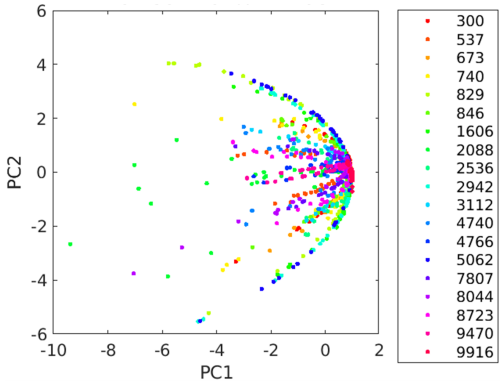


Figure 8: Principal Component Analysis (Semi-urban)

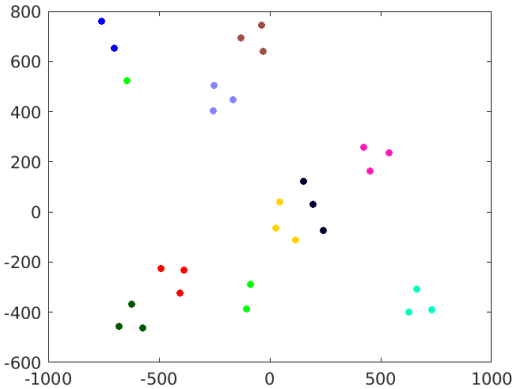


Figure 9: t-SNE (Indoor)

The clustergram function in MATLAB utilizes the linkage func-

tion and dendrogram that produces a hierarchical clustering result and is presented in a heatmap. The dendrogram in Figure 5 and Figure 6 show the cluster using a tree diagram that links the MPC with a close relationship in the lowest node and represents the high separation of the values with the upper branch. In contrast with the other tools in this work, this function is limited to presenting a heatmap of the reference cluster ID. The authors suggest creating a customized heatmap based on the computed cluster id and doing a dendrogram to compare the algorithm's performance and the hierarchical clustering.

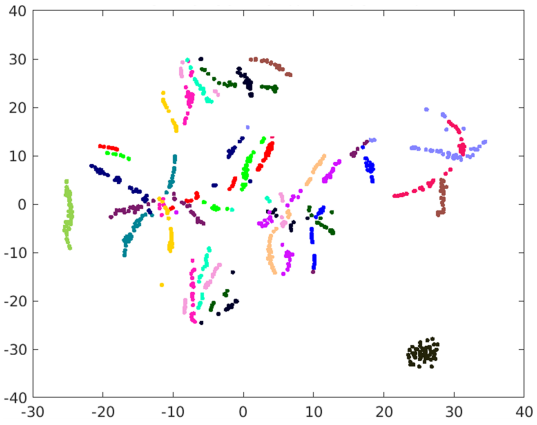


Figure 10: t-SNE (Semi-urban)

PCA outputs two principal components with the highest variance in the MPC dataset. MATLAB's pca function provides each principal component's scores with the two highest scores are projected to represent the dataset in a 2D scatterplot. These principal components are computed by matrix factorization using the singular value decomposition (SVD), including the eigenvectors. By truncating the matrix to the desired dimension, pca produces a low dimensional representation of the data using the principal components. The x-axis in Figure 7 shows the highest scored principal component that handles most of the variance. The increased value of data points in both axes shows the MPCs variance in the first and second principal components. The cluster with ID 247 indicates a low variance in both the principal components and can be interpreted as uncorrelated to the other clusters. On the other hand, the MPC in the semi-urban scenario clusters is grouped closely and signifies a high correlation as can be seen in Figure 8.

The tsne function in MATLAB from the extracted dataset, the cluster ID is removed from the dataset and later represents each cluster classification using colors. The features of the MPC, which is an $L \times 9$ matrix, is reduced to an $L \times 2$ matrix and plotted in a scatterplot. Given the pairwise distance of x_k and x_l MPC, the function computes the conditional probability that the kth MPC picks the lth MPC as its neighbor. The perplexity parameter value sets the number of neighboring points desired by the user. The t-SNE implementation can be varied by selecting the perplexity parameters. Furthermore, the number of iterations was set to the default value of the tsne function with 500 steps. In the indoor scenario in Figure 9, the perplexity was set to group the neighboring data points is maximized to 5 since the data set has 21 MPC. Based on visual observation, the grouping reflects the cluster properly, and each of the neighboring groups can be correlated to form a new cluster. In the semi-urban environment in Figure10, perplexity was set to 40 adjacent points, and observed overlapping clusters could be correlated with one another. Even so, the visual grouping by the t-SNE performs well and isolates clusters that have high dissimilarity with other clusters. The use of t-SNE for iteratively providing insight into the MPC clusters can help interpret the propagation environment's physical realism if the interacting objects are identified.

Due to the wide variety of data when measuring MPC in propagation scenarios, several algorithms are being investigated to correctly cluster the MPC. Using these visualization techniques, designers can quickly identify outliers produced by the algorithm and provide a physical realism within the data. It can be observed that the t-SNE can significantly identify natural clusters between MPC and can visually analyze the effectiveness of algorithms applied to measured MPC. With the built-in functions of MATLAB, the user can specify another column with another cluster ID that is the output of a clustering algorithm. These flexible parameters can compare the clusters from the ground truth of the data set.

Furthermore, visualization can be done as a pre-processing and post-process steps on clustering to observe the structure of data before using an algorithm and the effectiveness of the clustering algorithm with the MPC. Using t-SNE, which solves the crowding problem of the data set, the clusters in the plot can be easily used to identify specific scatterers with physical distance and have a high effect on the signals being transmitted. This work presents visualization to enhance the manual approach of clustering in parallel with the development and research of clustering algorithms.

The authors encourage the use of the variation and extensions of the visualization tools presented. The implementation of the visual techniques with algorithmic clustered datasets, both measured and synthetic, is also considered.

6 Conclusions

This study implements visualization techniques to wireless multipath clusters. It can be observed that these techniques aid the analysis in finding natural clusters and outliers of MPC visually. As measurement campaigns suggest increasing the parameters, MPC clustering then has more features than previously discussed. Since the optimal algorithm for clustering MPC is still an open topic for research, visual analysis can be drawn to enhance the clusters by visual representation further. The techniques used in this study considers the angular, delay, and relative power properties of the MPC and are visualized in one plot. The use of the parallel coordinate plot and t-SNE is effective in both the indoor and outdoor scenarios. The proper validation of clustering results in channel modeling can be improved by allowing users to visualize each feature of the data in the dimension provided by the parameters. The user domain knowledge and inputs based on the scenario can serve as a basis to improve the validity of these clusters, leading to an efficient and effective way of observing the effects of the wireless channel. Presenting the data in different possible ways aids cognition and infers knowledge in clustering the MPC. Furthermore, the execution of this work with more data analysis and the development of a GUI is considered for future work.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment De La Salle University is acknowledged for supporting this work.

References

[1] M. Chen, U. Challita, W. Saad, C. Yin, M. Debbah, "Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial," IEEE Communications Surveys Tutorials, **21**(4), 3039–3071, 2019, doi: 10.1109/COMST.2019.2926625.

[2] E. Bonek, "Chapter 2 - MIMO Propagation and Channel Modeling," in A. Sibille, C. Oestges, A. Zanella, editors, MIMO, 27 – 54, Academic Press, Oxford, 2011, doi:10.1016/B978-0-12-382194-2.00002-2.

[3] M. ITU-R, Minimum requirements related to technical performance for IMT-2020 radio interface (s), Nov, 2017.

[4] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. T. Brink, I. Gaspar, N. Michailow, A. Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago, F. Wiedmann, "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," IEEE Communications Magazine, **52**(2), 97–105, 2014, doi:10.1109/MCOM.2014.6736749.

[5] A. Africa, L. Bulda, E. Del Rosario, M. Marasigan, I. N. I. M. Z. Marasigan, "Radio Wave Propagation: Simulation of Free Space Propagation Path Loss," Int. J. Emerg. Trends Eng. Res, 2020.

[6] R. He, B. Ai, A. F. Molisch, G. L. Stuber, Q. Li, Z. Zhong, J. Yu, "Clustering Enabled Wireless Channel Modeling Using Big Data Algorithms," IEEE Communications Magazine, **56**(5), 177–183, 2018, doi:10.1109/MCOM.2018.1700701.

[7] L. Materum, J.-i. Takada, I. Ida, Y. Oishi, "Mobile Station Spatio-Temporal Multipath Clustering of an Estimated Wideband MIMO Double-Directional Channel of a Small Urban 4.5 GHz Macrocell," EURASIP Journal on Wireless Communications and Networking, **2009**(1), 804021, 2009, doi: 10.1155/2009/804021.

[8] A. A. M. Saleh, R. Valenzuela, "A Statistical Model for Indoor Multipath Propagation," IEEE Journal on Selected Areas in Communications, **5**(2), 128–137, 1987, doi:10.1109/JSAC.1987.1146527.

[9] K. Yu, Q. Li, D. Cheung, C. Prettie, "On the tap and cluster angular spreads of indoor WLAN channels," in 2004 IEEE 59th Vehicular Technology Conference. VTC 2004-Spring (IEEE Cat. No.04CH37514), volume 1, 218–222 Vol.1, 2004.

[10] N. Czink, P. Cera, J. Salo, E. Bonek, J. Nuutinen, J. Ylitalo, "A Framework for Automatic Clustering of Parametric MIMO Channel Data Including Path Powers," in IEEE Vehicular Technology Conference, 1–5, 2006, doi: 10.1109/VTCF.2006.35.

[11] M. R. Arias, "Analysis of Multipath Propagation based on Cluster Channel Modelling Approach," Polibits, (42), 77–82, 2010, doi:10.17562/PB-42-9.

[12] Z. Huang, R. Zhang, J. Pan, Y. Jiang, D. Zhai, "A Framework of Multipath Clustering Based on Space-Transformed Fuzzy c-Means and Data Fusion for Radio Channel Modeling," IEEE Transactions on Vehicular Technology, **69**(1), 4–15, 2020, doi:10.1109/TVT.2019.2947605.

[13] Y. Li, J. Zhang, P. Tang, L. Tian, "Clustering in the wireless channel with a power weighted statistical mixture model in indoor scenario," China Communications, **16**(7), 83–95, 2019, doi:10.23919/JCC.2019.07.007.

[14] A. Teologo, "Cluster-Wise Jaccard Accuracy of KPower Means on Multipath Datasets," International Journal of Emerging Trends in Engineering Research, **7**, 203–208, 2019, doi:10.30534/ijeter/2019/16782019.

[15] M. Roque, "Interface for the Factor-inclusion Weighting Approach in Determining the Number of Multipath Propagation Clusters," International Journal of Advanced Trends in Computer Science and Engineering, **6**, 1768–1776, 2019, doi:10.30534/ijatcse/2019/107842019.

[16] J. Li, B. Ai, R. He, M. Yang, Z. Zhong, Y. Hao, G. Shi, "On 3D Cluster-Based Channel Modeling for Large-Scale Array Communications," IEEE Transactions on Wireless Communications, **18**(10), 4902–4914, 2019, doi: 10.1109/TWC.2019.2930694.

[17] S. Liu, D. Maljovec, B. Wang, P. Bremer, V. Pascucci, "Visualizing High-Dimensional Data: Advances in the Past Decade," IEEE Transactions on Visualization and Computer Graphics, **23**(3), 1249–1268, 2017, doi:10.1109/TVCG.2016.2640960.

[18] K. Chen, L. Liu, "VISTA: Validating and refining clusters via visualization," Information Visualization, **3**(4), 257–270, 2004, doi:10.1057/palgrave.ivs.9500076.

[19] T. Metsalu, J. Vilo, "ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap," Nucleic acids research, **43**(W1), W566–W570, 2015, doi:10.1093/nar/gkv468.

[20] S. W. Marvel, K. To, F. A. Grimm, F. A. Wright, I. Rusyn, D. M. Reif, "ToxPi Graphical User Interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models," BMC bioinformatics, **19**(1), 1–7, 2018, doi: 10.1186/s12859-018-2089-2.

[21] M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, M. M. Charrad, "Package 'nbclust'," Journal of statistical software, **61**, 1–36, 2014.

[22] W. Chen, F. Guo, F. Wang, "A Survey of Traffic Data Visualization," IEEE Transactions on Intelligent Transportation Systems, **16**(6), 2970–2984, 2015, doi:10.1109/TITS.2015.2436897.

[23] M. Steinbauer, A. F. Molisch, E. Bonek, "The double-directional radio channel," IEEE Antennas and Propagation Magazine, **43**(4), 51–63, 2001, doi: 10.1109/74.951559.

[24] F. Ademaj, S. Schwarz, T. Berisha, M. Rupp, "A Spatial Consistency Model for Geometry-Based Stochastic Channels," IEEE Access, **7**, 183414–183427, 2019, doi:10.1109/ACCESS.2019.2958154.

[25] D. Z. He, Ruisi, Applications of Machine Learning in Wireless Communications, Institution of Engineering and Technology, 2019.

[26] R. Xu, D. Wunsch, "Survey of clustering algorithms," IEEE Transactions on Neural Networks, **16**(3), 645–678, 2005, doi:10.1109/TNN.2005.845141.

[27] J. Tang, J. Liu, M. Zhang, Q. Mei, "Visualizing large-scale and high-dimensional data," in Proceedings of the 25th international conference on world wide web, 287–297, 2016, doi:10.1145/2872427.2883041.

[28] D. A. Keim, "Information visualization and visual data mining," IEEE Transactions on Visualization and Computer Graphics, **8**(1), 1–8, 2002, doi: 10.1109/2945.981847.

[29] J. Heinrich, D. Weiskopf, "State of the Art of Parallel Coordinates," in Eurographics, 2013.

[30] T. Galili, A. O'Callaghan, J. Sidi, C. Sievert, "heatmaply: an R package for creating interactive cluster heatmaps for online publishing," Bioinformatics, **34**(9), 1600–1602, 2018, doi:10.1093/bioinformatics/btx657.

[31] I. T. Jolliffe, "Principal components in regression analysis," in Principal component analysis, 129–155, Springer, 1986.

[32] L. v. d. Maaten, G. Hinton, "Visualizing data using t-SNE," Journal of machine learning research, **9**(Nov), 2579–2605, 2008.

[33] J. F. Blanza, A. T. Teologo, L. Materum, "Datasets for Multipath Clustering at 285 MHz and 5.3 GHz Bands Based on COST 2100 MIMO Channel Model," in 2019 International Symposium on Multimedia and Communication Technology (ISMAC), 1–5, 2019, doi:10.1109/ISMAC.2019.8836143.

[34] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, P. D. Doncker, "The COST 2100 MIMO channel model," IEEE Wireless Communications, **19**(6), 92–99, 2012, doi:10.1109/MWC.2012.6393523.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1071-1081 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Review of Different Methods for Optimal Placement of Phasor Measurement Unit on the Power System Network

Ademola Abdulkareem, Divine Ogbe, Tobiloba Somefun*

Department of Electrical and Information Engineering, Covenant University, Ota, 112107, Nigeria

ARTICLE INFO

Article history:
Received: 06 August, 2020
Accepted: 11 November, 2020
Online: 14 December, 2020

Keywords:
Phasor Measurement Unit
Voltage Phasor
Current Phasor
Global Positioning System
Optimal PMU Placement
Power System Observability

ABSTRACT

Phasor Measurement Unit (PMU) is an integral device for tracking, protection, and regulation of the power network. PMU gives synchronised calculations of actual-time data for voltage phasor, current phasor, and the frequency. Placing PMU in every node to observe the power network is not realistic from an economic standpoint and even for big data management. Thus, raising the number of PMUs in the node network to achieve optimum monitoring of the power network is a mandatory issue. In tracking the problem of optimal PMU placement, various methods have been proposed by various researchers in recent times. To facilitate understanding among the various approaches, this study classified the existing methods into two headings; heuristic and mathematical(conventional)methods. The study discusses the different existing optimisation approaches used to solve the optimal PMU placement problems. The benefits, as well as the drawbacks of each approach, is thoroughly examined and wind up the discussion that the heuristic approach is a fast knowledge-based methodology, which was used as the optimization solver for the IEEE 14, 30 and Nigeria national grid 28 and 52 bus system.

1. Introduction

Phasor measurement unit is an electrical instrument that measures the phase angle and magnitude of current and voltage in the power network system, using a synchronised time source provided by global positioning system, which can provide accuracy up to one microsecond [1]. To decide the condition of a power network, the estimations of the state variables (phase angles and voltage magnitude), of all system nodes should be known. Realising these values helps to calculate the flow of real and reactive power in the power network. It will be very strenuous to consider system nodes one of after the other most especial a large system, this is one area in which intelligent devices find application [2] in this case PMU estimations solve the problem effortlessly.

At a particular period, the voltage phase angle and magnitude and current sinusoidal waveforms are expressed as real and imaginary numbers that are known as a phasor. The value of the magnitude is dependent on the amplitude of the sinusoidal waveform, while the phase angle is gotten relying on the time position as presented in the figures. Figure 1a shows the sinusoidal waveform is behind as for the cosine reference, as it is behind, the phase angle will be negative, though, in Figure 1b, the sinusoidal waveform is ahead as for the cosine reference. Subsequently, the phase angle is positive. Since the phasor measurement unit is provided with the global positioning system, every phasor estimation is time-tagged, consequently, permitting phasor estimations gotten from the phasor measurement units in different positions in a system to synchronise.

PMU firstly receives signals in analogue form. The signal being received contains estimated voltages and currents from current and voltage transformer. The anti-aliasing filter then used in eliminating high-frequency factors surpassing the sampling limit of Nyquist. It will, therefore, suppress the frequency factors that exceed the Nyquist sampling rate. The phase-locked oscillator divides the global positioning system one pulse each second into the needed number of pulses as regards the sampling of the waveform. Through the use of a 16-piece accuracy analogue to a digital converter, the input signals will be digitised at sampling moments when the time signals are sampled from the phase-locked oscillator before feeding it to the phasor microprocessor. The microprocessor resamples the signals from the digitised sample information and estimates the positive succession. Then the

*Corresponding Author: Tobiloba Somefun, km 10, Idiroko Road, Canaanland Covenant University, +2348037632337, Email: tobi.shomefun@covenantuniversity.edu.ng

approximate phasor is time-stepped before being moved using modems to a communication medium [3].

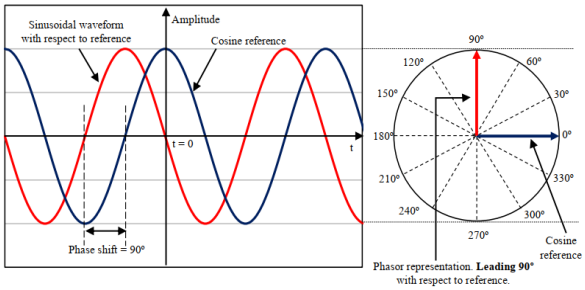


Figure 1: (a) phasor representation. -90º regarding cosine reference [3]

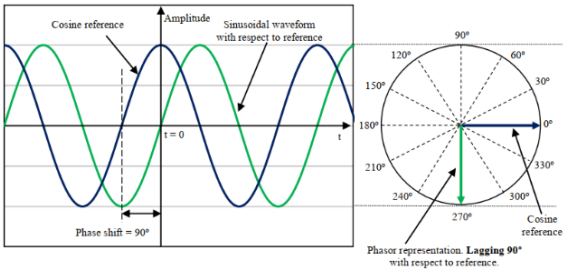


Figure 1: (b) phasor representation. 90º regarding cosine reference [3]

2. Optimal PMU Placement Problem Formulation

There are two main performances used in the analysis for observing power system; they are numerical observability and topological observability [4]. Numerical observability suffers from a lot of matrix computations; thus, it is not advisable to employ them for analysing power systems, especially large ones. Topological observability of a system is achieved when a full rank of traversing the tree is obtained. Some simplified and improved effective rules when topologically analysing a power system are illustrated below:

Rule 1: For a PMU-installed node, voltage phasor and the connecting branches current phasor are known as demonstrated in Figure 2. This measurement is known as direct measurement. In Figure 2, PMU is installed in node 1; consequently, the voltage phasor of the node is known, and the current phasor of the connecting line is also known following the first rule.

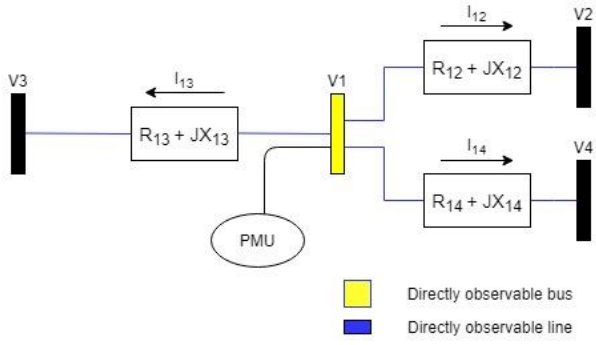


Figure 2: illustrating Direct measurement Rule 1

Rule 2: When the voltage phasor is known, and the current phasor at one end of the branch is also known, it guarantees that the other end is observable by determining the voltage phasor, as illustrated in Figure 3. This type of measurement is known as pseudo-measurement.

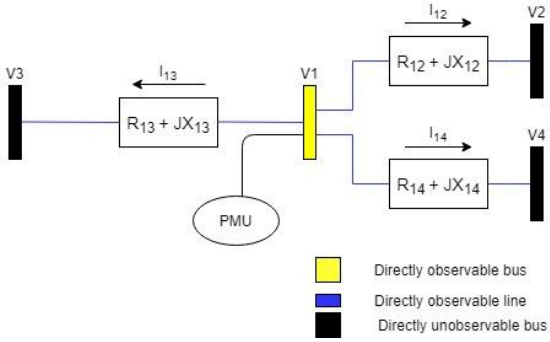


Figure 3: illustrating pseudo-measurement Rule 2

Since the value of the current phasors of the lines connected to node 1 is known, the value of the other nodes can be solved for by utilising Ohms law.

$$I_{12} = \frac{V_1 - V_2}{R_{12} + jX_{12}} \tag{1}$$

$$V_2 = V_1 - I_{12}(R_{12} + jX_{12}) \tag{2}$$

$$I_{13} = \frac{V_1 - V_3}{R_{13} + jX_{13}} \tag{3}$$

$$V_3 = V_1 - I_{13}(R_{13} + jX_{13}) \tag{4}$$

$$I_{41} = \frac{V_4 - V_1}{R_{14} + jX_{14}} \tag{5}$$

$$V_4 = V_1 + I_{41}(R_{41} + jX_{41}) \tag{6}$$

Rule 3: In a case where the voltage phasor for both ends is known, the current phasor of the connecting branch can be gotten as shown in Figure 4. Knowing the voltage phasor of node-1 and node-2, the current phasor of the connecting line can be calculated using Ohm's law.

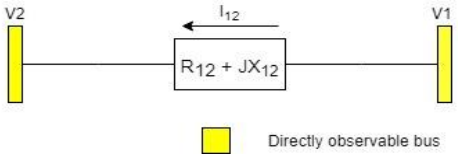


Figure 4: Illustrating observability of the branch current for rule 3

$$I_{12} = \frac{V_1 - V_2}{R_{12} + jX_{12}} \tag{7}$$

Rule 4: For a zero-injection node, knowing all the current of all connected branches, except one, calculating the current for the unknown branch can be achieved using Kirchhoff current law (KCL) this rule is demonstrated in Figure 5. node-3 is the zero-injection node, and current phasor of every branch is known except the current for the branch between node-3 and node-4.

Rule 5: For a zero-injection node where the voltage phasor is not known, if all the neighbouring nodes are known, the zero-injection node can be known through calculations, as shown in Figure 6. In the diagram, node-3 is the zero-injection node and the

voltages for the neighbouring nodes is known. By using node calculations, the voltage for node-3 is known.

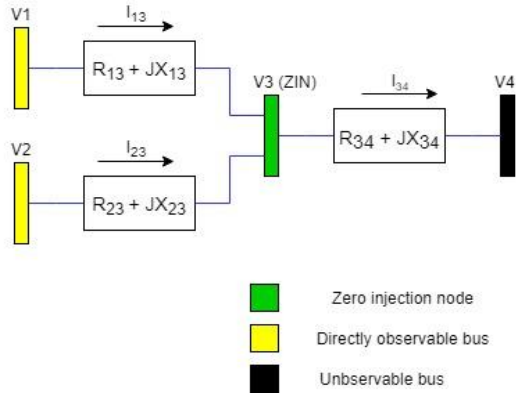


Figure 5: Model observability of zero-injection node for rule 4

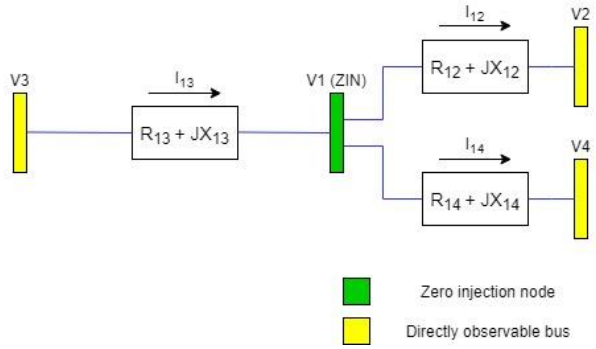


Figure 6 Modelling unobservable ZIN rule 5

$$V_3 = V_1 - I_{13}(R_{13} + jX_{13}) \quad (8)$$

$$V_3 = V_2 + I_{32}(R_{32} + jX_{32} \quad (9)$$

$$V_3 = V_4 + I_{34}(R_{34} + jX_{34}) \quad (10)$$

$$0 = I_{12} - I_{32} - I_{24} \quad (11)$$

Rule 6: For a group of neighboring zero-injection nodes that are not observable because the voltages are not known, but the neighboring nodes connected to the group of zero injection nodes are known. The unknown voltage phasors of the zero-injection nodes can be made known by using both Kirchhoff's current and voltage laws as illustrated in Figure 7. In this figure, the voltage of node-3 and node-4, but the voltages of the nodes connected to them are known.

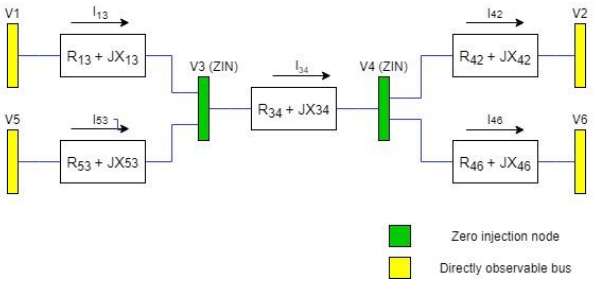


Figure 7 Modelling unobservable ZINs rule 6

Minimum number of PMU: One of the purposes of the optimal PMU placement problem is to resolve the minimum number of PMUs needed for a power system and their right allocation to make sure that the system is totally observable. This objective can be represented mathematically as shown:

$$\sum_{i}^{\text{N nodes}} A_{ij} \quad (12)$$

$$s.t\, B * A \geq 1 \quad (13)$$

$$I = [111 \ldots \ldots 1]^{T}_{N\,X\,1} \quad (14)$$

$$A(i) = \begin{cases} 1 & if\ PMU\ is\ installed\ on\ node\ i \\ 0 & if\ PMU\ is\ not\ installed\ on\ node\ i \end{cases} \quad (15)$$

$$B(i,j) = \begin{cases} 1 & if\ i = j \\ 1 & if\ nodes\ i\ and\ j\ are\ joined \\ 0 & if\ nodes\ i\ and\ j\ are\ not\ joined \end{cases} \quad (16)$$

3. Placement Methods

Several methods have been deployed for optimal location on power system network in different regards such as distributed generators and PMU [5,6]; in this section, several of the methods used to optimally site PMU are extensively reviewed.

3.1. Conventional Method

Two of the conventional methods that have been reported in previous studies are discussed under this subsection.

3.1.1. Integer Linear Programming (ILP)

Integer Linear Programming methods, also recognised as binary integer programming, considering both the system's injection and power measurement as well as PMU error calculation by state estimation. Integer linear programming algorithm is based on the individual vectors derived from the adjacency matrix of the transverse tree [7]. Upon decomposition, the PMUs are positioned ideally in the networks utilising the integer linear programming principle to reduce the deployment rate. Two types of ILP models are mixed integer linear programming (MILP) and binary integer linear programming (BILP) [8].

3.1.2. Integer Quadratic Programming (IQP)

Integer quadratic programming utilises the connectivity matrix that reflects the topology of the system to define the optimal PMU placement (OPP) problem. The quadratic function's objectives were structured, taking into account linear constraints and the vector of integer values. The method will decrease the number of PMUs by providing full network observability in standard operations as well as in outages [8].

In [9], the authors discussed factors that affect calculation precision, including measurements inconsistencies, number of PMU measurements, and quasi-power measurements implementation. Quasi-power uncertainty is determined by the propagation concept of uncertainty, and the optimum location of PMU is calculated by integer quadratic linear programming. Various types of state variables as well exhibited different sensitivities to different measurements [10].

3.1.3. Drawbacks of Conventional Methods

Usually, the conventional technique is utilised to model a network when the network is not so complex. Also, conventional

technique takes more computational time when compared with Heuristic techniques to solve a complicated network problem.

3.2. Heuristic Methods

Most of the Heuristic Methods that have been reported in previous studies are discussed under this subsection.

3.2.1. Genetic Algorithm

The foundation of genetic algorithm is to model natural selection which has no requirement for secondary functions like derivatives calculation. Some of the advantages of genetic algorithm which makes it more applicable to OPP problems are: the likelihood of the local minimum being trapped is reduced, Calculations are decreased from one state to the other, and the fitness assessment of each sequence directs the search [11].

In [12], they investigated optimally placing PMUs and communication links for distributed state estimation in distribution networks. They combined integer linear programming and genetic algorithm to find the least number of PMUs in a distribution network. A distributed state estimation system segmenting technique is created to minimise the burden it takes to compute and communicate [13]. The aim of the partitions is to balance the number of buses in subareas. By turning the network segmenting into the problem of assigning a centralised bus of each subarea, the scale-balanced partitioning scheme between subareas is obtained by applying genetic algorithms.

3.2.2. Particle Swarm Optimisation

Particle swarm optimisation is a similar technique to genetic algorithm, where population solutions are randomly assigned to a system firstly [14]. In [15], the authors described particles as entities that are hovering through multidimensions in space. For every particle, the best location is determined by the fittest position faced by that particle and its neighbouring particles. The process of the particle swarm optimisation begins with a primary point and velocity for every particle, where the velocity of each particle is restricted by reasons of not hovering in unusable space as well as overflowing forbiddance [16].

3.2.3. Tabu Search

The Tabu search algorithm is a combinatorial method, and it combines techniques such as conventional method and heuristic method. This method is majorly used in solving combinatorial optimisation problems in covering and scheduling. Tabu lists, which is a major element of tabu search, consists of numbers of states visited in recent times and unwanted states. Another major element of tabu search is the state definition and the encompassing area. The tabu search resets when it does not converge [17].

In [18], the authors solved the optimal PMU placement problem and achieved maximum observability of a system and maximum redundancy by utilising tabu search-based linear state estimation model [19]. This quick technique of topological observability analysis required a loss simulation function relying on the matrix of incidence to solve the optimal PMU placement problem. This technique also concerns the convenience and high speed to monitorable power network by manipulating integer numbers [20].

3.2.4. Differential Evolution

The method of differential evolution (DE) uses elements such as N-dimensional vectors to reduce progressing space functions. Mutation, fusion and choice are the key functions used in the global optimising process. Differential evolution approach is utilised extensively in various problems of cost function including multimodal, non-differentiable, and nonlinear functions. Certain advantages of this method include simultaneous computations, simple use, and strong synchronisation properties [21].

The researchers introduced multi-objective optimal PMU placement in [22] utilising a non-dominating filtering DE technique, that is an organic synthesis of the nondominated filtering procedure and DE technique Pareto. This theory put into consideration total measurement reliability and voluntary PMU failure to achieve complete observability of the network in addition to finding a solution to the optimal PMU placement problem [3] Using the differential evolution technique obtained from genetic algorithm contributed to the suggested method for the nondominated sorting differential evolution algorithm. It was stated that obtaining a particular and absolute Pareto front and having several Pareto ideal solutions was the improvement of this method.

3.2.5. Simulated Annealing

The simulated annealing method solves complex combinatorial optimisation where the existing solution is changed arbitrarily. The new changed solution is the worst change with a decreased probability as computing continues. The ideal outcome for a major problem of combinatorial optimisation requires an effective disturbance framework, cost function, space for the solution, and cooling plan for the simulated annealing to solve. Simulated annealing can be considered adequate by looking for a large-scale network and achieving good speed in seeking an ideal or near-ideal solution [23]

Authors in [24], investigated effective use of different PMU estimated values and exposure to extremely critical network data were put into consideration for the ideal deployment of PMUs to make the power system fully observable. A topological observability formulation based on the incidence matrix evaluated the original PMU arrangement to have a network with complete observability.

3.2.6. Ant Colony Optimization

A further concept used to present an optimisation problem solution is ant colony optimisation, which originally utilises the ant population. The function of the ant colony was to travel across neighbouring states of the problem by implementing an ideal controller for the stochastic local resolution, resulting in solving optimising problem. Evaporation of the pheromone trace and action of the daemon are procedures in ant colony optimisation. The use of ant colony optimisation to determine good paths across graphs will decrease computational issues [25].

Authors in [26], investigated on an enhanced ant colony optimization which was used to solve the ideal PMU placement problem for achieving a totally visible power network with the least number of PMUs and putting into consideration total measurement redundancy. The depth-first search as a conceptual

graph approach was used to create an estimation tree to evaluate if the power system was observable [27].

3.2.7. Iterated Local Search

The key viewpoint of iterated local search is to search for a subspace characterised by local optimal solutions rather than the entire solution space. By using an integrated heuristic, there is a series of solutions where the best option is achieved if repeated arbitrary tests of the heuristic were to be used [28].

The optimal PMU placement method introduced in [29] involves two phases which included an initial PMU dispensation to reach a visible network by using an iterated local search to decide the least number of PMUs required to make a system fully visible. Page rank placement formulation is utilised in this approach to determine the significance of every bus.

3.2.8. Mutual Information

Taking into account not only access to a system with maximum observability but also analysing irregularities in the network states that utilised mutual information between PMUs measurements values and system states, and information theoretical method to solve optimal PMU placement problems was addressed in [30].

3.2.9. Matrix Reduction

For the problems of placing PMUs, a distribution matrix occurs, whereby the graph indicates the distribution scope whenever a station is placed in different locations. The problem level is determined by the range of the distribution. The approach of matrix reduction aims to establish ideal placement by increasing the matrix of incidence. Using a formulation based on removing digital data, the matrix reduction technique was proposed to achieve the least number of PMU to reach observability of the network completely and reduce the computation process [31].

3.2.10. Imperialistic Competition Algorithm

Imperialistic competition technique is a new technique formulated to solve optimal placement problems. Just like other heuristic techniques, imperialistic competition algorithm begins with a first population named country which is a colony of two types. Competitiveness of the countries brings about the minimal of the problem. The effectiveness of the Imperialistic competition technique in wide area optimisation was evaluated on test buses and as confirmed [32].

3.2.11. Immune Genetic Algorithm

The Immune Genetic Algorithm was utilised to proffer solution to placing the PMU optimally utilising three impact vaccinations to create a topologically measurable power system. The two measures that occur in an immune genetic algorithm approach to defend against viruses and bacteria are vaccines and immunity alternatives. The vaccination basis is the combination of technical knowledge and previous data about the optimal PMU placement problem. Two operators, namely, crossover and mutation, were considered by the immune genetic algorithm, which was utilised to maximise the tests [33].

3.2.12. Biogeography Based Optimisation

The biogeography-based technique is a mathematical model for the movement of species from one region, which is known as a habitat to another region, rising and extermination conditions of species. This technique is used to provide solutions to the problem of optimisation through complex performance environments and to improve the efficiency of species migration and integration into the habitat [34].

3.2.13. Chemical Reaction Optimization

To find a solution to the OPP problems, a newly formed heuristic approach named population-based chemical reaction technique was implemented. Through modelling the behaviour and movement of gases in a chemical reaction, reaching a low energy steady state was the key goal of the chemical reaction techniques, which sought to achieve the least free energy level. High performance was demonstrated by the application of the chemical reaction technique to metrics and practical problems [35].

Using a new heuristic approach called chemical reaction technique and simplified chemical reaction technique model, optimal PMU placement problem solved to achieve a completely measurable power network considering the system with and without zero-injection nodes [36].

3.2.14. Artificial Bee Colony (ABC)

Artificial bee colony technique's key motivating criteria are food resources path, food resources range, and food resources performance information [37]. The developed formulation could be used in solving unimodal and multi-modal mathematical optimization problems. Also, a multiple objective optimal PMU placement algorithm, called binary-coded was implemented in solving the optimal PMU placement problem and obtain the least quantity of PMU and highest possible network redundancy. Consideration was given to eventuality constraint conditions like a one-line outage, and arrangements of the least quantity of needed PMUs were determined [37].

3.2.15. Bacteria Foraging Algorithm

Biological evolution propensity towards animal exclusion with weak feeding strategies that work to find, handle, and consume food and spread the genes of animal performance in feeding approaches as they are more probable to achieve productive results. Weak foraging approaches after several years are either discarded or revamped [4]. Four operators utilised in the developed bacterial foraging algorithm were chemotaxis, swarming, reproduction, and removal and distribution [38].

3.2.16. Cellular Learning Automata

Cellular learning automata technique is cantered on the use of training automaton to change cellular automaton state conversion likelihood. This approach begins by determining each cell's inner condition. Instead, on the basis of cellular learning automata law, the strengthening signal for every training automata is calculated. Maintaining this operation will achieve the intended state [39].

In [29], a CLA approach was used to examine the integration of two opposing objective functions like the least number of PMUs and total measurement reliability to achieve a completely measurable power network. This research put into consideration

the drawbacks of flexibility, including PMU/branch loss and traditional measurements and zero-injection nodes.

3.2.17. Hybrid Methods

Utilising a formulated hybrid genetic algorithm and simulated annealing, a solution was introduced in [38] to ideally place PMUs and remote terminal units for a large network of interconnected grids determined by a multiarea system state estimate. Using remote terminal units and conventional measurements, a PMU was incorporated into a power network to ensure the calculated state more precisely and reduce the expense of conventional measurement and cost of the remote terminal unit. Utility centre of one multiarea area to reach system measurements required a single PMU placement, as PMU measurement calculated the network node voltage. A criterion for the PMU was the node with the highest connecting line. The identification of faulty data was achieved by taking into account the vital measurement of each region.

A synthesis of least traverse tree technique with the optimised genetic algorithm is a hybrid method named MST-GA presented in [40] to achieve the least number of PMUs required to make a system fully observable and to consider optimising redundancy. This approach enhanced mutation operation taking into account grid topology information. As the key and side steps, fusion and mutation were used as an activity to create new entities, the outcome of the new analysis of this approach was a decrease in the number of PMUs required and a variety of solutions [41].

To easily capture the objectives and contributions of the previous studies with respect to methods/techniques deployed Table 1 is developed while Table 2 gives the comparison of conventional and heuristic methods with respect to four factors namely; the size of the problem, data requirement, accuracy, and effectiveness.

3.2.18. Drawbacks of Heuristic Methods

To efficiently implement the heuristic technique, it takes a specific level of expertise and experience [4,8]. Often these functionality specialists are difficult and costly to employ, particularly if they need domain knowledge. If the analysts are not included in the development, they may not be informed of the technical design constraints or why such design choices have been made [8]. Heuristic tests are poorly organized, and thus risk identifying one-time, low-priority issues. Lastly, heuristic assessments do not allow for an assessment of the performance of re-designs [3].

Table 1: Objectives and contributions of investigated studies

Author(s)/Reference Number	Objective	Technique	Considerations	Gaps	Benefits
B. Mallikarjuna, P. Gopakumar, M. J. B. Reddy and D. K. Mohanta [23]	Ideal positioning of PMUs to allow monitoring of incorrect data. simulated annealing technique is implemented with the stochastic new approach	Simulated Annealing	detection of critical measurements	Zero-injection nodes (ZIN)& Measurement Redundancy was not considered	Observability of the power network with critical measurements
C. Margarida, K. Xenia and V. Ana, [42]	Limiting the quantity of PMU allocation utilising stochastic simulated annealing algorithm	Stochastic Simulated Annealing	Detection of critical measurements is introduced as a consequence function	The algorithm was not Line loss, PMU the outage	Detection of critical measurements
Z. Zhida, Y. Hao, L. Peng, L. Peng, K. Xiangyu, W. Jianzhong and W. Chengshan [12]	Reducing the quantity of PMUs, and determining their geographical allocation, achieving a total network of measurable power. A genetic algorithm-based procedure is used, and PMUs are installed with current phasor measurements as the maximum quantity of simultaneous lines are highlighted in all network nodes	Genetic Algorithm	Connection of PMUs with the number of current phasors to be calculated	Quality of the result were not documented.	Needed channels for current in the optimisation problem
C. Peng, H. Sun and J. Guo. [43]	Limiting the quantity of PMUs and increasing the redundant measurements. In place of one optimal solution, Pareto-ideal solutions are presented	Non-Dominated SortingGenetic Technique	Zero-injection nodes & Measurement Redundancy	It was assumed that PMUs capacity is limitless. PMU channels limit was not considered	Offering Pareto-ideal facade forcontradictory objectives, solution repair
S. M. Mazhari, H. Monsef, H. Lesani and A. Fereidunian [39]	A multi-objective ideal allocation to minimise the quantity of PMUs and increase the accuracy of measurements. PMU allocation is also analysed by the implementation of a standardised observability function based on traditional non-synchronous	Cellular Learning Automata	Traditional measurements, PMU failure, Branch loss	This technique might not give the least quantity of PMUs to make the network observable	High performanceinelectrical net works on a large scale
A. A. Hany, Y. A. Almoataz and V. Mukherjee [25]	The ABC definition is used to attain the least quantity of PMUs to achieve a total measurable power network, fulfilling measurement redundancy	Artificial Bee Colony	One branch loss, zero-injection nodes	The pheromone variation span is restricted to interval [0, -1],	System viability and efficiency is shown by the comparison of the simulation results with the previous works

N. C. Koutsoukis, N. M. Manousakis, P. S. Georgilakis and G. N. Korres [19]	Two contending purposes involving the least quantity of PMU and maximum redundancy on the network	Tabu Search	Zero-injection nodes & Measurement Redundancy	Placement of PMU in ZINwas not allowed in the initialization.	Optimal problems solved with high precision and reduced calculation
R. Ramachandran and S. Karthick [44]	Limiting the quantity of PMU placement and developing a parallel tabu search formulation	Parallel Tabu Search	Communication restraint, Zero-injection nodes, and matrix condition of the state estimations	Traditional measurements, PMU failure, Branch loss was not considered	Reduced calculation period
M. Nazari-Heris and B. Mohammadi-Ivatloo [4]	Two contending purposes involving the least quantity of PMU and maximum redundancy	Clonal Algorithm	Total redundancy	ZIN enhancement of secondary voltage control operation was not considered	Maximum velocity of the procedure, achieving practical schemes
M. Zhou, V. Centeno, A. Phadke, Y. Hu, D. Novosel and H. Volskis [31]	Reducing the quantity of PMUs. Utilising pre-processing techniques and providing solutions using mathematical based methods	Matrix Reduction	Computer-generated data removal, Pre-processing technique and matrix reduction procedure, utilising Lagrangian relaxation	Quality assessment was not recorded because there was no other tested optimal placement set for the Brazilian network	Minimising the size of allocation model and the calculation exertion, employed in a large-scale network
S. Md, I. H. Md, M. Abido, A.-F. Taher and M. A.H [15]	Limiting the quantity of PMUs so as the power network can be fully observed. Developing a hybrid technique based on binary particle swarm optimisation	Binary Particle Swarm Optimisation	Increasing redundancy in the power network, one PMU and Multi-PMU failure	loss of PMU and branch loss were ignored	Maximum velocity of the procedure and reduced calculation period
A. A. Hany, Y. A. Almoataz and V. Mukherjee. [25]	Reducing the quantity of PMU in the system. Developing an improved ant colony technique	Ant Colony Optimisation	High redundant measurements in the system	The pheromone variation span is restricted to interval [0, -1],	Avoiding inactivity behaviour and elevated velocity of the procedure, employing a graph theoretic technique
S. S. Noureen, V. Roy and S. B. Bayne [8]	Reducing the quantity of PMU and remote terminal units with critical measurements free	Hybrid Genetic Algorithm and Simulated Annealing	Traditional measurement and RTU faulty data recognition, one-line outage	Fault monitoring, ZIN was ignored	Relevant to existing power networks observed utilising remote terminal units
S. Mahapatra, S. Pandab and S. C. Swaina [33]	Limiting the quantity of PMU placement and increasing measurements redundancy, taking considerations of traditional measurements	Bacterial Foraging Algorithm	Total redundancy & zero-injection nodes	The linear model could not correctly depict complex dynamics of the network, particularly in disturbances.	Appropriate for existing power networks due to modelling standard measurement
M. N. Seyyed and M. Javad [21]	Limiting means square error by achieving the least quantity of PMUs, with or without the presence of standard measurement	Differential Evolution	Normal measurement least square error of state estimate	Fault monitoring, Zero-injection nodes was ignored	The precise, fast and straightforward procedure, ability to employ in multiple objectives problem
B. Rohit and B. Biplab [45]	Limiting the cost of installation of Mismodelling dissimilar cost of PMUs for various nodes	Particle Swarm Optimisation	Dissimilar cost of PMU allocations	It was assumed that PMUs capacity is limitless. PMU channels limit was not considered	Taking consideration of real-time equipping cost of PMUs, limiting entire cost in place of the quantity of PMUs
B. Rohit and B. Biplab [46]	Reducing the number of meters and PMUs Taking consideration of one line/meter loss and one line/PMU loss	Biogeography Based Optimisation	Zero-injection nodes, line loss, SCADA meter loss, PMU the outage, PMU/branch loss	Traditional measurements, PMU failure, Branch loss was not considered	using simulated node reduction method for decreasing the size of the network
M. Hurtgen and J.-C. Maun. [28]	Reducing the overall quantity of PMUs utilising PageRank placement technique and iterated local search	Iterated Local Search	PMU outage	Zero-injection nodes (ZIN)& Measurement Redundancy was not considered	Taking consideration of eventualities, simple comprehension and implementation
S. S. Noureen, V. Roy and S. B. Bayne. [8]	Limiting the quantity of PMUs. Observing guide nodes needed for secondary voltage control	Branch and Bound	Zero-injection nodes enhancement of secondary voltage control operation	PMU outage, branch loss, measurements redundancy, considered	Observing guide nodes for boosting speed of the voltage control system
Y. Zhao, P. Yuan, Q. Ai and T. Lv [22]	Reducing the quantity of PMUs to monitor both the overall network and fault in the power network	Differential evolution	Fault monitoring, Zero-injection nodes	The constraints utilized ignored some optimal solutions	Utilising PMUs for power network monitoring and fault detection

S. S. Noureen, V. Roy and S. B. Bayne [8]	Limiting means square error by achieving the least quantity of PMUs	Differential evolution	Taking consideration of incessant alterations in the power network's topology	Fault monitoring, ZIN was ignored	Integrating PMU allocation problem into state estimate.
A. Ahmadi, Y. Alinejad-Beromi and M. Moradi [16]	Limiting the quantity of PMUs, recommending new the rule for the network to be observed topologically	Modified binary particle swarm optimisation	Zero-injection nodes, PMU/Branch loss	Measurement redundancy was not studied, and the problem of local least affected the results	Establishing new rules of topologically Observable valuation for limiting the quantity of the needed PMUs
H. Mohammadi, G. Khademi, D. Simon and M. Dehghani [34]	Two contending purposes involving the least quantity of PMU and maximum redundancy	Multi-objective biogeography-based optimisation	measurements redundancies, zero-injection nodes.	The large amount of cost evaluations utilized made a little difference in the solution	The proposed technique generates well-circulated Pareto-optimal solutions
M. Nazari-Heris and B. Mohammadi-Ivatloo [4]	Multi-objective model, reducing the quantity of PMUs and increasing measurements	Genetic algorithm	Considering single branch/single PMU loss considered	ZIN enhancement of secondary voltage control operation was not ignored	Very little population and iterations needed are less
L. Qiao, C. Tao, W. Yang, N. Rohit, F. Franz and M. D. Ilić [30]	Reducing the quantity of PMUs utilising anInfo-theoretic approach	Mutual information	PMU outage, standard measurement	ZIN was not considered	Modelling the improbabilities in the network states
A. Bashian, M. Assili, A. Anvari-Moghaddam and O. R. Marouzi [11]	Reducing the quantity of PMUs needed for a total observable network and increasing measurements redundancy. A topologically observable rule of zero-injection nodes is also presented	Binary imperialistic completion algorithm	PMU outage, branch loss, measurements redundancy,	Cost considering ZIN was ignored	Quick integration, less deviance, competence of locating global optimal and zero standard deviation
S. Li and Z. Meng [37]	Artificial bee colony technique is implemented to attain least Quantity of PMUs to achieve total observability of the power network, a fitting measurement redundancy	Artificial bee colony	One branch loss and Zero-injection nodes	The technique in this study was only tested on small scale network	Viability and performance of the technique validated by comparison of the simulation outcomes with the related previous research

Table 2: Comparison of Conventional and Heuristic Technique

S/N	Properties	Conventional Technique	Heuristic Technique
1	Size of the Problem	Usually, the conventional technique is utilised to model a network when the network is not so complex.	Heuristic technique is a form of computation which tries to mimic human or animal capability in a very simplified manner in order to proffer solution to the problem of PMU allocation
2	Data requirement	The conventional method utilises physical principles to obtain network relationships, which are typically rightfully generalised with several assumptions and need a previous understanding of the nature of the data relationships.	The heuristic technique focuses on data based on the data only to decide the framework and variables to solve a network problem, with fewer assumptions about the network's nature.
3	Accuracy	As compared to Heuristic technique, because of several factors implemented, the conventional technique does not have a better degree of correctness.	Owing to very little possibilities of failure, Heuristic technique has higher precision.
4	Effectiveness	The conventional technique takes more computational time when compared with Heuristic techniques to solve a complicated network problem.	The heuristic technique can be used effectively in performing routine and tasks that consume time.

4. Results

According to the comparison in Table 2, heuristic technique did better in terms of complexity of the network, data requirement, and level of correctness and effectiveness as regards to time. For this research genetic algorithm is selected the optimization solver for optimal placement of PMUs on IEEE14, 30 and Nigerian 330KV (28 and 52 bus system) networks. It was selected in place of other heuristic algorithms because it is one of the pioneer evolutionary algorithms, simplicity and ease of implementation, and logical reasoning behind the use of operators like selection, crossover and mutation. The mutation operator is executed bit by bit after offspring is produced, for every node in the chromosome expression, an evenly appropriated random number is created between zero and one. If the number produced is lower than the likelihood of the mutation, then the corresponding bit will be

modified from zero to one or one to zero, thereby putting or eliminating the PMU on the appropriate node. Once all the offspring have been generated, the current population and the created offspring are categorized based on the decreasing order of the objective function provided, and a tournament selection is carried out to determine the new generation. Again, a difference is made based on the size on the problem, the top 10 chromosomes are chosen to utilize the tournament selection scheme and inserted directly into the current generation. The suggested approach is implemented to find a solution to the problem by taking consideration of these cases: Normal operation; and Case considering Zero Injection (ZIN)

4.1. Normal operation

In this case, the solution for the optimal PMU placement problem is obtained by ignoring ZIN.

Table 3: Normal operation

	IEEE14	**IEEE 30**	**NG 28**	**NG 52**
Number of PMU	4	10	11	17
PMU location Bus number	2, 6,7, 9	1, 5, 8, 11, 12, 19, 23, 26, 29	1, 3, 4, 12, 13, 17, 20, 21, 23, 24, 27	3, 4, 6, 8, 10, 16, 19, 24, 26, 27, 32, 35, 39, 41, 42, 46, 51

Table 3 shows PMU placement result for the IEEE 14, IEEE 30, Nigeria bus 28 and 52 considering normal arrangement of the bus.

4.2. Case considering ZIN

In this case, the solution for the optimal PMU placement problem is obtained by considering ZIN.

Table 4: Locations of the ZINs

Bus	**Zero Injection location**
IEEE-14	7
IEEE-30	6, 9, 22, 25, 27,28
NG-28	11
NG-52	-

Table 4 shows the location of the zero-injection node for each bus considered.

Table 5: Considering Zero Injection Node

	IEEE14	**IEEE 30**	**NG 28**	**NG 52**
Number of PMU	3	7	10	N/A
Bus PMU location	2, 6, 9	1, 2, 10, 12, 19, 23, 27	1, 3, 4,12, 17, 20, 21, 23, 24, 27	N/A

Table 5 shows PMU placement results for the IEEE 14, IEEE 30, Nigeria bus 28, and 52 considering zero injection node. When a zero-injection node is considered, the number of PMUs needed to observe a system is reduced because that particular bus is merged with a connecting node.

Table 6: System of Redundancy Index (SORI)

Bus	GA	ILP [54]	IQP [55]
IEEE14	NPMU: 4 SORI: 19	NPMU: 4 SORI: 19	NPMU: 4 SORI: 19
IEEE 30	NPMU: 10 SORI: 52	NPMU: 10 SORI: 50	NPMU: 10 SORI: 50
NG 28	NPMU: 11 SORI: 46	N/A	N/A
NG 52	NPMU: 17 SORI: 71	N/A	N/A

Table 6 shows a benchmark with previous studies that utilized a mathematical method considering a normal operation. The result is favourably compared with the previous work carried out on IEEE 14 and IEEE 30 bus system. However, the result obtained in this study is superior to the mathematical methods since it gives the same number of PMU and a higher quality of SORI. Based on this fact, Genetic Algorithm is recommended and is adequately fit heuristic technique for PMU placement which improved resilience of the system and enhanced its reliability.

5. Conclusion

The problem of placing PMU optimally was solved using GA heuristic technique and the quality of the result was compared to results of previous research, from the comparison as shown in the result, it can be observed that the system of redundancy index for the heuristic algorithm used did better than the mathematical programming method when implemented on the IEEE 30 bus system, and achieved the same quality for the IEEE 14 bus system. Methods such as conventional and heuristic techniques, exist to solve the optimal placement problems; this study presented a thorough review study on both techniques. Stabilisation and control of Optimal PMU Placement (OPP) have become a power network necessity. We covered a magnificent set of techniques in this study. This would give researchers a structure for classifying these methods and choosing the appropriate solution prior to applying for PMU placement. Analysis of optimisation techniques in this research shows that minimising the OPP would be a framework and help identify optimisation conditions for projections in time to come.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research work and publication charge is fully funded by Covenant University Centre for Research, Innovation and Discovery (CUCRID).

References

[1] A.G. Phadke, "Synchronized Phase Measurement in Power System," IEEE Computer Applications in Power, **6**(2), 10–15, 1993, doi:10.1109/67.207465.

[2] P.O. Oluseyi, T.E. Somefun, O.M. Babatunde, T.O. Akinbulire, O.O. Babayomi, S.A. Isaac, D.E. Babatunde, "Evaluation of energy-efficiency in lighting systems for public buildings," International Journal of Energy Economics and Policy, **10**(6), 2020, doi:10.32479/ijeep.9905.

[3] N. Rahman, A.F. Zobaa, "Optimal PMU placement using topology transformation method in power systems," Journal of Advanced Research, **7**(5), 625–634, 2016, doi:10.1016/j.jare.2016.06.003.

[4] M. Nazari-Heris, B. Mohammadi-Ivatloo, Application of heuristic algorithms to optimal PMU placement in electric power systems: An updated review, Renewable and Sustainable Energy Reviews, 2015, doi:10.1016/j.rser.2015.04.152.

[5] T.E. Somefun, C.O.A. Awosope, A. Abdulkareem, A.S. Alayande, "Deployment of power network structural topology to optimally position distributed generator within distribution system," Journal of Engineering Science and Technology Review, **13**(1), 2020, doi:10.25103/jestr.131.2.

[6] T.S. Shomefun, A. Ademola, C.O.A. Awosope, A.I. Adekitan, "Critical review of different methods for siting and sizing distributed-generators," Telkomnika (Telecommunication Computing Electronics and Control), **16**(5), 2018, doi:10.12928/TELKOMNIKA.v16i5.9693.

[7] S.P. Singh, A.K. Thakur, S.P. Singh, "PMU Placement for Maximum Observability of Power System under Different Contingencies," in Energy Procedia, 2017, doi:10.1016/j.egypro.2017.05.208.

[8] S.S. Noureen, V. Roy, S.B. Bayne, "Phasor measurement unit integration: A review on optimal PMU placement methods in power system," in 5th IEEE Region 10 Humanitarian Technology Conference 2017, R10-HTC 2017, 2018, doi:10.1109/R10-HTC.2017.8288967.

[9] Y. Xueliang, Z. Hengxu, J. Cao, "Analysis of influencing factors in hybrid state estimation," in Asia-Pacific Power and Energy Engineering Conference, APPEEC, 2016, doi:10.1109/APPEEC.2016.7779892.

[10] N. Xie, F. Torelli, E. Bompard, A. Vaccaro, "A graph theory based methodology for optimal PMUs placement and multiarea power system state estimation," Electric Power Systems Research, 2015, doi:10.1016/j.epsr.2014.08.023.

[11] A. Bashian, M. Assili, A. Anvari-Moghaddam, O.R. Marouzi, "Co-optimal PMU and communication system placement using hybrid wireless sensors," Sustainable Energy, Grids and Networks, 2019, doi:10.1016/j.segan.2019.100238.

[12] Z. Zhao, H. Yu, P. Li, X. Kong, J. Wu, C. Wang, "Optimal placement of PMUs and communication links for distributed state estimation in distribution networks," Applied Energy, 2019, doi:10.1016/j.apenergy.2019.113963.

[13] P.P. Bedekar, S.R. Bhide, V.S. Kale, "Optimum PMU placement considering one line/one PMU outage and maximum redundancy using genetic algorithm," in ECTI-CON 2011 - 8th Electrical Engineering/ Electronics, Computer, Telecommunications and Information Technology (ECTI) Association of Thailand - Conference 2011, 2011, doi:10.1109/ECTICON.2011.5947933.

[14] R. Babu, B. Bhattacharyya, "Optimal allocation of phasor measurement unit for full observability of the connected power network," International Journal of Electrical Power and Energy Systems, 2016, doi:10.1016/j.ijepes.2016.01.011.

[15] M. Shafiullah, M.I. Hossain, M.A. Abido, T. Abdel-Fattah, A.H. Mantawy, "A modified optimal PMU placement problem formulation considering channel limits under various contingencies," Measurement: Journal of the International Measurement Confederation, 2019, doi:10.1016/j.measurement.2018.12.039.

[16] A. Ahmadi, Y. Alinejad-Beromi, M. Moradi, "Optimal PMU placement for power system observability using binary particle swarm optimization and considering measurement redundancy," Expert Systems with Applications, 2011, doi:10.1016/j.eswa.2010.12.025.

[17] C. Lu, Z. Wang, Y. Yu, "Optimal PMU placement for pessimistic dynamic vulnerability assessment," IET Generation, Transmission and Distribution, 2018, doi:10.1049/iet-gtd.2017.1261.

[18] Q. Xu, L. Yu, Y. Wang, X. Kong, X. Yuan, "Optimal PMU Placement for the Distribution System Based on Genetic-Tabu Search Algorithm," in 2019 IEEE PES Innovative Smart Grid Technologies Asia, ISGT 2019, 2019, doi:10.1109/ISGT-Asia.2019.8881465.

[19] N.C. Koutsoukis, N.M. Manousakis, P.S. Georgilakis, G.N. Korres, "Numerical observability method for optimal phasor measurement units placement using recursive tabu search method," IET Generation, Transmission and Distribution, 2013, doi:10.1049/iet-gtd.2012.0377.

[20] A. Yda Amira, G. Fathi, A. Amira, G.F.-M.-C. on Systems, & S., U. 2011, "Optimal PMU placement for full network observability case of the tunisian network," Ieeexplore.Ieee.Org, 1–5, 2011, doi:10.1109/ssd.2011.5986789.

[21] S.S.M. Nosratabadi, J. Modarresi, J.M.-2017 S.G. Conference, undefined 2017, "Phasor measurement units placement considering double contingency by differential evolution algorithm based on pareto method," Ieeexplore.Ieee.Org, **2018-Janua**, 1–6, 2018, doi:10.1109/SGC.2017.8308885.

[22] Y. Zhao, P. Yuan, Q. Ai, T. Lv, & T.L.-I.J. of E.P., U. 2015, "Optimal PMU placement considering topology constraints," Elsevier, **73**, 240–248, 2015, doi:10.1016/j.ijepes.2015.04.022.

[23] B. Mallikarjuna, P. Gopakumar, M.J.B. Reddy, D.K. Mohanta, "A case study on optimal phasor measurement unit placement for emerging Indian national smart grid," in International Conference on Signal Processing, Communication, Power and Embedded System, SCOPES 2016 - Proceedings, Institute of Electrical and Electronics Engineers Inc.: 1956–1960, 2017, doi:10.1109/SCOPES.2016.7955789.

[24] S. Akhlaghi, "Optimal PMU placement considering contingency-constraints for power system observability and measurement redundancy," in IEEE Power and Energy Conference at Illinois (PECI), Urbana, IL, USA, IEEE:, 2016, doi:10.1109/PECI.2016.7459251.

[25] A.Y. Abdelaziz, H.A. Abdelsalam, V. Mukherjee, "Optimal PMU Placement in a Distribution Network Considering Network Reconfiguration," Ieeexplore.Ieee.Org, 2015, doi:10.1109/ICCPCT.2014.7054997.

[26] B.K.S. Roy, A.K. Sinha, A.K. Pradhan, "Optimal phasor measurement unit placement for power system observability - A heuristic approach," in IEEE SSCI 2011 - Symposium Series on Computational Intelligence - CIASG 2011: 2011 IEEE Symposium on Computational Intelligence Applications in Smart Grid, 77–82, 2011, doi:10.1109/CIASG.2011.5953335.

[27] R. Salem, A.Y. Abdelaziz, A.M. Ibrahim, R.H. Salem, Optimal PMU Placement for Complete Observability Using Heuristic Methods, 2012.

[28] M. Hurtgen, J.C. Maun, "Optimal PMU placement using Iterated Local Search," International Journal of Electrical Power and Energy Systems, **32**(8), 857–860, 2010, doi:10.1016/j.ijepes.2010.01.028.

[29] S. Chakrabarti, E. Kyriakides, "Optimal Placement of Phasor Measurement Units for Power System Observability," IEEE TRANSACTIONS ON POWER SYSTEMS, **23**(3), 1433, 2008, doi:10.1109/TPWRS.2008.922621.

[30] Q. Li, T. Cui, Y. Weng, R. Negi, F. Franchetti, M.D. Ilic, "An information-theoretic approach to PMU placement in electric power systems," IEEE Transactions on Smart Grid, **4**(1), 446–456, 2013, doi:10.1109/TSG.2012.2228242.

[31] M. Zhou, V.A. Centeno, A.G. Phadke, Y. Hu, D. Novosel, H.A.R. Volskis, "A preprocessing method for effective PMU placement studies," in 3rd International Conference on Deregulation and Restructuring and Power Technologies, DRPT 2008, 2862–2867, 2008, doi:10.1109/DRPT.2008.4523897.

[32] A. Ghasemkhani, A. Anvari-Moghaddam, J.M. Guerrero, B. Bak-Jensen, "An Efficient Multi-objective Approach for Designing of Communication Interfaces in Smart Grids," Ieeexplore.Ieee.Org, Nov. 2020, doi:10.1109/ISGTEurope.2016.7856286.

[33] S. Mahapatra, S. Panda, S.C. Swain, "A hybrid firefly algorithm and pattern search technique for SSSC based power oscillation damping controller design," Ain Shams Engineering Journal, **5**(4), 1177–1188, 2014, doi:10.1016/j.asej.2014.07.002.

[34] H. Mohammadi, G. Khademi, D. Simon, M. Dehghani, Multi-Objective Optimization of Decision Trees for Power System Voltage Security Assessment, Nov. 2020, doi:10.1109/SYSCON.2016.7490524.

[35] M.H.F. Wen, J. Xu, V.O.K. Li, "Optimal Multistage PMU Placement for Wide-Area Monitoring," IEEE TRANSACTIONS ON POWER SYSTEMS, **28**(4), 2013, doi:10.1109/TPWRS.2013.2277741.

[36] J. Xu, M.H.F. Wen, V.O.K. Li, K.C. Leung, "Optimal PMU placement for wide-area monitoring using chemical reaction optimization," in 2013 IEEE PES Innovative Smart Grid Technologies Conference, ISGT 2013, 2013, doi:10.1109/ISGT.2013.6497845.

[37] S. Li, Z. Meng, "Optimal PMU placement based on improved binary artificial bee colony algorithm," in 2017 IEEE Transportation Electrification Conference and Expo, Asia-Pacific, ITEC Asia-Pacific 2017, Institute of Electrical and Electronics Engineers Inc., 2017, doi:10.1109/ITEC-AP.2017.8081008.

[38] P.K. Ghosh, "Complete and incomplete PMU observablity for hybrid state estimation," in 6th International Conference on Computation of Power, Energy, Information and Communication, ICCPEIC 2017, Institute of Electrical and Electronics Engineers Inc.: 867–870, 2018, doi:10.1109/ICCPEIC.2017.8290487.

[39] S. Mahdi Mazhari, S. Member, H. Monsef, H. Lesani, A. Fereidunian, "A Multi-Objective PMU Placement Method Considering Measurement Redundancy and Observability Value Under Contingencies," Ieeexplore.Ieee.Org, 2013, doi:10.1109/TPWRS.2012.2234147.

[40] B. Ozsoy, M. Gol, "A Hybrid State Estimation Strategy with Optimal Use of Pseudo-Measurements," in Proceedings - 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe, ISGT-Europe 2018, Institute of Electrical and Electronics Engineers Inc., 2018, doi:10.1109/ISGTEurope.2018.8571513.

[41] M.M. Rezaoui, D. Ridha, A.A. Laouid, A. Kouzou, R.D. Mohammedi, "Optimal PMUs Placement Using Hybrid PSO-GSA Algorithm," Ieeexplore.Ieee.Org, 25–27, 2019, doi:10.1109/ICPEA1.2019.8911178.

[42] M. Carvalho, X. Klimentova, A. Viana, "Observability of power systems with optimal PMU placement," Computers and Operations Research, **96**, 330–349, 2018, doi:10.1016/j.cor.2017.10.012.

[43] C. Peng, H. Sun, J. Guo, "Multi-objective optimal PMU placement using a non-dominated sorting differential evolution algorithm," International Journal of Electrical Power and Energy Systems, **32**(8), 886–892, 2010, doi:10.1016/j.ijepes.2010.01.024.

[44] R. Ramachandran, S. Karthick, "Optimal PMU Placement for Tamil Nadu Grid Under Controlled Islanding Environment," Procedia Technology, **21**, 240–247, 2015, doi:10.1016/j.protcy.2015.10.021.

[45] B. Rohit, B. Biplab, "Optimal Placement of Phasor Measurement Unit using Binary Particle Swarm Optimization in Connected Power Network," in 2015 IEEE UP Section Conference on Electrical Computer and Electronics (UPCON), IEEE, Allahabad, India: 1–5, 2015, doi:10.1109/UPCON.2015.7456695.

[46] K. Jamuna, Y.L. Selva, G. Gnanasekaran, "Placement of PMUs for power system observability using biogeography based optimization," in Proceedings - 2013 3rd International Conference on Advances in Computing and Communications, ICACC 2013, 355–358, 2013, doi:10.1109/ICACC.2013.76.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1082-1088 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Human Emotion Recognition Based on EEG Signal Using Fast Fourier Transform and K-Nearest Neighbor

Anton Yudhana[1,*], Akbar Muslim[1], Dewi Eko Wati[2], Intan Puspitasari[2], Ahmad Azhari[3], Murein Miksa Mardhia[3]

[1]*Faculty of Industrial Technology, Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, 55166, Indonesia*

[2]*Faculty of Teacher Training Education, Department of Teacher Training of Early Childhood, Universitas Ahmad Dahlan, Yogyakarta, 55166, Indonesia*

[3]*Faculty of Industrial Technology, Informatics Department, Universitas Ahmad Dahlan, Yogyakarta, 55166, Indonesia*

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 28 November, 2020
Online: 14 December, 2020

Keywords:
Electroencephalography (EEG)
Brainwaves
FFT
KNN

ABSTRACT

Human emotional states can transform naturally and are recognizable through facial expressions, voices, or body movements, influenced by received stimuli. However, the articulation of emotions is not practicable by every individual, even when feelings of joy, sadness, or otherwise are experienced. Biomedically, emotions affect brain wave activities, as the continuously functioning brain cells communicate through electrical pulsations. Therefore, an electroencephalogram (EEG) is used to capture input from brain signals, study impulses, and determine the human mood. The examination generally included observing a person's frame of mind in response to a given stimulus where the immediate results were inconclusive. In this study, the associated classifications were normal, focused, sad, and shocked. The raw brainwave data from 50 subjects were recorded by employing a single-channel EEG called the Neurosky Mindwave. Meanwhile, the assessments were performed while the candidates' minds were stimulated by listening to music, watching videos, or reading books. The Fast Fourier Transform (FFT) method was utilized for feature extractions, along with the K-nearest neighbours (K-NN) for classifying brain impulses. The parameter k had a value of 15, and the average classification accuracy was 83.33%, while the highest accuracy for the focused emotional state was 93.33%. The Neurosky Mindwave in conjunction with the FFT and KNN techniques is potential analytical solutions to facilitate the enhanced identification of human emotional conditions.

1. Introduction

The signal of the brainwaves has specific traits and characteristics of the person. The brainwaves cannot be imitated or interpreted by individuals so that there is no chance of resemblance. Identity recognition is essential to identify the features of the person [1].

Identity development is done using biometric technology, which aims at personality-implantation through the use of parts of the body or social behavior—application of Identity Recognition using Biometrics Technology for self-recognition by using parts of the body or human behavior [2-6]. In [7], the author has presented machine-learning algorithms to categorize EEG dynamics according to subject self-reported emotional states during music listening. Support Vector Machine was employed to classify four emotional states and obtained an accuracy of 82.29% across 26 subjects. Vanitha [8] has used Hilbert-Huang Transform (HHT) to remove artifacts and relevant time-frequency features from the EEG signal, and the extracted features are manipulated to detect stress levels using Support Vector Machine (SVM) Classifier on real-time stress detection.

EEG is used as a diagnostic tool by medicine. In comparison, a broader application of the EEG is to analyze a person's mental state, thought patterns, or emotional state [9]. The EEG signal's visual observation is real, given the very low amplitude of the EEG signal, and the pattern is very complex. In addition, EEG signals are strongly influenced by various variables, including emotional conditions, health, respondent activity, environment, electrical disturbances from other organs, and external stimuli.

*Corresponding Author: Anton Yudhana, eyudhana@ee.uad.ac.id

[10]. Conducting the introduction of the EEG signal requires a pattern of brain activity that is prominent and constant. This pattern emerged due to specific stimulation from volunteers. EEG signals are shown to provide more insight into the subject's emotional processes and responses. The recorded EEG signals will get the waveform, the duration, the orientation of the signal, and the signal rhythm. [11, 12].

The EEG signal pattern data's shape and position are not yet known because the signal pattern and response speed response of each individual's brain to the received stimuli are different [13]. Therefore, we need a signal processing method EEG that can pattern specific brain activity [14].

In this study, the Fast Fourier Transform (FFT) based spectrum analysis method converts the EEG signal from the time domain to frequency domain. The FFT results were classified by K-Nearest Neighbor (KNN) as a training data set to characterize the brain wave output from the EEG as a material for analyzing differences in brain waves in each emotional state studied.

2. Data Acquisition

2.1. Neurosky Mindwave Mobile

Neurosky Mindwave Mobile is a brain-computer interface (BCI) device that can be used wirelessly and connected to several devices such as computers and smartphones via Bluetooth connection to collect input from the human brain wave activity [15].

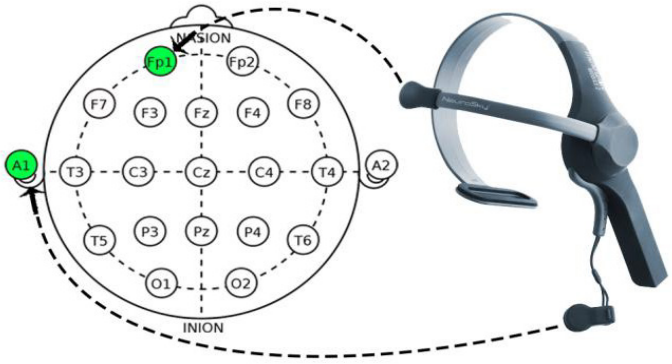


Figure 1: Neurosky Mindwave Mobile and electrode's position

Mindwave provided single-channel EEG in the form of an electrode and placed in the forehead or Fp1 position to measure the brain's electrical voltage and ear-clip on the left ear or A1 position as ground. With Neurosky's ThinkGear and eSense technologies, connected devices can receive a real-time visual representation of brain waves generated from the electrodes [16, 17].

2.2. Subjects

In this study, EEG data were recorded from 50 subjects aged around 20 - 22 years old who were in healthy condition. All of the subjects were undergraduate students and were informed about the purpose of this experiment.

2.3. Stimuli

The use of appropriate stimuli is needed to get better brain wave recording results where a psychologist has validated the stimulus used for this study. Respondents were given stimuli in the form of listening to music, watching videos, and mathematical calculations to influence brain wave activity in the subject to produce an emotional condition that was following the object to be studied [18, 19].

2.4. Procedure of Data Collecting

Before starting data collection, make sure that the respondent is in good health. The data collection process is carried out in a quiet and calm place to make subjects feel comfortable and reduce disturbances affecting data collection. We use the computer laboratory at Ahmad Dahlan University for this research.

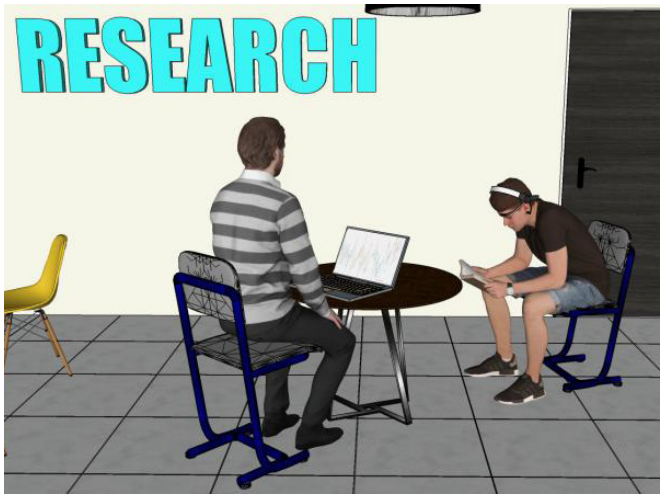


Figure 2: Illustration of the data collecting process

Data collection will be needed as much as 200 data with 50 respondents and carried out four times with a span of at least two days. The recording time for data collection is 7 minutes. When the recording begins, the respondent will be given a stimulus in instrumental music, video, and math questions [20].

3. Methods

3.1. Fast Fourier Transform (FFT)

FFT is applied in various fields,from digital signal processing and solving partial differential equations to algorithms to multiplicate integers in large numbers. FFT is an algorithm for calculating discrete Fourier transforms (DFT) quickly and efficiently. Because many signals in a communication system are continuous in the time domain, we use the Fourier transform to represent the frequency domain [21].

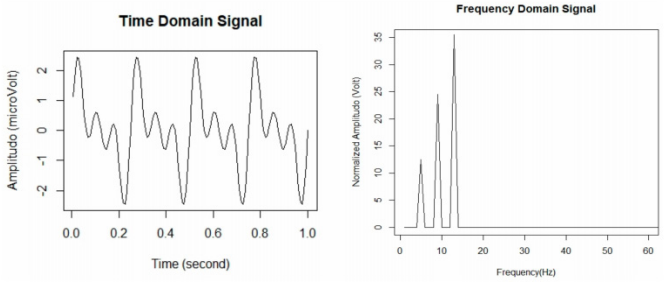


Figure 3: Time-domain signal and frequency domain of EEG signals

FFT is a reliable method for signal processing in the form of a sine wave as EEG signals, and it was proven in previous studies

related to real-time applications because of superior speed compared to other available methods. However, it is not suitable for the analysis of short EEG signals [22]. The FFT formula is defined in (1).

$$S(f) = \int_{-\infty}^{\infty} s(t) e^{-j2\pi ft} dt \quad (1)$$

$e^{-j2\pi ft}$

FFT algorithm is used to reduce the complex transformations performed by DFT. For comparison, when using DFT, then the transformation complexity is N2. Meanwhile, with the FFT algorithm and faster transformation times, the complexity of the transformation is too decreased to (N) log (N).

3.2. K-nearest neighbor

K-Nearest Neighbor is a method for classifying objects based on training data closest to the object. The data is then presented with a distance measure to be processed into a mathematical calculation. In this classification, the calculation is done by using Euclidean Distance. The training data with the closest distance are said to be neighbors and then sorted from the closest distance to the farthest distance. Each neighbor is different from one another or the like. Similar neighbors with the largest number among *k* neighbors correspond to the object being classified [23]. To find the neighbor, it uses distance metrics like euclidean distance as given in equation (2) .

$$D(x, y) = \Sigma_{k=1}^{N} \sqrt{(x_k - y_k)^2} \quad (2)$$

where *D (x, y)* is the distance between *x* and *y* facilities, x_k-y_k is the *x* and y coordinates of variable *k* (*k* = 1,3,5,… n). The value of k must be an odd number, and the number of training data is even to avoid ambiguity. This value is necessary to facilitate the achievement of classification results from the number of closest neighbors. When there is a class with the most neighbors, the test data will get that class result [24].

3.3. Hardware Design

The block diagram above shows that Neurosky Mindwave Mobile is a major component in this hardware system. Neurosky captures input from brain waves via electrodes attached to the FP1 area of the respondent's head and sends an EEG signal via a Bluetooth connection, which acts as a communication protocol between Mindwave and PC / Laptop.

Figure 4: Hardware system diagram

3.4. Software Design

The block diagram above explains that when raw EEG data is obtained from recording brain waves using Neurosky Mindwave on respondents who have been given a stimulus. Raw data taken in the signal pre-processing stage is cut 2 seconds early to minimize noise at the beginning of the recording, due to external interference such as a Bluetooth connection between the recording device and the application, or due to the imperfect installation of the device on the respondent's head. The next stage of feature extraction is carried out using the FFT algorithm to convert the EEG signal in the time domain into the frequency domain and extract all important frequency components from the EEG signal such as Delta waves (0.1 - 3 Hz), Theta (4 - 7 Hz.), Alfa (8 - 12 Hz), Beta (12 - 30 Hz), and Gamma (30 - 100 Hz) [25]. From the results of feature extraction, a huge amount of data is obtained, and data simplification must be done by not including data that is below the average value so that the classification results are more accurate for each emotion. Furthermore, grouping training data and test data are carried out using the KNN classification method in mathematical calculations with the Euclidean formula to calculate the closest distance to the results of the FFT feature extraction. The final step is to determine the accuracy of the KNN classification results on waves using the Confusion Matrix.

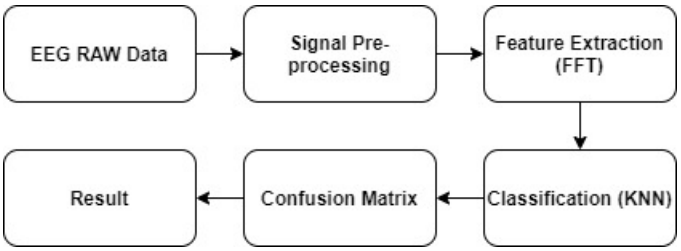


Figure 5: System design diagram

4. Result

In this research, several experimental stages were carried out. The first step is signal pre-processing to reduce noise in raw data, then feature extraction to change the form of RAW data from brain waves to the FFT spectrum. Second, monitoring the FFT results in each emotional state, followed by taking all the highest data from the FFT spectrum with a range of 0.2 Hz. Finally, KNN is used to classify emotions (normal, focused, sad, and shocked) and calculate accuracy using a confusion matrix.

4.1. Feature Extraction of Raw Data

As previously explained, FFT feature extraction with a sampling frequency of 512 Hz is used to convert the time domain signal from raw data to the frequency domain. We use MATLAB2016b software to do this process. The results of running the program can be seen in Figure 6 below.

The picture above shows the feature extraction program results, which produces two different forms of signal output with the same raw data input from subjects who have been given a stimulus in the form of listening to instrumental music. It can be seen that the signal in the time domain has data along with 2.13×10^5 ms with the highest amplitude of 631, which is in the range of 1.32×10^5 ms, while the signal in the frequency domain has data with a range of 0 - 50 Hz where the highest amplitude value is at a frequency of 2.26 Hz. These results will show a different value for each respondent due to different human brain wave activity against the given stimulus.

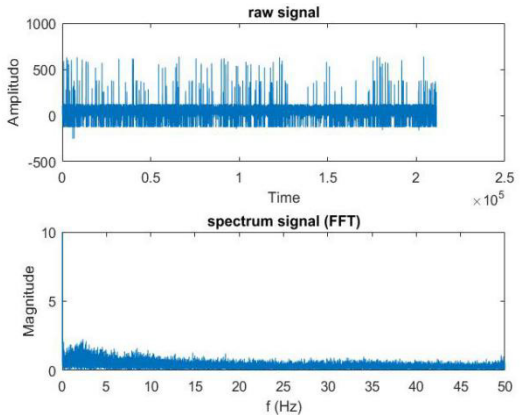


Figure 6: Feature extraction result of raw data

4.2. FFT Spectrum Comparison

In this stage, a comparison of four signal combinations resulted from FFT feature extraction that represented the subject's emotional condition. The results of running programs are shown in Figure 7.

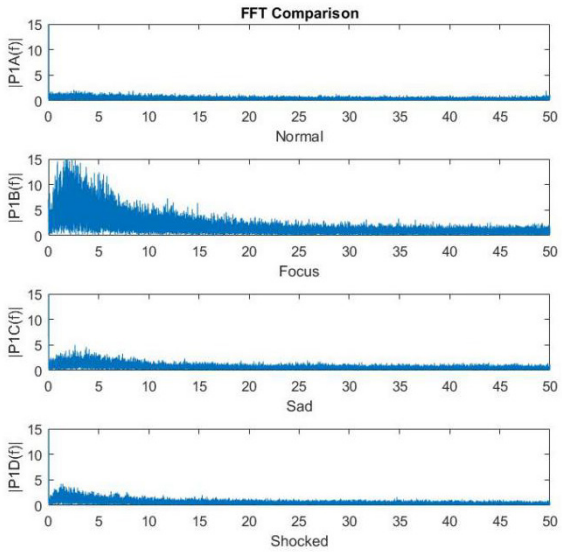


Figure 7: FFT spectrum of all emotional state

It can be seen that the spectrum of the FFT signal in four emotional states, namely Normal, Focus, Sad, and Shocked. When the spectrum is observed, there is a significant difference from the shape of the spectrum wherein the Focused and Normal emotional state, the difference in the form of the spectrum is very recognizable compared to the Sad and Shocked emotional states.

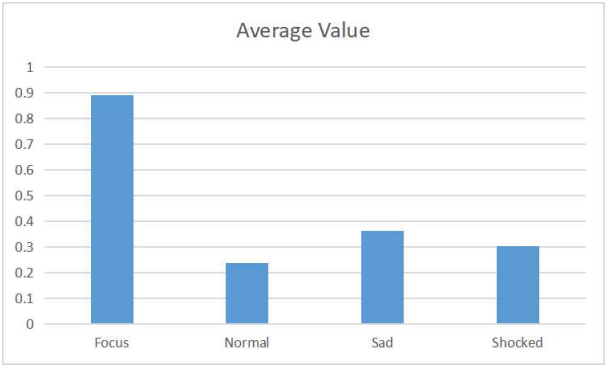


Figure 8: The average value of amplitude in the FFT spectrum

Then the amplitude value is compared to prove the truth of the observations. However, each spectrum with a frequency range of 0 - 50 Hz produces hundreds of thousands of data so that only the average value from data will be taken. Normal emotions get an average score of 0.2386, Focus 0.8904, Sad 0.3625, and Shocked 0.3044. Figure 8 shows the average value of the FFT spectrum.

From a comparison of the average values on each FFT spectrum, focus emotional state has the highest amplitude and lowest for Normal emotions, between Sad and Shocked emotions there is a small difference in the average results so that it can be proven the truth of observation result on the FFT spectrum.

4.3. Collecting Data Based on Spectrum Peaks

A reduction process is carried out at this stage at a frequency of 50 Hz to 10 Hz because the fluctuation of data values in the FFT spectrum always occurs at frequency range 0-10 Hz. Figures 9 to 12 show the results of the reduction process.

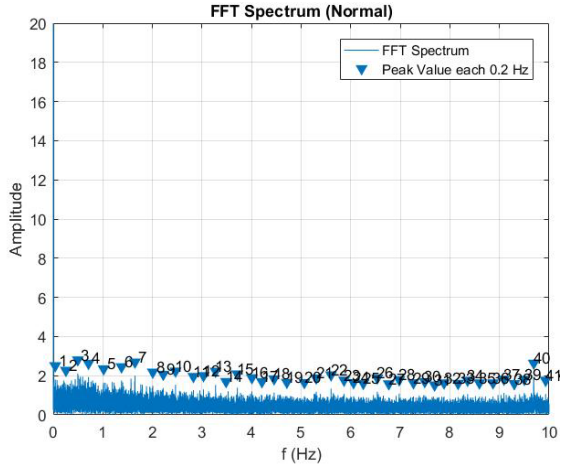


Figure 9: FFT Spectrum of Normal emotion

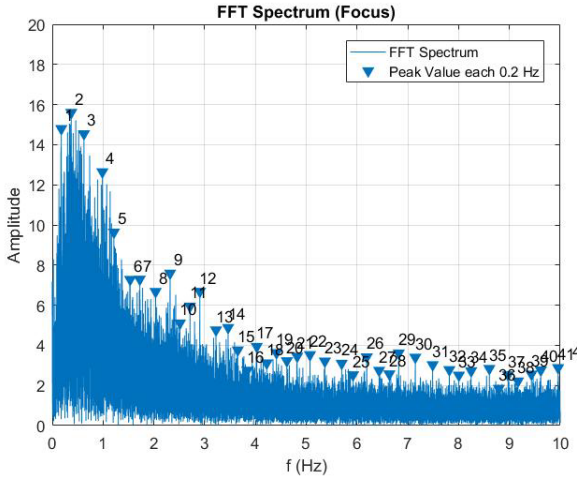


Figure 10: FFT spectrum of Focus emotion

The Values at the highest peaks of the marked frequency spectrum will be collected from the figure above. The frequency in the range 0 - 10 Hz produces 40 data by looking for the highest peak value every 0.2 Hz. Then, from 40 data, each 10 data's average value will be sought to produce four average values used

as material for the classification process. The results of the average value are shown in table 1.

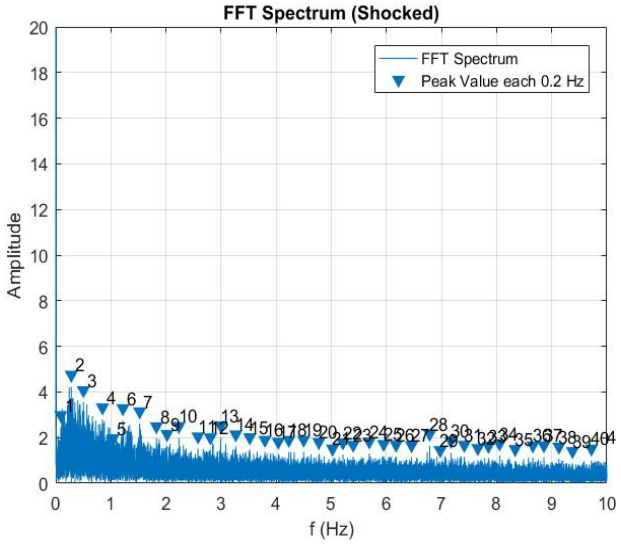


Figure 11: FFT spectrum of Shocked emotion

Table 1: Frequency average value

Focus	Normal	Sad	Shocked
8.891733187	1.908573527	4.199081623	3.059306615
3.371044125	1.2510798	2.052021058	1.633249441
2.851979474	1.124556395	1.583934658	1.250179144
2.597699756	1.202698894	1.554512291	1.136366638

If the reduction process is not carried out in the FFT spectrum, then based on the previous discussion (4.2), it is explained that the frequency in the range 50 Hz produces hundreds of thousands of EEG data and a minimum average value, which can complicate the classification process. We get between the results of previous observations and the current calculations to get very similar results from this process. It can be proven that the results of recording brain waves with the stimulus used to produce good data quality.

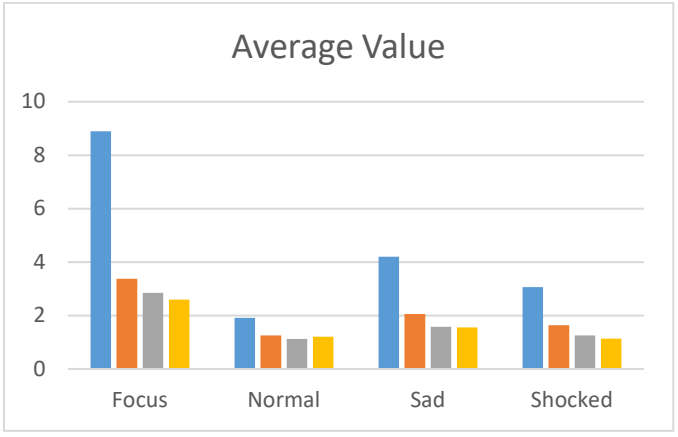


Figure 12: Average value of FFT based on spectrum peaks

4.4. KNN Classifier

In this process, 50 respondents studied and recorded brain waves four times obtained 200 data with 140 training data and 60 testing data. The KNN classification is then carried out to determine the results of the subject's actual emotional state from the 60-testing data.

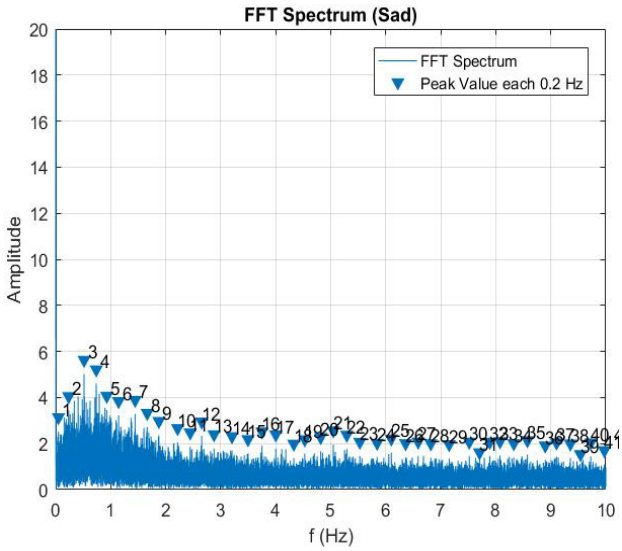


Figure 13: FFT spectrum of Sad emotion

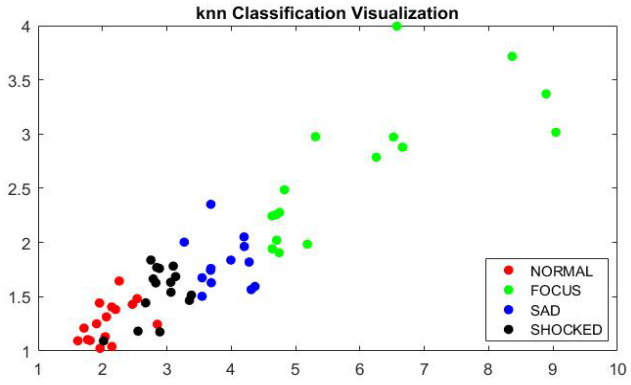


Figure 14: Visualization of KNN classification

The KNN classification results run through MATLAB2016b are displayed in a scatter chart and the coordinate points. It can be seen that each color at the coordinate point is a representation of the emotional condition of the subject, which forms a group based on the value obtained from 60 testing data. If observed, there are several color points whose coordinate values are closer to other neighboring groups. This indicates a change in the status of the color point. more clear results are shown in Table 2 below

Table 2: Result of KNN Classifier

	Match	Error	Match	Error	Match	Error
***k* Values**	**7**		**15**		**23**	
Normal	11	4	**13**	**2**	13	2
Focus	13	2	**14**	**1**	12	3
Sad	9	6	**12**	**3**	11	4
Shocked	12	3	**11**	**4**	12	3
Total	45	15	**50**	**10**	48	12

The table above shows that the KNN classification results using the k = 15 value are the best choice among k = 7 or k = 23. Sixty

test data produces 50 data that corresponds to the original emotional state status, and 10 data does not match. Then perform accuracy testing using the confusion matrix as in table 3 below.

Table 3: Accuracy result using confusion matrix

	Accuracy		
***k* Values**	**7**	**15**	**23**
Normal	73.33%	**86.66%**	86.66%
Focus	86.66%	**93.33%**	80.00%
Sad	60.00%	**80.00%**	73.33%
Shocked	80.00%	**73.33%**	80.00%
Average	74.99%	**83.33%**	79.99%

From the table above, the highest average accuracy is obtained from the results of the KNN classification using a value of k = 15 with a result of 83.33%, where the accuracy value of the focused emotional state gets the highest result of 93.33%, followed by a normal emotional state of 86.66%, 80% sad emotional state, and shocked get the lowest accuracy of 73.33%. Differences in accuracy results using the right stimulus dramatically affects the accuracy obtained. Regulations about sad emotions and focus are needed to help improve the accuracy of the results obtained.

5. Conclusion

This study aims to identify human emotions towards a given stimulus by applying the FFT method as feature extraction from the raw data generated by recording brain waves using the Neurosky Minwave mobile with the KNN method using the variable k = 15 to classify the FFT spectrum. Accuracy in each emotional state studied was performed using a confusion matrix where the best accuracy was obtained for Focus emotions as much as 93.33%, followed by Normal emotions 86.66%, Sad 80%, and Shocked getting the lowest accuracy results of 73.33%. The results of the average accuracy in the classification of brain waves were 83.33%. This research found that FFT and KNN can be used as alternatives in the EEG signal analysis method.

Acknowledgment

The research is funded by Ministry of Research and Technology / National Agency for Research and Innovation Fundamental Research with contract number: 1627.9/LL5/PG/2020

References

[1] A. Yudhana, S. Mukhopadhyay, I.R. Karas, A. Azhari, M.M. Mardhia, S.A. Akbar, A. Muslim, F.I. Ammatulloh, "Recognizing Human Emotion patterns by applying Fast Fourier Transform based on Brainwave Features," Proceedings - 1st International Conference on Informatics, Multimedia, Cyber and Information System, ICIMCIS 2019, 249–254, 2019, doi:10.1109/ICIMCIS48181.2019.8985227. This paper is an extension of work originally presented in International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS).

[2] I. Jayarathne, M. Cohen, S. Amarakeerthi, "BrainID: Development of an EEG-based biometric authentication system," 7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference, IEEE IEMCON 2016, (October), 2016, doi:10.1109/IEMCON.2016.7746325.

[3] Z. Pang, J. Li, H. Ji, M. Li, "A new approach for EEG feature extraction for detecting error-related potentials," 2016 Progress in Electromagnetics Research Symposium, PIERS 2016 - Proceedings, 3595–3597, 2016, doi:10.1109/PIERS.2016.7735380.

[4] B. Trowbridge, C. Rodriguez, J. Prine, M. Holzemer, J. McCormack, and R. Integlia, "Gaming, fitness, and relaxation," in Games Entertainment Media Conference (GEM), 2015 IEEE, 1–4, 2015.

[5] D. Bright, A. Nair, D. Salvekar, S. Bhisikar, "EEG-based brain controlled prosthetic arm," Conference on Advances in Signal Processing, CASP 2016, 479–483, 2016, doi:10.1109/CASP.2016.7746219.

[6] A. Yudhana, Sunardi, Priyatno, "Development of Door Safety Fingerprint Verification using Neural Network," Journal of Physics: Conference Series, **1373**(1), 2019, doi:10.1088/1742-6596/1373/1/012053.

[7] Y.P. Lin, C.H. Wang, T.P. Jung, T.L. Wu, S.K. Jeng, J.R. Duann, J.H. Chen, "EEG-based emotion recognition in music listening," IEEE Transactions on Biomedical Engineering, **57**(7), 1798–1806, 2010, doi:10.1109/TBME.2010.2048568.

[8] V. Vanitha, P. Krishnan, "Real time stress detection system based on EEG signals," Biomedical Research (India), 2016(Special Issue 2), S271–S275, 2016.

[9] S. Al Irfan, A. Yudhana, S.C. Mukhopadhyay, I.R. Karas, D.E. Wati, I. Puspitasari, "Wireless Communication System for Monitoring Heart Rate in the Detection and Intervention of Emotional Regulation," Proceedings - 1st International Conference on Informatics, Multimedia, Cyber and Information System, ICIMCIS 2019, 243–248, 2019, doi:10.1109/ICIMCIS48181.2019.8985210.

[10] P.D. Purnamasari, P. Yustiana, A.A. Putri Ratna, D. Sudiana, "Mobile EEG Based Drowsiness Detection using K-Nearest Neighbor," 2019 IEEE 10th International Conference on Awareness Science and Technology, ICAST 2019 - Proceedings, (June), 1–5, 2019, doi:10.1109/ICAwST.2019.8923161.

[11] M. Soleymani, S. Asghari-Esfeden, Y. Fu, M. Pantic, "Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection," IEEE Transactions on Affective Computing, **7**(1), 17–28, 2016, doi:10.1109/TAFFC.2015.2436926.

[12] V. Starcevic, D. Berle, K. Viswasam, A. Hannan, D. Milicevic, V. Brakoulias, E. Dale, "Specificity of the Relationships Between Dysphoria and Related Constructs in an Outpatient Sample," Psychiatric Quarterly, **86**(4), 459–469, 2015, doi:10.1007/s11126-015-9344-8.

[13] D. Nie, X.W. Wang, L.C. Shi, B.L. Lu, "EEG-based emotion recognition during watching movies," 2011 5th International IEEE/EMBS Conference on Neural Engineering, NER 2011, 667–670, 2011, doi:10.1109/NER.2011.5910636.

[14] R.A. Surya, A. Fadlil, A. Yudhana, "Identification of Pekalongan Batik Images Using Backpropagation Method," Journal of Physics: Conference Series, 1373(1), 2019, doi:10.1088/1742-6596/1373/1/012049.

[15] E.C. Djamal, M.Y. Abdullah, F. Renaldi, "Brain computer interface game controlling using fast fourier transform and learning vector quantization," Journal of Telecommunication, Electronic and Computer Engineering, **9**(2–5), 71–74, 2017.

[16] S. Sengupta, S. Biswas, S. Nag, S. Sanyal, A. Banerjee, R. Sengupta, D. Ghosh, "Emotion specification from musical stimuli: An EEG study with AFA and DFA," 2017 4th International Conference on Signal Processing and Integrated Networks, SPIN 2017, 596–600, 2017, doi:10.1109/SPIN.2017.8050019.

[17] A. Azhari, L. Hernandez, "Brainwaves feature classification by applying K-means clustering using single-sensor EEG," International Journal of Advances in Intelligent Informatics, **2**(3), 167–173, 2016, doi:10.26555/ijain.v2i3.86.

[18] R. Das, E. Maiorana, P. Campisi, "EEG Biometrics Using Visual Stimuli: A Longitudinal Study," IEEE Signal Processing Letters, **23**(3), 341–345, 2016, doi:10.1109/LSP.2016.2516043.

[19] N. Masood, H. Farooq, "Investigating EEG patterns for dual-stimuli induced human fear emotional state," Sensors (Switzerland), **19**(3), 1–23, 2019, doi:10.3390/s19030522.

[20] E.C. Djamal, D.P. Pangestu, D.A. Dewi, "EEG-Based Recognition of Attention State Using Wavelet and Support Vector Machine," 2016 International Seminar on Intelligent Technology and Its Applications (ISITIA), 139–144, 2016, doi: 10.1109/ISITIA.2016.7828648.

[21] C.S. Huang, C.L. Lin, L.W. Ko, S.Y. Liu, T.P. Su, C.T. Lin, "Knowledge-based identification of sleep stages based on two forehead electroencephalogram channels," Frontiers in Neuroscience, **8**(SEP), 1–12, 2014, doi:10.3389/fnins.2014.00263.

[22] R. Jaswal, "Brain Wave Classification and Feature Extraction of EEG Signal by Using FFT on Lab View," International Research Journal of Engineering and Technology, 1208–1212, 2016.

[23] A.E. Putra, C. Atmaji, F. Ghaleb, "EEG-Based Emotion Classification Using Wavelet Decomposition and K-Nearest Neighbor," Proceedings - 2018 4th International Conference on Science and Technology, ICST 2018, **1**, 1–4, 2018, doi:10.1109/ICSTC.2018.8528652.

[24] S. Shukla, R. Chaurasiya, Emotion Analysis Through EEG and Peripheral Physiological Signals Using KNN Classifier, Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB), 2019, doi:10.1007/978-3-030-00665-11

[25] B. Senzio-Savino, M.R. Alsharif, C.E. Gutierrez, K. Yamashita, "Synchronous emotion pattern recognition with a virtual training environment," Proceedings of the 2015 International Conference on Artificial Intelligence, ICAI 2015 – WORLDCOMP, 650–654, 2019.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1089-1095 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

5G, Vehicle to Everything Communication: Opportunities, Constraints and Future Directions

Boughanja Manale*, Tomader Mazri

Advanced Systems Engineering, Electrical Engineering, Networks and Telecommunications System, Ibn Tofail Science University, Kenitra, 14070, Morocco

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 14 November, 2020
Online: 14 December, 2020

Keywords:
5G
V2X
Security
Challenges

ABSTRACT

The 5G, as a new source of telecommunication infrastructure technology, has attracted many stakeholders to promote the progress of its standards and the development of its technology industry. The 5G reinforces new technologies and delivers vehicles for everything services (V2X) to drivers and passengers. It also offers several advantages. On the other hand, due to pervasive network connectivity, there are significant trust, security, and privacy issues for vehicles this affects the overall performance of 5G V2X. This paper provides a comprehensive review; and particularly; an overview of the mobile communication system and introduces the Next Generation (NG). additionally, the paper concentrates on the fifth-generation structure and its use cases. The 5G-V2X as a new field of communication requires more attention and focus, moreover, it aims to present the problems facing this new technology as well, and some future directions for research have been highlighted.

1. Introduction

Telecommunication changed radically in recent years, from traditional wired telephony oriented services to data-based services, from non-shrewd devices to smart handheld and versatile mobile computers.

Mobile ad hoc networks (MANET) formed a network without infrastructure because the mobile nodes in the network dynamically establish a path to transmit data packets. The strength of this network is its rapid development, robustness, flexibility, and mobility. With the rapid growth of this network and the massive focus on data creation through mobile phones and devices. Previous generations become unable to manipulate this data because they achieve their limits. For this purpose, a fifth-generation has been developed. The appearance of 5G has given several advantages to mobile involvement and some new technologies appear with it [1] to enhance communication between devices:

- Millimeter waves
- Small cells
- Massive multi-input multi-output (MIMO)
- Beamforming
- Full-duplex.

The 5G named also next-generation (NG), is emerging the network. Due to their capacity to handle large and diverse sets of devices, the NG is expected to bring a true wireless worldwide web (WWWW) which will have the ability to support massive data rates and a huge number of user applications and services. Moreover, it will reduce latency and will improve energy efficiency.

The 5G aims to create a networked society, where anyone can view and share information anytime from anywhere [2]. It will bring wireless connectivity for everything that is connected. Due to this fact, 5G faces some issues that can be summarized as follows:

- The number of connected devices has increased considerably;
- Traffic has increased significantly;
- Wide variety of applications with various characteristics and demands.

Our goal is to provide a detailed review where we will show the development brought to 5G and present some use cases and application of 5G, especially we will discuss the relation between 5G and V2X communication and application. This paper is an extension of an original work presented at the international conference of computer science and renewable energies (ICCSRE) [3]. In the previous article, we have shown that with the application of 5G the generation of data will increase. For that, it is essential to implement a dedicated storage method to manage this large

*Corresponding Author: Boughanja Manale, Salé,Morocco, 2126 68 81 32 32 & boughnja.manale@gmail.com

quantity of data. In this article, we will focus more on NG itself; explain the difference between 5G and prior generation, and some applications that allow generating this data. For this reason, our paper is organized as follows: In the first section, we will present an overview of the development of mobile networks. Then, we will introduce the 5G new radio access technologies in section two. We will provide some applications for 5G in section four. Further, in section five, we will introduce the 5G V2X application, and we will present the security issues that face the 5G V2X services. To address this security issue, we present some future directions for further studies to promote greater efforts on trust, confidence, assurance, and trustworthiness in 5G V2X services. Finally, we will conclude our paper.

2. Overview

So far, four generations of cellular communication systems have already been adopted [4]. Starting with the 1G that relies on an analog FM cellular system to 4G. With the huge interest in data creation and connected devices, the past generation achieved their limits. To deal with this problem the 5G has been created to solve the problem of the frequency spectrum cutting [5]. The 5G performances will develop later on with the progressive introduction of new technologies: use of millimeter waves, small cell, massive MIMO (or smart micro antennas), beamforming, and full-duplex [6], we will briefly explain those new advances.

- Massive MIMO: the concept was created in [7]. The purpose is to permit the system to achieve higher throughput. The massive MIMO can provide multiple services to various customers simultaneously in all directions.
- Millimetric waves: mobile wireless systems have been always in the band 300MHz - 3GHz. However, the spectrum becomes congested with the huge increase in mobile data traffic and the number of connected devices and the frequency band is coming to an end, creating a signal cutoff [8]. The proposed solution was to use millimeter waves. This latter refers to the wave between 30 and 300 GHz, and their wavelength varies between 1 and 10 mm [9].
- Small cells: are considered to be low-power radio switching equipment controlled by the administrator [10]. We can describe a small cell as a little power base station that will be very nearby to the usual towers and transmit the signal without having any break of connection or attenuation of the signal bypassing all obstacles.
- Beamforming: given the number of antennas used in 5G, the signal will be transmitted from the base station to the devices in an unmonitored manner [11]. This technique permits the orientating of the signal.
- Full duplex (FD): this technique brings to the device the capability to Send and get the data on the identical frequency band. FD improves the efficiency of the network.

The figure 1 summarizes those advances.

3. 5G new radio access technology

5G wireless is conceived to provide a connected environment, where information can be accessed and shared anywhere and anytime, by anyone and deliver a wireless connection for anything. However, with this requirement, the 5G will face several challenges like the volume of traffic the large variety of applications, and so on. To overcome this difficult task, 5G wireless access demands not just new functionality, but also a much more spectrum and larger frequency bands. The millimeter waves provide a huge spectral range, wide transmission bandwidth for the 5G [12]. The figure 2 shows the frequency spectrum.

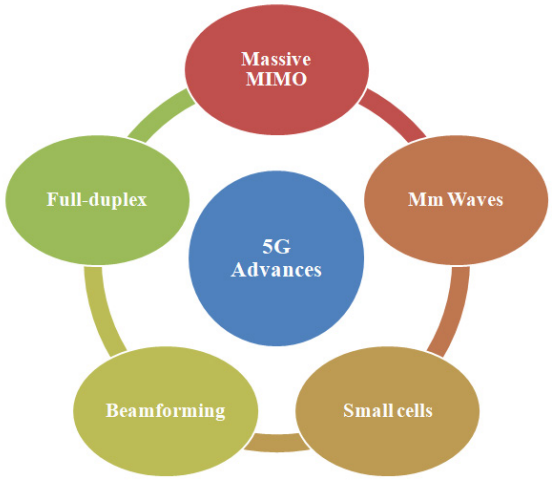


Figure 1: 5G Advances

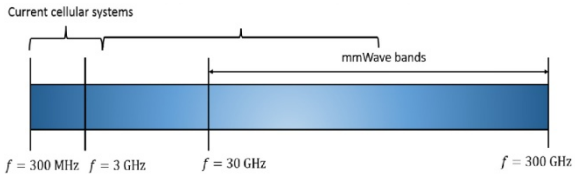


Figure 2: The frequency spectrum of the current and NG communication system

The past generations of cellular mobile systems were operating below 6 GHz frequencies, but with the implementation of the mm waves frequencies, much more spectrum is available and can deal with the huge amount of connected devices With the use of this principle in both 5G and Long Term Evolution (LTE), will have covered to have wireless access [13]. The appearance of 5G did not stop the evolution of LTE. LTE continues its evolution to meet the growing demand for 5G. This evolution is called the LTE evolution [14]. Figure 3 shows the LTE and 5G radio access.

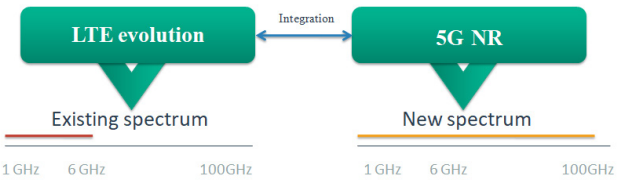


Figure 3: 5G radio access

3.1. Spectrum for 5G

Due to the lack of spectrum in the conventional mobile wireless band regulator started to envisage alternative approaches to manage the spectrum availability. For this reason, one of the major adjustments between the past generation of mobile communication and the fifth generation is the use of a higher frequency spectrum in the millimeter-wave interval [15]. It is about technological evolution and revolution, mixed to the spectrum. This approach is

made to make the spectrum more available with a higher frequency to exploit the frequency above 6 GHz and provide more bandwidth for mobile applications. The regulators have taken some measures for the proper functioning of the new generation among them:

- Spectrum sharing: technologies that will allow wireless broadband applications to operate bands assigned to multiple sites.
- Spectral efficiency: to minimize the bandwidth needed to supply a given quality of service level.
- Advanced MIMO: for higher modulation and lower signaling throughput.
- Densification: 5G communication system aims to get mobile customers closer to base stations, particularly by implementing a dense network of small cells.

All these developments reduce the amount of spectrum required for a single user, which indirectly leads to a greater spectrum availability to serve more users with an increased Quality of Services (QoS).

3.2. Use cases for 5G

5G will need to bring solutions for a variety of equipment running within a completely heterogeneous environment. The new human-centric applications include several fields: augmented reality, virtual reality, and gaming. This requires an extremely higher flow rate and reduced delay. In other words, we can differentiate two segments for machine-model communication:

- The Massive Internet of things (MIoT): which is categorized by a large number of low-cost device connections, in this case, a small amount of data supported by each device with extended battery lifetime and a wide range of coverage. This type is used more in the field of intelligent building, transport logistics, and agriculture.
- The critical internet of things (CIoT): shaped with highly dependable, extremely low-latency connectivity. For example, it can support autonomous vehicles, robotic surgery, and road safety.

Unlike previous products, 5G aims to bring improved support for a diversity of distinct services. 5G target three sets of use cases with very unique features:

- Enhanced Mobile Broadband (eMBB): This kind of use case is a direct extension of the conventional mobile broadband connection scenario based on the current mobile telecommunication standards. This set of boxes addresses broadband access everywhere, and in highly inhabited areas. This later provides the user with multiple services such as multimedia and 3D video [16].
- Large scale machine to machine communication (mMTC): with high latency and stability demands. This type of use case mainly relates machine to machine communication (MTC). It enables mobile communication with vehicles, this type requires simultaneous broadband and high mobility capacity. The problem with the mMTC that requires increased reliability and decreased latency [17].
- Ultra-Reliable Low Latency Communication (URLLC): the development of IoT is leading to a multiplication of wireless attached devices carrying TCM network traffic. It is intended that the expected number of these devices will soon surpass. The goal of mMTC is to provide the ability to connect a large set of equipment that may occasionally send a small quantity of data. mMTC devices are categorized by having long battery life and can be deployed remotely [16].

3.3. The key feature of 5G

The 5G was defined by the 3rd generation partnership project (3GPP) as a collection of key functionalities with associated parameters. In addition to the technological and protocolling environment and architecture inherited from the 4G system. We can summarize those technologies as follows:

- 3D MIMO: to ensure high performance in terms of spectral effectiveness. MIMO was firstly implemented in the LTE system was improved than the third generation [18]. The old system was composed of traditional antennas and composed of 8 to 10 elements stacked in a column, but with the evolution, restraint adjustments have been made and provide a downward tilt angle for network convergence. In this situation, the identical radio frequency (RF) pulses are delivered to everyone on the antenna. In such a manner, with the implementation of 3D MIMO, the signal will be simultaneously monitored in both the horizontal and vertical levels.
- Software-Defined Air Interface (IA): the fifth generation intended to handle various use scenarios and a broad range of the spectrum with a signal AI. For this reason, a unified framework was defined to handle different usage scenarios.
- User-Centric Network: to ensure the most suitable operator environment on the entire network, based on user requirements and the network conditions. Therefore, to provide a better user-centric experience the 5G network should handle a user context awareness where the network can serve the user more effectively when there is more information or expertise which means that the more information is loaded onto the network, the better the customer is served. Also, it should have the MEC functionality that includes traffic, security, transport isolation computing, and storage resources. Last and not least it should have a core network perspective to plan the use of the base stations.
- SDN/NFV: This technology enables the management and coordination of various network tranches on a shared physical infrastructure [19].

4. 5G applications

4.1. 5G and Mobile Ad Hoc Network

Mobile Ad Hoc Network (MANET) refers to a signal type of cellular network that contains mobile wireless nodes for communication [20]. These nodes are dynamically organized in random and volatile topologies. The utility to use the 5G for wireless communication in the field of MANET is to give more efficiency to the network. With the implementation of the NG, the MANET network achieves high levels of flexibility and intelligence. Figure 4 illustrates an example of MANET using the 5G connection.

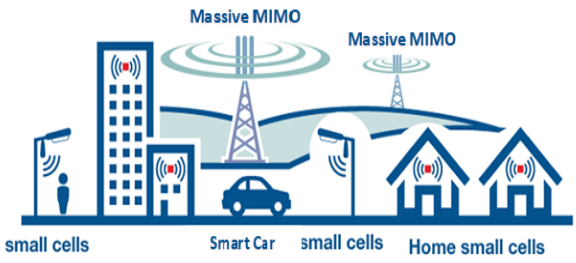


Figure 4: An example of MANET applied in 5G

The NG allows the connection of several devices and provides communication between multiple types of equipment. This communication can be classified into two different parts the device to device communication (D2D) and the vehicle for everything communication (V2X).

D2D communication: also known as proximity service communication (ProSe). The main objective of this functionality is to allow mobile devices to explore the presence of any other close devices within a radius of 500 meters and establish direct communication with it. The basic D2D procedures consist of three main steps:

- Device detection (ProSe discovery): detects if there are other devices nearby.
- Link parameters: establish a link between the associated devices.
- Data communication (ProSe communication): send or receive data via a set of links.

Figure 5 shows the D2D communication.

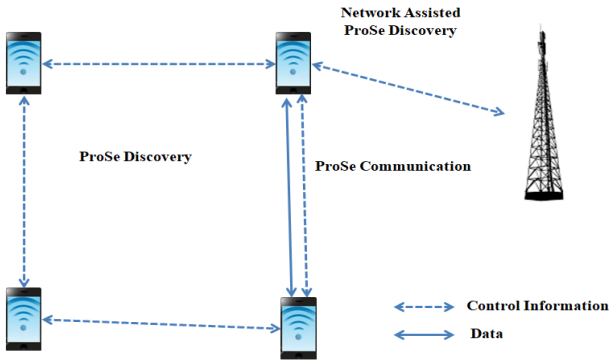


Figure 5: Basic ProSe procedure

V2X communication: are the exchanges between the vehicle and any device that may impact the vehicle, and the opposite. The X can for instance be on a different vehicle (V), an area of infrastructure (A/I), a pedestrian (P), a domicile (D), or any entity that communicates with the vehicle. We can summarize the cases of use of the vehicle with X as follows:

- Vehicular internet: user experience similar to that provided by their home and network.
- Pre-collision detection and mitigation: detect impending collisions and exchange relevant data from the vehicles involved.
- Cooperative vehicles: V2V and V2I communication to safely operate vehicles such as a train of autonomous cars on a highway.
- V2V information exchange: allows the vehicle to communicate on road to avoid any congestion.

5. 5G and V2X

The 5G mobile communication system will continue to emerge to meet unprecedented requirements beyond the capabilities of the past generations. The NG will not only be used for human interaction but also will be implemented in-vehicle communication. Therefore, the 5G has to support:

- Connection density: to handle million of demands and links across the square kilometer.
- Continuous latency from end to end.
- The density of road traffic volume: several dozens of terabytes created each second per square km.
- Mobility and maximum data rate.
- Network efficiency and high network capacity.

Continuous improvement and the increase in the number of applications have led to the development of new usages, in particular, that of connected objects. Vehicular communications or V2X communications allow vehicles to communicate in real-time with their environment to enhance road control, traffic performance, and energy savings. Figure 6 illustrates the different modes of communication, grouped under the term V2X.

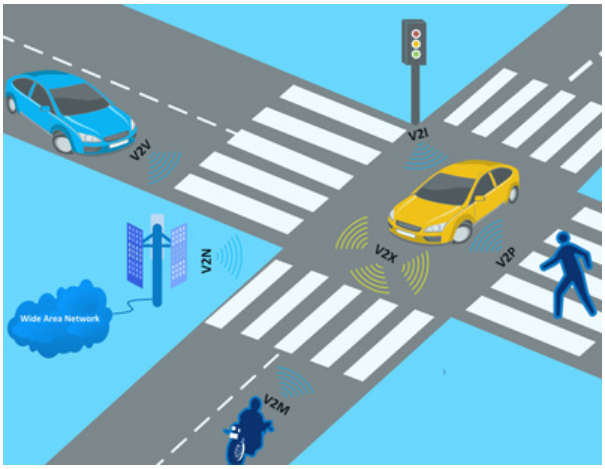


Figure 6: V2X communication modes

Embedded 5G will reform the experience of users in the transport sector (vehicles, subways, planes, trains, etc.), in particular by developing internet connectivity, which will make it possible to create an environment connected to the vehicle. V2X (Vehicle to Everything) communication technology will enable enormous changes the almost instantaneous exchange can be summarized as follows:

- Vehicle-to-vehicle (V2V): that permit inter-vehicle communication.
- Vehicle to infrastructure (V2I): it another type of communication that allows the vehicle to communicate with

its surrounding like the communication between the roadside unit and the vehicle.

- Vehicle to pedestrian (V2P); that permits the exchange between the vehicles and the pedestrian into the road.
- Vehicle to motorcycle (V2M): Communication between vehicles and motorcycles to decrease the high rate of collisions and deaths in motorcycle crashes.

The concept of a connected vehicle is based on sensors that operate in line-of-sight such as radars, cameras, or laser remote sensing. The 5G will play a key role in making road transport more efficient and safe. Indeed, the automotive industry will require a more robust, stable, and low-latency network.

5.1. 5G V2X architecture

The structure of the fifth generation and vehicle to everything architecture is based on four network levels:

- 5G Access Network;
- Network Edge;
- 5G Core Network;
- Data Network.

The 5G Access Network is composed of a Next Generation Radio Access Network (NG-RAN) or/and a non 3GPP access network, which interconnects the 5G Core Network and the User Equipment (UE). Figure 6 shows the 5G V2X architecture:

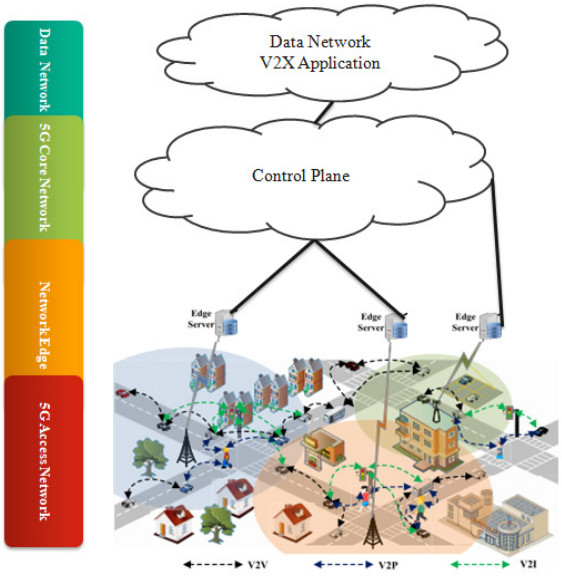


Figure 7: 5G V2X architecture

5.2. Security issues in 5G V2X services: problems and risks

Even though 5G guidelines are as yet being defined it will confront a few difficulties the moment of its implementation. The new radio (NR) will require a steady improvement in security. In this specified situation, portable security and oversaw security services are one of the most escalated areas regarding advancement, to react to the difficulties in the field of digital security inborn to computerized change for the communication service providers and their clients. Moreover, 5G vows to reshape services and advance development by providing an expanded transfer speed, multitenant edge processing, and a few creative services, high-data transmission networks, and high-bandwidth connectivity to machine learning and artificial intelligence (AI). The insurance of services in 5G V2X requires ground-breaking, versatile, and coordinated security. The implementation of 5G V2X brings with it several security threats. The attack on the vehicle can lead to serious problems and can affect all sides of the security requirement

- Attacks on availability: the attacks that focus on the availability of data are the riskiest in V2X communication since they have a real impact on the basic circumstances of well-being. We can find attacks such as the jamming attack which is a radio interference attack that can be carried out across any communication based on wireless technology.
- Attack on data integrity: where the attacker intercepts the data circulated into the network. the attacker can launch attacks such as GPS spoofing or even replay attacks.
- Attack on confidentiality: as an important security requirement in V2X communication and all sent data in the V2X should be secured. An attack such as eavesdropping and location tracking are launched in the network to break through the V2X confidentiality
- Attack on authenticity: Assuring authenticity consists of the process to verify that nodes are granted access to the network according to their identities using certificates and digital signatures. Some threats that infringed this requirement are certificate replication where the nodes use replicated certificates that were added to the blacklist. Moreover, the Sybil attacks where one node (the attacker node) pretends to be many identities in the road network.
- Attack on the non-repudiation: Non-repudiation is concerned with the ID of the effective node that is performing a particular behavior.

The table below shows some attacks that impact the security requirement:

Table 1: Threats of security for 5G V2X

Security Requirement	Attack	External	Internal
availability	Blackhole Grey hole	✓	×
	Flooding	✓	×
	Jamming	×	×
	Coalition	✓	×
data integrity	Messages Alteration	✓	×
	False messages injection	✓	×
	Replay	✓	✓
	GPS Spoofing	×	×
confidentiality	Eavesdropping	✓	×

	Location tracking	×	×
authenticity	Certificate replication	✓	✓
	Sybil	✓	✓
non-repudiation	Masquerading impersonation	✓	✓

5.3. Security solutions for 5G V2X communication

Vehicular communication experiences numerous difficulties, hence, guaranteeing safe communication is viewed as a complicated process due to the lack of safety features in both empowering innovations of the vehicle to everything interchanges, an additive safety scheme is required to secure vehicle networks. The fields of security for 5G V2X was not deeply discussed. In this section, we will try to gather the mean full strategy that has been implemented to secure the 5G V2X communication. In [21], they proposed a lightweight and protected navigation system for VANET. Each vehicle implements a cryptographic and digital signature to manage the secured correspondence process from the vehicle to the roadside unit (RSU). The system dealt with the replay attack and the man-in-the-middle attack. Another solution was proposed by [22], consisting of a trusted and smart forwarding protocol using double ciphering for data packets and implemented strong verification schemes to gauge the trust of the node. In [23], they proposed a new technique depends on a one-time Identity-Based confirmed asymmetric gathering key consent to make cryptographic blend areas that oppose malicious Eavesdroppers. There are several types of research have proposed potential solutions for 5G from a machine learning (ML) perspective. In [24], the authors suggested a monitoring system employing a Support Vector Machine (SVM) learning algorithm to prevent the replay attack. Also in [25] an anti-jamming methodology for auxiliary users to choose the correspondence channel and mobility. Another researcher [26], based on deciding the arrangements of conceivable interfacing neighboring vehicles, and design the boundaries reserving the joint V2V network. In [27], they have introduced a trusted and standard secure signature solution for collecting plain text for vehicle system applications. The authors in [28], have suggested an algorithm for VANET that delivers authenticate delivered messages. He gave oversight roles to some specified nodes referred to as verification nodes. The authors in [29], suggested a model framework as an IDS based on compartmental attacks. it addressed various attacks in VANET such as the black hole, wormhole, and Sybil attacks.

6. Discussion

This article presents the state of the art on 5G technology applied to transportation to make roads safer. The security of 5G V2X communication presents a great challenge, for this reason, a security layer should be added to strengthen security, and the communication must respect the CIA principle (confidentiality, integrity, and availability):

- Confidentiality concerns the condition under which the vehicle user has the right to know what information is sent by the On-Board Unit (OBU) and for how long this information exists.
- Integrity to verify that all data has been transferred.
- Availability, to ensure that the protocols must remain operational even in the event of a failure, and remains fault-tolerant and stabilize once the failure has been eliminated.

In addition to two other aspects: privacy, this technique implies the preservation of the privacy of the driver and the vehicle. Also, the authentication which consists of ensuring that the intended recipient receives messages from a real sender.

As we have already presented, many applications and solutions have been implemented to ensure the security of 5G V2X communication to preserve and maintain the security of users. We can summarize these in three categories:

- Identity-Based Cryptography (IBC): The concept of IBC is a cryptographic scheme very suitable for devices with limited computing resources. It has been presented in many articles, to encrypt and effectively sign the messages interchanged.
- Behavior-Based: The concept is quite simple this solution is based on the behavior of the attached vehicles communicating in the network and the messages transmitted by the RSU.
- Machine Learning-Based (ML): This provides threat and attack prediction based on suspicious network activities and anomaly detection. The implementation of ML enables automation, adaptability, and enables efficient orchestration and dynamic provisioning for scaled network resources cognition and prediction of mobility

Last and not least the security of such a network can be done depending on four parameters:

- Arrangement: shows if the model is mounted on a flat or cluster network.
- Structure: specifies if the model has used structured, distributed, or hybrid.
- Purpose: specify if the goal of the system is an IDS, the implementation of a safe route, or the protection of the message content.
- Targeted attacks: specifies the attacks handled in the model.

7. Open issues and future orientations

Open issues and future perspectives were highlighted in this section. Securing the 5G V2X is a quite difficult challenge, For this reason, both security solutions and possible cures for identified and unidentified menaces rely on 5G-V2X implementation approaches. We can divide two approaches for 5G V2X. The first one the V2X is implemented with non-autonomous-5G (NA-5G), possible assaults on LTE-V2X are real and can leverage 5G-V2X capabilities. However, with the second implementation Autonomous-5G (A-5G), the attack window diminishes and safeguarding from threats can be augmented. Network deployment plays a critical factor in the 5G-V2X security policy decision. for this purpose, we have encountered the security needs at two levels in the 5G V2X field:

- Secure network caching at the network edge: Storing commonly demanded data in the cache at the edges of storage is an important and effective way to decrease the delay of V2X operations. Yet, it is difficult to maintain the privacy of stored and buffered content, from which opponents can determine which are the relevant services for a given target. To ensure

network caching is secure, these key issues need to be resolved: knowing when and where to cache the desired data for a given vehicle, choosing trusted peripheral devices for data storage, and ensuring the confidentiality of buffered data when peripheral nodes are under attack. For this reason, cache placement could be well planned.

- Security-enhanced network slicing: The security and confidentiality of the grid segmentation have not yet gained a lot of attention. Therefore, it is essential to perform both authentication and authorization for a specified part of a network. For this reason, it is crucial to guarantee authentication with low latency to V2X services.

8. Conclusion and perspectives

Five generation and V2X services have expanded tremendously and have benefited our day-to-day lives. While providing a range of advantages to conductors and passengers on the road, 5G V2X's connectivity has as well raised several concerns about security, trust, and privacy. In this article, we have presented the fifth generation and its characteristics. Besides, we have identified the key challenges that face this new generation. We also briefly outlined the main issues in securing the 5G V2X by determining its issues of reliability, safety, and privacy. Finally, we brought some open issues and future orientations in this field. This latter should have more attention in the future our perspective is to build a system that will be able of securing the communication between the vehicle and everything basing on the 5G connection.

References

[1] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfröjd, T. Svensson, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," IEEE Communications Magazine, **52**(5), 44–51, 2014, doi:10.1109/MCOM.2014.6815892.

[2] E. Dahlman, G. Mildh, S. Parkvall, J. Peisa, J. Sachs, Y. Selén, "5G radio access," Ericsson Review (English Edition), **91**(1), 42–47, 2014.

[3] B. Manale, M. Tomader, "A survey of different storage methods for NGN mobile networks," in Proceedings of 2019 International Conference of Computer Science and Renewable Energies, ICCSRE 2019, 2019, doi:10.1109/ICCSRE.2019.8807526.

[4] M. Iwamura, "NGMN view on 5G architecture," IEEE Vehicular Technology Conference, **2015**, 2015, doi:10.1109/VTCSpring.2015.7145953.

[5] T.L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," IEEE Transactions on Wireless Communications, **9**(11), 3590–3600, 2010, doi:10.1109/TWC.2010.092810.091092.

[6] A. Gupta, R.K. Jha, "A Survey of 5G Network: Architecture and Emerging Technologies," IEEE Access, **3**, 1206–1232, 2015, doi:10.1109/ACCESS.2015.2461602.

[7] S. SEN, J. CLARK, J. TAPIADOR, Security Threats in Mobile Ad Hoc Networks, 2010, doi:10.1201/ebk1439819197-9.

[8] M.Á. Vázquez, L. Blanco, A. Pérez-Neira, "Sparse beamforming for mmWave spectrum sharing systems," European Signal Processing Conference, **2018-Septe**(1), 1517–1521, 2018, doi:10.23919/EUSIPCO.2018.8553148.

[9] E. Al Abbas, A.T. Mobashsher, A. Abbosh, "Polarization reconfigurable antenna for 5G cellular networks operating at millimeter waves," Asia-Pacific Microwave Conference Proceedings, APMC, 772–774, 2017, doi:10.1109/APMC.2017.8251562.

[10] M. Tayyab, X. Gelabert, R. Jantti, "A simulation study on handover in LTE ultra-small cell deployment: A 5G challenge," IEEE 5G World Forum, 5GWF 2019 - Conference Proceedings, 388–392, 2019, doi:10.1109/5GWF.2019.8911616.

[11] T. Varum, A. Ramos, J.N. Matos, "Planar microstrip series-fed array for 5G applications with beamforming capabilities," 2018 IEEE MTT-S International Microwave Workshop Series on 5G Hardware and System Technologies, IMWS-5G 2018, 2018, doi:10.1109/IMWS-5G.2018.8484697.

[12] E. Dahlman, S. Parkvall, "NR - The new 5G radio-access technology," IEEE Vehicular Technology Conference, **2018-June**, 1–6, 2018, doi:10.1109/VTCSpring.2018.8417851.

[13] X. Lin, A. Grovlen, K. Werner, J. Li, R. Baldemair, J.F.T. Cheng, S. Parkvall, D.C. Larsson, H. Koorapaty, M. Frenne, S. Falahati, "5G New Radio: Unveiling the Essentials of the Next Generation Wireless Access Technology," IEEE Communications Standards Magazine, **3**(3), 30–37, 2019, doi:10.1109/MCOMSTD.001.1800036.

[14] O. Teyeb, G. Wikström, M. Stattin, T. Cheng, S. Faxér, H. Do, "Evolving LTE to Fit the 5G Future," Ericsson Technology Review, 1–17, 2017.

[15] S. Parkvall, E. Dahlman, A. Furuskär, M. Frenne, "WIRELESS AND RADIO COMMUNICATION NR : The New 5G Radio Access Technology," IEEE Communications Standards Magazine, **1**(December), 24–30, 2017, doi:10.1109/MCOMSTD.2017.1700042.

[16] M.A. Abu-Rgheff, 5G Physical Layer Technologies, 2019, doi:10.1002/9781119525547.

[17] B. Badic, C. Drewes, I. Karls, M. Mueck, Rolling out 5G: Use cases, applications, and technology solutions, 2016, doi:10.1007/978-1-4842-1506-7.

[18] G. Liu, Y. Huang, F. Wang, J. Liu, Q. Wang, "5G features from operation perspective and fundamental performance validation by field trial," China Communications, **15**(11), 51–61, 2018, doi:10.1109/CC.2018.8543047.

[19] J. Ordonez-Lucena, P. Ameigeiras, Di. Lopez, J.J. Ramos-Munoz, J. Lorca, J. Folgueira, "Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges," IEEE Communications Magazine, **55**(5), 80–87, 2017, doi:10.1109/MCOM.2017.1600935.

[20] R. Bruzgiene, L. Narbutaite, T. Adomkus, "MANET Network in Internet of Things System," Ad Hoc Networks, (May 2017), 2017, doi:10.5772/66408.

[21] G. Li, M. Ma, C. Liu, Y. Shu, "A lightweight secure VANET-based navigation system," 2015 IEEE Global Communications Conference, GLOBECOM 2015, 2015, doi:10.1109/GLOCOM.2014.7417462.

[22] S.K. Bhoi, P.M. Khilar, "A Road Selection Based Routing Protocol for Vehicular Ad Hoc Network," Wireless Personal Communications, **83**(4), 2463–2483, 2015, doi:10.1007/s11277-015-2540-x.

[23] L. Zhang, "OTIBAAGKA: A New Security Tool for Cryptographic Mix-Zone Establishment in Vehicular Ad Hoc Networks," IEEE Transactions on Information Forensics and Security, **12**(12), 2998–3010, 2017, doi:10.1109/TIFS.2017.2730479.

[24] Q.G. Fan, L. Wang, Y.N. Cai, Y.Q. Li, J. Chen, "VANET Routing Replay Attack Detection Research Based on SVM," MATEC Web of Conferences, **63**, 4–7, 2016, doi:10.1051/matecconf/20166305020.

[25] G. Han, L. Xiao, H.V. Poor, "Two-dimensional anti-jamming communication based on deep reinforcement learning," ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2087–2091, 2017, doi:10.1109/ICASSP.2017.7952524.

[26] L.T. Tan, R.Q. Hu, "Mobility-aware edge caching and computing in vehicle networks: A deep reinforcement learning," IEEE Transactions on Vehicular Technology, **67**(11), 10190–10203, 2018, doi:10.1109/TVT.2018.2867191.

[27] M.S.I. Mamun, A. Miyaji, H. Takada, "A multi-purpose group signature for vehicular network security," Proceedings - 2014 International Conference on Network-Based Information Systems, NBiS 2014, 511–516, 2014, doi:10.1109/NBiS.2014.93.

[28] U. Khan, S. Agrawal, S. Silakari, "Detection of Malicious Nodes (DMN) in vehicular ad-hoc networks," Procedia Computer Science, **46**(Icict 2014), 965–972, 2015, doi:10.1016/j.procs.2015.01.006.

[29] H. Sedjelmaci, S.M. Senouci, "An accurate and efficient collaborative intrusion detection framework to secure vehicular networks," Computers and Electrical Engineering, **43**, 33–47, 2015, doi:10.1016/j.compeleceng.2015.02.018.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 1096-1102 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Development of Secondary Processing Data Methods under Single Point Thunderstorm Activity Monitoring

Anatoly Panyukov[*,1], Alexander Bogushov[2]

[1]*School of Electronic Engineering and Computer Science, South Ural State University, 454080, Russian Federation*

[2]*Institute of Natural Sciences and Mathematics, South Ural State University, 454080, Russian Federation*

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 25 November, 2020
Online: 14 December, 2020

Keywords:
single-point lightning sensor
pre-storm radiation
equivalent electric dipole
thunderstorm simulation
passive monitoring
visualization

ABSTRACT

Forecasting the occurrence and development of thunderstorms near insured mobile objects, such as aircraft carriers, oil tankers, requires the usage of appropriate single-point monitoring systems (SPMS) of thunderstorm activity. These systems must be small and included in the equipment of the objects. A SPMS based on EH location finder meets the requirements of small dimensions. But the primary processing algorithms for them used to determine the location of the radiation source lead to errors in the calculation of the bearing per atmospheric discharge. This inevitably occurs due to the presence of an anomalous component of the magnetic field of an arbitrarily oriented dipole. The error depends on the orientation of the equivalent dipole of the radiation source. If the dipole moment is vertical, the error is minimal, otherwise the direction finding error is completely determined by the horizontal projection of the equivalent dipole and can be equal to π radians. Information of a thunderstorm activity collected by SPMS can be used by secondary processing algorithms to identify a real location of the thunderstorm foci. It is especially important when processing pre-storm radiation which consist of intra- and inter-cloud discharges. Secondary processing algorithms reduce errors by analyzing the parameters of all registered lightning discharges over a certain time interval. The proposed algorithms divide the entire space into a set of three-dimensional cells and determine the their grade of membership to a thundercloud or its projection. The presented methods for analyzing the location parameters of radiation sources defined across the entire set of received signals and ways to display them reduce the uncertainty of determining the coordinates of the thunderstorm location. Processing of pre-thunderstorm radiation is essential for predicting the development of thunderstorms.

1 Introduction

Predicting the occurrence and development of a thunderstorm near the insured object requires the development of appropriate single-point systems for monitoring thunderstorm activity. For mobile insured objects, such as aircraft carriers, oil tankers, these systems must be small and included in the equipment of the objects [1]. Primary processing algorithms which used to determine location of the radiation source can lead to errors in calculation the bearing of discharge. It inevitably occurs in a single-point monitoring systems (SPMS). The error depends on the orientation of the equivalent dipole of the radiation source. If the dipole moment is vertical then the error is minimal, otherwise the direction finding error is completely determined by the horizontal projection of the equivalent dipole and can be equal to π radians.

Information of a thunderstorm activity collected by a SPMS can be used by secondary processing algorithms to identify a real location of the thunderstorm foci. It is particularly critical when processing pre-storm radiation which consist of intra- and inter-cloud discharges. Secondary processing algorithms can reduce the uncertainty of determining the coordinates of the thunderstorm location by analyzing the location parameters of radiation sources that found for a time interval. This approach has been announced at "2019 Russian Open Conference on Radio Wave Propagation (RWP), Kazan, Russia" [2] in the field of study that deals with electromagnetic methods of lightning detection. This paper is an original extension of this work.

Each cloud and ground lightning flash consist of many individual physical processes. These processes can be associated with

[*]Corresponding Author: Anatoly Panyukov, Email: anatoly.panyukov@gmail.com

characteristic electric and magnetic fields. Lightning emits significant electromagnetic energy in the radio-frequency range from below 1 Hz to near 300 MHz. This electromagnetic radiation can be detected by lightning locating systems. Lightning data have utility in different areas of importance meteorological service organizations. Typically, meteorological organizations use this data to protect life and property. Commercial organizations use lightning data to provide weather warnings, forecasts to aviation, agriculture, energy, mass media.

There are plenty of lightning locating systems that uses different radio-frequency ranges, for example, the very low-frequency (VLF) range from 3 to 30 kHz, the low-frequency (LF) range from 30 to 300 kHz. Accurate lightning locating systems necessarily employ multiple sensors. But single-point passive monitoring system are also popular as they are inexpensive, mobile and do not require communication. Many companies are engaged in the production of such devices [3]–[6].

A challenging problem which arises in this domain is processing pre-thunderstorm radiation (that is before the first lightning flash). During the project ISTC 1822 [7] were recorded 2.5 million atmospherics and less than 10% were defined as a lightning discharge, the rest radiation were classified as pre-thunderstorm radiation. Algorithms used in multiple sensors systems require registration of the atmospherics by several sensors, it is not always possible. Otherwise, single-point systems generally filtered out this data or it processed incorrectly.

Algorithms to determine location of the radiation source uses the magnetic direction finding (MDF) technique to calculate the bearing of discharge. It works really well for multistation systems especially with combination with time of arrival (TOA) technique. But using MDF technique in SPMS leads to significant errors for the discharges whose equivalent dipole has an orientation different from vertical.

In general, the results of the project showed possibility of creating a new generation of passive thunderstorm detection systems to extend their functionality. One way to overcome these problems is not only revision of the technical characteristics of sensors, but also development of new mathematical models and algorithms for analyzing thunderstorm activity. The problem of identifying the position of thunderstorm foci is practically important to forecasting the development of thunderstorm. The registration of the cloud radiation by a single-point lightning protection system, for the purposes of forecasting thunderstorm development, can be reached by expansion of receiving hardware dynamic range and development of mathematical and software.

The standard model for the lightning discharge is based on a vertical dipole that cannot adequately represent all discharges. Modern methods of analyzing the electromagnetic radiation source (EMRS) field allow to determine the parameters of the EMRS location by many algorithms based on the dipole as a model of the EMRS [1], [8]–[10]. A more correct approach is to simulate a lightning discharge as a arbitrarily oriented dipole and it possible by using special a group of $E-H$ methods that allow arbitrary orientation of dipoles [11, 12]. These algorithms allow us to estimate the position parameters of this dipole: the distance to the radiation source and the pseudo bearing, and can be used for processing pre-thunderstorm radiation. In this case, signals from various sources of electromagnetic radiation are registered, the appropriate mathematical model of which is an arbitrarily oriented electric dipole. Processing signals from dipoles with a predominant horizontal projection leads to uncertainty in the estimation of the bearing on the [2, 12] discharge, which significantly reduces the quality of dipoles location detection. One way to reduce this uncertainty is to analyze the information received from all dipole sources.

The aim of the work is describing approach for the secondary processing of monitoring results of thunderstorm activity in VLF range, allowing to get plausible estimates of the location of a thunderstorm center based on its pre-storm radiation.

A review of the status of passive storm monitoring systems and the demonstration of the use of lightning-position detection systems for the passive radar of hazardous meteorological phenomena are presented in [8]. Modern methods for analyzing the field, allowing to determine the parameters of the EMRS, characterizing its location and orientation are presented in the works [12].

2 Primary processing of EMRS signal

Figure 1 shows the geometry of the problem: the earth's surface is plane Oxy, the observation point is origin O, the polar coordinates of the radiation source are r, ψ, θ , and the angular coordinates of the radiation source dipole moment are ψ_0, θ_0.

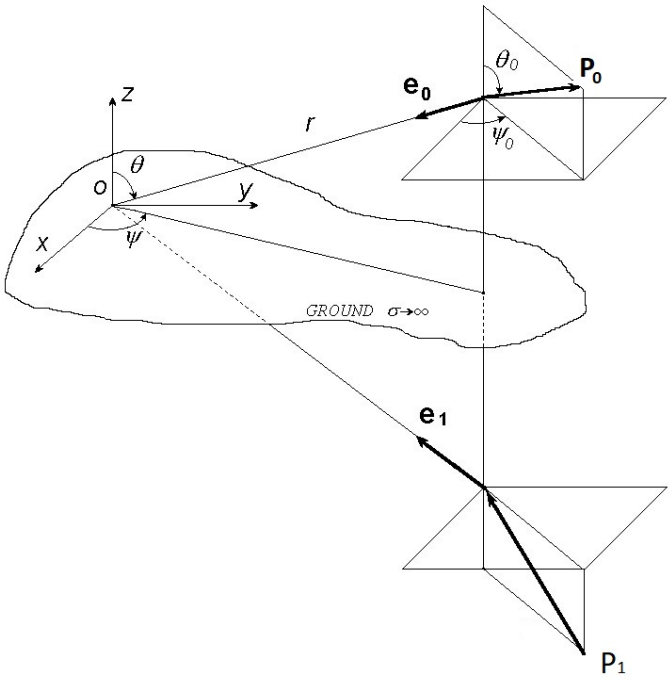


Figure 1: Geometry of the problem

There are a dipole $\mathbf{P_0}$ and an imaginary dipole $\mathbf{P_1}$ which is the mirror image of the dipole $\mathbf{P_0}$. $\mathbf{P_0}$ results a field $(\mathbf{E}, \mathbf{H})$ consisting of the field ($\mathbf{E}(\mathbf{P_0})$, $\mathbf{H}(\mathbf{P_0})$) of the source $\mathbf{P_0}$, and the reflected field represented by the field ($\mathbf{E}(\mathbf{P_1})$, $\mathbf{H}(\mathbf{P_1})$) of the $\mathbf{P_1}$.

The electric $\mathbf{E}$ and magnetic $\mathbf{H}$ components of the field in terms of the Cartesian system of coordinates with origin at the observation

point O and the O_z axis being the normal of the bounding plane have representation [9]

$$E_x = E_y = H_z = 0, \tag{1}$$

$$E_z = \frac{1}{2\pi\epsilon_0 c^2 r}\left\{p''[\cos\theta\sin\theta\sin\theta_0\cos(\psi-\psi_0) - \sin^2\theta\cos\theta_0] + \right.$$
$$\left. +(p'\alpha + p\alpha^2)[3\sin\theta\cos\theta\sin\theta_0\cos(\psi-\psi_0)]\right\} =$$
$$= \frac{w}{2\pi\epsilon_0 c^2 r}\left[p''u + (p'\alpha + p\alpha^2)(3u-2v)\right], \tag{2}$$

$$H_x = \frac{1}{2\pi c^2 r}(\sin\theta\cos\theta_0\sin\psi - \cos\theta\sin\theta_0\sin\psi_0)\,(p'' + p'\alpha) =$$
$$= \frac{w\sin\varphi}{2\pi c^2 r}\,(p'' + p'\alpha), \tag{3}$$

$$H_y = \frac{1}{2\pi c^2 r}(-\sin\theta\cos\theta_0\cos\psi + \cos\theta\sin\theta_0\cos\psi_0)(p'\alpha + p'') =$$
$$= -\frac{w\cos\varphi}{2\pi c^2 r}(p'\alpha + p''), \tag{4}$$

$$\alpha = \frac{c}{r}, \quad u = \sin\theta\cos(\phi-\psi), \quad v = \frac{\sin(\phi-\psi_0)}{\sin(\psi-\psi_0)}, \tag{5}$$

$$\tan\varphi = \frac{\sin\theta\cos\theta_0\sin\psi - \cos\theta\sin\theta_0\sin\psi_0}{\sin\theta\cos\theta_0\cos\psi - \cos\theta\sin\theta_0\cos\psi_0}, \tag{6}$$

c is speed of light in vacuum, ϵ_0 is permittivity of vacuum.

The values of the parameters r, φ, u, v are measured as a result of primary processing, for each pulse signal received from a radiation source [9]. The quality of estimating the location of an individual source is quite low. Really, a vertical dipole (that is when $\theta_0 = 0$) have

$$\varphi = \psi, \quad \sin\theta = u = \frac{1}{v}.$$

In other cases (that is for an inclined dipole) there is an uncertainty in determining the coordinates ψ and θ. As shown in the article in [9] they are bounded by

$$\arcsin u \le \theta \le \frac{\pi}{2}, \tag{7}$$

$$\varphi - \arccos u \le \psi \le \varphi + \arccos u. \tag{8}$$

The probability density for angles ψ and θ are

$$f_\psi(\eta) = \frac{s}{2\widehat{\psi}\,[s^2\cos^2(\varphi-\eta) + \sin^2(\varphi-\eta)]},$$
$$-\arccos u \le \eta - \varphi \le \arccos u, \tag{9}$$

$$f_\theta(\zeta) = \frac{su\cos\zeta\sin\zeta}{\widehat{\psi}\,(s^2u^2 - u^2 + \sin^2\zeta)\cdot\sqrt{\sin^2\zeta - u^2}},$$
$$\arcsin u \le \zeta \le \frac{\pi}{2}. \tag{10}$$

where

$$s = \left|\frac{1-vu}{uv}\right|, \quad \widehat{\psi} = \arctan\left(\frac{1}{s}\tan\arccos u\right). \tag{11}$$

It's easy to see that for $s < 1$, that is for θ_0 approaching the vertical,

$$\varphi = \arg\max_{\eta:\,-\arccos u\le\eta-\varphi\le\arccos u} f_\psi(\eta), \tag{12}$$

$$\arcsin u = \arg\max_{\zeta:\arcsin u\le\zeta\le\frac{\pi}{2}} f_\theta(\zeta), \tag{13}$$

in other words, the maximum likelihood estimates are the same as for the vertical dipole: $\psi = \varphi$ and $\theta = \arcsin u$; and for $s \ge 1$, that is for orientation with a significant horizontal component, poor angular resolution is gained.

An adequate mathematical model of the pre-thunderstorm EMRS is an inclined electric dipole with a predominance of the horizontal component. Processing signals from dipoles with a predominant horizontal projection leads to uncertainty in the estimation of the angle coordinates of the [2, 9] discharge, which significantly reduces the quality of dipoles location detection.

The multipoint system of passive monitoring of thunderstorm activity not only removes the uncertainty of location [9], but also allows to determine the orientation of the dipole [13, 14]. Nevertheless, the improvement of forecasting methods using single-point systems remains relevant [15, 16].

One way to reduce the uncertainty of thunderstorm location is to analyze the information received from all dipole sources for time T.

The purpose of secondary processing is to estimate the location of the thunderstorm center by information about all dipoles location parameters for time T.

3 Secondary processing of EMRS signals from thunderstorm foci

It is known that the Cartesian coordinates x, y of the possible position of the EMRS satisfy the condition [9]:

$$x\cos\varphi + y\sin\varphi = r\sin\theta\cos(\varphi-\psi) = ru, \tag{14}$$

which determines the OXY projection of the EMRS possible positions set. The coordinates of the possible location EMRS belong to the set of solutions of the system of equations

$$x\cos\varphi + y\sin\varphi = ru, \qquad x^2 + y^2 + z^2 = r^2. \tag{15}$$

Let S be set of all solutions of (15) system. In general, S is multiple.

Entire space divided into $4N^3$ cells:

$$C(i,j,k) = \left\{ (x,y,z) \;\middle|\; \begin{array}{l} i\Delta \le x \le (i+1)\Delta \\ j\Delta \le y \le (j+1)\Delta \\ k\Delta \le z \le (k+1)\Delta \end{array} \right\}, \quad k = 0,\,1,\,\ldots,\,N,$$
$$i,\,j = -N,\,-N+1,\ldots,-1,\,0,\,1,\,\ldots N. \tag{16}$$

The OXY plane is divided into flat cells

$$C(i,j) = \{(x,y):\; i\Delta \le x \le (i+1)\Delta,\; \Delta \le y \le (j+1)\Delta,$$
$$i,\,j = -N,\,-N+1,\ldots,-1,\,0,\,1,\,\ldots N\}, \tag{17}$$

and OZ axis of the coordinate system $OXYZ$ to be divided into segments

$$C(k) = \{z:\; k\Delta \le z \le (k+1)\Delta\}, \quad k = 0,\,1,\,\ldots,\,M.$$

In the future we will use the following notation:

$$C_{OXYZ}(r,u,\varphi) = \{(i,j,k) \mid C(i,j,k) \cap S \neq \emptyset\}$$

is set of cells in which the EMRS can be locate with a non zero probability;

$$C_{OXY}(r,u,\varphi) = \{(i,j) \mid \exists k:\ C[i,j,k] \cap S \neq \emptyset\}$$

is OXY projection of set $C_{OXYZ}(r,u,\varphi)$;

$$C_{OZ}(r,u,\varphi) = \{k \mid \exists C(i,j):\ C(i,j,k) \cap S \neq \emptyset\}$$

is OZ projection of set $C_{OXYZ}(r,u,\varphi)$.

For each cell containing solutions of the (15) system, we find the probability that the EMRS belongs to this cell when some hypothesis is fulfilled. In the future, we will use the found probabilities to build a three-dimensional map of the membership level of cells that are part of the thunderstorm center.

3.1 Hypothesis H_1: Uniform probability distribution on $C_{OXY}(r,u,\varphi)$

It is following from (15) that on the OXY plane, the valid points farthest from the origin of the coordinates are $(x_1,\ y_1)$ and $(x_2,\ y_2)$

$$x_{1,2} = r(u\cos\varphi \mp \sqrt{1-u^2}\sin\varphi), \tag{18}$$

$$y_{1,2} = r(u\sin\varphi \pm \sqrt{1-u^2}\cos\varphi, \quad z_{1,2} = 0. \tag{19}$$

The coordinates x, y for all possible ERMS positions belong to segment $[(x_1,y_1),\ (x_2,y_2)] \subset OXY$ between the points (x_1,y_1) and (x_2,y_2). The all possible radiation source locations can be represented in a parametric form as the set of points:

$$\mathcal{L}(r,u,\varphi) = \left\{ \begin{array}{c} x(\lambda) = ru\cos\varphi + \\ +r(2\lambda-1)\sqrt{1-u^2}\sin\varphi, \\ y(\lambda) = ru\sin\varphi + \\ +r(1-2\lambda)\sqrt{1-u^2}\cos\varphi, \\ z(\lambda) = 2r\times \\ \times\sqrt{\lambda(1-\lambda)(1-u^2)} \end{array} \middle| \ 0 \le \lambda \le 1 \right\}. \tag{20}$$

Let

$$\Lambda = \left\{0,\ \frac{1}{N},\ \frac{2}{N},\ \ldots,\ \frac{N-1}{N},\ 1\right\}.$$

It is following from (20) that points $(x(\lambda), y(\lambda)),\ \lambda \in \Lambda$ are evenly distributed on the segment $[(x_1,y_1),\ (x_2,y_2)]$, and

$$[(x_1,y_1),\ (x_2,y_2)] \quad \subset \bigcup_{(i,j)\in C_{OXY}(r,u,\varphi)} C(i,j).$$

If the hypothesis on uniform probability distribution on $C_{OXY}(r,u,\varphi)$ is accepted, there are equalities

$$C_{OXYZ}(r,u,\varphi) = \left\{(i,j,k)\ :\ i = \left\lfloor\frac{x(\lambda)}{\Delta}\right\rfloor,\ j = \left\lfloor\frac{y(\lambda)}{\Delta},\right\rfloor\ k = \left\lfloor\frac{z(\lambda)}{\Delta}\right\rfloor,\ \lambda\in\Lambda\right\},$$

$$C_{OXY}(r,u,\varphi) = \left\{(i,j)\ :\ i = \left\lfloor\frac{x(\lambda)}{\Delta}\right\rfloor,\ j = \left\lfloor\frac{y(\lambda)}{\Delta}\right\rfloor,\ \lambda\in\Lambda\right\}.$$

Let

$$N(i,j,k) = \mathrm{card}\,\{\lambda\in\Lambda\ :\ (x(\lambda),\ y(\lambda),\ z(\lambda)\in C(i,j,k))\},$$

$$N(i,j) = \sum_{k\in C_{OZ}(r,u,\varphi)} N(i,j,k), \quad .$$

Then, the probability of EMRS in cell $C[i,j,k]$ is equal to

$$\mathrm{P}_{OXY}(i,j,k) = \frac{N(i,j,k)}{N(i,j)\cdot N}.$$

3.2 Hypothesis H_2: Uniform probability distribution on $C_{OZ}(r,u,\varphi)$

It follows from (15) that the set of points of all possible locations of EMRS can be represented in a parametric form other than (20) :

$$\mathcal{L}(r,u,\varphi) = \left\{ \begin{array}{c} x(z) = ru\cos\varphi \pm \sqrt{r^2(1-u^2)-z^2}\sin\varphi, \\ y(z) = ru\sin\varphi \mp \sqrt{r^2(1-u^2)-z^2}\cos\varphi, \\ z:\ \ 0\le z\le \sqrt{r^2(1-u^2)} \end{array} \right\}. \tag{21}$$

Let

$$M = \left\lfloor\frac{\sqrt{r^2(1-u^2)}}{\Delta}\right\rfloor, \quad Z = \{0,\ \Delta,\ 2\Delta,\ \ldots,\ M\cdot\Delta\}. \tag{22}$$

It is following from (22) that points $z\in Z$ are evenly distributed on the segment

$$[0,\ \sqrt{r^2(1-u^2)}] \quad \subset \bigcup_{k:\in C_{OZ}(r,u,\varphi)} C[k].$$

If the hypothesis on uniform probability distribution on $C_{OZ}(r,u,\varphi)$ is accepted, there are equalities

$$C_{OXYZ}(r,u,\varphi) = \left\{(i,j,k)\ :\ i = \left\lfloor\frac{x(z)}{\Delta}\right\rfloor,\ j = \left\lfloor\frac{y(z)}{\Delta},\right\rfloor\ k = \left\lfloor\frac{z}{\Delta}\right\rfloor,\ z\in Z\right\},$$

$$C_{OXY}(r,u,\varphi) = \left\{(i,j)\ :\ i = \left\lfloor\frac{x(z)}{\Delta}\right\rfloor,\ j = \left\lfloor\frac{y(z)}{\Delta}\right\rfloor,\ z\in Z\right\}.$$

Let

$$M(i,j,k) = \mathrm{card}\,\{z\in Z\ :\ (x(z),y(z),z)\in C(i,j,k)\},$$

$$M(k) = \sum_{[i,j]\in C_{OXY}(r,u,\varphi)} M(i,j,k), \qquad k:\ k\cdot\Delta\in Z,$$

then

$$\mathrm{P}_{OZ}(i,j,k) = \frac{M(i,j,k)}{M(k)\cdot M}.$$

3.3 Hypothesis H_3: Uniform probability distribution on $C_{OXYZ}(r, u, \varphi)$

Since (20) and (21) give different representations of set $\mathcal{L}(r, u, \varphi)$, there are also two ways to represent a set of cells with a non-zero probability of having the EMRS:

$$C_{OXYZ}(r, u, \varphi) = \{(i, j, k) : \ \mathrm{P}_{OXY}(i, j, k) > 0\} = \\ = \{(i, j, k) : \ \mathrm{P}_{OZ}(i, j, k) > 0\}.$$

If the hypothesis on uniform probability distribution on $C_{OXYZ}(r, u, \varphi)$ is accepted, we have

$$\mathrm{P}_{OXYZ}(i, j, k) = \frac{1}{\operatorname{card} C_{OXYZ}(r, u, \varphi)}, \quad (i, j, k) \in C_{OXYZ}(r, u, \varphi).$$

3.4 Three-dimensional map of the level belonging of thunderstorm foci to space cells

So, for each cell containing solutions of the (15) system, we found the probability that the EMRS belongs to this cell when hypothesis $H \in \{H_1, H_2, H_3\}$ is fulfilled. Let us use the found probabilities to build a three-dimensional map of the membership level of cells that are part of the thunderstorm center when hypothesis H is fulfilled.

The set of discharges registered for the time period T is defined as $R(T)$. The cell grade membership to thunderstorm of the radiation source defined as the sum of all of radiation during this period.

$$\operatorname{grade}_H^{R(T)}(i, j, k) = \frac{\sum\limits_{d \in R(T)} \mathrm{P}_H(i, j, k)(d)}{\max\limits_{(l,m,n)} \sum\limits_{d \in R(T)} \mathrm{P}_H(l, m, n)(d)}. \qquad (23)$$

Further we demonstrate the possibility to estimate the thunderstorm center location by (23) formula applying simulation.

4 Simulation

Let's define two artificial thunderstorms $R_1(T)$ and $R_2(T)$ for modeling purposes. All discharges of the thunderstorms will have the same location $R = 60$ km, $\Delta = 10$ km, $\psi = \pi/4$, $\theta = \pi/3$.

Figures 3–8 demonstrate the observation point ($x = 0$, $y = 0$, $z = 0$) and the thunderstorm location cell is $C(i = 3,\ j = 3,\ k = 3)$ (marked with a blue square). Cell color shows the cell grade membership level to the thunderstorm (see figure 2).

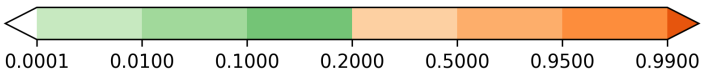


Figure 2: Scale of the level of belonging of a cell to a thunderstorm center

The $R_1(T)$ emits 256 pulse signals from dipole radiation sources with orientation $\psi_0 \in \{\pi k/16 : \ k = 1, 2, \ldots, 32\}$, $\theta_0 \in \{\pi k/8 : k = 0, 1, \ldots, 8\}$, that is EMRS are different incline dipoles, which is typical for a thunderstorm center. Location parameters values r, u, v, φ are calculated by (5)-(6) formulas.

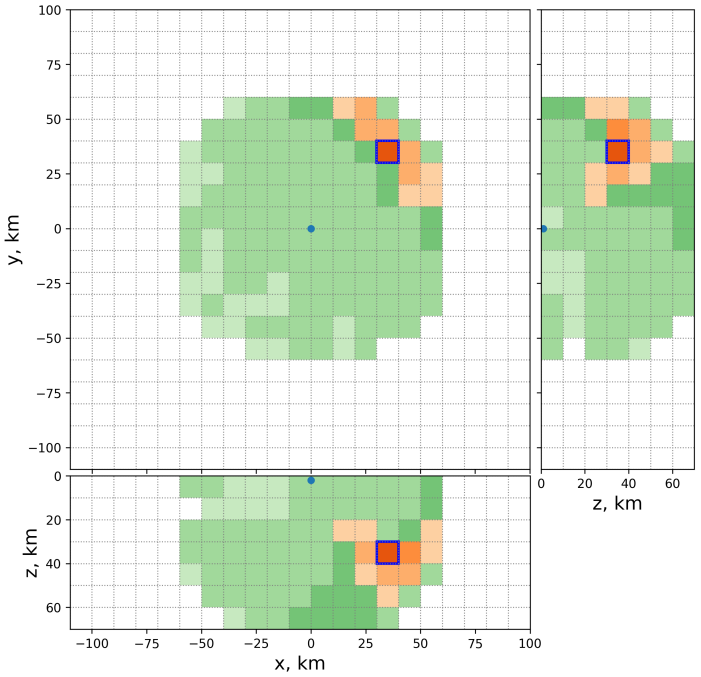


Figure 3: Map of $R_1(T)$ under hypothesis H_1

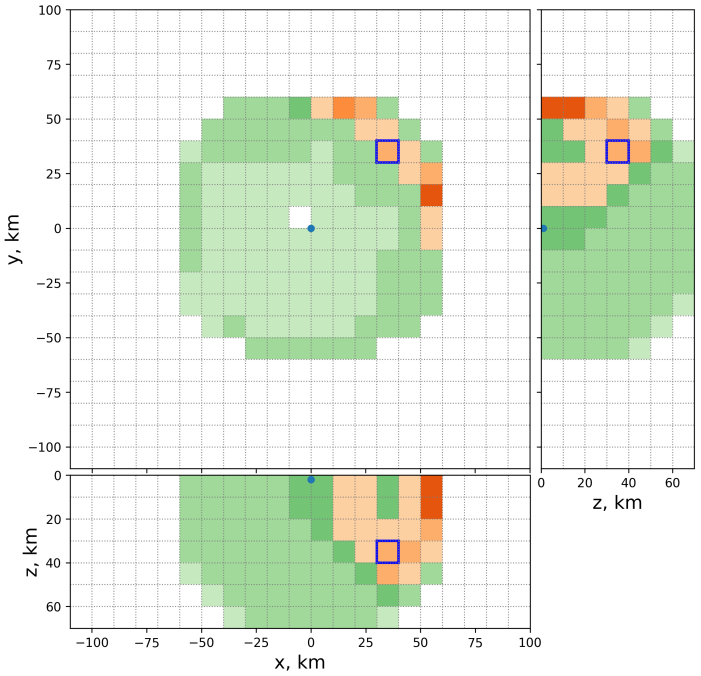


Figure 4: Map of $R_1(T)$ under hypothesis H_2

Figures 3–8 demonstrate three projections of the three-dimensional map of thunderstorm foci $R_l(T)$, $l \in \{1, 2\}$ level belonging to space cells under hypothesis $H \in \{H_1, H_2, H_3\}$

$$\operatorname{grade}_H^R(T)(i, j) = \sum_k \operatorname{grade}_H^R(T)(i, j, k), \qquad (24)$$

$$\operatorname{grade}_H^R(T)(i, k) = \sum_j \operatorname{grade}_H^R(T)(i, j, k), \qquad (25)$$

$$\text{grade}_H^R(T)(j,k) = \sum_i \text{grade}_H^R(T)(i,j,k). \quad (26)$$

Figure 3 demonstrates the belonging of desired cell to cell with maximum level of membership. Figure 4 demonstrates that the desired cell does not belong to cells with maximum level of membership. Figure 5 demonstrates the belonging of desired cell to cell with maximum level of membership.

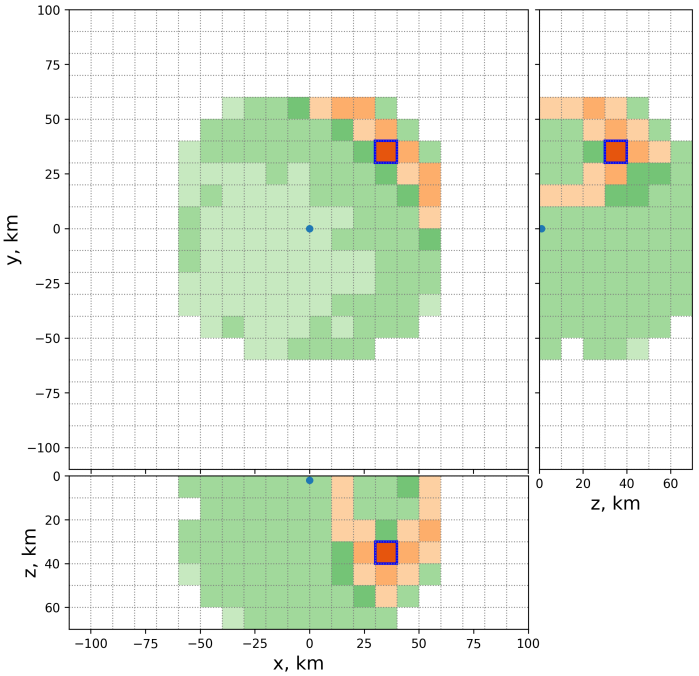


Figure 5: Map of $R_1(T)$ under hypothesis H_3

Figures 3, 5 make it possible to identify the desired cell $C(i = 3, j = 3, k = 3)$. Figure 4 does not allow to identify the desired cell at all.

Table 1 shows cells with grade membership level to the thunderstorm $R_1(T)$ no less 0.9 and similar to visual analysis of the projections hypothesis H_1 and H_3 found the desired cell.

Table 1: Cells with grade membership level to the thunderstorm $R_1(T)$ no less 0.9

Hypothesis	$C_{OXYZ}(r,u,\varphi) \cap \{(i,j,k) : \text{grade}_H^{R_1(T)}(i,j,k) > 0.9\}$
H_1	$\{(3,3,3),\ (3,2,3),\ (4,3,3)\}$
H_2	$\{(1,5,0),\ (1,5,1),\ (3,3,4),\ (5,1,0),\ (5,1,1)\}$
H_3	$\{(3,3,3)\}$

But the best hypothesis is hypothesis H_3, because it will allow to identify the cell with a thunderstorm center uniquely and accurately.

The $R_2(T)$ has the same location as $R_1(T)$ but emits only 100 pulse signals from dipole radiation sources with orientation $\psi_0 \in \{2\pi k/100 :\ k = 1,2,\ldots,100\}$, $\theta_0 = \pi/2$, that is EMRS are different horizontal dipoles, which is typical for a pre-storm center. Location parameters values r, u, v, φ are calculated by (5)-(6) formulas.

Figures 6–8 demonstrate three projections of the three-dimensional map of thunderstorm foci $R_2(T)$ level belonging to space cells under hypothesis $H \in \{H_1, H_2, H_3\}$.

Figure 6 demonstrates that the desired cell does not belong to cells with maximum level of membership.

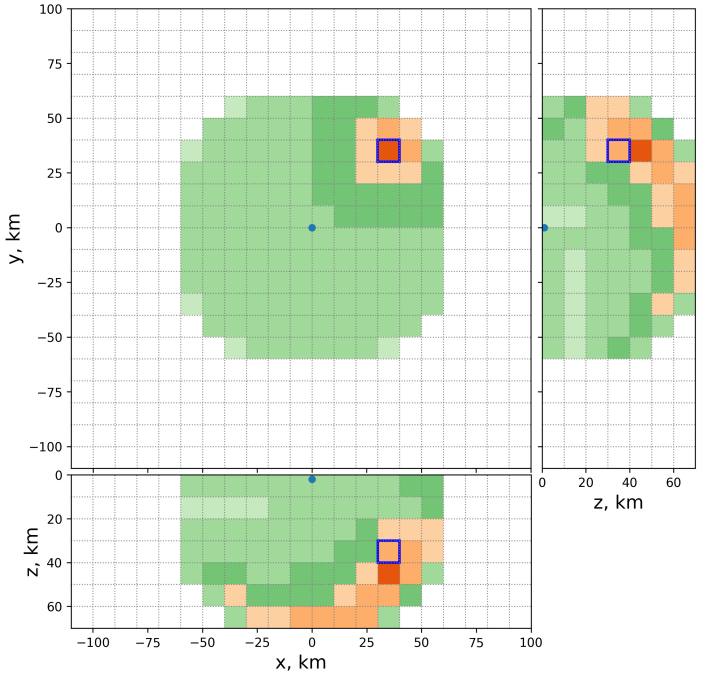


Figure 6: Map of $R_2(T)$ under hypothesis H_1

Figure 7 demonstrates that the desired cell does not belong to cells with maximum level of membership.

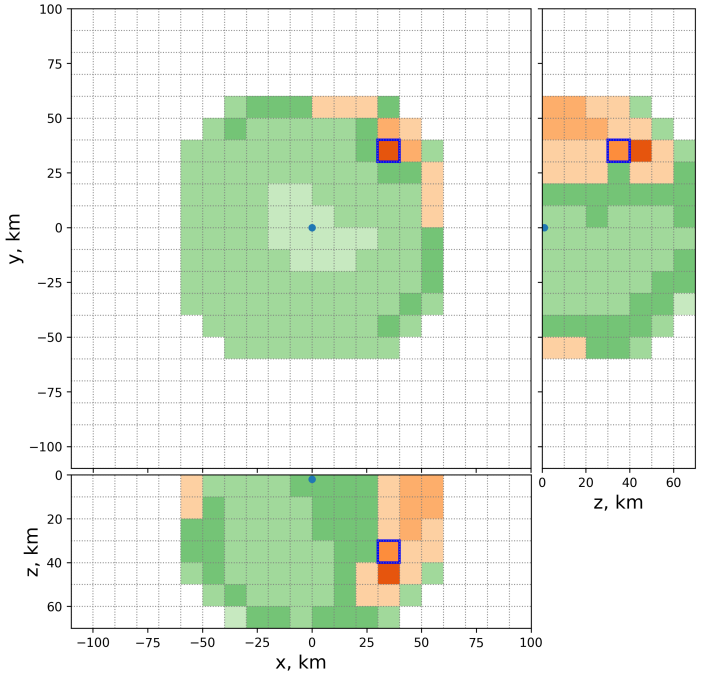


Figure 7: Map of $R_2(T)$ under hypothesis H_2

Figure 8 demonstrates that the desired cell belong to cells with maximum level of membership only.

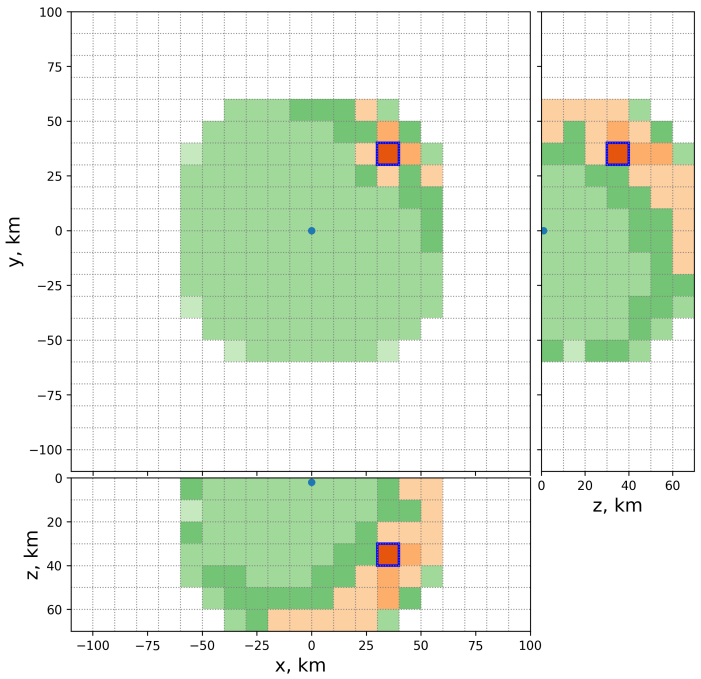


Figure 8: Map of $R_2(T)$ under hypothesis H_3

Table 2 also shows the best hypothesis is H_3. Hypotheses H_1 and H_2 maps do not contains the desired cell and does not allow to unambiguously establish the location of the foci.

Table 2: Cells with grade membership level no less 0.9 to the thunderstorm $R_2(T)$

Hypothesis	$C_{OXYZ}(r,u,\varphi) \cap \{(i,j,k):\ \text{grade}_H^{R_2(T)}(i,j,k) > 0.9$
H_1	$\{(3,3,4)\}$
H_2	$\{(3,3,4)\}$
H_3	$\{(3,3,3),\ (3,3,4)\}$

5 Conclusion

These findings provide a potential new technique to processing of the results of thunderstorm activity monitoring in the VLF range. The processing of all data from the SPMS during thunderstorm activity introduces a lot of uncertainty and makes it difficult to determine the real location of the thunderstorm center. The methods of analysis and visualization of the entire set of received signals presented in this paper reduce the uncertainty in determining the coordinates of the thunderstorm location. Processing of pre-thunderstorm radiation is significantly important to forecasting the development of thunderstorm.

The artificial single-point thunderstorm model described in the paper is a very simple, which is very convenient for research, however, a real thunderstorm can have a more complex structure, therefore it is important to carry out modelling with more complex models of thunderstorms, for example, check the proposed approaches for their ability to identification of multiple foci. This is an interesting topic for future work.

In future work, investigating another hypothesis of probability that the EMRS belongs to cell might be important.

References

[1] A. V. Panyukov, A. K. Bogushov, "Single-point system of positioning storms in close zone," RU Patent 2,599,775, 2016.

[2] A. V. Panyukov, A. K. Bogushov, "A Method For Processing Data of Thunderstorm Activity Monitoring," in 2019 Russian Open Conference on Radio Wave Propagation (RWP), 556–559, 2019, doi:10.1109/RWP.2019.8810345.

[3] "Thunderstorm Electric Field Mill EFM550," 2020.

[4] "Thunderstorm Advanced Total Lightning Sensor LS7002," 2020.

[5] "Boltek LD-350 Lightning Detector," 2020.

[6] "Biral BTD-300 Thunderstorm Detector," 2020.

[7] "The International Science and Technology Center. Project; 1822. Development and Investigation of the Single-Point System for Lightning Location in the Range of SuperLong Waves," 2000.

[8] I. Kononov, I. Yusupov, N. Kandaratskov, "Analysis of One-Point Methods for Lightning-Discharge Passive Location," Radiophysics and Quantum Electronics, **56**, 788–800, 2014, doi:10.1007/s11141-014-9481-1.

[9] A. V. Panyukov, "Estimation of the Location of an Arbitrary Oriented Dipole under Single-point Direction Finding," Journal of Geophysical Research, **101**, 14997–14982, 1996, doi:10.1029/96JD00067.

[10] B. Taibin, "An Approach to Define Parameters for Localization of Thunderstorm," IEEE Antennas and Propagation Magazine, **48**, 48–54, 2006, doi:10.1109/MAP.2006.1650816.

[11] A. V. Panyukov, A. K. Bogushov, "The Spectral Statistical Method for Determining the Location Parameters of a Dipole Source of Electromagnetic Radiation," Radiophysics and Quantum Electronics, **59**, 278–288, 2016, doi:10.1007/s11141-016-9696-4.

[12] A. V. Panyukov, "Inverse Problem for an Electrical Dipole and the Lightning Location Passive Monitoring System," in M. S. Kandelousi, editor, Electric Field, chapter 13, IntechOpen, Rijeka, 2018, doi:10.5772/intechopen.74264.

[13] S. He, M. Popov, V. Romanov, "Explicit full identification of a transient dipole source in the atmosphere from measurement of the electromagnetic fields at several points at ground level," Radio Science, **35**(1), 107–117, 2000, doi:https://doi.org/10.1029/1999RS002198.

[14] M. Popov, S. He, "Identification of a transient electric dipole over a conducting half space using a simulated annealing algorithm," Journal of Geophysical Research, **105**, 20821–20831, 2000, doi:10.1029/2000JD900261.

[15] A. Leal, V. Rakov, E. Alves, M. Lopes, "Estimation of –CG lightning distances using single-station E-field measurements and machine learning techniques," 1–8, 2019, doi:10.1109/SIPDA47030.2019.9004484.

[16] "Vaisala. ALARM Automated Lightning Alert and Risk Management System," 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1103-1108 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Evaluation of the Physico-Chemical Properties of Soil and Apple Leaves (Malus Domestica) in Beni Mellal-Khenifra Region, Morocco

Berrid Nabyl[1], Lougraimzi Hanane[2], El-Khabbazi Houda[3], Abidli Zakaria[3], Hamidi Otman[1], Keltoum Rahali[1], El Mahjoub Aouane[1,*]

[1]*Laboratory of Biotechnology, Quality and Environment, Faculty of Science, Ibn Tofail, University, Kenitra, Morocco*

[2]*Laboratory of Nutrition, Health and Environment, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco*

[3]*Laboratory of Genetic and Biometric, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco*

ARTICLE INFO

Article history:
Received: 01 September, 2020
Accepted: 26 November, 2020
Online: 14 December, 2020

Keywords:
Physico-chemical properties
Soil
Apple leaves
Mineral elements
Organic matter
Morocco

ABSTRACT

The apple tree (Malus domestica) is an agricultural species of great importance in Morocco. Currently, it occupies about a quarter of the surface of fruit-bearing rosaceae. Apple tree production has evolved rapidly, stimulated by a buoyant market, a varietal range which is tending to diversify and a dynamic profession. The present study aims to determine the physico-chemical properties of soil and apple leaves in Beni Mellal-Khenifra region, Morocco. Data collection was consisted to sample the soils at different depths, three Pearson correlation coefficients were determined: i) Correlation between organic matter and soil mineral elements, ii) Correlation between clay content and soil mineral elements and correlation between CaCO3 content and soil mineral elements. The analytical results of soils were compared to reference values. The results show that there is no correlation between these mineral elements and the rate of CaCO3 and clay fraction. Analysis of the main components showed that the apple leaf contains more sodium, copper and zinc and less magnesium, phosphorus and potassium. The coefficient of variation (CV) of manganese content expressed as a percentage is significantly higher than that of the other elements, while the potassium content is the lowest, in particular for magnesium and sodium the coefficient of variation is zero. In apple orchards in the Middle Atlas, there are great variations in terms of yield levels, fertilizer inputs, soil richness in nutrients and their concentration at leaf level. These variations are due to differences in cultivation practices, in particular fertilization, because the relationship between apple yield and nutrient content at soil and leaf level is significant.

1. Introduction

The agricultural sector plays a key role in economic activities in Morocco. Likewise, it represents a strong support to the development of the country because it constitutes a higher percentage of 20% of the gross domestic product and employs about 40% of the active population [1, 2].

In Morocco, the apple tree (*Malus domestica*) is the species, which occupies the largest surface in terms of fruit rosaceae. Currently, orchards planted with apples total 32,000 ha, or about a quarter of the surface area of fruit-bearing rosaceae. Annual production is estimated at 560-600,000 tonnes, based on an average yield of around 20 T/ha [3]. However, Intensive agriculture and the search for new fertile land put pressure on the ecosystem, which in turn leads to reduced soil fertility [4, 5]. As a result, yields are relatively low and land productivity decreases. There are several types of soils in Morocco which have developed on different geological substrates and in various climatic, topographical and biotic contexts. These soils must be protected as a natural resource because it contains substances essential to the maintenance of ecosystems and human populations, such as essential minerals. this study is carried out in the Beni Mellal-Khenifra region which is characterized by a variety of soils and by the predominance of fertile types rich in organic matter which allows an important agricultural activity, in particular for apple trees [6]. So far, interpretation of soil and leaf analyzes for

*Corresponding Author, El Mahjoub Aouane, University Campus BP 133 Kenitra, Morocco. Email: aouane_mahjoub@yahoo.fr

Moroccan apple orchards has been limited. For this reason, the main objective of our study is to determine the physicochemical properties of soil analysis and apple leaves in the Beni Mellal-Khenifra region in Morocco. More specifically, i) Determining the correlation coefficients between organic matter and soil mineral elements, ii) Correlation between clay content and soil mineral elements and correlation between CaCO3 content and soil mineral elements.

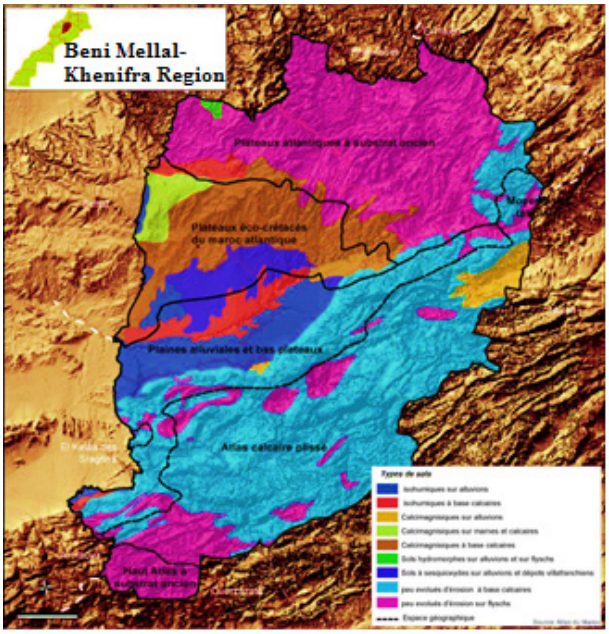


Figure 1: Geographical location and soil types in Beni Mellal-Khenifra region, study site [7].

2. Materials and methods

2.1. Presentation of the study area

The Beni Mellal-Khenifra region is a major contributor to national agriculture. From field crops to local products, its environmental diversity allows it to have a very varied agriculture. Between mountains, foothills and plains, the diversity of climates and soils offers the Beni Mellal-Khenifra region many agricultural opportunities. In addition to its environmental and natural wealth, this region has the advantage of having significant water resources unlike other regions of the Kingdom. The study area is part of the perimeter of the Beni Mellal-Khenifra region, which is made up of four large distinct geographic areas, which maintain relations of exchange, complementarity and forms of solidarity, particularly in the field of agriculture and extensive farming [7]. It has a total surface area of 2.8 million hectares, of which 960,000 hectares of useful agricultural area (205,000 hectares under irrigation and 700,000 hectares of rangelands). Moreover, the region has adopted a specific agricultural plan designed to remove the constraints that hamper agricultural development. It is a specific region, inserted between the Atlas Mountains and the Atlantic coastal provinces, with a much-contrasted physiognomy and a marked character. However, these contrasts, however clear-cut they may appear, present undeniable advantages for the implementation of mutually beneficial solidarity, the rich natural heritage (hydraulic and forest) which is highly diversified, the agricultural potential of the plain which is being reconverted (Beni Amir and Beni Moussa), the opportunities offered by mining activity, the major structuring projects (connecting infrastructures and agro-industrial and commercial platforms) and the hinge position in the national system [7].

2.2. Soil types in the area

The soils in the region are diverse due to the nature of the source rocks. They are shallow, shallow, and low in humus. Their water reserves are sometimes limited. Moreover, the dominance of holm oak, a very plastic species from an ecological point of view, confirms this situation when the ecological atmosphere is difficult. The substrates are colonized by thermo-xerophilic (*Aleppo pine*, thuja and red juniper) and xerophilic (*Thuriferous juniper*) species (Figure 1). Even at altitude, the soils generally remain shallow to shallow, humus and humid for a fairly long period of the year, allowing the establishment of friendly forest species (zene oak, cedar and holm oak) [7].

2.3. Soil sample collection

The study covered 19 observation sites (10 samples by site). Soil samples were taken at different depths (0-60 cm) using an auger on plots of apples. These samples were air dried overnight or longer and sieved through a 2 mm sieve to obtain a fine soil. The results of the soil analyzes were compared to limit values and reference standards. We also measured the electrical conductivity and pH of the main elements using the same techniques as below:

Table 1: Analysis technique.

Analysis	Methods
Granulometry	Sieving (2 mm to 45 µm sieve)
pH 1/ 2.5	pH to put
CE 1/5 25 °C	Granulometry
Cation exchange capacity CEC and exchangeable shares	Cationic exchange capacity CEC and exchangeable shares
Na+ and K+ Cations	Flame spectrophotometer
Ca++ and Mg++ Cations	Volumetric titration
Chlorides Cl-	MOHR method (silver nitrate precipitation in the presence of chromates)
Sulphate SO4--	Precipitation at BaCl2

2.4. Sampling of plant material

The foliar analysis of the apple tree (or foliar diagnosis of the apple tree), is the best way to determine the quantity and the type of fertilizer to give to the fruit trees. Thirty samples of healthy and homogeneous trees were chosen (70 to 90 days after full flowering (stage F2)) and the samples were representative of the plot. We chose the woody branches representative of the vigor and the average growth of the tree (eliminate the greedy ones), the leaf in the middle of the branch with its stem.

In total 200 sheets. The collected leaves are placed in new perforated plastic bags and awaiting transport to the analysis

laboratory; they are stored cool or cold at 4-5 ° C so that they retain all their characteristics. The analysis is carried out the day after harvest at the latest. It focuses in particular on the nitrogen content expressed as a percentage of dry matter (% DM). The results of soil analyzes were compared to values thresholds and reference standards specific to apple production (Table 2).

Table 2: Reference thresholds [8].

	N (% MS)	P (% MS)	K (% MS)	Ca (% MS)	Mg (% MS)
Apple	2.2-2.4	0.22	1.13 à 1.75	1.25	0.22 à 0.3
Prunes	2.4 -3	0.14-0.25	1.6 - 3	1.5 – 3	0.3 - 0.8

2.5. Statistical methodology

Statistical analysis was performed using SPSS software. Descriptive statistics were used and analyzes of variance and comparison of means were performed at a probability level of $p<0.05$.

3. Results

3.1. Physico-chemical characteristics of the soil studied for each site

The results of the particle size analysis of the studied soils are presented in table 3. On the samples analyzed, the texture tests did not reveal significant clay contents except for site 2, which showed values barely greater than 45%. The clay content of the studied soils varies between 26% and 46.57%. Under these conditions, the deep migration of nutrients is slow and leaching losses are low. According to the Texture Triangle (FAO), which is used to determine the textural class of soils, the study area mainly has loamy clay soils. Considering also that apple trees prefer healthy and deep loamy-clayey-loamy soils with a permeable subsoil, it can be assumed that the studied soils are favorable for this apple crop.

Table 3: Physical constituents of the soils studied.

Observation sites	Depth	Granulometry (%)		
		Clay	Silt	Sand
1	20-50	35	30	35
2	0-20	35	30	35
3	0-20	45	25	30
4	20-50	45	30	25
5	30-60	26	18.77	55.23
6	0-30	40.73	25.98	33.29
7	30-60	39.91	26.1	33.99
8	0-30	26.92	47.52	25.56
9	30-60	27.13	46.69	26.18
10	0-30	31.97	39.2	28.83
11	30-60	32.1	38.99	28.91
12	0-30	44.81	43.24	11.95
13	30-60	45.02	43.45	11.53
14	0-30	46.49	42.54	10.97
15	30-60	46.96	41.99	11.05
16	0-30	40.46	35.71	23.83
17	30-60	41	35.5	23.5
18	40-60	42.7	44.05	12.95
19	0-40	41	43.5	15.5

The pH is an important parameter of soil dynamics, the degree of acidity or basicity plays a very essential role in the absorption of nutrients by the plant. The average pH of the soil samples studied is eight, which shows that the majority of sites are basic with an alkaline tendency (Table 4).

Table 4: General Statistics of different physico-chemical characteristics of the samples studies.

Characteristics	Min	Max	Median	Average		Standard deviation
				Statistique	Erreur std	Statistique
MO	0.48	3	2.00	1.71	.187	0.772
pH	8	9	8.00	8.06	.063	0.250
CaCO3	6	41	10.00	12.11	1.838	8.013
Clay	26	47	41.00	38.58	1.596	6.955
Silt	19	48	37.50	36.44	2.033	8.624
Sand	12	55	26.00	26.11	2.567	10.889
Fe mg/kg	1	8	6.50	5.83	.452	1.917
Bm g/kg	0	0	0.00	0.00	0.000	0.000
Zn mg/kg	0	6	3.00	2.72	0.523	2.218
Cu mg/kg	0	1	1.00	0.65	0.119	0.493
P2O5 mg/kg	2	47	5.00	10.40	2.953	11.438
K2O meq/100g	0	551	340.00	301.94	38.147	157.283
CaO meq/100g	12	6850	1388.00	3217.27	689.230	2669.376
MgO meq/100g	0	380	320.00	269.53	36.272	140.479

The soils studied exhibited an organic matter content in the range 0.48-3%. Analysis of the elements of Iron, zinc and copper revealed average contents of 6.5 mg/kg, 3 mg/kg and 1 mg/kg respectively. For all the soils studied and sampled at different depths, three Pearson correlation coefficients were determined: one between the different soil types and CaCO3 content, the second between the different soil types and the percentage of organic matter, and the third between the different soil types and the clay content. The coefficients of determination obtained show

that variation of soil types and their sampling at different depths explains 12% of the variation of clay content, 11% of CaCO3 and 8% of the material content. In terms of physico-chemical characteristics, descriptive statistics are the set of methods and techniques that allow us to present, describe and summarize, numerous and varied data; they allow us to know the characteristics of our sample.

For all variables, the standard deviation is more or less low, so the dispersion of the measurements around the mean is narrow. The results are therefore consistent, except for the last three variables, namely the K2O, CaO and MgO content, where the standard deviation is very high, which shows that there is a wide dispersion of the results for these contents (Table 5,6 and 7).

Table 5: Correlation between organic matter and soil mineral elements

		Fe mg/kg	B mg/kg	Zn mg/kg	Cu mg/kg	P2O5 mg/kg	K2O meq/100g	CaO meq/100g	MgO meq/100g
MO %	Correlation Pearson	0.402	.a	-0.073	0.239	-0.018	0.573*	0.433	0.285
	Sig.(bilatérale)	0.123		0.788	0.391	0.951	0.026	0.094	0.285
	N	16	5	16	15	14	15	16	16

***The correlation is significant at the 0.01 level (bilateral).*
**The correlation is significant at the 0.05 level (bilateral).*
at. Calculation impossible because at least one variable is a constant

Table 6: Correlation between clay content and soil mineral elements

		Fe mg/kg	B mg/kg	Zn mg/kg	Cu mg/kg	P2O5 mg/kg	K2O meq/100g	CaO meq/100g	MgO meq/100g
Arg %	Correlation Pearson	-0.142	.a	-0.108	-0.135	-0.054	-0.329	0.369	-0.312
	Sig.(bilateral)	0.575		0.671	0.605	0.850	0.197	0.132	0.207
	N	18	6	18	17	15	17	18	18

***The correlation is significant at the 0.01 level (bilateral).*
**The correlation is significant at the 0.05 level (bilateral).*
at. Calculation impossible because at least one variable is a constant.

Table 7: Correlation between CaCO3 content and soil mineral elements

		Fe mg/kg	B mg/kg	Zn mg/kg	Cu mg/kg	P2O5 mg/kg	K2O meq/100g	CaO meq/100g	MgO meq/100g
CaCO3	Correlation Pearson	0.224	.a	0.377	0.375	-0.201	-0.010	0.171	0.321
	Sig.(bilateral)	0.372		0.123	0.138	0.473	0.970	0.499	0.194
	N	18	6	18	17	15	17	18	18

***The correlation is significant at the 0.01 level (bilateral).*
**The correlation is significant at the 0.05 level (bilateral).*
at. Calculation impossible because at least one variable is a constant.

In order to study the relationship between the mineral element content and the CaCO3, MO and clay fraction, we calculated the Pearson correlation coefficient (Figure 2). The results show that there is no correlation between these mineral elements and the rate of CaCO3 and clay fraction. The results for the organic matter content show that the latter explains 57% ($p<0.05$) of the variation in K2O content.

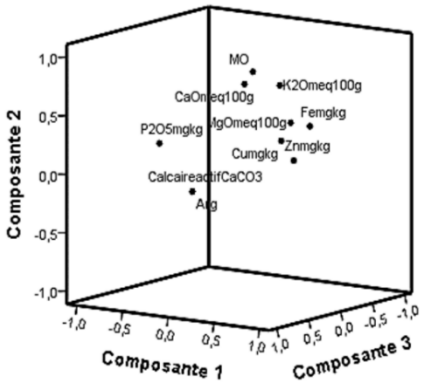


Figure 2: Principal component analysis of the physico-chemical properties of the soils sampled

The ANOVA test application, to compare the average characteristics of 19 sites (without distinction of depths) is presented in the table above. It shows that in the different sites studied, the difference in mean is insignificant for the parameters pH, organic matter, CaCO3 content, clay content, B, CaO; it is also insignificant for the silt and sand variables, so the null hypothesis H0 is accepted as long as there is no difference in the means. For the variables Fe, Zn, Cu, P2O5, K2O, MgO the difference of the means is significant and, in this case, the null hypothesis H0 is rejected.

3.2. Physico-chemical properties of apple leaf soil

The differences in the average composition of the three varieties are shown in the table 8. The apple leaf contains more Na, Cu and zinc and less Mg, P and K. The coefficient of variation (CV) of the manganese content expressed as a percentage is much higher than that of the other elements, and the coefficient of variation (CV) of the potassium content is the lowest, especially for magnesium and sodium the coefficient of variation is zero (Figure 3).

Table 8: Global apple leaf nutrient content statistics.

	Average	Standard deviation	CV	Min	Max
Magnesium, Mg (%)	0.1900	0	0	0.19	0.19
Phosphorus, P (%)	0.2200	0.01732	7.873	0.20	0.23
Potassium, K (%)	0.553333	0.0152753	2.761	0.5400	0.5700
Calcium, Ca (%)	1.4333	0.25580	17.846	1.19	1.70
Manganese, Mn (ppm)	24.033	6.5577	27.286	20.0	31.6
Zinc, Zn (ppm	32.733	6.0343461	18.435	25.8000	36.8000
Copper, Cu (ppm)	129.33	12.662	9.790	118	143
Boron, B (ppm)	26.333	2.5482	9.677	23.4	28.0
Sodium, Na (%)	249.00	0.000	0	249	249

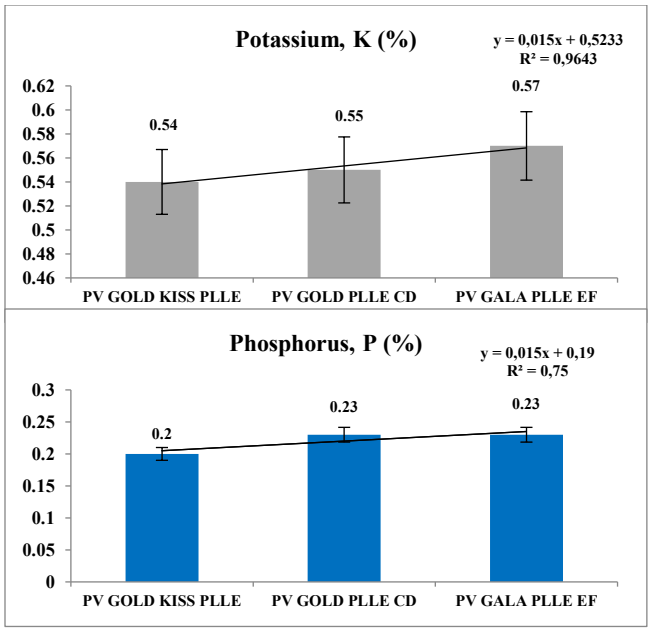


Figure 3: Phosphorus and potassium content in different apple varieties.

The differences in the average composition of the four varieties are shown in the table 9 and figure 4. Apple leaf contains more Na, Cu and zinc and less Mg, P and K. The coefficient of variation (CV) of the copper content expressed as a percentage is significantly higher than that of other elements, and the coefficient of variation (CV) of boron content is the lowest.

Table 9: Global statistics on nutrient levels in apple fruit twigs.

	Average	Standard Deviation	CV	Min	Max
Azote, N (%)	0.42000	0.151217	36.004	0.2100	0.560
Magnesium, Mg (%)	0.612500	0.4090130	66.778	0.2500	1.100
Phosphorus, P (%)	0.060000	0.0529150	88.192	0.0100	0.130
Potassium, K (%)	2.4800	0.84573	34.102	1.31	3.26
Calcium, Ca (%)	1.8000	1.07316	59.620	0.60	3.10
Manganese, Mn (ppm)	4.5100	1.98741	44.067	2.10	6.30
Zinc, Zn (ppm	11.9675	7.93202	66.280	2.60	19.70
Copper, Cu (ppm)	2.2225	2.05596	92.506	1.02	5.30
Boron, B (ppm)	4.9475	0.94026	19.005	4.12	6.30
Sodium, Na (%)	0.2100	0.07659	36.473	0.15	0.31
Chlorine, Cl (%)	0.1525	0.0377492	24.754	0.1100	0.200

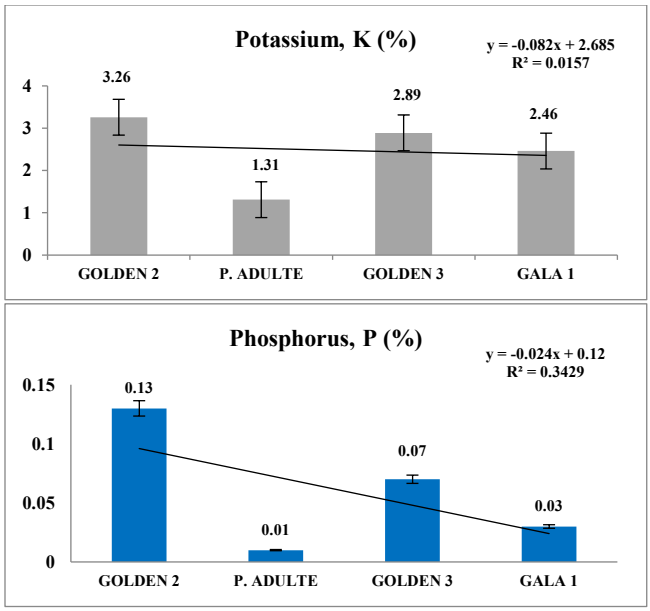


Figure 4: Phosphorus and potassium content in different varieties of apple fruit.

4. Discussion

The apple tree represents in Morocco an agricultural species of great importance. Growing apples requires a long period of vegetative rest to meet its cold needs; it also requires very specific conditions with regard to the various production factors, in particular for the soil / irrigation water combination. The main areas of apple plantations in Morocco are; Meknes, Midelt, Khenifra, Haouz Marrakech, Fes, Ouarzazate. Our study was carried in Beni Mellal-Khenifra region. The results of the granulometric analysis of the studied soils showed that the samples have a silty clay texture, which is often very favourable for apple tree cultivation. A good yield could therefore be obtained if the mineral elements provide the essential and important quantities for the good development of the crop. According to the literature review, the physicochemical properties of the soil play a major role in the yield of apples, in particular the pH [10,11]. It should be noted that apple trees generally prefer slightly acidic soils with a pH between 6.5 and 6.8 [12]. However, the soil samples analyzed recorded rather average alkaline pH values of up to 8.85. Despite the measured pH, qualified as alkaline, the yields obtained on the farm are very satisfactory. Indeed, even if we often talk about an optimal pH for a particular plant, in fact the pH does not meet a physiological need of the plant. It is the action of pH on the physical, chemical and biological properties of the soil that creates a more favourable environment for mineral nutrition and plant growth. In fact, plants are most often able to cope with a pH difference of up to one unit. Most plants grow very well in soil with a pH of 6.5 [12]. Doucet in 2006 states that agricultural species that grow at pH 6 most often prefer a pH of 6.5 and that all plants that grow in neutral soil with a pH of around 7 can tolerate a pH of 6.5 [13]. Regarding active limestone, apple tree fears soils with a high content of active limestone. This is not the case for our study soil with variable proportions which can sometimes be very high or even twice the tolerated standards. This can easily lead to iron deficiency. Furthermore, the more abundant the total CaCO3 rate in the soil, the greater the risk of blocking trace elements and phosphorus, hence the need to take these constraints into account

when developing fertilization programs. For a good interpretation of the rate of organic matter, it is necessary to take into account the quantity, but also the quality of this organic matter, the ventilation, humidity and temperature conditions. According to the guide reasoned fertilization in fruit arboriculture, presented by the Chamber of Agriculture of Tarn-et-Garonn, the value of the desirable organic matter content according to a clay content> 30 % must be between 2.5 and 3%. However, the organic matter content in the present study is generally low, with contents varying between 0.48% and 3 %. Nevertheless, it is therefore necessary to add MO to the soil. As long as the OM content is high, the production potential of the soil is great. In addition, it will improve the physical properties (structure, porosity, water and air circulation, water retention capacity) and chemical (cation exchange capacity, reversible anionic bond, release of trace elements. by decomposition in the soil) [15]. Compared to the interpretation standards given by the INRA (National Institute of Agronomic Research) for Contribution to the establishment of standards of fertilization of apple trees in the Middle Atlas. According to Dr. Rachid Razouk (researcher, CRRA Meknes), the optimal levels of nutrients found in clay-loam soils (the case of the study) are classified as low for phosphorus, in the “sufficient” class for potassium and in the "high" class of Ca and Mg [16]. However, for K2O, the values found correspond in fact to the optimum intakes for a high apple yield. Internationally, experts agree that 1.5% is the theoretical critical limit, below which fertility declines rapidly. It is often desirable to aim to maintain a minimum rate of 2.5% and even 3.5 to 4% in heavy soils [13]. A similar study was carried out to assess the quality of agricultural soils in the northeastern area of the plain Tadla (Morocco).The results indicated that 1.12% of the study area has poor suitability for sustainable intensive agriculture due to their unsuitable texture and low salinity, while about 77% of cultivated soils are adapted to agricultural production [17]. According to Barakat et al, 2017 the diagnostic of the quality of agricultural land in Beni-Moussa, Tadla plain (Morocco), showed that about 1.69% (1324.38 ha) of studied area are classified in the clas ‘excellent suitable’, 17.14% (13449.81 ha) in ‘good suitable’, 75.70% (59393.30 ha) in ‘medium suitable’ and 5.47% (4293.11 ha) in ‘poor suitable’ for sustainable intensive agriculture. Reading the results of the leaf analysis and their comparison with international standards for the nutrient content of apple leaves shows that for the elements phosphorus, calcium and magnesium, their concentration levels in the leaves are too satisfactory. On the other hand, the nutrient potassium has levels in the leaves below the desired limits compared to the required reference standards. A slight potassium deficiency is probably due to excessive nitrogen fertilization, as nitrogen-induced vegetative growth generates high potassium requirements. In general, leaf levels of N, P, K, Ca, and Mg are significantly correlated with their levels in the soil. However, foliar analysis is an effective tool for assessing the nutritional status of trees and for readjusting fertilizer requirements taking into account factors that can affect nutrient availability and root uptake.

5. Conclusion

The soils studied and sampled at different depths (0-60 cm) have a clay-loam texture. They are relatively poor in organic matter, the content of organic matter and major elements are not sufficient to optimize the optimum yield of the apple tree culture. The study shows a correlation between apple yield and nutrient content in the soil and leaves. The nutrient deficiencies observed can be corrected by the application of fertilization and adequate maintenance. However, further studies are needed to fully understand other variables such as the nature of rootstocks, the varieties cultivated and the method of irrigation which strongly influences apple tree production and yield. A prospect may be distant; indeed, it would be very interesting to generalize our paradigm to other sectors related to apple growing. As well as tackling research on phytosanitary techniques in order to move from the conventional vision to the integrated vision. Further research could also focus on other crops.

References

[1] R. Moussadek Impacts of conservation agriculture on the properties and productivity of vertisols in Central Morocco. Afrika focus. **25** (2) 147-151, 2012.

[2] N. Akesbi. Evolution and perspectives of agriculture Moroccan. In Report 50 years of human development and perspectives for 2025. Casablanca. Maghrebi. **85**-198, 2006.

[3] Agricultural credit in Morocco: Accessed on November 09, 2020. Available at: https://www.fellah-trade.com/fr/ plant-sector / technical-sheets / apple tree.

[4] P. Milleville, G. Serpantie. Agrarian dynamics and the question of the intensification of farming in the Sahelian and savanna zones of Africa. Accounts of the French Academy of Agriculture **80**, 149-161, 1994.

[5] B. Diwediga, K. Hounkpe, K. Wala, K. Batawila, T. Tatoni, Akpagana, K. Off-season agriculture on the banks of the oti and its tributaries. African Crop Sci. J. **20**, 613-624, 2012.

[6] R. Moussadek. Impacts of conservation agriculture on the properties and productivity of vertisols in Central Morocco. Afrika focus. **25**(2), 147-151, 2012.

[7] Regional Directorate of Agriculture of Beni Mellal-Khenifra, agricultural monograph of the Beni Mellal-Khenifra region. 2016. Accessed on March 26, 2020.

[8] P. Martin –Prevel, J. Gagnard, P. Gautier. Plant analysis in the control of the diet of temperate and tropical plants. 131, 1984.

[9] A. Finigue. Water resources in the direction of the Beni Mellal Atlas: management issues and local development. 2018. University of Angers Doctoral School DEGEST Angers Thesis. Accessed on March 26, 2020. Available on https://tel.archives-ouvertes.fr/tel-01708289/document.

[10] Ministry of Agriculture and Maritime Fisheries. Technology transfer in agriculture. 2008. Consulted on March 26, 2020. Available at https://www.agrimaroc.net/bulletins/btta_165.pdf.

[11] K. Aggelopoulou, D. Wulfsohn, S. Fountas, G. Nanos, T. Gemtos, S. Blackmore. Spatial variability of yield and quality in an apple orchard. Precision Agriculture **11**(5), 538–556, 2010.

[12] K. Aggelopoulou, S. Fountas, D. Pateras, G. Nanos, Gemtos, T. Soil spatial variability and site-specific fertilization maps in an apple orchard. Precision Agriculture **12**(1), 118–12, 2011.

[13] K. Pontikis, Applied Pomology. Stamoulis Publications, Athens, Greece, 74.

[14] R. Doucet. Climate and agricultural soils. ed. Berger, Eastman, Quebec. xv, 443, 2003.

[15] Tarn and Garonne. Chamber of Agriculture. Guide To Reasoned Fertilization In Fruit Arboriculture. 2019. Accessed March 27, 2020. Available at: [https://mag.inrameknes.info/? P = 1400.

[16] R. Razouk Contribution to the establishment of apple tree fertilization standards in the Middle Atlas. 2018. Accessed March 27, 2020. Available at: https://mag.inrameknes.info/?p=1400.

[17] W. Ennaji, A. Barakat, M. El Baghdadi, H. Oumenskou, M. Aadraoui, L. A. Karroum, & A. Hilali. GIS-based multi-criteria land suitability analysis for sustainable agriculture in the northeast area of Tadla plain (Morocco). Journal of Earth System Science, **127**(6), 79, 2018.

[18] A. Barakat, W. Ennaji, A. El Jazouli, R. Amediaz, & F. Touhami. Multivariate analysis and GIS-based soil suitability diagnosis for sustainable intensive agriculture in Beni-Moussa irrigated subperimeter (Tadla plain, Morocco). Modeling Earth Systems and Environment, **3**(1), 3, 2017.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1109-1124 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Multi-Criteria Decision Analysis Coupled with GIS and Remote Sensing Techniques for Delineating Suitable Artificial Aquifer Recharge Sites in Tafilalet Plain (Morocco)

Aicha Ousrhire[*,1], Hassane Oulidi Jarar[2], Abdessamad Ghafiri[1]

[1]*Laboratory of geomatics applied to geosciences and the environment, Department of geology, Hassan II university, Casablanca, 20660, Morocco*

[2]*SGEO research team, Department of mathematics, computer science and geomatics, Hassania school of public works, Casablanca, 20230 ,Morocco*

ARTICLE INFO

Article history:
Received: 12 October, 2020
Accepted: 17 November, 2020
Online: 14 December, 2020

Keywords:
AHP
Artificial aquifer recharge
Aquifer depletion
GIS
Remote sensing
Tafilalet plain
Morocco

ABSTRACT

Despite that groundwater is an important and vital water resource, it is not well managed; depletion of aquifers around the world due to overexploitation is of serious concern especially in arid regions where the situation is much more alarming. Tafilalet plain in Morocco which belongs to this type of environment is certainly no exception and is viewing its groundwater disappearing. Artificial aquifer recharge (AAR) is found to be appropriate to such an urgent issue. Thus, the objective of this paper is the exploration of suitable sites to process Artificial aquifer recharge in Tafilalet plain by the joint use of remote sensing (RS), geographic information systems (GIS), and the analytic hierarchy process (AHP) method. For doing so, eight parameters were considered as groundwater influencing parameters such as slope, soil, geology, land Cover/land Use, depth to the water table, aquifer transmissivity, electrical conductivity and drainage density. The laters were integrated and processed in a GIS, their thematic layers were created, and their relative weights were generated by The AHP method based on their significance in recharging the aquifer. Afterward, thematic layers were reclassified and assigned their weights, so the GIS overlay tool was used for inferring artificial groundwater recharge potential regions in the study area. 47 % were identified as suitable while only 12 % were identified as unsuitable. Such studies facilitate groundwater management for stakeholders and water managers because important decisions may be taken in a record time which will preserve water resources and prevent them from being in an alarming situation.

1. Introduction

Groundwater is a valuable natural resource and constitutes a significant freshwater reservoir as it accounts for 30 % of the earth's freshwater while surface water accounts for just 0.3 % [1, 2]. Besides, groundwater offers a multitude of benefits in different forms. In fact, groundwater meets the need for drinking water to relatively half of the world's population [3, 4] and around 38 % of the world's irrigated lands are supplied by groundwater [5], not to mention its contribution to the industry sector and its paramount role for the environment. Nevertheless, there is growing evidence that groundwater is under threat of numerous difficulties such as climate change impacts, contamination, pollution in addition to depletion [6–10]. Actually, many aquifers throughout the world suffer from groundwater depletion [11–17] and this situation is much more serious in arid regions since groundwater constitutes their main source of water supply. Mismanagement and namely over-exploitation are considered as its major cause alongside many other factors. In fact, rapid population growth, agricultural and industrial activities that require a considerable availability of water together with climatic conditions notably of arid environments consisting of high evapotranspiration rates, low precipitation, and low vegetation cover are placing greater pressure on limited aquifer resources which has led to their intensive extraction in many parts of the world [9, 12, 17–25]. Moreover, If this abstraction is greater than the natural groundwater recharge, it is referred to as an overdraft; so, groundwater levels will drop sharply over time, triggering depletion [8, 9, 14, 17, 25–30], which can make the situation complicated as the replenishment of the aquifers

*Corresponding Author: Aicha Ousrhire, Email: aicha.ousrhire@gmail.com

to the original levels takes a lot of time [3] and it is more complicated in cases of fossil or compacting aquifers because of their irreversible and non-renewable character [14, 31]. In such a way, countless negative impacts and many political, social, environmental, and economic issues, such as conflicts, land subsidence, degradation of groundwater dependent ecosystems, deterioration of water quality, aquifers drying and obviously water scarcity will take place particularly in countries which are groundwater dependent [10, 11, 14, 23–25, 27, 28, 32–39].

The above situation has been a major point of concern of water researchers, water managers, and water decision-makers, seeing that in many parts of the world a lot of measures have been taken to properly manage their groundwater resources and to prevent the worsening of the situation. The implementation of monitoring networks to control groundwater levels, the establishment of groundwater laws, and the setting up of exploitation rules for regulating its use, in addition to the organization of awareness-raising companies to urge on the importance of water and also education on water conservation are a good illustration of the undertaken measures [9,37,38,40,41], but even so, the decline of groundwater resources levels has been noticed and is still ongoing in many places around the globe. This is especially due to the violations of the established rules and the intensive mining of underground water notably in case of droughts. Added to that, the natural recharge of aquifers by streams, precipitation, rivers, pond seepage, reservoirs, irrigation canals, and lakes has become overwhelmed by the intensity of groundwater withdrawals [25] and has become difficult to occur especially in urbanized areas [42], because the presence of impervious surfaces such as roads, sidewalks, driveways and parking lots has led to many modifications to natural processes like impeding the infiltration of water. Thereby, the natural recharge of aquifers has proved to be insufficient to fill the quantity of water coming out of the aquifers in many parts of the world. Given these trends, water researchers, water managers, and water decision-makers have been prompted to think about solutions and innovative scientific analyses in order to bring better and sound management strategies that may address this groundwater issue and guarantee its sustainable use. If decreasing discharge and regulating abstraction has not brought important results, then this depletion can only be alleviated by increasing and boosting recharge.

Thus, artificial or managed aquifer recharge is a promising solution to this problem. Actually, artificial aquifer recharge (AAR) or managed aquifer recharge (MAR) is defined as an intentional human intervention arranged to enhance groundwater resources through processes implemented to heighten and improve the replenishment of depleted aquifers or depleting aquifers [22, 43–47]. The (AAR) has many gains and benefits, the major ones are the restoration of groundwater balance and the revival of the overexploited aquifers [11, 19, 25, 39, 48, 49], the other ones are enhancement and raising the water table which will cut down its operating costs, improving groundwater quality and prevention of seawater intrusion in the case of coastal aquifers [28, 46, 49, 50], alleviating floods adverse impacts given that during strong storms aquifers can be managed to play the role of a water storage reservoir which will mitigate the magnitude of flooding [22, 45, 51–53], besides (AAR) can limit the loss of land covered with surface water storage structures and limit the loss of water by evaporation in dams, ponds, and other surface reservoirs by storing it in aquifers [12, 39]. In order to artificially recharge aquifers, treated wastewater, floodwater, or water pumped from another aquifer can be used and redirected across the land surface. So, a variety of ways and systems have been developed such as infiltration basins, canals, ponds, ditches, sprinkler systems, rills, and irrigation furrows where the surface water is put on and subsequently infiltrates into the soil and moves downward to recharge aquifers or simply artificial recharge can be achieved by injecting water directly into the aquifer through injection wells especially in case of confined aquifers and each of these ways is implemented according to specific conditions. Applying the (AAR) needs some criteria to be met notably, the presence of highly permeable geological formations allowing percolation of recharge water, endowed with high capacity of water storage (unsaturated zone thickness should be deep enough to increase the piezometric level) and have high transmissivity. Moreover, the presence of a water source ensuring volumes of water needed for the recharge such as dams, stormwater runoff, wastewater treatment station, and desalination station besides, the water intended for artificial groundwater recharge as well as those of receiving aquifer must be of good quality and meet a certain number of standards and also good knowledge of the climatic characteristics of the region is essential in (AAR) projects (temperature, evaporation, rainfall, frequency of floods, periods of drought, snowmelt, etc.). Consequently, performing the (AAR) requires a suitable place whose selection depends on solid scientific analysis. Actually, selecting optimal sites for (AAR) involves integrating several parameters proved to have an influence on the infiltration of water to aquifers. To do this, a combination of geospatial technologies and multi-criteria decision analysis (MCDA) is used in this study.

Tafilalet plain was chosen as the study area in the present study as it is one of the Moroccan regions calling for urgent intervention and requiring important measures to be taken at all levels, including the environmental one; it suffers from enormous environmental difficulties mainly scarcity and poor water quality, due essentially to long periods of drought that has affected the region in recent years, combined to severe nature of the climate which makes groundwater resources the main source of water supply in low-flow periods. In 1967, the Hassan Addakhil dam was constructed in the north of Errachidia city and located at a distance of about 80 km from the Tafilalet plain, it was built on wadi Ziz in order to regulate floodwaters and, simultaneously, it serves a variety of goals, ranging from irrigation to water supply for domestic and industrial uses and remarkably it helps to establish a certain balance of groundwater of Tafilalet aquifer both quantitatively and qualitatively, but in front of considerable population growth and an agrarian-based economy resulting in increasing the number of wells notably prived ones, the exploitation of groundwater resources is proceeding at a very fast pace which has lead to their overdraft, so the implementation of artificial aquifer recharge in this region is essential to recover its overexploited groundwater.

Thus, the prime focus of the current paper is to delineate the artificial groundwater recharge potential regions in Tafilalet plain (Morocco), using the analytic hierarchy process (AHP) method which is a forme of Multiple-criteria decision analysis (MCDA) and geospatial technologies represented here by remote sensing (RS) and geographic information systems (GIS).

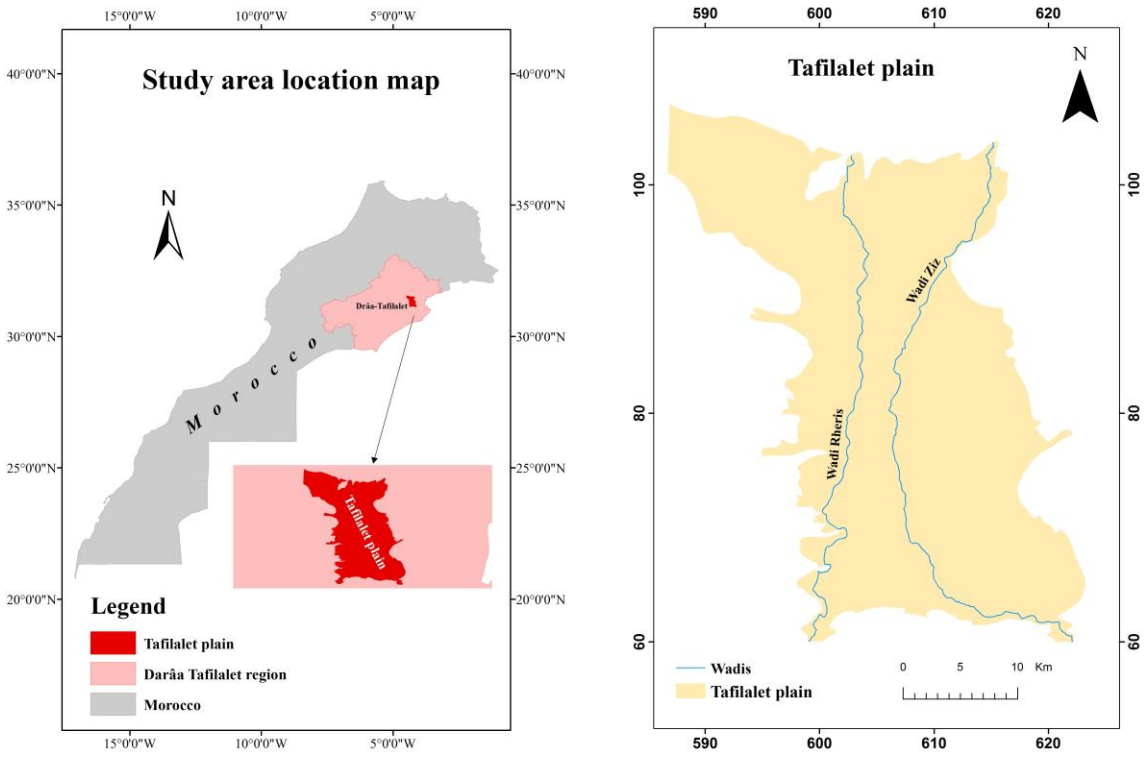


Figure 1: Study area location map

After this introduction, the following parts are organized in five sections, the first one is an overview of the study area, the second one is a presentation of the materials and definition of the methods used, the third one is a description of the major results, the fourth one is a discussion about the outcomes and the last one is a conclusion of the main ideas of this study.

2. Study Area

Tafilalet plain belongs to the pre-Saharan region located in the extreme southeast of the kingdom of Morocco, particularly in the Drâa-Tafilalet region, at an altitude of 700 to 800 m, it lies in between the coordinates of projection system Lambert zone Morocco (Zone I) X = 580000 m to X = 630000 m and Y = 50000 m to Y = 120000 m (Figure 1), covers an area of about 700 km², and the majority of its population are rural. Besides, Tafilalet plain is part of the quaternary basin of Tafilalet; it is an erosional depression arising out of the secondary and tertiary coverage clearing during the quaternary era, the bottom of this depression was, afterward, covered by the alluvium of wadis Ziz and Rheris. In fact, it constitutes the zone of gathering and accumulation of alluvial deposits, conveyed by surges of the Ziz and Rheris wadis and their tributaries. Moreover, Tafilalet plain extends over three different geological formations, it is enclosed by the High Atlas in the north, it crosses the pre-Africain furrow and is surrounded at the south by the old massif of the Anti-Atlas domain as well as overlying an impermeable Palaeozoic and pre-Cambrian substratum.

Concerning the climate, Tafilalet plain is dominated by an arid to semi-arid climate, typified by a great rainfall scarcity as the average annual precipitation barely reaches 70 mm all over the area, this is mainly due to the presence of Atlas reliefs in the north which constitute a barrier to oceanic influences. The climate is also characterized by a great irregularity of the temperature, rising to very high values in summer (about 40 degree C ° especially in July) and dropping to very low values in winter (surroundings 0 degree C ° especially in January), this great difference in temperature brings out two seasons: a hot season from May to October and a cold season from November to April so, in the Tafilalet plain, The dry periods last almost all the year with extremely hot summers and highly cold winters which is mainly due to its openness on the Sahara. In addition, The mean annual evapotranspiration is close to 500 mm which is higher than the mean annual rainfall, this means that Tafilalet plain suffers from water deficit all over the year. Adding to that, the dominant winds are commonly known as Chergui, they are characterized by their hotness and dryness and they blow from the southeast to the northeast.

Regarding the water resources, Tafilalet plain is, on the one hand, underlain by confined and unconfined aquifers, the north of the region encloses the infra Cenomanian aquifer which is a totally confined aquifer, whereas the entire area is almost occupied by an unconfined, shallow and very extensive aquifer named as "Tafilalet aquifer" which is the aquifer targeted in this study. In fact, this aquifer is not very thick and the water below the ground surface is 2 m to more than 30 m in some places, its shape matches the slightly convex shape of the plain and its slope is very weak. Consequently, this makes the water of this aquifer sensitive to hydrological and atmospheric conditions such as pollution and evapotranspiration. The spatial distribution of Tafilalet aquifer transmissivities indicates dispersed values ranging from 10^{-2} m²/s

to 10^{-4} m²/s and groundwater flows mainly from the north-east to south-west, south, and south-east characterized by an average hydraulic gradient that declines from upstream to downstream [54]. From a lithological point of view, the base of the aquifer is constituted by conglomerates and lacustrine limestones including gravelly levels and at the top of the aquifer, there are sands, gravel, and pebbles, covered with unevenly distributed powerful silts. Moreover, the main sources of groundwater recharge of the Tafilalet aquifer are essentially from the High Atlas mountains surrounding the plain in the north, from surface runoff including rivers, mainly Ziz and Rheris wadis, as well as irrigation water which infiltrates to the aquifer, and without forgetting precipitation and dams contribution. Conversely, the Tafilalet aquifer discharges mostly by evapotranspiration, by extraction from wells, and by Khettarats and springs. On the other hand, Tafilalet plain is crossed by many ephemeral wadis, the major ones are wadi Ziz in the center, wadi Rheris in the west. Besides, the region holds some springs resulting from the emergence of groundwater resources to the surface, such as Aïn Dolla-Attrouss and Aïn Chorfa.

Furthermore, the heart of the Tafilalet plain is made up of a very large palm grove, which occupies most of its surface so, a great part of the plain is devoted to agriculture, especially irrigated agriculture, making it the main economic activity in this region and is the main source of income for a large population. Because of climate aridity, the majority of crops are water-demanding which implies a strong mobilization of water notably during the months of June and July and due to the scarcity of surface waters in the region, growers are turning to groundwater exploitation to fill this water gap and meet their water needs. This has made the groundwater balance in aquifers of Tafilalet plain quickly in deficit.

3. Materials and Methods

The overall process followed in this study of the artificial aquifer recharge site selection has been schematically presented in (Figure 2).

3.1. Influencing factors Selection reason

The prediction of suitable sites for artificial aquifer recharge depends on a number of parameters that have an influence in either a direct or indirect way on the process of aquifer recharge. The selection of these factors was based firstly on a deep review of studies of artificial recharge carried out in other areas in the world, secondly on a complete understanding of the characteristics and specificity of the study area "Tafilalet Plain" and thirdly on the point of view of experts. Thus, eight factors were revealed to have control over the recharge of the Tafilalet aquifer, including slope, soil, geology, land cover/ land use, depth to the water table, aquifer transmissivity, electrical conductivity and drainage density as mentioned in (Table 1).

- Slope: The slope is one of the essential factors influencing the artificial aquifer recharge because the slope directly affects the penetration of water into the soil. Actually, the flatter the ground, the more time the water has to sink in, unlike the steep slopes where the water runs over the ground rather than entering it [25, 55–57]. However, the penetration of water is not guaranteed even if the slope is suitable because it depends on the soil's nature.
- Soil: Soil is a factor of significant importance in aquifer recharge because soil's nature determines the percolation ability of the soil as well as the water infiltration quantity and eventually governs the recharging capacity of the bedrock [58]. Moreover, soil acts as a purificator of water [29] that is why it should contain high rates of loam and sand to avoid groundwater contamination and to ensure a high infiltration rate.
- Geology: Taking into consideration geology in artificial aquifer recharge site selection is of major importance because geological formations which must be selected should enjoy high permeability so that water can flow seamlessly downward towards the groundwater besides, they should have adequate transmissivity and high water storage capacity.
- Depth to the water table: Depth to the water table or unsaturated zone thickness is the distance from the ground surface to the groundwater, it is another major factor to consider in artificial groundwater recharge site selection since, the deeper the water table, the greater the amount of groundwater storage, and the more efficient the artificial aquifer recharge will be Moreover, the deeper the water table, the more purified the water will be, because it will pass through a very large layer of soil, which is called soil aquifer treatment (SAT) [19, 29]. Acceptable depth is about 3 m consequently, sites, where the depth is less than 3 m, should not be considered [59].
- Transmissivity: transmissivity is an important hydrodynamic parameter of an aquifer which makes its consideration paramount in the selection of suitable sites for artificial groundwater recharge. Actually, transmissivity determines the rate of water flowing per unit width of the fully saturated thickness of an aquifer under a unit hydraulic gradient [25, 60–62] so, to avoid the excessive formation of groundwater mounds that interfere with the infiltration process, aquifers should be unconfined and sufficiently transmissive to allow the lateral flow of infiltrated water well away from the recharge zone [49].
- Drainage density: Drainage density is defined as the total length of streams of all orders per unit area and provides an indication of the watershed draining state. Drainage density was heeded as another factor of artificial aquifer recharge site selection because it indirectly signals the suitability of a site. In fact, high drainage density allows less penetration and more surface runoff and areas with great infiltration capacity are endowed with low drainage density [25, 39, 51, 58, 63–68] hence appropriate to artificial aquifer recharge.
- Electrical Conductivity: Groundwater quality is a paramount parameter that should be investigated in the artificial aquifer recharge projects because it is a determinant factor of the future usage of that water. Usually, the water intended for artificial aquifer recharge is of good quality but when it penetrates across a geological formation and mingles with different dissolved materials in it, its quality may get altered. Electrical conductivity (EC) is considered in this study because its variations have similar trends to those of total dissolved solids (TDS) as shown in the following formula :

$$EC = TDS / 0.65 \quad (1)$$

In fact, the higher the electrical conductivity, the poorer the water quality and vice versa [69]. Thus, the quality of groundwater will be assessed through the EC in the artificial aquifer recharge site selection.

- Land Cover / Land Use (LCLU) : (LCLU) is one of the important aquifer recharge influencing factors, it affects both the infiltration capacity and the water restraining of the land. Actually, areas with an important density of vegetation cover are appropriate for undertaking artificial aquifer recharge projects because the denser the vegetation cover, the more the runoff is slowed down and the more water infiltration takes place and also the less evapotranspiration and soil surface erosion will occur. In addition, organisms present in the soil and plant roots form openings in the subsoil, allowing water to infiltrate easily, unlike in areas without vegetation and in built-up areas, where the soil is mostly impervious, which favors runoff and prevents infiltration.

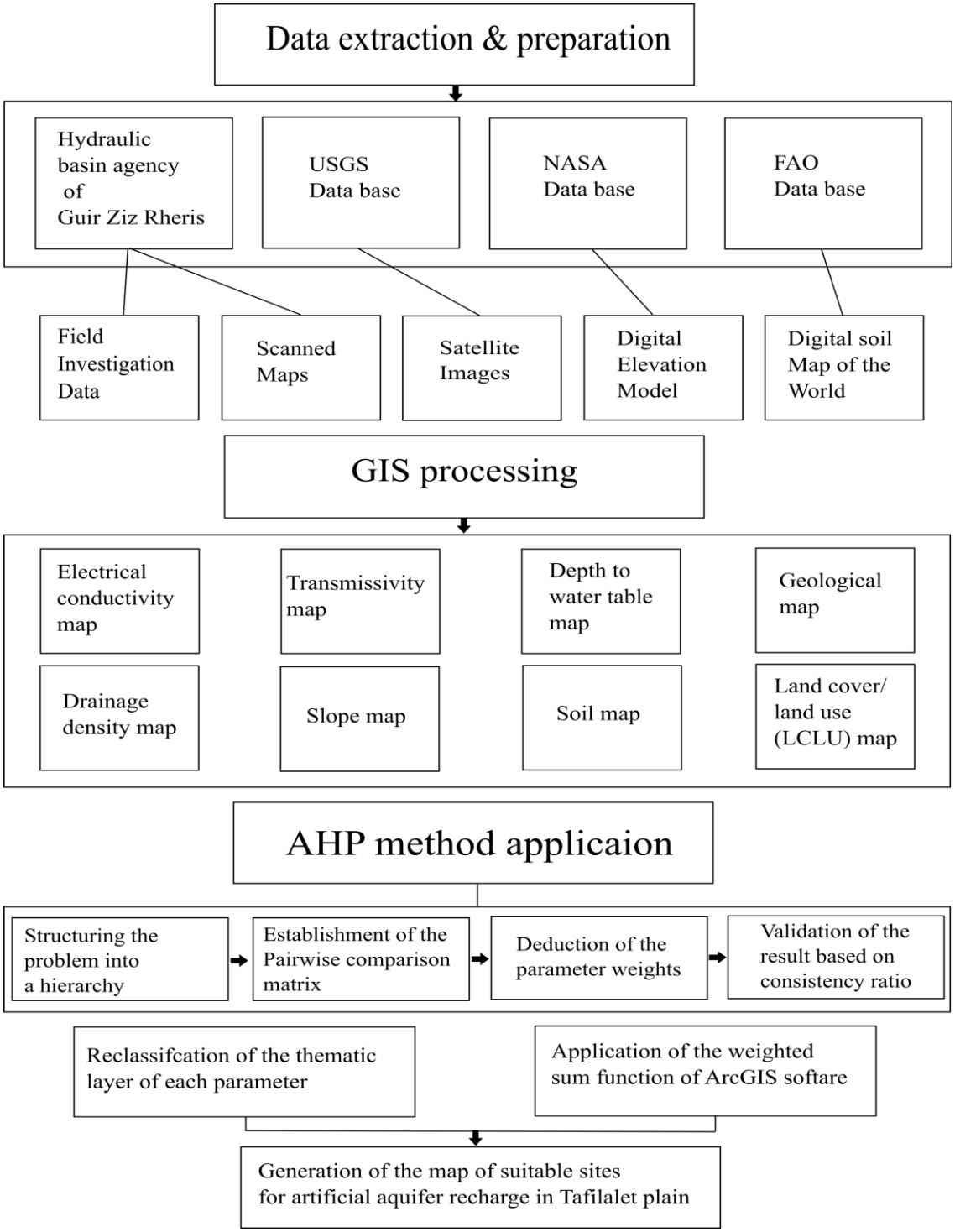


Figure 2: Methodology workflow

Table 1: Data Extraction and Preparation

Data	Origin format	Data source	Acquisition mode	Output
Slope	ASTER DEM at 30 m	National Aeronautics and space administration (NASA) website	Extraction	Slope map of Tafilalet plain
Soil	Digital soil map of the world	Food and Agriculture Organization of the United Nations (FAO) website	Extraction	Soil map of Tafilalet plain
Geological data	geological map of Tafilalet-Taouz (1/200000 scale)	Hydraulic basin agency of Guir Ziz Rheris	Digitalization	Geological map of Tafilalet plain
land Cover/land Use (LCLU)	satellite image (Landsat 8)	United states geological survey (USGS) institute website	Classification	LCLU map of Tafilalet plain
Depth to the water table	Excel sheet of field investigation data	hydraulic basin agency of Guir Ziz Rheris	Interpolation	Depth to the water table map of Tafilalet aquifer
Transmissivity	Excel sheet of field investigation data	hydraulic basin agency of Guir Ziz Rheris	Interpolation	Transmissivity map of Tafilalet aquifer
Electrical conductivity	Excel sheet of field investigation data	hydraulic basin agency of Guir Ziz Rheris	Interpolation	Electrical conductivity map of Tafilalet aquifer
Drainage density	ASTER DEM at 30 m	National Aeronautics and space administration (NASA) website	Extraction of the hydrographic network then application of Line Density	Drainage density map of Tafilalet plain

3.2. Data extraction, preparation, and GIS processing:

Given the interdisciplinary nature of groundwater, using remote sensing (RS) and geographic information systems (GIS) techniques are required because RS is a technology that provides important and information-rich geospatial data, especially for unreachable and very large areas. Moreover, GIS is commonly known for its usefulness to handle, analyze, integrate, and process various thematic geospatial data.

Therefore, in order to get the geographic data related to the above-mentioned factors, various data sources were used: FAO database, NASA database, USGS database, and other data were given by the hydraulic basin agency of Guir-Ziz-Rheris as detailed in Table 1. Furthermore, in order to process and to analyze those data, the geographic data processing functions of the ArcGIS software were used. In fact, as indicated in Table 1, slope and hydrographic network were extracted from ASTER DEM at 30 m, then the drainage density was formed by applying the Line Density function to the hydrographic network, the soil was extracted from the digital soil map of the world taken from the FAO database, geology was digitized from the geological map of Tafilalet-Taouz (1/200000 scale), electrical conductivity, transmissivity and depth to the water table were taken from field surveys carried out with the hydraulic basin agency of Guir-Ziz-Rheris and land cover/land use (LCLU) was classified from satellite image (Landsat 8) drawn from USGS database. Thus, a geospatial database was constructed and filled with the aforementioned data. Afterward, using ArcGIS software a thematic layer of each factor was created (Figure 3) and each thematic layer has been converted to raster format and reclassified into five classes pertaining to its importance for artificial groundwater recharge (Figure 4), each class is translated to a category and each category has assigned a score as it is detailed in Table 2. Actually, the five categories adopted are very good, good, moderate, poor, and very poor in accordance with a specific evaluation scale (Table 3).

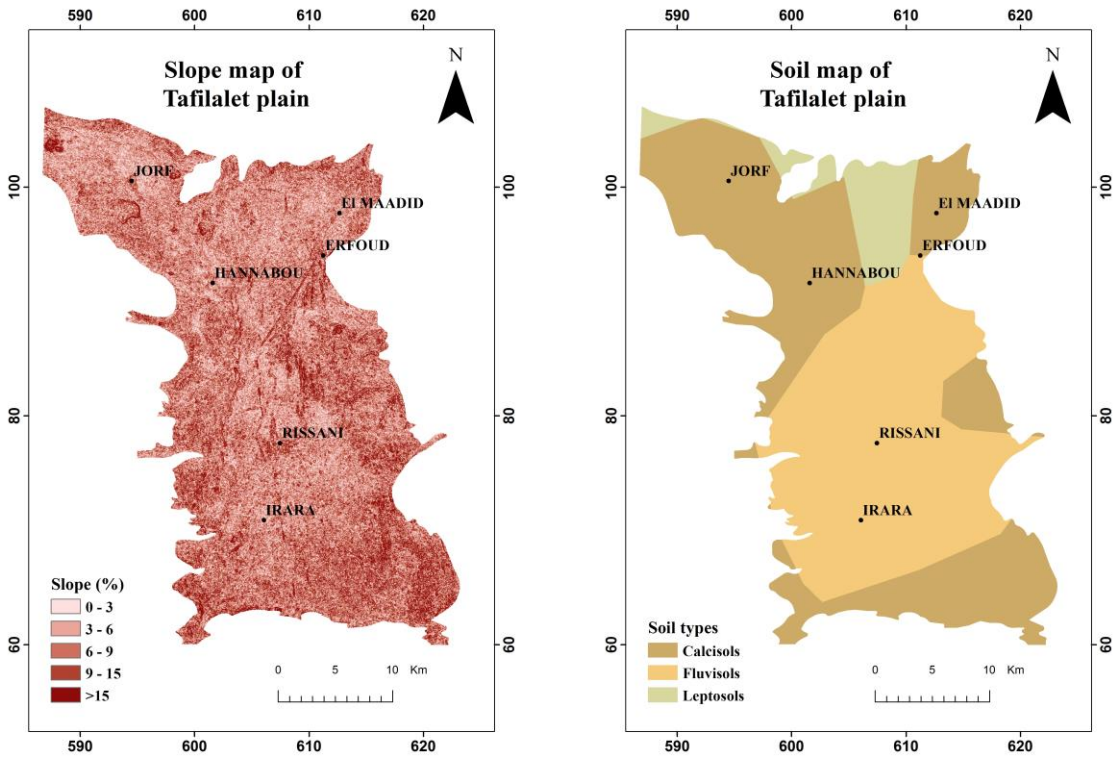
3.3. AHP method application

After reclassification of all the thematic layers, appropriate weights were generated by the analytic hierarchy process method (AHP).

Indeed, AHP is a multi-criteria decision analysis method and a mathematically-based technique [70], the mathematical underpinnings and the philosophy of which were developed in the seventies of the last century by the mathematician Thomas L. Saaty [71], AHP involves breaking down a convoluted multi-criteria decision problem into a hierarchy also called decision modeling in order to organize and better understand the problem and analyze the decision to be made.

a) Slope map

b) Soil map



c) Geological map

d) LCLU map

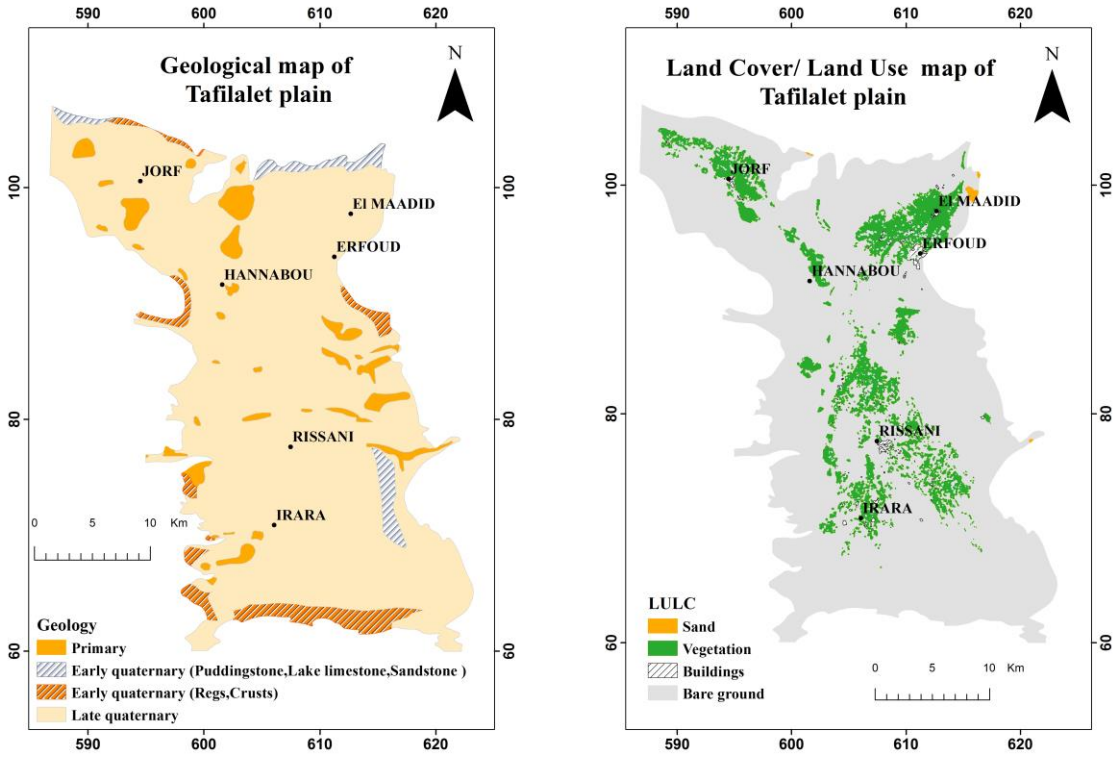
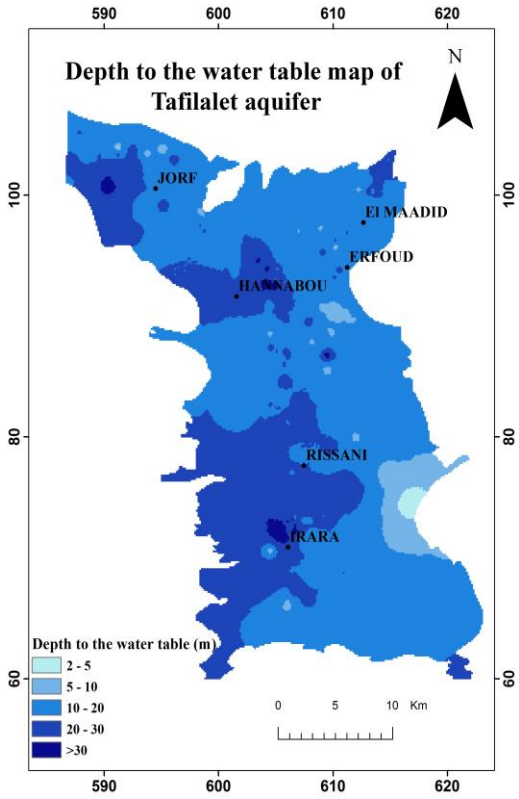
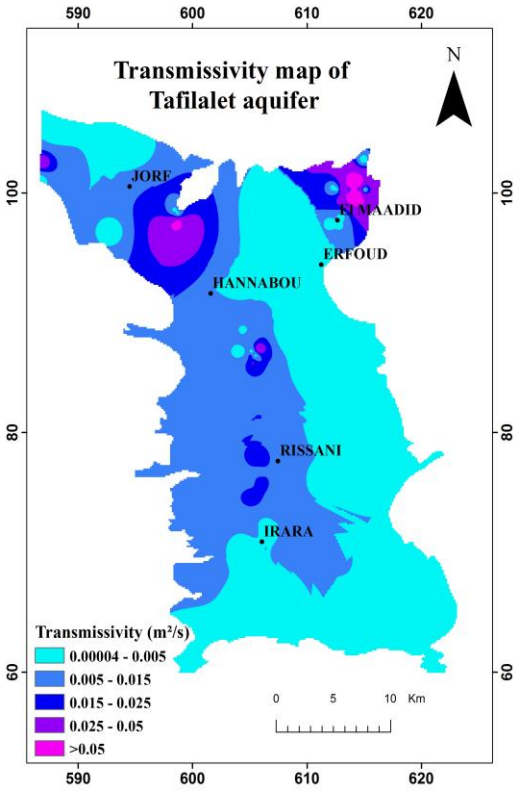


Figure 3: Thematic maps

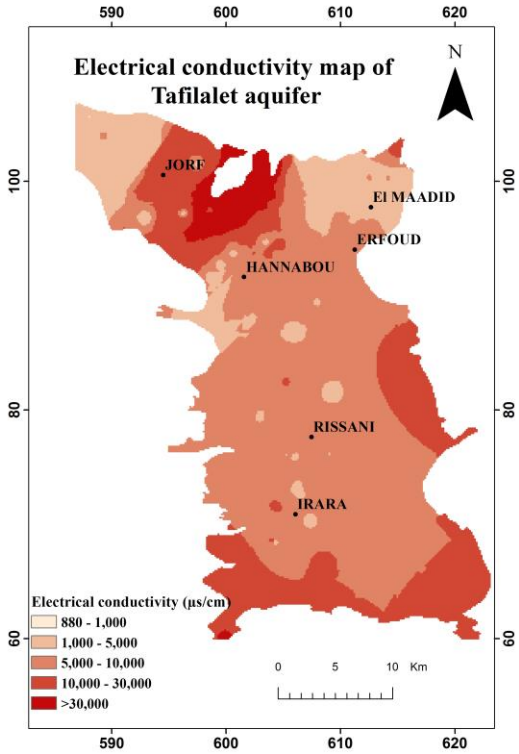
e) Depth to the water table map



f) Transmissivity map



g) Electrical Conductivity map



h) Drainage Density map

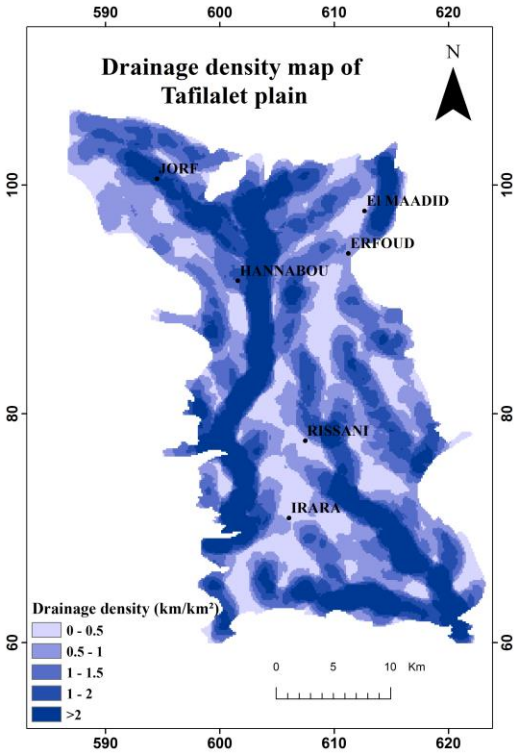
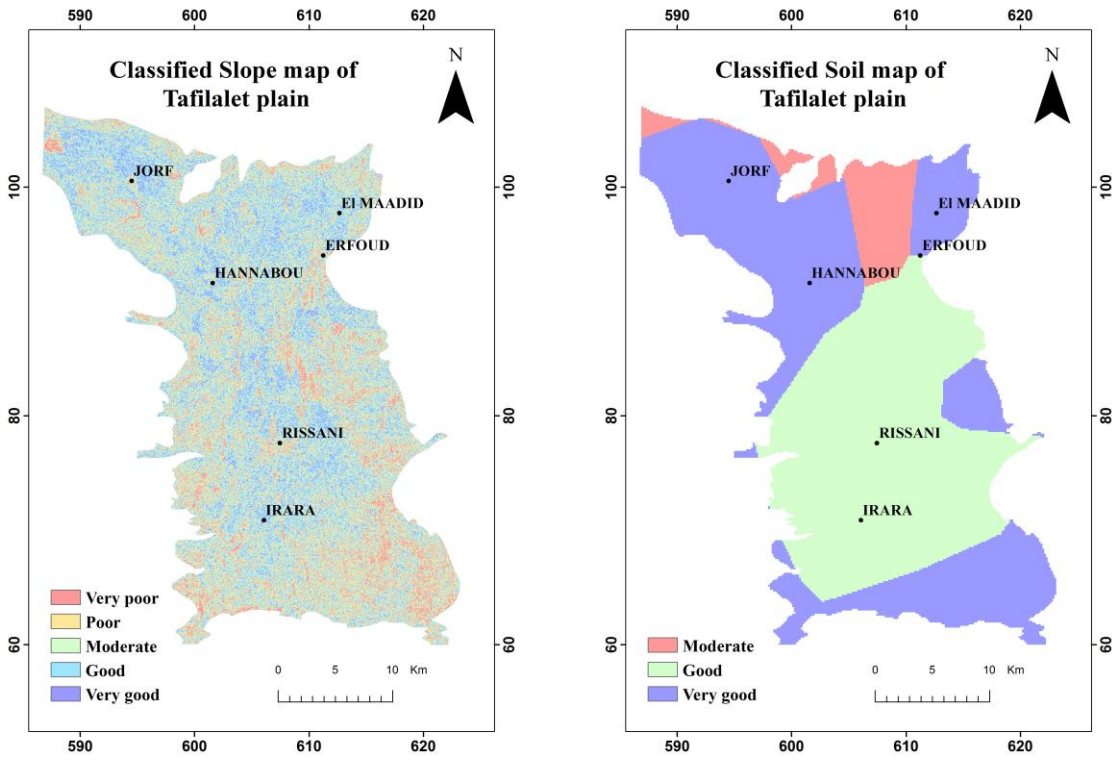


Figure 3: (Continued)

a) Classified Slope map

b) Classified Soil map



c) Classified Geological map

d) Classified LCLU map

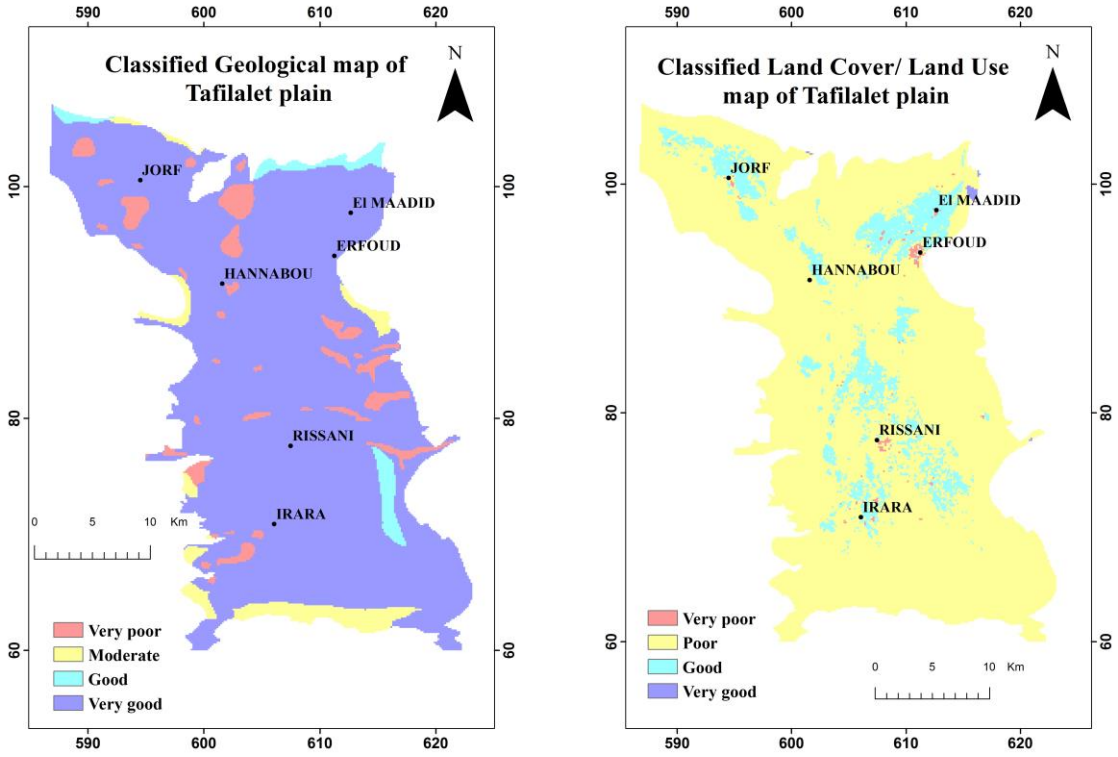
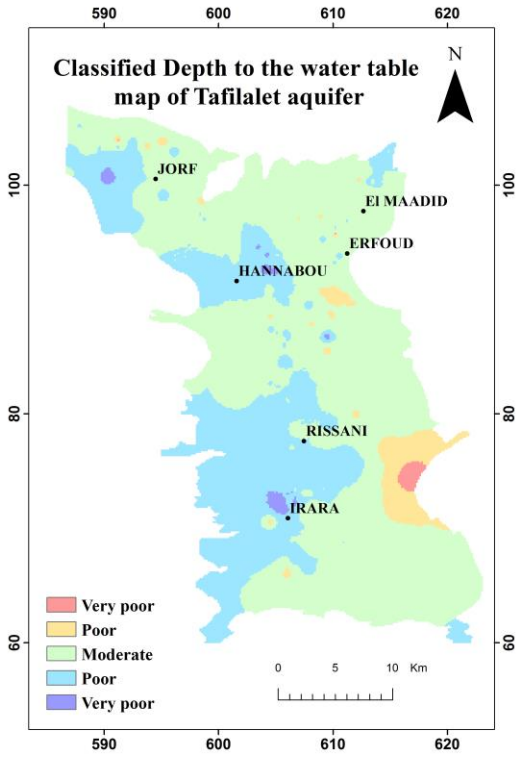
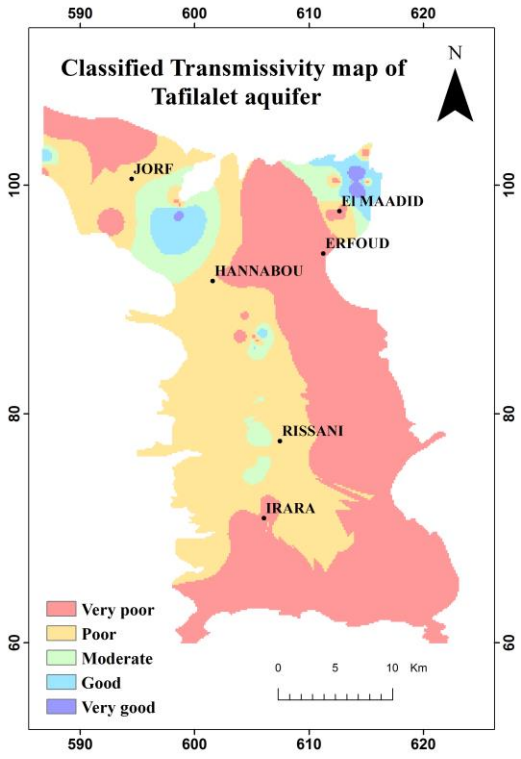


Figure 4: Classified thematic maps

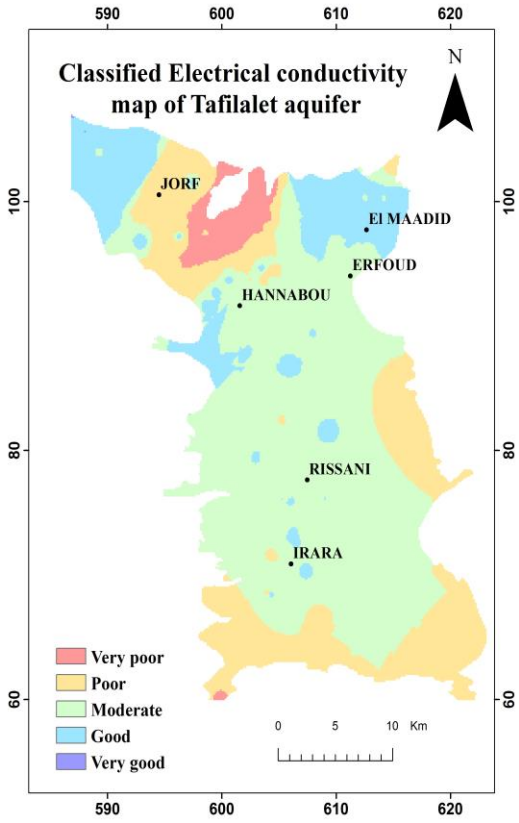
e) Classified Depth to the water table map



f) Classified Transmissivity map



g) Classified Electrical Conductivity map



h) Classified Drainage Density map

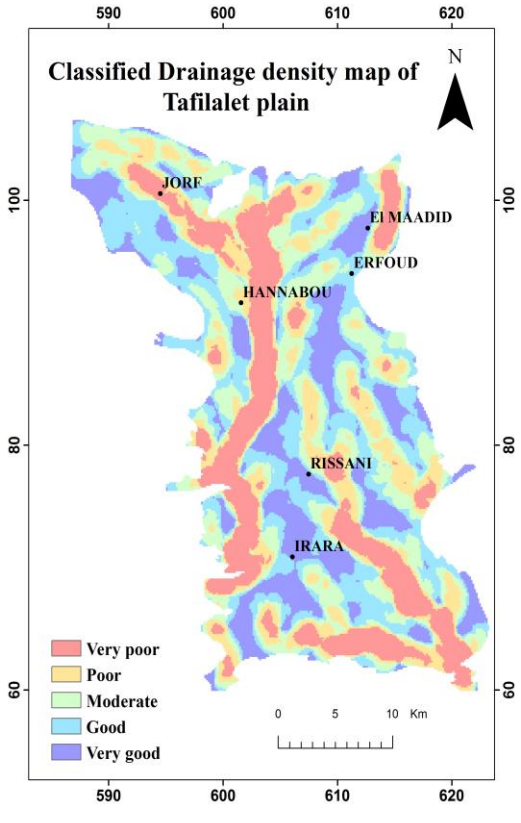


Figure 4: (Continued)

Table 2: Classes of each parameter and assigned categories

Parameter	Class	Category	Score
Slope (%)	0 - 3	Very good	9
	3 - 6	Good	8
	6 - 9	Moderate	5
	9 - 15	Poor	2
	> 15	Very poor	1
Soil	Calcisols	Very good	9
	Fluvisols	Good	8
	Leptosol	Moderate	5
Geology	Primary	Very poor	1
	Early quaternary (Regs, crusts)	Moderate	5
	Early quaternary (Puddingstone, Lake Limestone, Sandstone)	Good	8
	Late quaternary	Very good	9
LCLU	Sand	Very good	9
	Vegetation	Good	8
	Buildings	Very poor	1
	Bare ground	Poor	2
Depth to the water table (m)	2 - 5	Very poor	1
	5 - 10	Poor	2
	10 - 20	Moderate	5
	20 - 30	Good	8
	> 30	Very good	9
Transmissivity (m^2/s)	0.00004 - 0.005	Very poor	1
	0.005 - 0.015	Poor	2
	0.015 - 0.025	Moderate	5
	0.025 - 0.05	Good	8
	> 0.05	Very good	9
Electrical conductivity (µs/cm)	880 - 1000	Very good	9
	1000 - 5000	Good	8
	5000 - 10000	Moderate	5
	10000 - 30000	Poor	2
	> 30000	Very poor	1
Drainage density (km/km^2)	0 - 1	Very good	9
	1 - 2	Good	8
	2 - 3	Moderate	5
	3 - 4	Poor	2
	> 4	Very poor	1

After structuring the problem in such a way, the second step is to evaluate, compare and prioritize in a pairwise comparison matrix, based on decision-makers' judgments each criterion against the others pertaining to their importance in the problem decision making. To represent the magnitude of judgments, Saaty's fundamental scale should be used (Table 4), in doing so, relative weights of each criterion is deduced using Saaty's eigenvector technique and the generated weights must be validated because judgments are subjective and can typically be made with errors, so Saaty defined an index for measuring the degree of inconsistency which is the consistency ratio C. R. and it should be computed.

Table 3: Adopted evaluation scale for reclassification

Category	Very poor	Poor	Moderate	Good	Very good
Score	1	2	5	8	9

Actually, C. R. is a way of measuring the number of "errors" created when making the judgments, the rule-of-thumb is that if the C.R. is below 0.1, the errors are relatively small, thus the final estimate is acceptable [70,72–76].

The above-explained concept of AHP could be explained mathematically as follows:

Suppose $c_1, c_2, \ldots, c_n$ are n criteria influencing a decision with respect to a given problem, the pairwise comparisons matrix $B=[b_{ij}]$ of that problem is an $n \times n$ matrix and it may be represented as follows :

$$\begin{array}{c} \\ c_1 \\ \vdots \\ c_n \end{array} \begin{array}{c} \begin{array}{ccc} c_1 & \cdots & c_n \end{array} \\ \begin{bmatrix} 1 & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ 1/b_{1n} & \cdots & 1 \end{bmatrix} \end{array}$$

comparing the n criteria in pairs should comply with the following conditions:

$b_{ij}>0$ and $b_{ij}=\frac{1}{b_{ji}}$ *for all i and j* which means that B should be a positive and a reciprocal matrix. Besides, the values that can be attributed to b_{ij} *for all i and j* must be on the scale of intensities defined by Saaty (Table 4).

Table 4: The fundamental Scale

Intensity of importance	Definition	Explanation
1	Equal importance	Two criteria contribute equally to the objective
3	Moderate importance	one criterion is slightly favored over another
5	Strong importance	one criterion is strongly favored over another
7	Very strong importance	one criterion is highly favored over another
9	Extreme importance	one criterion is extremely favored over another
2,4,6,8	Intermediate values	when compromise is needed
Reciprocals	if criterion i has one of the above values assigned to it when compared with criterion j, then j has the reciprocal value when compared with i.	

Deriving the weights $w_1, w_2, \ldots, w_n$ relative to the n criteria $c_1, c_2, \ldots, c_n$ can be done by Saaty's eigenvector technique when judgments are inconsistent. Saaty's eigenvector technique stipulate that weights can be computed as the principal right eigenvector of the matrix B:

$$BW = \lambda_{max} W \tag{2}$$

where λ_{max} is the maximum eigenvalue of the matrix B, or

$$w_i = \frac{\sum_{j=1}^{n} b_{ij} w_j}{\lambda_{max}} \quad for\ all\ i = 1, 2, \ldots, n \tag{3}$$

After generating the weights, the consistency of the matrix should be measured by C.R. as explained above but first consistency index C.I. should be computed by the following formula:

$$C.I = \frac{\lambda_{max} - n}{n - 1} \tag{4}$$

because C.R. is the ratio of C.I. to R.I. :

$$C.R = \frac{C.I}{R.I} \tag{5}$$

with:

R.I. is the random index and is defined for each size of the matrix which is illustrated in Table 5.

C.R. must be less than 0.1 to consider the result as acceptable, larger values require the decision-maker to revise judgments in order to reduce the inconsistencies.

Therefore, for the current study, application of the AHP gives firstly the hierarchy structure illustrated in Figure 5, secondly, the pairwise matrix (Table 6) was developed by comparing pairs of factors with the help of expert point of views and using the fundamental scale of Saaty (Table 4) to translate judgments into values. Comparisons were done with regard to the relative influence of each factor on aquifer recharge and on implementing the (AAR) technique in Tafilalet plain. R.I. used is 1.41 because the dimension of the pairwise matrix of the current study is 8×8 (Table 6) so, after computation, the obtained C.R. is equal to 0.041 which is less than 0.1 and according to the AHP method this result is satisfactory and the generated weights are acceptable.

Based on the generated weights, geology is the highest eigenvector while drainage density is the lowest (Table 6). In fact, geology, soil type, depth to the water table, and transmissivity are the factors that were given the highest weights. This was expected because, according to the analysis that was made during pairwise comparisons, the first thing to look out for in a project for artificial recharge of an aquifer suffering from overexploitation is the geological formations, which must be sufficiently permeable and transmissive, the soil, which must have a high infiltration capacity and the groundwater which must be deep enough to provide sufficient space for water storage to allow the water table to regenerate its level.

3.4. AAR index generation

After generating the relative weights for each factor by the AHP method, the reclassified thematic layers of the factors were multiplied by their weights and finally added up using the weighted sum function of ArcGIS software.

Thus, the potential sites for the artificial aquifer recharge map was produced by calculating the AAR index according to the following formula:

$$I_{AAR} = \sum_{i=1}^{n} F_i w_i \tag{6}$$

where:

-I_{AAR} is the AAR index

-n is the number of factors influencing (AAR) which is eight in the current study

-F_i stands for factor

-w_i stands for weight

Table 5: Random inconsistency index

n	1	2	3	4	5	6	7	8	9	10
R.I.	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

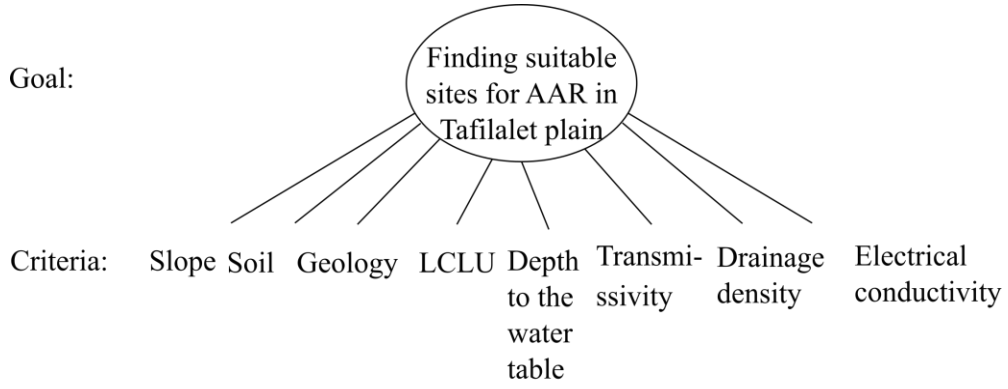


Figure 5: Hierarchy structure of the current study

Table 6: Pairwise matrix developed for the current study

Criteria	Slope	Soil	Geology	LCLU	Depth to the water table	Transmissivity	Electrical conductivity	Drainage density	Generated Weights
Slope	1	1/2	1/5	1/3	1/2	1/2	1	5	0.07
Soil	2	1	1/2	2	1	1	3	6	0.15
Geology	5	2	1	4	2	3	2	6	0.28
LCLU	3	1/2	1/4	1	1/2	1/2	2	5	0.10
Depth to the water table	2	1	1/2	2	1	1	3	7	0.16
Transmissivity	2	1	1/3	2	1	1	3	6	0.15
Electrical conductivity	1	1/3	1/2	1/2	1/3	1/3	1	5	0.07
Drainage density	1/5	1/6	1/6	1/5	1/7	1/6	1/5	1	0.02

4. Result

The map of potential sites for artificial aquifer recharge which is illustrated in Figure 6 was produced as explained above and then reclassified according to five categories which are: very poorly suitable, poorly suitable, moderately suitable, suitable, and very suitable. It indicates that the most suitable areas for the artificial groundwater recharge of Tafilalet aquifer are located notably in the northeast and in the west of the plain. Statistically, About 47 % of the study area is suitable for artificial aquifer recharge and about 40 % are moderately suitable. Thus, those results can be used to formulate an efficient groundwater management plan for sustainable utilization of these limited resources because identified areas will be prioritized in the implementation of artificial aquifer recharge projects.

Based on the produced map of potential sites for artificial recharge of Tafilalet aquifer (Figure 6), most suitable zones are underlain by high permeable geological formations that mostly belong to the late quaternary and covered with soils with high infiltration capacity most often constituted by alluvium and basic aeolian reliefs rich in meteoric materials and located in areas where the water table is very deep more than 20 m deep in addition the slope of the plain is generally not too steep and the above-mentioned description of the Tafilalet aquifer revealed that it is an alluvial and unconfined aquifer so, the artificial aquifer recharge technique most conducive with regard to the aforementioned conditions is infiltration basins nevertheless, the climatic characteristics of the region known by very high evapotranspiration rates will constitute a big challenge because more water will be lost by evaporation so less water will infiltrate unless the artificial aquifer recharge is conducted just during wet seasons when evapotranspiration rate decreases. Thus, several other factors must be taken into account when deciding whether a specific (AAR) system is suitable to be implemented, even if the appropriate locations are identified.

5. Discussion

The present-day situation of Tafilalet aquifer should urge water managers to work hard to avoid its worsening so, great efforts in risk mitigation need to be greatly accelerated and they should be centered on a range of scientific and engineering-based solutions. The current study presents one of those solutions which is the (AAR) and how to identify, in a scientific way, the most suitable locations for its implementation. In fact, A GIS-based multi-criteria analysis was performed, using eight thematic maps in order to produce an appropriate map for the identification of sites for (AAR) in Tafilalet plain. The resulting map shows that the most suitable areas are identified in the northeast and in the west of the region while inappropriate locations are limited to small portions in the northern, southern, and eastern parts of the study area. This result is extremely important because it will help taking appropriate decisions for mitigating the overexploitation problem of the Tafilalet aquifer that is why the scientific approach should be taken into consideration in such important projects because it gives nearly accurate results and it brings time savings.

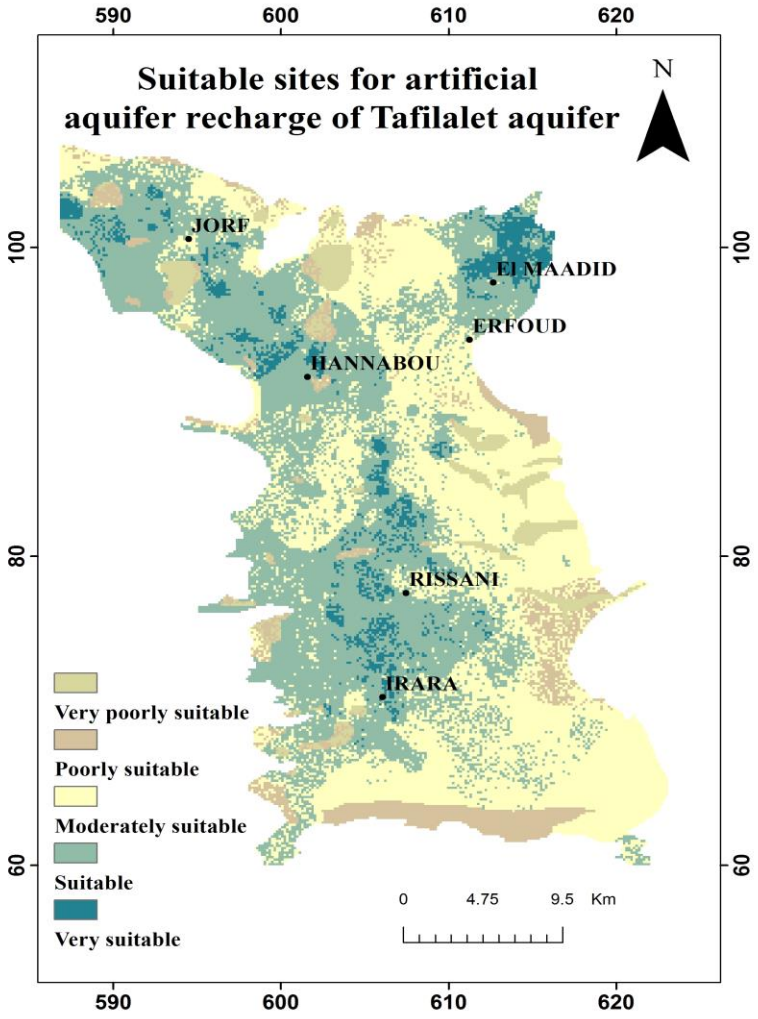


Figure 6: Potential artificial aquifer recharge sites map

The (AAR) will certainly sort out the problem of depletion of groundwaters in the Tafilalet aquifer as it was proved in some other aquifers in Morocco and in other countries. In fact, the (AAR) has already initiated and applied to some aquifers suffering from the same problem in Morocco such as the Charf El Akab aquifer in the region of Tangier, the Haouz Mejjate aquifer in the region of Marrakech, the Souss aquifer in the region of Agadir and the Berrechid aquifer located in the south of Casablanca, those aquifers have already started to reap its benefits. Even if the (AAR) practices in Morocco are still very much in their infancy, the technology is beginning to be recognized at the national level as having a value in tackling groundwater depletion issues.

Some attempts have been made in Tafilalet plain by the hydraulic basin agency of Guir-Ziz-Rheris in this regard and artificial recharging techniques have already begun to emerge on the plain, such as the construction of dikes or sills in the bed of the wadi Ziz in order to slow down the speed of the water and increase the wetted areas along the bed of the wadi, however, if the working field is extended to other locations that must be well chosen, these locations must be identified by means of scientific decision-making like GIS and multi-criteria analysis taking into consideration several criteria such as geological, pedological, hydrogeological, and many other factors to make sure that they are suitable for artificial recharge, in doing so, other artificial recharge techniques can be performed and can better and quickly restore the groundwater level situation of Tafilalet aquifer to its normal stage.

6. Conclusion

The aridity of the climate together with the excessive need for water is putting pressure on groundwater resources of Tafilalet plain so, the water level of the Tafilalet aquifer is declining to a threateningly low level which may undermine its productivity and causing its depletion, problems which are more likely to have irreversible adverse effects may occur like imperiling groundwater-dependent ecosystems, human activities as well as economic development of the region. Thus, the core objective of this study is to investigate suitable places for artificially recharge the Tafilalet aquifer to restore its levels to normal and fill the gap caused by overexploitation. For this purpose, thematic layers of factors that control groundwater recharge like depth to the water table, drainage density, electrical conductivity, slope, aquifer transmissivity, geology, soil, LCLU were prepared mapped, and reclassified in a GIS (ArcGIS software), after that, AHP method

was applied to compute weights of each factor and the map of appropriate sites of artificial groundwater recharge was generated by the application of weighted sum tool of ArcGIS software. The result has revealed that around 47% of the study area has been identified as a high potential zone for groundwater recharge which is a very important outcoming. However, identifying suitable places is essential but not sufficient for implementing (AAR) because as mentioned above even if infiltration basins are estimated as the appropriate (AAR) system to implement in Tafilalet plain, there are other factors to consider like climatic conditions of the region, need for periodic maintenance of the basin in order to sustain infiltration rates over the long term because organic matter build-up, silting and/or clogging can make the surface of the basin impervious and the capacity of infiltration will decrease over time in addition to the existence of a good quality water source nearby that will ensure the supply of water to be put into the basin and many other conditions that if not met will render the project impractical. Lastly, the present study has shown an example of the benefits of coupling multi-criteria analysis with GIS and remote sensing techniques in taking relevant decisions to better manage and protect groundwater resources.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors wish to record their deep sense of gratitude and profound thanks to all the employees of hydraulic basin agency of Guir-Ziz-Rheris for their warm welcome and their support to bring this work to fruition.

References

[1] I.A. Shiklomanov, "World water resources: a new appraisal and assessment for the 21st century; 1998," 40, 1998.

[2] Gleick Peter H, World fresh water resources, 1993.

[3] T. Oki, Shinjiro. Kanae, "Global Hydrological Cycles and World Water Resources," Science, **313**(5790), 1068–1072, 2006, doi:10.1126/science.1128845.

[4] S. Tushaar, Burke J, Villholth K, Angelica, M, Custodio, E, Daibes, F, Hoogesteger, J, Giordano, Mark, Girman, J, van der Gun, J, Kendy, E, Kijne, J, Llamas, R, Masiyandima, Mutsa, Margat, J, Marin, L, Peck, J, Rozelle, S, Sharma, Bharat R, Vincent, L, Wang, J, Groundwater: a global assessment of scale and significance, 395-423., 2007.

[5] S. Siebert, V. Henrich, K. Frenken, J. Burke, Update of the digital global map of irrigation areas to version 5., Institute of Crop Science and Resource Conservation, University of Bonn, Germany; Land and Water Division, Food and Agriculture Organization of the United Nations, Rome, Italy, 2013, doi:10.13140/2.1.2660.6728.

[6] T. Havril, Á. Tóth, J.W. Molson, A. Galsa, J. Mádl-Szőnyi, "Impacts of predicted climate change on groundwater flow systems: Can wetlands disappear due to recharge reduction?," Journal of Hydrology, **563**, 1169–1180, 2018, doi:10.1016/j.jhydrol.2017.09.020.

[7] E. Sener, A. Davraz, "Assessment of groundwater vulnerability based on a modified DRASTIC model, GIS and an analytic hierarchy process (AHP) method: the case of Egirdir Lake basin (Isparta, Turkey)," Hydrogeology Journal, **21**(3), 701–714, 2013, doi:10.1007/s10040-012-0947-y.

[8] M. Giordano, "Global Groundwater? Issues and Solutions," Annual Review of Environment and Resources, **34**(1), 153–178, 2009, doi:10.1146/annurev.environ.030308.100251.

[9] A.S. Qureshi, P.G. McCornick, A. Sarwar, B.R. Sharma, "Challenges and Prospects of Sustainable Groundwater Management in the Indus Basin, Pakistan," Water Resources Management, **24**(8), 1551–1569, 2010, doi:10.1007/s11269-009-9513-3.

[10] A. Jódar-Abellán, J.A. Albaladejo-García, D. Prats-Rico, "ARTIFICIAL GROUNDWATER RECHARGE. REVIEW OF THE CURRENT KNOWLEDGE OF THE TECHNIQUE," 12, 2017.

[11] S. Kaliraj, N. Chandrasekar, N.S. Magesh, "Identification of potential groundwater recharge zones in Vaigai upper basin, Tamil Nadu, using GIS-based analytical hierarchical process (AHP) technique," Arabian Journal of Geosciences, **7**(4), 1385–1401, 2014, doi:10.1007/s12517-013-0849-x.

[12] P.K. Naik, M. Mojica, F. Ahmed, S. Al-Mannai, "Storm water injection in Bahrain: pilot studies," Arabian Journal of Geosciences, **10**(20), 2017, doi:10.1007/s12517-017-3232-5.

[13] A. Rawluk, A. Curtis, E. Sharp, B.F.J. Kelly, A.J. Jakeman, A. Ross, M. Arshad, R. Brodie, C.A. Pollino, D. Sinclair, B. Croke, M.E. Qureshi, "Managed aquifer recharge in farming landscapes using large floods: an opportunity to improve outcomes for the Murray-Darling Basin?," Australasian Journal of Environmental Management, **20**(1), 34–48, 2013, doi:10.1080/14486563.2012.724785.

[14] L.F. Konikow, E. Kendy, "Groundwater depletion: A global problem," Hydrogeology Journal, **13**(1), 317–320, 2005, doi:10.1007/s10040-004-0411-8.

[15] M. Giordano, K.G. Villholth, eds., The agricultural groundwater revolution: opportunities and threats to development, CABI, Wallingford, 2007.

[16] D. Molden, al, eds., Water for food, water for life: a comprehensive assessment of water management in agriculture, Earthscan, 2007.

[17] M.I. Al-Hamdi, Competition for scarce groundwater in the Sana'a Plain, Yemen : a study on the incentive systems for urban and agricultural water use, CRC Press / Balkema, Netherlands, 2000.

[18] Yoshihide Wada, Ludovicus P. H. van Beek, Cheryl M. van Kempen, Josef W. T. M. Reckman, Slavek Vasak, Marc F. P. Bierkens, "Global depletion of groundwater resources," GEOPHYSICAL RESEARCH LETTERS, 37, 2010.

[19] F. Pedrero, A. Albuquerque, H. Marecos do Monte, V. Cavaleiro, J.J. Alarcón, "Application of GIS-based multi-criteria analysis for site selection of aquifer recharge with reclaimed water," Resources, Conservation and Recycling, **56**(1), 105–116, 2011, doi:10.1016/j.resconrec.2011.08.003.

[20] A.S. Qureshi, P.G. McCornick, M. Qadir, Z. Aslam, "Managing salinity and waterlogging in the Indus Basin of Pakistan," Agricultural Water Management, **95**(1), 1–10, 2008, doi:10.1016/j.agwat.2007.09.014.

[21] N. Iqbal, F. Hossain, H. Lee, G. Akhter, "Satellite Gravimetric Estimation of Groundwater Storage Variations Over Indus Basin in Pakistan," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, **9**(8), 3524–3534, 2016, doi:10.1109/JSTARS.2016.2574378.

[22] P. Pavelic, K. Srisuk, P. Saraphirom, S. Nadee, K. Pholkern, S. Chusanathas, S. Munyou, T. Tangsutthinon, T. Intarasut, V. Smakhtin, "Balancing-out floods and droughts: Opportunities to utilize floodwater harvesting and groundwater storage for agricultural development in Thailand," Journal of Hydrology, 55–64, 2012, doi:10.1016/j.jhydrol.2012.08.007.

[23] W. Martín-Rosales, J. Gisbert, A. Pulido-Bosch, A. Vallejos, A. Fernández-Cortés, "Estimating groundwater recharge induced by engineering systems in a semiarid area (southeastern Spain)," Environmental Geology, **52**(5), 985–995, 2007, doi:10.1007/s00254-006-0541-5.

[24] Y.B. Katpatal, A.M. Pophare, B.R. Lamsoge, "A groundwater flow model for overexploited basaltic aquifer and Bazada formation in India," Environmental Earth Sciences, **72**(11), 4413–4425, 2014, doi:10.1007/s12665-014-3342-2.

[25] A. Singh, S.N. Panda, K.S. Kumar, C.S. Sharma, "Artificial Groundwater Recharge Zones Mapping Using Remote Sensing and GIS: A Case Study in Indian Punjab," Environmental Management, **52**(1), 61–71, 2013, doi:10.1007/s00267-013-0101-1.

[26] Y.S. Prasad, B.V. Rao, "Groundwater depletion and groundwater balance studies of Kandivalasa River Sub Basin, Vizianagaram District, Andhra Pradesh, India," Groundwater for Sustainable Development, **6**, 71–78, 2018, doi:10.1016/j.gsd.2017.11.003.

[27] C.J. Nevill, "Managing Cumulative Impacts: Groundwater Reform in the Murray-Darling Basin, Australia," Water Resources Management, **23**(13), 2605–2631, 2009, doi:10.1007/s11269-009-9399-0.

[28] L.J. Frookh, L. Froukh, "Artificial Aquifer Recharge with Treated Wastewater in North Gaza," Riyadh,Saudi Arabia: 9, 2019.

[29] R. Spandre, Artificial Groundwater Recharge, EOLSS: 20, 2009.

[30] D. Das, Drainage and lineament analysis towards artificial recharge of groundwater, Springer Berlin Heidelberg, Berlin, Heidelberg: 37–44, 2011, doi:10.1007/978-3-642-24076-8_5.

[31] K.G. Villholth, E. Lopez-gunn, J. Van Der Gun, A. Garrido, K. Conti, Advances in Groundwater Governance, CRC Press, 2019.

[32] M.R. Llamas, P. Martínez-Santos, "Intensive Groundwater Use: Silent Revolution and Potential Source of Social Conflicts," Journal of Water Resources Planning and Management, **131**(5), 337–341, 2005, doi:10.1061/(ASCE)0733-9496(2005)131:5(337).

[33] S.B. Gingerich, C.I. Voss, A.G. Johnson, "Seawater-flooding events and impact on freshwater lenses of low-lying islands: Controlling factors, basic management and mitigation," Journal of Hydrology, **551**, 676–688, 2017, doi:10.1016/j.jhydrol.2017.03.001.
[34] W.-C. Hung, C. Hwang, J.-C. Liou, Y.-S. Lin, H.-L. Yang, "Modeling aquifer-system compaction and predicting land subsidence in central Taiwan," Engineering Geology, 147–148, 78–90, 2012, doi:10.1016/j.enggeo.2012.07.018.
[35] T.J. Burbey, "The Influence of Geologic Structures on Deformation due to Ground Water Withdrawal," Ground Water, **46**(2), 202–211, 2008, doi:10.1111/j.1745-6584.2007.00395.x.
[36] Q. Feng, G. Liu, L. Meng, E. Fu, H. Zhang, K. Zhang, "Land subsidence induced by groundwater extraction and building damage level assessment — a case study of Datun, China," Journal of China University of Mining and Technology, **18**(4), 556–560, 2008, doi:10.1016/S1006-1266(08)60293-X.
[37] M.J. Abdulrazzak, P.D. Oikonomou, N.S. Grigg, "Transboundary Groundwater Cooperation among Countries of the Arabian Peninsula," Journal of Water Resources Planning and Management, **146**(1), 05019023, 2020, doi:10.1061/(ASCE)WR.1943-5452.0001140.
[38] R.L. Nelson, "Assessing local planning to control groundwater depletion: California as a microcosm of global issues: LOCAL PLANNING TO CONTROL GROUNDWATER DEPLETION," Water Resources Research, **48**(1), 2012, doi:10.1029/2011WR010927.
[39] R.A. Freeze, J.A. Cherry, Groundwater, Prentice-Hall, Englewood Cliffs, N.J, 1979.
[40] E.A. Varouchakis, P.G. Theodoridou, G.P. Karatzas, "Decision-Making Tool for Groundwater Level Spatial Distribution and Risk Assessment Using Geostatistics in R," Journal of Hazardous, Toxic, and Radioactive Waste, **24**(1), 04019031, 2020, doi:10.1061/(ASCE)HZ.2153-5515.0000464.
[41] C.K. Singh, Y.B. Katpatal, "A GIS Based Design of Groundwater Level Monitoring Network Using Multi-Criteria Analysis and Geostatistical Method," Water Resources Management, **31**(13), 4149–4163, 2017, doi:10.1007/s11269-017-1737-z.
[42] W.B. Cutter, "Valuing Groundwater Recharge in an Urban Context," Land Economics, **83**(2), 234–252, 2007.
[43] P. Dillon, "Future management of aquifer recharge," Hydrogeology Journal, **13**(1), 313–316, 2005, doi:10.1007/s10040-004-0413-6.
[44] P. Dillon, P. Stuyfzand, T. Grischek, M. Lluria, R.D.G. Pyne, R.C. Jain, J. Bear, J. Schwarz, W. Wang, E. Fernandez, C. Stefan, M. Pettenati, J. van der Gun, C. Sprenger, G. Massmann, B.R. Scanlon, J. Xanke, P. Jokela, Y. Zheng, R. Rossetto, M. Shamrukh, P. Pavelic, E. Murray, A. Ross, J.P. Bonilla Valverde, A. Palma Nava, N. Ansems, K. Posavec, K. Ha, et al., "Sixty years of global progress in managed aquifer recharge," Hydrogeology Journal, **27**(1), 1–30, 2019, doi:10.1007/s10040-018-1841-z.
[45] S. Rahimi, M. Shadman Roodposhti, R. Ali Abbaspour, "Using combined AHP–genetic algorithm in artificial groundwater recharge site selection of Gareh Bygone Plain, Iran," Environmental Earth Sciences, **72**(6), 1979–1992, 2014, doi:10.1007/s12665-014-3109-9.
[46] A.K. Bhattacharya, "ARTIFICIAL GROUND WATER RECHARGE WITH A SPECIAL REFERENCE TO INDIA," 8, 2010.
[47] T. Asano, Artificial Recharge of Groundwater, Butterworth Publishers, Boston, 1985.
[48] A. Mahdavi, S.H. Tabatabaei, M. Nouri, R. Mahdavi, "Identification of groundwater artificial recharge sites using Fuzzy logic: A case study of Shahrekord plain, Iran," 2012.
[49] H. Bouwer, "Artificial recharge of groundwater: hydrogeology and engineering," Hydrogeology Journal, **10**(1), 121–142, 2002, doi:10.1007/s10040-001-0182-4.
[50] E. Widad, "Etude de la zone favorable pour la recharge artificielle de la nappe phréatique Mnasra," 55, 2014.
[51] H. Nasiri, A.D. Boloorani, H.A.F. Sabokbar, H.R. Jafari, M. Hamzeh, Y. Rafii, "Determining the most suitable areas for artificial groundwater recharge via an integrated PROMETHEE II-AHP method in GIS environment (case study: Garabaygan Basin, Iran)," Environmental Monitoring and Assessment, **185**(1), 707–718, 2013, doi:10.1007/s10661-012-2586-0.
[52] H.A. Faraji Sabokbar, H. Nasiri, M. Hamze, S. Talebi, Y. Rafiei, "Identification of suitable areas for artificial groundwater recharge using integrated ANP and pair wise comparison methods in GIS environment, (case study: Garbaygan Plain of Fasa)," Geography and Environmental Planning, **44**(4), 2012.
[53] J. Ghayoumian, B. Ghermezcheshme, S. Feiznia, A.A. Noroozi, "Integrating GIS and DSS for identification of suitable areas for artificial recharge, case study Meimeh Basin, Isfahan, Iran," Environmental Geology, **47**(4), 493–500, 2005, doi:10.1007/s00254-004-1169-y.
[54] I. Bouaamlat, A. Larabi, M. Faouzi, "Assessment of Water Resources in the Tafilalet Oasis System by a Mathematical Model (South East of Morocco)," 1, 2012.
[55] I.P. Senanayake, D.M.D.O.K. Dissanayake, B.B. Mayadunna, W.L. Weerasekera, "An approach to delineate groundwater recharge potential sites in Ambalantota, Sri Lanka using GIS techniques," Geoscience Frontiers, **7**(1), 115–124, 2016, doi:10.1016/j.gsf.2015.03.002.
[56] J. Ghayoumian, M. Mohseni Saravi, S. Feiznia, B. Nouri, A. Malekian, "Application of GIS techniques to determine areas most suitable for artificial groundwater recharge in a coastal aquifer in southern Iran," Journal of Asian Earth Sciences, **30**(2), 364–374, 2007, doi:10.1016/j.jseaes.2006.11.002.
[57] A. Chowdhury, M.K. Jha, V.M. Chowdary, "Delineation of groundwater recharge zones and identification of artificial recharge sites in West Medinipur district, West Bengal, using RS, GIS and MCDM techniques," Environmental Earth Sciences, **59**(6), 1209–1222, 2010, doi:10.1007/s12665-009-0110-9.
[58] C. william Carlston, Drainage density and streamflow, Professional Paper, 1963.
[59] Guide on artificial recharge to groundwater, central ground water board ministry of water resources, new delhi, 2000.
[60] Z. Şen, Basic Porous Medium Concepts, Elsevier: 43–97, 2015, doi:10.1016/B978-0-12-800075-5.00002-9.
[61] J. Bear, Hydraulics of groundwater, McGraw-Hill, Inc., New York, 1979.
[62] G.P. Kruseman, N.A. de Ridder, Analysis and evaluation of pumping test data, 2. ed. (compl. rev.), repr, International Institute for Land Reclamation and Improvement, Wageningen, 1994.
[63] W. Luo, D.T. Pederson, "Hydraulic conductivity of the High Plains Aquifer re-evaluated using surface drainage patterns: HYDRAULIC CONDCTIVITY RE-EVALUATED," Geophysical Research Letters, **39**(2), n/a-n/a, 2012, doi:10.1029/2011GL050200.
[64] J.J. De Vries, "The groundwater outcrop-erosion model; evolution of the stream network in The Netherlands," Journal of Hydrology, **29**(1–2), 43–50, 1976, doi:10.1016/0022-1694(76)90004-4.
[65] R.E. Horton, "EROSIONAL DEVELOPMENT OF STREAMS AND THEIR DRAINAGE BASINS; HYDROPHYSICAL APPROACH TO QUANTITATIVE MORPHOLOGY," Geological Society of America Bulletin, **56**(3), 275, 1945, doi:10.1130/0016-7606(1945)56[275:EDOSAT]2.0.CO;2.
[66] J. Krishnamurthy, A. Mani, V. Jayaraman, M. Manivel, "Groundwater resources development in hard rock terrain - an approach using remote sensing and GIS techniques," International Journal of Applied Earth Observation and Geoinformation, **2**(3–4), 204–215, 2000, doi:10.1016/S0303-2434(00)85015-1.
[67] B. Rambert, Recherche sur la signification hydrogéologique de la densité du drainage: Application à la détermination et à la cartographie de l'écoulemment souterrain., Bureau de recherche géologique et minières, France, 1973.
[68] T. Dunne, Chapter 1. Hydrology mechanics, and geomorphic implications of erosion by subsurface flow, Geological Society of America: 1–28, 1990, doi:10.1130/SPE252-p1.
[69] J.A. Sandoval, C.L. Tiburan, "Identification of potential artificial groundwater recharge sites in Mount Makiling Forest Reserve, Philippines using GIS and Analytical Hierarchy Process," Applied Geography, **105**, 73–85, 2019, doi:10.1016/j.apgeog.2019.01.010.
[70] B.L. Golden, E.A. Wasil, P.T. Harker, eds., The Analytic Hierarchy Process: Applications and Studies, Springer Berlin Heidelberg, Berlin, Heidelberg, 1989, doi:10.1007/978-3-642-50244-6.
[71] T.L. Saaty, "A scaling method for priorities in hierarchical structures," Journal of Mathematical Psychology, **15**(3), 234–281, 1977, doi:10.1016/0022-2496(77)90033-5.
[72] T.L. Saaty, "How to make a decision: The Analytic Hierarchy Process," European Journal of Operational Research, 48, 1990.
[73] R.W. Saaty, "The analytic hierarchy process—what it is and how it is used," Mathematical Modelling, **9**(3–5), 161–176, 1987, doi:10.1016/0270-0255(87)90473-8.
[74] T.L. Saaty, "Decision making with the analytic hierarchy process," International Journal of Services Sciences, **1**(1), 83, 2008, doi:10.1504/IJSSCI.2008.017590.
[75] E. Mu, M. Pereyra-Rojas, Understanding the Analytic Hierarchy Process, Springer International Publishing, Cham: 7–22, 2017, doi:10.1007/978-3-319-33861-3_2.
[76] T.L. Saaty, What is the Analytic Hierarchy Process?, Springer Berlin Heidelberg, Berlin, Heidelberg: 109–121, 1988, doi:10.1007/978-3-642-83555-1_5.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1132-1142 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Experimental Analysis of Thin Layer Drying of Ginger Rhizome in Convective Environment

Gbasouzor Austin Ikechukwu[*,1], Sabuj Mallik[2], Jude Ejikeme Ebem Njoku[3], Joshua Depiver[2]

[1]*Department of Mechanical Engineering, Chukwuemeka Odumegwu Ojukwu University, P. M. B. 02 Uli.Anambra State, 234, Nigeria*

[2]*Mechanical and Manufacturing Engineering and Built Environment, School of Engineering University of Derby/Britannia Markeaton Street Derby, DE22 3AW, UK*

[3]*Faculty of Engineering and Science, Department of Engineering and Science, University of Greenwich Medway Campus, Central Avenue Chatham Maritime Kent ME4 4TB, UK*

ARTICLE INFO

Article history:
Received: 17 July, 2020
Accepted: 03 November, 2020
Online: 14 December, 2020

Keywords:
Blanched
Moisture content
Peeled
Thermal conductivity
Unblanched
Unpeeled

ABSTRACT

This research presents the outcomes of thin layer convective drying of ginger rhizomes with different treatments using the ARS-0680 environmental chambers for the drying process. TD 1002A – Linear Heat Conduction Experimental equipment was used in measuring the thermal conductivity of the ginger samples at varied temperature levels ranging from 10°C – 60°C and drying time of 2 – 24 hours. The result of this experiment shows that the average moisture content for 2 hours drying at 60°C was 70.6% while at 24 hours drying was an average of 7.55% which is close to the range of 4 – 7%. The drying rate at drying times of 24 hours was 0.889/°C and 0.4437/°C for 2 hours drying, giving 50% by moisture reduction rate. While the lowest moisture content (5.98%) is obtained for unpeeled ginger and the highest (9.04%) at 24 hours drying at 60°C, the analysis shows the variation in the temperature varied as the ginger shrinks when subjected to heat.

1. Introduction

Thin layer drying can be utilized in removing volatile liquid from porous materials such as foodstuffs, ceramic products, wood, and so on. Porous materials have microscopic capillaries and pores which allow for the simultaneous transfer of both moisture and air when subjected to heating or cooling. The drying of moist porous solids involves concurrent exchange of heat and mass. Moisture is eliminated simply by evaporation into an unsaturated gaseous phase. Drying is fundamentally essential for the preservation of crops for prospect usage. Crops are usually conserved by removing sufficient moisture from them to evade decay and deterioration. Reducing the moisture content of ginger rhizome to a very low level will discourage the growth and multiplication of microorganism which could decompose the rhizome. Nevertheless, the drying process could cause the enzymes present in ginger rhizomes to be destroyed.

Thin-layer drying studies provide the basis for understanding the unique drying characteristics of any particular food material. The results of such studies have been extensively employed in the simulation of dryers under deep-bed drying conditions and for quantifying parameters for the design of specialized drying equipment. In thin-layer drying, the moisture content of a bio-material exposed to a stream of drying air of known relative humidity, velocity, and the temperature is observed over a while. Several mathematical models have been developed to simulate moisture movement and mass transfer during the drying of many agricultural products [1]. The porous material that will be used for this study is ginger. The physicothermal properties of ginger will be obtained and used in the resulting mathematical equations.

Ginger rhizomes are popular in most countries throughout the world [2]. Ginger rhizomes are edible spices of tropical origin, though it can thrive in humid regions. It is harvested in different seasons depending on the region. In Nigeria, the harvesting period is usually in July, while in areas like Hawaii and the USA harvesting season is in December. Gingers grows up to 125cm high under conducive environments and it is an annual crop [3, 4].

Figure 1: Raw materials for the experiments (Ginger Rhizomes) [1]

*Corresponding Author: Gbasouzor Austin Ikechukwu, Email: unconditionaldivineventure@yahoo.com

It is grown for its pungently aromatic underground stem or rhizome which is an important export crop valued for its powder, oil, and oleoresin, all of which have both food and medicinal value [5].

Ginger is an herbaceous perennial plant known as *Zingiberofficinale*, which belongs to the order *scitamineae* and the family *zingriberaceae*

Nigeria is presently the fifth top producer of ginger in the world and one of the principal exporters of ginger [6] The most important form in which ginger enters international trade is as a dried product; next in importance is as preserved product, and least important is as fresh product [7].

The quality of fresh ginger produced in Nigeria is the best in the world. However, it has been observed that the quality of its dried ginger has been declining due to the low level of mechanization of ginger production and processing [8]; with the attendant mold growth and loss of some important ginger qualities because of which Nigerian ginger attracts the cheapest price in the world market [9].

The aforementioned issues formed the basis for this work. Attention will be directed towards the use of a thin-layer drying process to determine the drying characteristics of thin layered ginger in a convective environment.

The drying of the porous material was conducted experimentally under free and forced convection environmental conditions. Ginger rhizomes used were peeled, unpeeled, split, and then cut into slices before drying at elevated temperatures in the environmental chamber [10]. The heat and mass transfers were studied using available correlations of boundary layer equations.

The economy of Nigeria had since the mid-1960s of oil boom deviated from agriculture to petroleum. This has placed undue pressure on the oil reserves in the volatile Niger Delta region and agriculture and its produce have been neglected. Therefore, any adverse influence on oil both locally and internationally affects the economy of Nigeria drastically. The present administration of President Muhammad Buhari has promised Nigerians better days, change in all sectors, and improved agricultural outputs. Nigeria needs to process its agriculture produce to derive the desired benefits available in the international markets. This research seeks to find solutions to the prevailing low quality of dried ginger in Nigeria.

Most researches done on ginger rhizomes were focused on the effects of cutting, solar-drying, and particle size of Ginger (Zingiber officinale Roscoe) pigment and powder [11]; configuration of volatile oil [9]; biochemical variations in ginger during storage [12]; and advances of ginger processing technologies [8, 13, 14, 15-17]. The goal of this work is to analyze experimentally the drying characteristics of thin layered ginger rhizome in a convective environment.

This study on the thin layer drying characteristics of ginger rhizomes produced in Nigeria will be experimental and analytical. It will not delve into production methods, harvesting techniques, and marketing strategies. For this study; four different treatments were carried out on the ginger. They are blanched, unblanched, peeled, and unpeeled treatments [18].

The ginger rhizomes obtained for the study required minimum duration of six to eight months of planting and will be dried to 7 – 15% moisture content. The ginger samples used in this study were obtained from one region. It is assumed that most ginger produced in Nigeria have similar quality and characteristics.

2. Materials and Method

The ginger samples utilized in this study were supplied from Kachia in Southern Kaduna in the Kaduna State of Nigeria and were preserved at room temperature before being used for the experimentations. The experiments were carried out at the Electronic Manufacturing Engineering Laboratory (ERMERG) Hawkes building, University of Greenwich, Maritime Kent, UK.

Variously treated ginger samples (peeled, blanched, unpeeled and unblanched) were used in the experiments. The variously treated ginger samples were cut into slices of 30mm diameter and 18mm thickness by scoopers designed for this purpose. ARS – 0680 Temperature and Humidity Chamber were used to dry the samples. The chamber houses the samples and temperature is controlled by an electric heater. The ginger drying experiment was conducted according to ASAE Standard S352.2. Before the commencement of the experiment, the whole apparatus was operated for at least 15-30 minutes to stabilize the humidity, air temperature, and velocity in the dryer.

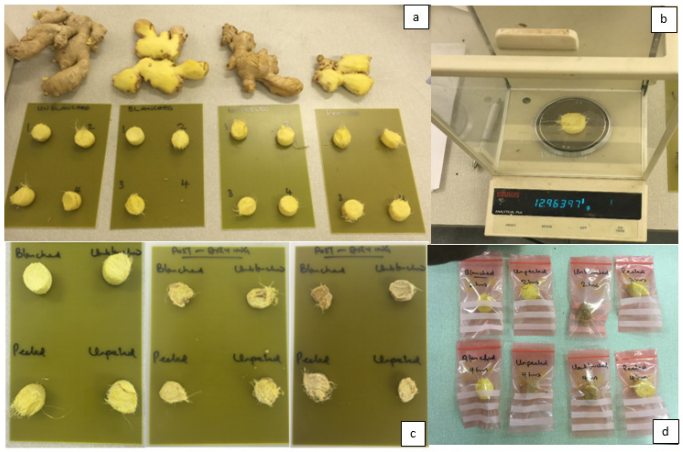


Figure 2: Schematic representation of ginger samples (a) Samples of fresh ginger (b) Sample in the electronic balances (c) Samples after drying; and (d) Bagged ginger samples after drying.

Drying started by 08:00 hours and lasted till the specimen attended the final moisture content. The weight loss of the sample in the environmental chamber was recorded during the drying period of 2 and 24 hours with an electronic balance (EK-200g, Max 200±0.01g) see figure 2c. At the end of drying, the thermal conductivity of the dried sample was measured with the aid of linear heat conduction equipment, TD1002A - Linear Heat Conduction Experiment Unit (LHTEU) [19].

3. Mathematical Modeling and Theoretical Principle

3.1. Determination of the most suitable model for drying

Thin-layer drying always requires a good understanding of the regression and correlation analysis. Linear and non-linear regression analysis are used to ascertain the relationship between variables MR and t in thin layer drying for selected drying models.

The recommended models chosen for applications were further validated using correlation analysis, standard error of estimate (SEE) and root means square error (RMSE) analysis respectively. The major indicator for selecting the best models is the determination coefficient (R^2). The highest determination coefficient and lowest standard error of estimate and RMSE values are used to determine the goodness of fit [20, 21, 22]. The determination coefficient (R^2); standard error of estimate (SEE) and root mean square error (RMSE) calculations can be performed using the following equations [23]:

$$R^2 = \frac{\sum_{i=1}^{N}(MR_i - MR_{pre,i})\sum_{i=1}^{N}(MR_i - MR_{exp,i})}{\sqrt{\left[\sum_{i=1}^{N}\left(MR_i - MR_{pre,,i}\right)^2\right] - \left[\sum_{i=1}^{N}\left(MR_i - MR_{exp,i}\right)^2\right]}} \quad (1)$$

$$SEE = \frac{\sum_{i=1}^{N}\left(MR_{exp,i} - MR_{pre,i}\right)^2}{d_f} \quad (2)$$

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}\left(MR_{pre,i} - MR_{exp,i}\right)^2\right]^{1/2} \quad (3)$$

where N is the number of observations, $MR_{pre,}\ I\ ith$ predicted moisture ratio values, $MR_{exp,i}\ ith$ experimental moisture ratio values, and d_f is the number of degree of freedom of the regression model.

3.2. Moisture Content (%) Calculation Formula

The moisture content of the materials can be calculated by using two methods: wet or dry basis

i. The wet basis is calculated as follows:

$$M_{wb} = \frac{w(i) - w(j)}{w(i)} \quad (4)$$

where

$M_{wb} = Moisture\ Content, wet\ basis\ (\%)$
$w(i) = mass\ of\ the\ sample\ before\ drying\ (g)$
$w(j) = mass\ of\ the\ sample\ after\ drying\ (g)$

The dry basis is calculated as follows:

$$M_{db} = \frac{w(t) - d}{d} \quad (5)$$

Moisture content, dry basis $\boldsymbol{M_{db}}$, is the amount of water per unit mass of dry solids (bone dry) existing in the sample

Where

$M_{db} = \text{Moisture Content, dry basis (\%)}$
$w(t) = \text{mass of wet materials at instant t (g)}$
$w = \text{mass of wet material (g)}$
$d = \text{mass of dry material (g)}$

Note that the two moisture contents are related by the following equation:

$$M_{db} = \frac{M_{wb}}{1 - M_{wb}} \quad (6)$$

3.3. Determination of the effective diffusivity and activation energy

The effective diffusivity of agricultural products can be determined using Fick's Second law for slab geometry [17, 18]. The common geometries were considered during the analytical solution of Fick's Second law for the infinite slab is expressed as:

$$MR = \frac{8}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}exp\left(-\frac{(2n+1)^2\pi^2 D_{eff}t}{4L^2}\right) \quad (7)$$

where n is a positive integer, L is the half-thickness of samples (m), D_{eff} is the effective moisture diffusivity (m^2/s), t is time (s), MR is the fractional moisture ratio.

Eq (7) can be modified in a logarithmic form as:

$$ln\ (MR) = \ln\left(\frac{8}{\pi^2}\right) - \left(\frac{\pi^2 D_{eff}t}{4L^2}\right) \quad (8)$$

The effective moisture diffusivity can be obtained by plotting $ln(MR)$against drying **time**; this gives a straight line with a slope (K) expressed as:

$$K = -\left(\frac{\pi^2 D_{eff}}{4L^2}\right) \quad (9)$$

The relationship between the effective diffusivity on temperature is described by the Arrhenius equation as [17-19]:

$$D_{eff} = D_0 exp\left(-\frac{E_a}{RT}\right) \quad (10)$$

Eq. (10) can be expressed in the logarithmic form as:

$$\ln D_{eff} = \left(-\frac{E_a}{RT}\right) + \ln D_0 \quad (11)$$

where D_0 is the pre-exponential factor of Arrhenius equation (m^2/s); E_a is the activation energy in (kJ/mol); R is the universal gas constant $8.314 J/molK$, and T is absolute air temperature (K).

From equation (11), plotting of $\boldsymbol{\ln D_{eff}}$ against $\boldsymbol{(T)^{-1}}$ would lead to the evaluation of activation energy for the diffusion of moisture during drying and $\boldsymbol{E_a}$ isobtainedas:$-(\boldsymbol{slope} \times R) = \boldsymbol{E_a}$, where ($-E_a/R$) is the slope of equation (18).

The average drying rate of the ginger in terms of initial and final moisture content could be evaluated using following equation [24].

$$\left(\frac{dM}{dt}\right)_{avg} = \frac{M_o - M_t}{t} \quad (12)$$

where $dm/dt\ is$ the average drying rate (kg/kg), t is the time (min), M_o is the initial moisture content ant and M_t is final moisture content [24].

The air velocity and temperature effect on the drying rate could be determined using appropriate statistical method. The overall drying rate of the ginger could be evaluated considering the total drying time as:

$$\left(\frac{dM}{dt}\right)_o = \frac{M_o - M_F}{t_1} \quad (13)$$

The evaluation of moisture content considering the dry basis could yield values more than 100% because the excessive moisture that could be present in some ginger rhizomes. The dry basis might be used to approximate the percentage moisture content as the moisture-free material. However, the wet basis moisture content is usually recommended for usage [25].

4. Results and Discussion

Tables 1 – 13 show the experimental results obtained during the whole experiment. This study considered two important features of thin-layer drying of ginger rhizomes:

- moisture content characteristics
- thermal conductivity of each sample at varying drying time and temperature with the aid of linear heat conduction's experimental unit in a convective chamber [26].

Table 1 shows the results obtained for moisture ratios and thermal conductivities of the variously treated ginger at a drying temperature of 10°C. The moisture ratios of the variously treated ginger samples decrease as the drying time increases. This implies that moisture is eliminated from the samples over time. Blanched treated samples recorded the least moisture ratio of 41.13. Also, the thermal conductivities for the various treatments decrease with the drying time. As expected, a decrease in moisture will not encourage faster moisture reduction.

Tables 2-6 present the results obtained for moisture ratios and thermal conductivities of the variously treated ginger at drying temperatures of 20°C, 30°C, 40°C, 50°C, and 60°C respectively. Both moisture ratios and thermal conductivities decrease with drying time. The least values were recorded at a drying time of 24 hours. The results of the variation of moisture ratios with drying time were plotted in figures 3 to 8 for various drying chamber temperatures,

The drying behavior of variously treated ginger rhizomes were studied using experimental drying curves. Drying curves for drying period of 2 –24hours at drying temperatures of 10°C to 60°C with step of 10 were considered. The results were employed to investigate the thermal conductivity content for the various treatments. The effect of moisture ratio on thermal conductivity with time was examined

4.1. Experimental Results at Temperatures 10°C

Table 1 shows the results obtained for moisture ratios and thermal conductivity of the variously treated ginger samples at a drying temperature of 10°C

Table 1: Moisture ratio (%) and thermal conductivity (W/m.K) at drying temperature of 10°C

Unblanched				**Peeled**			
Sample S/Nos.	Time (Hour)	Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$	Sample S/Nos.	Time (Hour)	Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$
1	2	88.64	0.4064	1	2	88.74	0.3768
2	4	84.75	0.3188	2	4	82.95	0.3004
3	8	78.28	0.2657	3	8	77.29	0.2623
4	10	73.67	0.2303	4	10	63.58	0.2115
5	14	65.22	0.1834	5	14	60.65	0.1919
6	16	51.1	0.1727	6	16	57.47	0.1658
7	24	49.55	0.1607	7	24	55.91	0.1449
Blanched				**Unpeeled**			
Sample S/Nos.	Time (Hour)	Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$	Sample S/Nos.	Time (Hour)	Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$
1	2	84.58	0.3290	1	2	91.08	0.3397
2	4	78.58	0.2878	2	4	83.91	0.3093
3	8	63.21	0.1993	3	8	82.07	0.2657
4	10	62.45	0.1901	4	10	73.41	0.2329
5	14	53.42	0.1699	5	14	68.77	0.2205
6	16	47.02	0.1558	6	16	64.68	0.2093
7	24	41.13	0.1400	7	24	62.22	0.1713

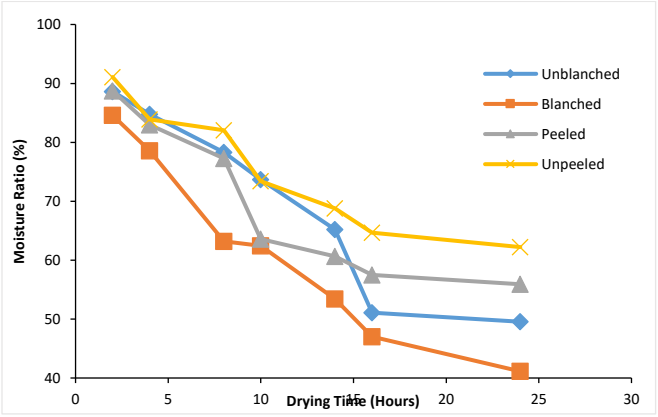


Figure 3: Variation of Moisture Ratio with Drying Time at 10^0C

Figure 3 shows the variation of moisture ratio with drying time at a temperature of 10°C for variously treated ginger samples. The unblanched ginger sample had a moisture ratio of about 88.64% before drying commenced, it lost moisture smoothly until 14hours where the moisture ratio reduced to 65.22%. It then drops abruptly to 51.1% within two hours and continued smoothly till 24hours. The blanched ginger sample had an initial moisture ratio of 84.58%; it lost moisture gradually till the end of the drying process with a final moisture ratio of 41.13%. It lost about 43.45% of moisture within twenty-four hours. Peeled and unpeeled treated samples lost 32.87% and 28.86% of moisture respectively within the twenty-four hours of drying. From figure 3, it could be deduced that the blanched ginger sample lost the highest amount of moisture with time while the unpeeled lost the least amount of moisture within the drying duration. Generally, the moisture ratio decreases with time for the variously treated ginger samples.

4.2. Experimental Results at Temperature of 20°C

Table 2 shows the results obtained for moisture ratios and thermal conductivity of the variously treated ginger samples at a drying temperature of 20°C

Table 2: Moisture ratio (%) and thermal conductivity (W/mK) at the drying temperature of 20°C

Unblanched				**Peeled**			
Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$	Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$
1	2	86.35	0.4064	1	2	87.85	0.3768
2	4	77.07	0.3188	2	4	77.18	0.3238
3	8	71.67	0.2382	3	8	72.83	0.2839
4	10	70.92	0.1974	4	10	66.39	0.2115
5	14	55.60	0.1901	5	14	50.00	0.1818
6	16	49.87	0.1658	6	16	47.71	0.1594
7	24	47.81	0.1491	7	24	37.49	0.1391
Blanched				**Unpeeled**			
Sample	Time	Initial Moisture	Thermal Conducti	Sample	Time	Initial Moisture	Thermal Conducti

S/Nos.	(Hour)	Ratio (%)	vity $\left(\frac{W}{m.K}\right)$	S/Nos.	(Hour)	Ratio (%)	vity $\left(\frac{W}{m.K}\right)$
1	2	86.29	0.2919	1	2	86.17	0.3454
2	4	77.81	0.2527	2	4	81.82	0.3343
3	8	67.75	0.2228	3	8	76.63	0.2839
4	10	65.19	0.1742	4	10	64.34	0.2329
5	14	43.29	0.1570	5	14	60.97	0.2205
6	16	38.58	0.1449	6	16	53.16	0.1802
7	24	34.26	0.1312	7	24	48.36	0.1713

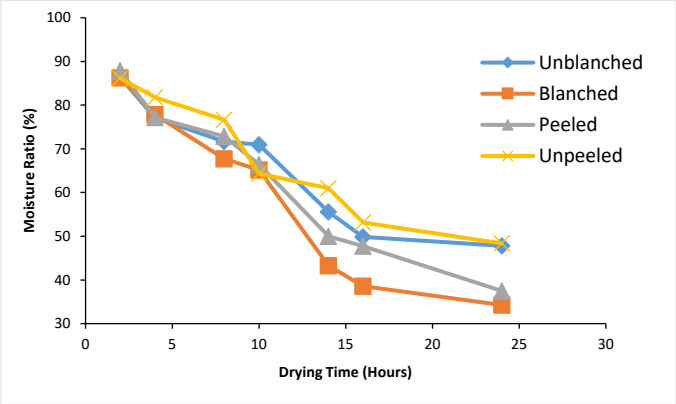


Figure 4: Variation of Moisture Ratio with Drying Time at 20^0C

Figure 4 presents the variation of the moisture ratio of the variously treated ginger sample with drying time at a temperature of 20°C. All the samples had almost the same initial moisture ratio of about 86%. The unblanched treated ginger sample lost about 38.54% of moisture, the blanched treated sample lost about 52.03%, the peeled treated sample lost about 50.36% and the unpeeled treated sample lost about 37.81%. As previously observed at drying temperature at 10°C, also at a drying temperature of 20°C blanched treated ginger sample lost the highest amount of moisture and the unpeeled treated ginger sample lost the least amount of moisture.

4.3. Experimental Results at Temperature of 30°C

Table 3 presented the results gotten for moisture ratios and thermal conductivity of variously treated ginger samples at a drying temperature of 30ºC

Table 3: Moisture ratio (%) and thermal conductivity (W/mK) at the drying temperature of 30°C

Unblanched				Peeled			
Sample S/No	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$	Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$
1	2	87.34	0.1074	1	2	87.95	0.1459
2	4	80.81	0.0996	2	4	80.12	0.1132
3	8	76.18	0.0987	3	8	74.83	0.0909
4	10	72.15	0.0955	4	10	66.33	0.0776
5	14	47.60	0.0809	5	14	45.73	0.0715
6	16	45.05	0.0785	6	16	38.35	0.0693
7	24	39.55	0.0677	7	24	27.76	0.0652
Blanched				Unpeeled			
Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$	Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$
1	2	86.65	0.1006	1	2	87.71	0.1126
2	4	78.52	0.0913	2	4	81.89	0.1021
3	8	65.23	0.0810	3	8	74.17	0.0810
4	10	62.35	0.0800	4	10	68.97	0.0740
5	14	30.50	0.0761	5	14	48.42	0.0658
6	16	24.49	0.0732	6	16	43.42	0.0630
7	24	17.48	0.0689	7	24	31.15	0.0611

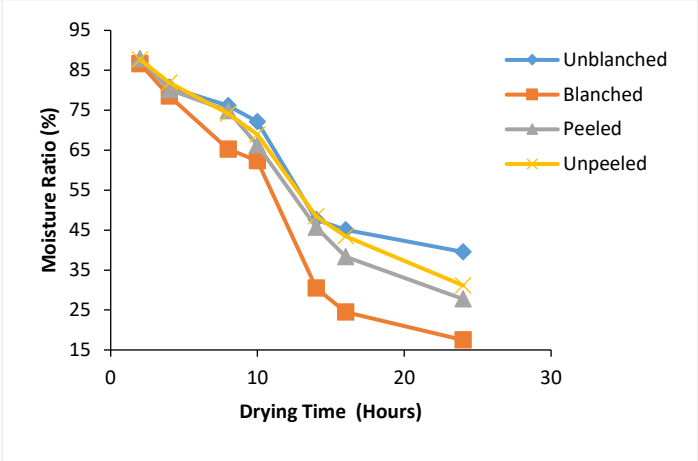


Figure 5: Variation of Moisture Ratio with Drying Time at 30^0C

Figure 5 presents the variation of the moisture ratio of the variously treated ginger sample with drying time at a temperature of 30°C. All the samples had almost the same initial moisture ratio of about 87%. The unblanched treated ginger sample lost about 47.79% of moisture, the blanched treated sample lost about 69.17%, the peeled treated sample lost about 60.16% and the unpeeled treated sample lost about 56.56%. The blanched treated ginger sample has the highest amount of loss while the unblanched treated ginger sample has the least amount of loss.

4.4. Experimental Results at Temperature of 40°C

Table 4 shows the results obtained for moisture ratios and thermal conductivity of the variously treated ginger samples at a drying temperature of 40ºC

Table 4: Moisture ratio (%) and thermal conductivity (W/m. K) at a drying temperature of 40°C

Unblanched				**Peeled**			
Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$	Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$
1	2	79.32	0.0756	1	2	75.93	0.0717
2	4	67.33	0.0691	2	4	60.47	0.0710
3	8	54.64	0.0660	3	8	49.39	0.0662
4	10	44.36	0.0638	4	10	47.08	0.0624
5	14	41.33	0.0608	5	14	30.59	0.0590
6	16	37.03	0.0581	6	16	27.01	0.0548
7	24	30.12	0.0557	7	24	23.92	0.0516
Blanched				**Unpeeled**			
Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$	Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$
1	2	70.11	0.0707	1	2	81.46	0.0717
2	4	54.37	0.0662	2	4	70.55	0.0658
3	8	40.6	0.0648	3	8	57.36	0.0611
4	10	27.84	0.0636	4	10	38.15	0.0572
5	14	23.65	0.0606	5	14	35.26	0.0560
6	16	18.83	0.0574	6	16	32.5	0.0557
7	24	17	0.0562	7	24	26.3	0.0543

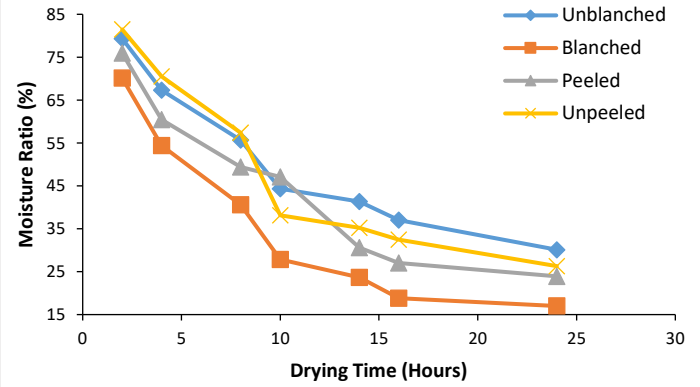


Figure 6: Variation of Moisture Ratio with Drying Time at 40^{0}C

Figure 6 presents the variation of the moisture ratio of the variously treated ginger sample with drying time at a temperature of 40°C. The unblanched treated ginger sample lost about 49.20% of moisture, the blanched treated sample lost about 53.11%, the peeled treated sample lost about 52.01% and the unpeeled treated sample lost about 55.16%. It can be observed that the unpeeled treated ginger sample experienced the highest moisture loss, but the blanched treated ginger sample attended the lowest moisture ratio of 17%.

4.5. Experimental Results at Temperature of 50℃

Table 5 shows the results obtained for moisture ratios and thermal conductivity of the variously treated ginger samples at a drying temperature of 50ºC

Table 5: Moisture ratio (%) and thermal conductivity (W/M. K) at a drying temperature of 50°C

Unblanched				**Peeled**			
Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$	Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$
1	2	71.65	0.0715	1	2	65.5	0.0759
2	4	58.38	0.0698	2	4	58.91	0.0695
3	8	42.51	0.0675	3	8	39.79	0.0634
4	10	37.55	0.0652	4	10	27.99	0.0571
5	14	28.44	0.0582	5	14	18.68	0.0555
6	16	25.53	0.0563	6	16	16.5	0.0543
7	24	17.95	0.0541	7	24	13.21	0.0519
Blanched				**Unpeeled**			
Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$	Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$
1	2	66.64	0.0730	1	2	67.85	0.0776
2	4	51.57	0.0650	2	4	57.92	0.0710
3	8	44.89	0.0626	3	8	41	0.0622
4	10	31.78	0.0610	4	10	39.92	0.0596
5	14	14.42	0.0584	5	14	32.2	0.0540
6	16	12.79	0.0581	6	16	23.71	0.0465
7	24	10.25	0.0556	7	24	15.49	0.0460

Figure 7 presents the variation of the moisture ratio of the variously treated ginger sample with drying time at a temperature of 50°C. The unblanched treated ginger sample lost about 53.70% of moisture, the blanched treated sample lost about 56.39%, the peeled treated sample lost about 52.29% and the unpeeled treated sample lost about 52.36%. It can be observed that the blanched treated ginger sample experienced the highest moisture loss and also attended the lowest moisture ratio of 10.25%.

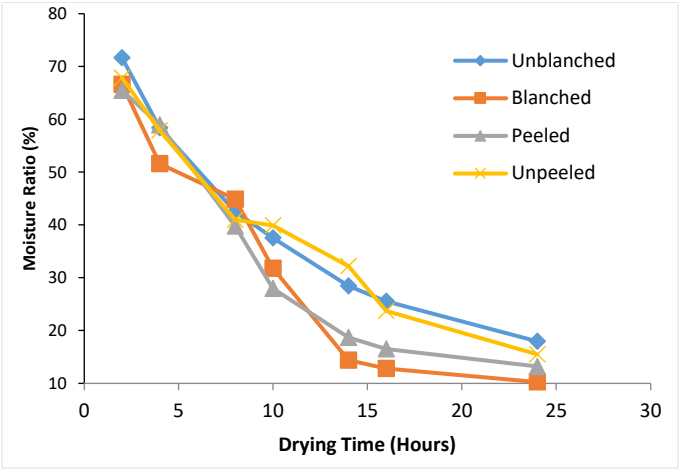


Figure 7: Variation of Moisture Ratio with Drying Time at 50^{0}C

4.6. Experimental Results at Temperature of 60℃

Table 6 shows the results obtained for moisture ratios and thermal conductivity of the variously treated ginger samples at a drying temperature of 60ºC

Table 6: Moisture ratio (%) and thermal conductivity (W/mK) at the drying temperature of 60°C

Unblanched				**Peeled**			
Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$	Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$
1	2	74.16	0.0762	1	2	70.75	0.0791
2	4	52.92	0.0720	2	4	46.68	0.0727
3	8	42.16	0.0695	3	8	29.89	0.0664
4	10	33.33	0.0691	4	10	24.17	0.0611
5	14	16.49	0.0652	5	14	13.82	0.0557
6	16	14.88	0.0644	6	16	11.54	0.0534
7	24	6.63	0.0553	7	24	8.56	0.0483
Blanched				**Unpeeled**			
Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$	Sample S/Nos.	Time (Hour)	Initial Moisture Ratio (%)	Thermal Conductivity $\left(\frac{W}{m.K}\right)$
1	2	63.11	0.0836	1	2	74.36	0.0776
2	4	47.27	0.0762	2	4	59.27	0.0689
3	8	26.49	0.0732	3	8	46.56	0.0622
4	10	17.71	0.0576	4	10	31.13	0.0596
5	14	14.15	0.0566	5	14	24.49	0.0540
6	16	10.32	0.0536	6	16	13.69	0.0465
7	24	9.04	0.0516	7	24	5.98	0.0460

Figure 8 presents the variation of the moisture ratio of the variously treated ginger sample with drying time at a temperature of 60°C. The unblanched treated ginger sample lost about 67.53% of moisture, the blanched treated sample lost about 54.07%, the peeled treated sample lost about 62.19% and the unpeeled treated sample lost about 68.38%. It could be observed that the unpeeled treated ginger sample experienced the highest moisture loss.

The respective ginger rhizome samples were dried at temperatures of 10°C to 60°C, at step of 10. Figures 3-8 present

the drying curve behavior of the variously treated ginger rhizomes at drying time of 2 - 24 hours and the effect of temperature on the different treatments. It could be observed that as the drying time increases, the moisture ratio decreases. Nevertheless, caution must be exercised to avoid excessive heating resulting from high temperature. Drying of ginger at excessive temperature could result in severe quality loss and shrinkages. Drying at above 60°C could result to quality loss and minor discoloration of the samples.

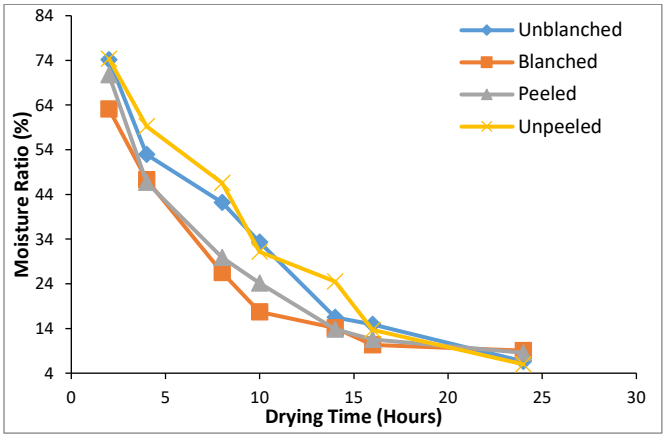


Figure 8: Variation of Moisture Ratio with Drying Time at 60^0C

The thermal conductivity decreases with the moisture ratio as the drying time progresses as shown in Tables 1-6. In Figures 3-8 the moisture ratio of the blanched ginger samples decreased the most when compared to other ginger samples while the unpeeled decreased the least. This could be as a result of the morphology of the ginger samples.

4.7. Drying Rate of Ginger Rhizome

Equation 12 shows that the average drying rate for ginger samples could deduct from Figures 3-8. The drying rate is the gradient of the graph of moisture ratio versus drying time. The values of the rate of drying for the various samples at different temperatures level were presented in Table 7; also, the averages for the various treatments of the samples were presented in Table 7. The negative value for the drying rate is a result of a loss of mass as the water content of the ginger escapes during drying. The blanched samples recorded the highest average drying rate indicating that it dries faster than the other samples, while the unblanched samples recorded the least average drying rate.

Table 7: Average Drying Rate for Variously Treated Ginger Samples

Samples	**dM/dt**						
	10^0C	20^0C	30^0C	40^0C	50^0C	60^0C	**Average**
Unblanched	-1.967	-1.839	-2.433	-2.177	-2.37	-2.98	**-2.2943**
Blanched	-2.026	-2.568	-3.495	-2.365	-2.683	-2.366	**-2.5838**
Peeled	-1.598	-2.341	-2.956	-2.359	-2.547	-2.614	**-2.4025**
Unpeeled	-1.351	-1.844	-2.768	-2.554	-2.339	-3.126	**-2.3303**

The sliced treated sample exhibited the highest drying rate, which might be attributed to diffusion behavior of the surface. The moisture travelled faster to the surface and evaporated than other treatments. Similar research showed that blanching increases the drying rate [27]. The various treatments exhibited different drying behaviors at lower drying temperatures; but as the drying temperature increased to 50°C, the difference becomes minimal. These variances could arise since blanching moderately exposed the sample to hot water and some cells might be slackened; which could cause the moisture diffusion to be higher and as consequently, resulting to higher the drying rate. The effect of this becomes conspicuous at drying temperature of 40°C - 50°C. Similar results were presented for red chili [28] and pear fruit [29]. External and internal factor have been reported to be responsible for controlling drying mechanisms in agricultural products. [30, 31]. The results indicated that the use of environmental chamber for drying purposes has minimized the drying time when compared with available reports on drying of peeled and unpeeled ginger which hitherto took 11 days to attain a moisture content of 17% using the open sun drying and 7.8% with the solar dryer [32]. Although samples in the solar dryer dry faster than those in the open-air sun, the convective drying methodology is a better time-saving measure. This observation agrees with some published reports [33-35].

4.8. Effect of Drying Time on Thermal Conductivity

Drying time is an important factor in the agro-based industrial process. Most agriculture products come in wet conditions and need to be dried to required standard moisture content at a given time interval. Variations in thermal conductivity of the variously treated ginger samples at various drying temperatures for drying time are shown in tables 8 to 13. Figures 9 to 14 present the effect of drying time on the thermal conductivities of the variously treated ginger samples.

Table 8: Variations in Thermal Conductivity of the Variously Treated Ginger Samples at Drying Temperature of 10°C

Time (Hour)	Thermal Conductivity (unblanched) $\left(\frac{W}{m.K}\right)$	Thermal Conductivity (blanched) $\left(\frac{W}{m.K}\right)$	Thermal Conductivity (unpeeled) $\left(\frac{W}{m.K}\right)$	Thermal Conductivity (peeled) $\left(\frac{W}{m.K}\right)$
2	0.4064	0.329	0.3397	0.3768
4	0.3188	0.2878	0.3093	0.3004
8	0.2657	0.1993	0.2657	0.2623
10	0.2303	0.1901	0.2329	0.2115
14	0.1834	0.1699	0.2205	0.1919
16	0.1727	0.1558	0.2093	0.1658
24	0.1607	0.14	0.1713	0.1449

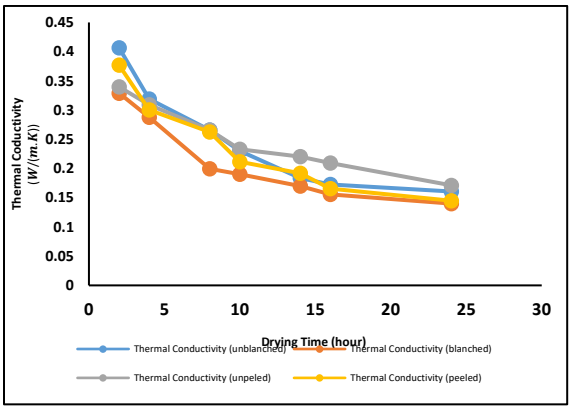


Figure 9: Variations of Thermal Conductivities of the Ginger samples at a drying temperature of 10^0C

Figure 9 presents the variation of thermal conductivity with drying time at a drying temperature of 10°C. The best fits to the data were logarithmic and polynomial of the second-order trend. The thermal conductivities for the variously treated samples decrease with time, which implies that as time progresses, less amount of moisture is lost. The thermal conductivity of the unblanched treated sample reduced from 0.4064W/mK to 0.1607W/mK within the twenty-four drying time. The thermal conductivity of the blanched treated sample reduced from 0.3397W/mK to 0.1713W/mK within the twenty-four drying time. The thermal conductivity of the unpeeled treated sample reduced from 0.329W/mK to 0.14W/mK within the twenty-four drying time. The thermal conductivity of the peeled treated sample reduced from 0.3768W/mK to 0.1449W/mK within the twenty-four drying time.

Table 9: Variations in Thermal Conductivity of the Ginger Samples at Drying Temperature of 20°C

Time (Hour)	Thermal Conductivity (unblanched) $\left(\frac{W}{m.K}\right)$	Thermal Conductivity (blanched) $\left(\frac{W}{m.K}\right)$	Thermal Conductivity (unpeeled) $\left(\frac{W}{m.K}\right)$	Thermal Conductivity (peeled) $\left(\frac{W}{m.K}\right)$
2	0.4064	0.2919	0.3454	0.3768
4	0.3188	0.2527	0.3343	0.3238
8	0.2382	0.2228	0.2839	0.2839
10	0.1974	0.1742	0.2329	0.2115
14	0.1901	0.157	0.2205	0.1818
16	0.1658	0.1449	0.1802	0.1594
24	0.1491	0.1312	0.1713	0.1391

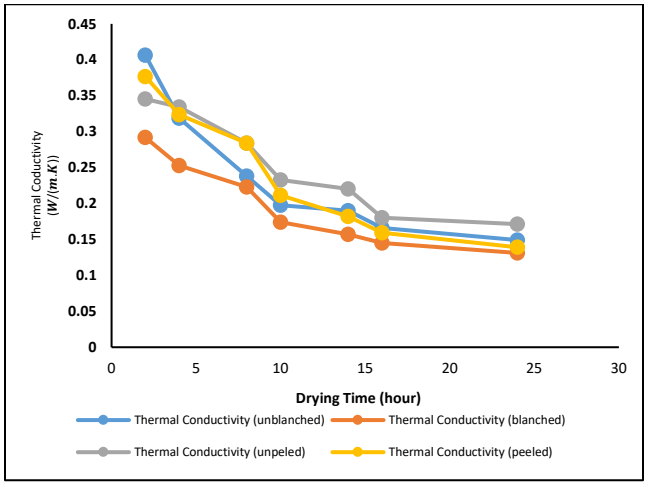


Figure 10: Variations of Thermal Conductivities of the Ginger samples at a drying temperature of 20^0C

Figure 10 presents the variation of thermal conductivity with drying time at a drying temperature of 20°C. Also, the best fits to the data were logarithmic and polynomial of a second-order trends. Also, the thermal conductivities for the variously treated samples decrease with time. It could be seen that as time increases, the thermal conductivity of the blanched treated sample is reduced as low as 0.1312W/mK.

The variation of thermal conductivity with drying time at a drying temperature of 30°C is shown in figure 11. Also, the best fits to the data were logarithmic and polynomial of a second-order trend. As expected the thermal conductivities for the variously treated samples decrease with time. It could be seen that as time increases, the thermal conductivity of the unpeeled treated sample is reduced as low as 0.0611W/mK.

Table 10: Variations in Thermal Conductivity of the Ginger Samples at Drying Temperature of 30^0C

Time (Hour)	Thermal Conductivity (unblanched)	Thermal Conductivity (blanched)	Thermal Conductivity (unpeeled)	Thermal Conductivity (peeled)
2	0.1074	0.1006	0.1126	0.1459
4	0.0996	0.0913	0.1021	0.1132
8	0.0987	0.081	0.081	0.0909
10	0.0955	0.08	0.074	0.0776
14	0.0809	0.0761	0.0658	0.0715
16	0.0785	0.0732	0.063	0.0693
24	0.0677	0.0689	0.0611	0.0652

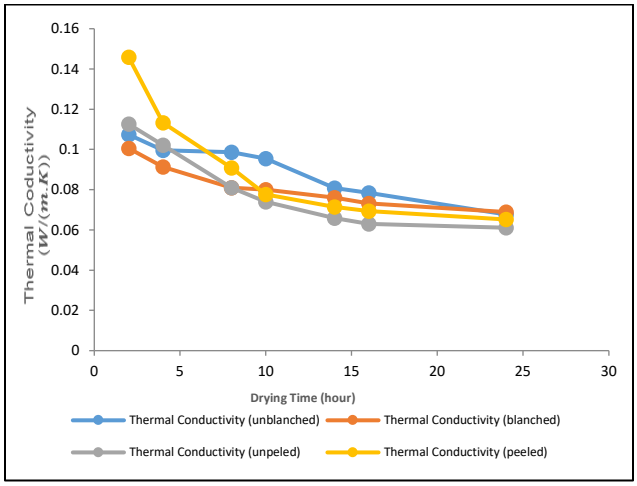


Figure 11: Variations of Thermal Conductivities of the Ginger samples at a drying temperature of 30^0C

Table 11: Variations in thermal conductivity of the ginger samples at drying temperature of 40^0C

Time (Hour)	Thermal Conductivity (unblanched)	Thermal Conductivity (blanched)	Thermal Conductivity (unpeeled)	Thermal Conductivity (peeled)
2	0.0756	0.0707	0.0717	0.0717
4	0.0691	0.0662	0.0658	0.071
8	0.066	0.0648	0.0611	0.0662
10	0.0638	0.0636	0.0572	0.0624
14	0.0608	0.0606	0.056	0.059
16	0.0581	0.0574	0.0557	0.0548
24	0.0557	0.0562	0.0543	0.0516

The variation of thermal conductivity with drying time at a drying temperature of 40^0C is shown in figure 12. The best fits to the data were logarithmic and polynomial of a second-order trend. As expected the thermal conductivities for the variously treated samples decrease with time. It shows that as time increases to twenty fours, the thermal conductivity of the peeled treated sample reduces to 0.0516W/mK.

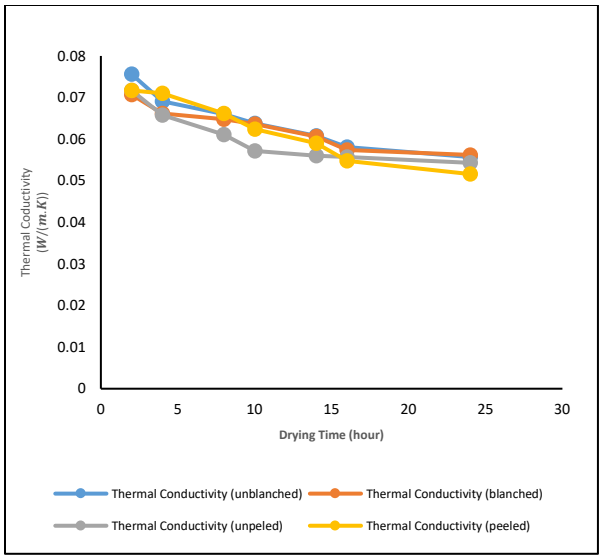


Figure 12: Variations of Thermal Conductivities of the Ginger samples at a drying temperature of 40^0C

Table 12: variations in thermal conductivity of the ginger samples at the drying temperature of 50^0C

Time (Hour)	Thermal Conductivity (unblanched)	Thermal Conductivity (blanched)	Thermal Conductivity (unpeeled)	Thermal Conductivity (peeled)
2	0.0715	0.073	0.0776	0.0759
4	0.0698	0.065	0.071	0.0695
8	0.0675	0.0626	0.0622	0.0634
10	0.0652	0.061	0.0596	0.0571
14	0.0582	0.0584	0.054	0.0555
16	0.0563	0.0581	0.0465	0.0543
24	0.0541	0.0556	0.046	0.0519

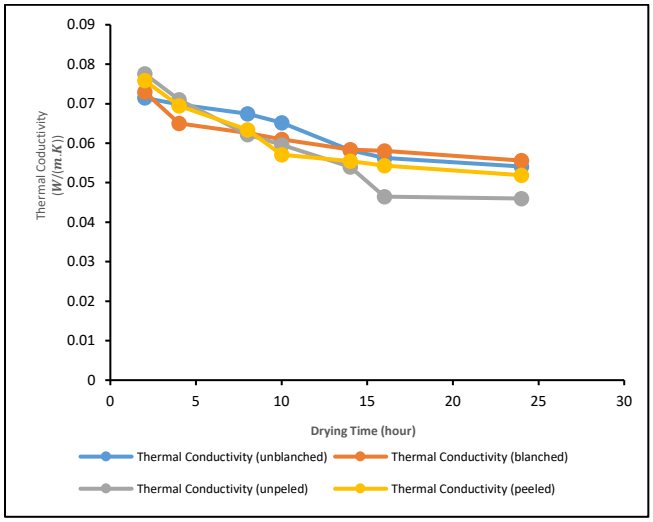


Figure 13: Variations of Thermal Conductivities of the Ginger samples at a drying temperature of 50^0C

Figure 13 present the variations of thermal conductivity with drying time at a drying temperature of 50°C. The best fit to the data was found to be logarithmic and polynomial of a second-order trend. The thermal conductivities for the variously treated samples decrease with time, while the unpeeled treated sample exhibited the least value of thermal conductivity.

Table 13: variations in thermal conductivity of the ginger samples at the drying temperature of 60^0C

Time (Hour)	Thermal Conductivity (unblanched)	Thermal Conductivity (blanched)	Thermal Conductivity (unpeeled)	Thermal Conductivity (peeled)
2	0.0762	0.0836	0.0776	0.0791
4	0.072	0.0762	0.0689	0.0727
8	0.0695	0.0732	0.0622	0.0664
10	0.0691	0.0576	0.0596	0.0611
14	0.0652	0.0566	0.054	0.0557
16	0.0644	0.0536	0.0465	0.0534
24	0.0553	0.0516	0.046	0.0483

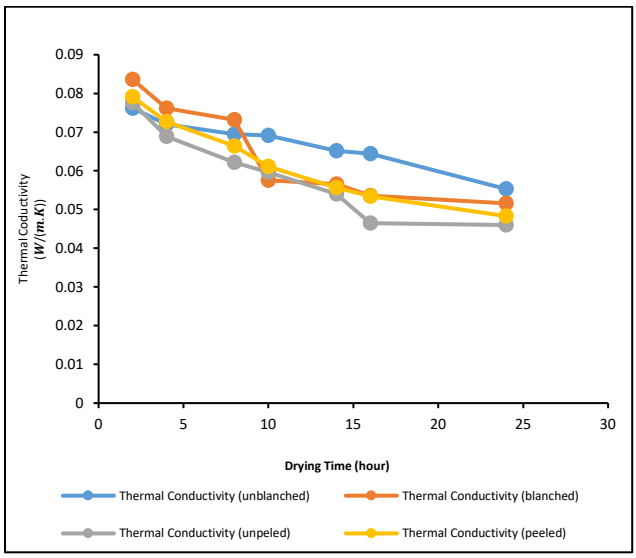


Figure 14: Variations of Thermal Conductivities of the Ginger samples at a drying temperature of 60^0C

Figure 14 presents the variation of thermal conductivity with drying time at a drying temperature of 60°C. The best fits to the data were found to be logarithmic and polynomial of a second-order trend. The thermal conductivities for the variously treated samples decrease with time. Also, the initial thermal conductivities for the various treatments decrease with temperature. The thermal conductivity value recorded was highest at drying time of 2 hours and then drops meaningfully until it reaches a drying time of 24 hours. This trend was prominent at drying temperatures of 10°C - 20°C and became insignificantly at drying temperatures of 30°C - 60°C for the variously treated samples.

The drying characteristics of Nigerian ginger rhizomes investigated shown that the drying route employed might be hastened only during the initial stages of the drying process where the movement of mass and air and are external factors that influence the rate of drying. The unpeeled and blanched treated ginger rhizomes are similarly influenced by the presence of moisture barriers like the unpeeled skins. As the studies reveal, drying of ginger rhizomes at low temperatures of 10°C-20°C does not have much significance on drying behavior as it maintains high initial moisture content and high thermal conductivity. However, as the temperature increases above 60°C ginger rhizomes becomes sensitive to temperature both in texture and color. The author opines that the drying of ginger rhizomes could be accomplished at temperatures between 50°C to 60°C to maintain the desired drying criteria.

5. Conclusion

The following conclusions were drawn from this study:

- It was gathered that ginger rhizomes might be dried at various temperatures. But ginger rhizomes dried at lower temperature might not reduce the effects of pest and bacterial infections, while drying at high temperatures say 60°C will considerably reduce the effects of pest and bacterial infecting connected with moist ginger rhizomes.
- The drying rate at drying time of 24 hours was 0.889/°C and 0.4437/°C for 2 hours drying time, giving about 50% reduction in moisture content. The interception which theoretically gives the initial moisture content of 0°C is lower at 24 hours drying (59.33%) compared to 95.12% on dry bases at 2 hours drying time, as expected. The average drying time for the variously treated ginger sample is about 2.4hours.
- The result of this study shows that the lowest moisture content of 5.98% was obtained for unpeeled ginger while the highest value of 9.04% was gotten for blanched sample for drying time of 24 hours and drying temperature of 60°C.
- The average moisture contents were 70.6% and 7.55% for 2 hours drying time at 60°C and 24hours drying time at 60°C respectively. The moisture content of 7.55% is close to the range of 4-7% desired for this research. This is an improvement to the result of 22.54% obtained at 50°C under blanched condition drying for 32 hours. [36, 37] as reported earlier the principal processing of ginger rhizomes involve sorting, washing, soaking, splitting, or peeling and drying to moisture content of 7-12%.
- The importance of drying ginger for a long time at temperatures close to 60°C has been shown in this work. At higher temperatures, ginger shrinkage and surface discoloration may occur. As can be seen, good results are attainable at a temperature of 60°C to maintain the quality of the products.
- The average effective moisture diffusivity and the average activation energy for the variously treated ginger rhizome samples are $5.49 \times 10^{-10} m^2/s$ and 22.63kJ/mol respectively.

References

[1] A. I. Gbasouzor, M. Sabuj, and J. E. Dara, "Determination of thin layer drying characteristics of ginger rhizome slices at varied temperatures," in October 2019 World Congress on Engineering and Computer Science, 468-476, 2019.

[2] E.K. Akpinar and S. Toraman " Determination of drying kinetics and convective heat transfer coefficients of ginger slices," Heat and Mass Transfer, **51**(10), 1-11, 2015, doi:10.1007/s00231-015-1729-6.

[3] M. Nishina, D. Sato, W. Nishijima, R. Mau, "Ginger root production in Hawaii", Hawaii Cooperative Extension Service, **6**(92), 3-6, 2013.

[4] A.I Gbasouzor, Joshua Depiver and J E.Njoku, "A Temperature distribution and boundary condition on heat transfer from discretized element of dried ginger rhizome using Matlab PDES for optimal preservation" World Congress on Engineering and Computer Science, 712-718, 2018, doi:10.1007/978-981-15-6848-0_8.

[5] NEPC, "Product Profile of Ginger" Nigerian Export Promotion Council b/k 312, Kumba St. Wuse, Zone-II, Abuja, 1- 6, 1999.

[6] FAO, The state of food and agriculture, Food and Agricultural Organization of the United Nations, 2008.

[7] Edwards David, Market for Dried, Preserved and Fresh Ginger, Natural Resource Institute, U. K., 1975.

[8] L. I. Onu, G. I.Okafor, "Effect of physical and chemical factor variations on the efficiency of mechanical slicing of Nigerian ginger (Zingiber Officinale Rose)", Journal of Food Engineering, **56**(1), 43–47, 2003, doi: 10.1016/S0260-8774(02)00146-2.

[9] O. Ekundayo, I. Laakso, R. Hiltumen, "Composition of ginger (Zingiber officinale roscoe) volatile oils from Nigeria", Flavour and Fragrance Journal, **3**(2), 85-90, 1988, doi.org/10.1002/ffj.2730030207.

[10] A. I. Gbasouzor, S. N. Omenyi "Convective Drying of Ginger Rhizomes" in Transactions in Engineering Technologies, Springer Nature Singapore Netherlands, 2019.

[11] G.I. Okafor and J.N. Okafor, "Effects of pricking, sun-drying and sieving on Ginger (Zingiber officinale Roscoe) colour and powder," Nigerian Food Journal, **25**(1), 155-160, 2007, doi: 10.4314/nifoj.v25i1.33664.

[12] E. Oti, P. Okwuowulu, Ohiori V. and G. Chijioke "Biochemical changes in ginger (Zinginger officinale roscoe) rhizomes stored under river sand and under dry grasses in pits in the humid tropics," Trorical Science, **28**(2), 87-94, 1988.

[13] C. G. Arocha and K. J. Simonyan, "Development of a motorized ginger rhizomes peeling machine," Nigerian Journal of Technology, **38**(3), 769 – 776, 2019, doi.org/10.4314/njt.v38i3.32.

[14] G. Nwandikom and B. Njoku, "Design-related physical properties of Nigerian ginger," Proceedings of First National Ginger Workshop, 101-107, 1988.

[15] G. Akomas and E. Oti, "Developing a technology for the processing of Nigerian ginger (Zingiber Officinale Roscoe)," Proceedings of the First National Ginger Workshop, 93-100, 1988

[16] L. Onu, "Design and fabrication of manual ginger slicing machine," Annual Research Report, Federal Institute of Industrial Research, Oshodi, Lagos, Nigeria, **34**(1), 1997.

[17] C.N. Egbuchua and E.C. Enujeke, "Growth and yield responses of ginger (Zingiber officinale) to three sources of organic manures in a typical rainforest zone, Nigeria," Journal of Horticulture and Forestry, **5**(7), 109-114, 2013, doi.10.5897/JHF2013.0302.

[18] A.I. Gbasouzor, Prediction of Thin Layer Drying Characteristics of Ginger Rhizome Slices in Convective Environment. Ph.D. Thesis Department of Mechanical Engineering, Nnamdi Azikiwe University, Awka, Nigeria, 2019.

[19] A.I. Gbasouzor, S.N. Omenyi and S. Mallik, "Simulation and discretized element of ginger on thin layer Drying using MATLAB PDEs for optimal programming," Journal of Engineering and Applied Sciences, **16**(1), 2020.

[20] E.K. Akpinar, "Mathematical modelling of thin layer drying process under open sun of some aromatic plants," Journal of Food Engineering, **77**(4), 864–870, 2006, doi.org/10.1016/j.jfoodeng.2005.08.014.

[21] Z. Erbay and F. Icier, "A review of thin layer drying of foods: theory, modeling, and experimental results," Critical reviews in food science and nutrition, **50**(5), 441–464, 2010, DOI: 10.1080/10408390802437063.

[22] L.R, Verma, R.A. Bucklin, J.B. Endan and F.T. Wratten, "Effects of Drying Air Parameters on Rice Drying Models,"Transactions of the American Society of Agricultural Engineers, **28**(1), 0296–0301, 1985, doi: 10.13031/2013.32245.

[23] A.I. Gbasouzor and S.N. Omenyi, "Modelling and Prediction of Thin Layer Drying Characteristics of Ginger Rhizome Slices in Convective Environment," Proceedings of the World Congress on Engineering and Computer Science, 477-486, 2019.

[24] R. Hasibuan and M. Bairuni, "Mathematical modeling of drying kinetics of ginger slices," AIP Conference Proceedings 1977, 020047, doi.org/10.1063/1.5042903.

[25] O.A. Aregbesola, B. Ogunsina, A.E. Sofolahan and N.N. Chime, "Mathematical modeling of thin layer drying characteristics of dika (Irvingia gabonensis) nuts and kernels," Nigerian Food Journal, **33**(1), 83–89, 2015, doi: 10.1016/j.nifoj.2015.04.012.

[26] M.M. Alam, M.Z. Hossain, M. Aktar, M.S. Islam and Z. Hasan, "Drying and Dehydration Kinetics of Ginger," International Journal of Innovation and Applied Studies, **9**(2), 956-967, 2014 .

[27] K.A. Bala, Drying and Storage of Cereal Grains, second edition, Wiley Blackwell, 2016.

[28] M.A. Hossain and B.K. Bala, "Drying of hot chilli using solar tunnel drier," Solar Energy, **81**(1), 85-92, 2007, doi.org/10.1016/j.solener.2006.06.008.

[29] S. Lahsasni, M. Kouhila, M. Mahrouz and J.T. Jaouhari, "Drying kinetics of prickly pear fruit (Opuntia ficus indica)," Journal of Food Engineering, **61**(2), 173-179, 2004, doi.org/10.1016/S0260-8774(03)00084-0.

[30] J.K.Gigler, W.K.P.van Loon, I.Seres, G.Meerdink and W.J.Coumans, "PH - Postharvest Technology: Drying Characteristics of Willow Chips and Stems," Journal of Agricultural Engineering Research, **77**(4), 391-400, 2000, doi.org/10.1006/jaer.2000.0590.

[31] O.V. Ekechukwu, "Review of solar-energy drying systems I: an overview of drying principles and theory," Energy Conversion and Management, **40**(6), 593-613, 1999, doi.org/10.1016/S0196-8904(98)00092-2.

[32] J. I. Eze and K.Agbo, "Comparative studies of sun and solar drying of peeled and unpeeled ginger," American Journal of Scientific and Industrial Research, **2**(2), 136-143, 2011, doi:10.5251/ajsir.2011.2.2.136.143.
[33] N.W. Desrosier and J.N. Desrosier, The Technology of Food Preservation, AVI Publishing Company, 1977.
[34] Y. Wang, M. Zhang and A.S. Mujumdar, "Convective Drying Kinetics and Physical Properties of Silver Carp (Hypophthalmichthys molitrix) Fillets," Journal of Aquatic Food Product Technology, **20**(4), 361-378, 2011, doi:10.1080/10498850.2011.575536.
[35] T. Koyuncu, Y.Pinar and F. Lüle, "Convective drying characteristics of azarole red (Crataegus monogyna Jacq.) and yellow (Crataegus aronia Bosc.) fruits," Journal of Food Engineering, **78**(4), 1471–1475, 2007, doi: 10.1016/j.jfoodeng.2005.09.036.
[36] R. Shalini, A. Ranjan and N. Kumar, "Studies on the Drying Characteristics of Apple pomace on tray drier," 16th International Drying Symposium (IDS 2008), 1636-1640, 2008.
[37] C.A. Biju, Drying characteristics of New Zealand chestnuts, Thesis, Master of Science MSc, University of Waikato, Hamilton, New Zealand, 2007.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1143-1152 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

A Novel Approach of Smart Logistics for the Health-Care Sector Using Genetic Algorithm

Marouane EL Midaoui*, Mohammed Qbadou, Khalifa Mansouri

Signals, Distributed Systems and Artificial Intelligence ENSET Mohammedia, University Hassan II of Casablanca, 20000, Morocco

ARTICLE INFO

Article history:
Received: 26 September, 2020
Accepted: 26 November, 2020
Online: 14 December, 2020

Keywords:
Health-Care Facilities Clustering
Genetic Algorithm
Vehicle Routing
Hospital Logistic
MDVRP

ABSTRACT

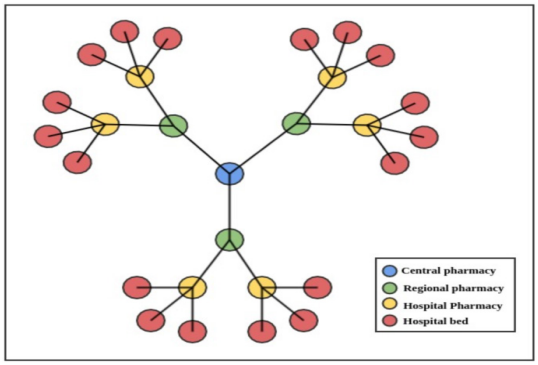
The heath-care sector has confronted significant difficulties in the past few years due to several issues such as insufficient human resources, budget cuts, and shortage of equipment and drugs. The logistics in the health-care sector take an extensive part of the budget, especially since it is the main axis to provide the hospital's pharmacies with goods (drugs and materials).The delay in the delivery of drugs, inappropriate transport conditions, and quality are the main challenges for the health-care supply chain.

These issues will affect the user and hospital satisfaction and product quality during transportation/storage. The Geo-Distributed storages play a considerable role to provide equipment and drugs faster, thus, routing and scheduling method and also clustering method for geographic data is needed for defining regions and areas to be served by distributed storage. Having multiple storages to severe multiple health-care facilities it can be presented as Multiple Depot Vehicle Routing Problem (MDVRP) which can be solved using metaheuristics, however, the MDVRP cannot take into account the nature of the context (health-care), priority, and capacity of the facility to serve which is in this context the hospital bed capacity (HBC).

This paper presents a new smart logistic method for the health-care sector to provide a better drug pulling system, by using the clustering method to provide a suitable location for new pharmacies and using genetic algorithm to perform routing and scheduling for Multi-Depot Vehicle Routing Problem (MDVRP) in the health-care context to feed the hospital's pharmacies with goods. This approach considers planned routes from the depot (central/regional pharmacy) to hospitals, as chromosomes, and groups the chromosome based on geographic distance and hospital bed capacity as a cluster. This method provides results with minimal computing power and resources and less calculation time.

1. Introduction

Logistics play a major role in sustaining hospitals and health-care facilities as motioned in [1], from operating rooms to patient bed and health-care units, they are all supplied by hospital pharmacies with drugs and equipment. As shown in figure 1, those pharmacies are supplied by health-care warehouses or in other words central pharmacies, in the same case, the goods can be sent to another warehouse (regional pharmacy) to provide a more accessible warehouse to the hospital. Each regional pharmacy is responsible for serving hospitals and facilities in a region; those Hospitals vary greatly geographically and demographically. This has a significant impact on how health-care logistics practices can be applied in each region. As a result, the drug takes time to be delivered and low quality of transport, which causes the patient to be unsatisfied, and that affects the service quality. Also, the quality can be affected by other factors such as the drug needs and bed capacity in each hospital (HBC).



*Corresponding Author: Marouane EL Midaoui, SSDIA, ENSET Mohammedia, University Hassan II OF Casablanca, Morocco, marouane.elmidaoui@gmail.com

Figure 1: Drug's supply chains in health-care

The regional pharmacy needs to store a specific number of drugs based on hospital needs. Therefore, a drug pulling system is needed. The idea to build regional pharmacies for each region is unrealistic, due to the limit of the construction budget, human resource, and infrastructural factor. That is why an alternative solution is building a regional pharmacy that can serve multi-region hospitals. In this case, the system needs to cluster health-care facilities based on each region's needs to determine the suitable set or regions that can be served by the selected pharmacy, with the proposition of the nearby position for pharmacy construction. However, in many cases such as pandemic disease period and urgent situation, the region pharmacy has to serve multiple hospitals at once or even serve other hospitals in other nearby region's hospitals, which can be presented as routing and scheduling issues of the drugs to hospitals from regional pharmacies. In this case, an intelligent transportation system (ITS) approach needs to be applied to provide routing and scheduling solutions.

Nowadays, researchers use clustering methods besides Artificial Intelligence to solve similar issues related to routing and scheduling problems. The genetic algorithm is an optimization technique based on a parallel search mechanism which makes it more efficient than other classical optimization techniques. It reflects the process of natural selection, using the genetic operation as crossover and mutation to escape the local optimum by selecting the fittest individuals to reproduce a new population.

This paper will discuss at first section Multiple Depot Vehicle Routing Problem and the mathematical formulation, in the second section will present the approach of clustering region by geographic distance and hospital bed capacity, with a comparison between three well-known clustering methods. The third section is dedicated to the routing and scheduling method using genetic algorithm for drug delivery between Regional pharmacies and hospitals based on drug demand

2. Literature Review

Multi-Depot Vehicle Routing Problem is considered one of the well-known problems in the research field currently for solving the delivery of goods in multi-depot instead of one depot as VRP, the goal is to optimize cost and exchange between multiple sources and destinations. As proposed in [2], the author presents an approach using a model MDGVRP (MD green VRP) based on an ant colony with a solution to route the optimal path while considering the capacity of the vehicle and fuel capacity.

An ordinary algorithm can resolve vehicle routing problems. However, Multi depot vehicle routing problems cannot be resolved without consuming high computing power and resources, and the precision is still questionable. To get better precision, the genetic algorithm (GA) can be used in this case, because of the stochastic properties and efficiency of GA, also MDVRP can be resolved by using those techniques. In [3] author measure the MDVRP efficiency with and without GA, the result can be concluded to good impact on MDVRP while using the GA in term of performance, Also in [4] the author described an approach to resolve the MDVRP based on clustering algorithms, the authors proposed two algorithms, first one is the 'Minimum Distance', after sorting the distance from suppliers to depots, the supplier with the minimum distance is selected for the nearest depot without exceeding the depot capacity, and repeating the process until all supplier have been assigned. The second algorithm is 'Minimum Distance with Balancing Depot ', the first supplier located nearest to the depot will be chosen to be assigned, after. each unassigned depot will be assigned to the closest supplier sequentially provided that their supply does not exceed the depot's capacity. In [5] the author developed a new method based on machine learning with a combination of GA for solving problems related to VRP, this method learn to classify the coordinates into regions which drive to divide to routing calculations into regions, this method provides many benefits over previous research and traditional solutions used by companies by 25.68% of routing cost, and 8.1% over classical GA methods[4], also the ML side of this algorithm can predict of future customer regions which give new behavior to resolve the problem.

Clustering has been used to solve problems in diverse fields as bioinformatics, data mining, several methods, and clustering algorithms have been proposed to define hospital clusters. William Thomas has contributed several times for clustering and identifying hospitals with a methodology of clustering communities and hospitals simultaneously, which led to a geographically connected area of populations served by a common cluster of hospitals [1-6]. The State of Michigan has applied the Thomas methodology. In [7] the author explains an approach based on disease-specific hospital networks with considering similarity among disease symptoms, k-mean is among the most popular clustering algorithms [8], K-means is an unsupervised algorithm for non-hierarchical clustering, making it possible to group in K distinct clusters the observations of the data set. Thus, similar data will be in the same cluster. Moreover, an observation can only be found in one cluster at a time (exclusivity of belonging). The same observation, therefore, cannot belong to two different clusters. In [9] the author audited similarity in geographic location and patient utilization to create health-care facilities clus0ters by using two steps k-means clustering. In [10] the author uses cluster analysis to group health-care facilities based on the distribution of health-care spending across service categories with k-means as a clustering technique. On the other hand, by comparing centralized and decentralized pharmacies structure. In [11] author used the Weighted k-Means Algorithm to cluster Text, in [12] author showed that in the case of centralized pharmacies, the cost of transport is reduced, which makes the logistics chain less expensive. The transportation of goods — which, in this case, are drugs and equipment — are linked to the routing and scheduling, these two factors can give satisfaction for the consumer, in our cases, it is the hospitals and patients. Many studies have been proposed to solve the routing and scheduling problem, in 1995 [13] the author introduced the MDVRP. In [14] the author proposed two solutions for the MDVRP problem based on Genetic algorithms. Both authors in [15] and [16] tried to solve the problem with exact methods. However, the author of [17] demonstrated that both techniques are not effective. In [18] proposed an approach to solve for Priced MDVRP, which the vehicle visits customer only if it is profitable, they propose two methods: the heuristic method to solve the MDVRP and mixed-integer linear programming (MILP), and conclude that heuristics give better results. In [19] author

solved the MDVRP with separate delivery by proposing a heuristic based on integer programming.

3. Positioning the Contribution

As mentioned above, transportation is a main part of health-care. Costs invested in logistics can exceed 50% [20] of the hospital budget. In addition to the coasts, there are infrastructural issues. Those kinds of issues can cause slow delivery and affect the patient's satisfaction. The good's quality can also be affected by several aspects, such as the delivery condition especially if the central pharmacy or regional pharmacies are located in another region or distant region, caused by the lack of structure and construction of pharmacy, or pharmacies that are not well distributed or centralized in some specific region based on a political decision. Besides infrastructural issues, the organization is also involved, especially for routing and scheduling, to distribute drugs, regional pharmacy needs to prepare and send a vehicle with the quantity or equipment of specific drug without exceeding the vehicle capacity, but also there are some cases when other regional pharmacies can serve other region's hospitals. Especially if the vehicle capacity and the distance (the hospital in the same pharmacy region are near to other region's hospitals) can afford the delivery or in urgent cases. This challenge can't be resolved by the traditional VRP (Vehicle Routing Problem), because in this case, we are facing a multi-depot routing (Multi-regional pharmacies).

Our contribution/model allows solving the above issues and challenges, by proposing locations for the construction of future regional pharmacies, based on the number of newest pharmacies, geographic distance, and hospital bed capacity in each cluster (a group of hospitals). Also, the model improves the routing and scheduling for drug delivery, while allowing regional pharmacies to serve other regional hospitals in specific cases without overlap between pharmacies for serving the targeted hospital. The proposed contribution has been applied to solve a recent challenge pointed by the Ministry of the health of Morocco. In this paper, we propose a solution to improve the creation of new centralized pharmacies and determine the methodology of good's (drugs and equipment) delivery to regions, by clustering regions who will be served by regional pharmacies as the first step based on Performance indicators shared by the same ministry [21]. The second step is to project the MDVRP problem to the logistic chain of health-care by making the hospital (or set of hospitals) as a customer and centralized pharmacies as a depot.

4. Vehicle Routing With Several Pharmacies

4.1. Background

The distribution of goods based on routing paths is one of the most important challenges of logistics currently. In general, the cost is associated with the distance. Costs invested in transportation exceed 50% of the total cost of a supply chain [20]. The objective then is to minimize the cost of combined routes on Multi-depot vehicle routing problems (MDVRP) by minimizing the distance traveled by a group of vehicles exposed to different requirements. However, MDVRP is a non-deterministic polynomial-time (NP-Hard) for deciding the courses of several vehicles of several depots to a group of customers simultaneously and return to the same deposit without surpassing the capacity limit of each vehicle.

The Vehicle Routing Problem (VRP) as illustrated in figure 2, is a class of operational search and combinatorial optimization problems. The aim is to determine the routes of a fleet of vehicles to deliver a list of customers or to carry out tours of interventions (maintenance, repairs, and checks) or visits (medical visits, commercial, etc.). The main idea is to minimize the cost of delivering goods. This problem is a classic extension of the traveling salesman problem. It is an employee who is responsible for distributing products (or services) on behalf of one or more employers.

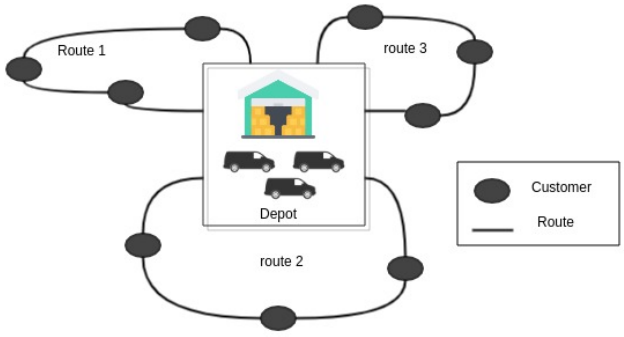


Figure 2: Illustration of Traditional VRP

MDVRP is a VRP with Multi depots from which the vehicle can serve its customers. If customers are grouped around repositories, then the distribution problem must be modeled as a set of independent VRPs. However, if clients and repositories are intertwined or for specific needs the delivery must be from another depot, the problem of multi-depot vehicle routing needs to be resolved. As shown in figure 3, the MDVRP processes several numbers of deposits. Each vehicle or set of vehicles is assigned to a single depot, which is usually both the origin and destination of the vehicle route

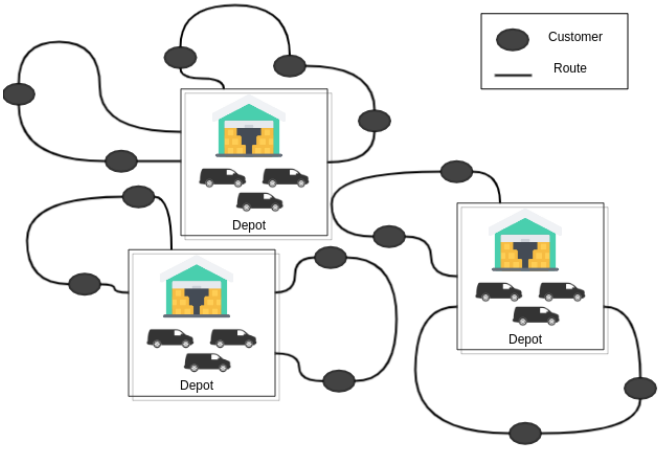


Figure 3: Illustration of MDVRP

We can group the approaches to solve the VRP into two doctrines: Exact algorithms and heuristics algorithms. Obviously, the MDVRP is more complex to solve than a simple VRP. Consequently, propose an optimal solution for an NP-hard

problem takes a long time, and maybe impossible with an exact algorithm. Both authors in [15] and [16] tried to solve the problem with exact methods. However, the author in [17] demonstrated that both techniques are not effective. In [18], the author proposes an approach to solve for Priced MDVRP, which the vehicle visits the customer only if it is profitable, they propose two methods: heuristic method to solve the MDVRP, and mixed-integer linear programming (MILP), and conclude that heuristics give better results.

It is unrealistic to use Linear Programming (LP) to solve large instances of the problem [22]. The intelligent algorithm is more widely applied, so the proposition is GA as a heuristic algorithm and one of the most promising hybrid intelligent methods for optimization problems. Distribution logistics encompasses all activities related to the supply of finished products and goods to customers. The central point of distribution logistics is the transportation of goods from the manufacturer to the consumer.

In our case, the manufacturers are a set of centralized pharmacies, and consumers are presented as hospitals' pharmacies. The idea of this contribution is to project the MDVRP model into the need for pharmacy logistics, representing centralized pharmacy as a single depot and hospitals as customers (or set of customers in cases of delegated hospitals to another hospital or health center). In some cases, one central pharmacy (CP) is not enough for some regions, also grouping/clustering phase is needed for our proposition, the grouping approach focuses on grouping hospitals and provides a suitable geographical place to consider building regional pharmacy while taking into account both distance and bed availability as a performance indicator in each region and hospitals which will be described in the next section.

4.2. Mathematical formulation

We can present our MDVRP mathematically as follows: $G = (V, E)$ is a graph where V is the set of vertices partitioned into two collections.

$V_c = \{V_1, V_2, \ldots, V_n\}$Represent the list of the hospitals.

$V_p = \{V_{n+1}, V_{n+2}, \ldots, v_{n+m}\}$ represent the collection of centralized pharmacies,

E set of edges connecting two points.

A Cost matrix $D = (D_{ij})$ is the length of arc (v_i, v_j) (corresponding to distance)

Expecting D symmetry, and the distance between sets of three points satisfies the triangle inequality.

Each centralized pharmacy $V_{n+m} \in V_d$ has the same model. k_m(The number of vehicles at the CP) are identical in term of vehicle (model, capacity Q, shape).

We consider the following:

Each vehicle starts from a centralized pharmacy (CP), serving a set of hospitals, at the end of the route or when Q is not enough to serve the next Hospital, the vehicle must return to the starting CP to refill.

Every hospital is visited only once by one vehicle, we consider in this example that a route can be done only by one vehicle.

The total demand for each route does not exceed the vehicle capacity Q

The vehicle cannot exceed the vehicle capacity Q, and the number provided by each CP is limited.

Centralized Pharmacies cannot serve each other's (CP cannot serve another CP).

The main objective for MDVRP, in this case, is to link hospitals into the proper CP with minimization of distance traveled by vehicle across the network.

n hospitals are grouped to form m clusters. Each cluster consists of $n_1, n_2, \ldots, n_m$ number of hospitals.

The number of vehicles based at a CP is k_m.

$k_j$is the group of hospital linked to a vehicle

Based on [23], MDVRP formulation for finding x which is the minimizes can be described as following:

Decision variables X and Y:

$$x_{ijkm} = \{\text{hospital j else } 0\}$$

$$y_{ikm} = \{\text{CP m else } 0\}$$

$$\text{MIN} = \sum_{p=1}^{m} \sum_{q=1}^{k_p} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} x_{ijqp} \quad (1)$$

Subject to:

$$\sum_{i=1}^{n} q_i y_{iqp} \leq Q \quad (2)$$

$$0 \leq n_{jq} \leq n_j \quad (3)$$

$$\sum_{q=1}^{k_j} n_{jq} = n_j \forall j = 1 \text{ to } m \quad (4)$$

$$\sum_{j=1}^{m} n_j = n \quad (5)$$

$$\sum_{p=1}^{m} \sum_{q=1}^{k_p} y_{iqp} = 1 \quad (6)$$

$$x_{ijqp} = 1 \vee 0 \quad (7)$$

$$y_{iqp} = 1 \vee 0 \quad (8)$$

The objective function according to Equation (1) minimizes the total delivery distance and cost of each vehicle within a CP. (2) limits capacity Q of the vehicle, (3) indicates the hospitals served by every vehicle must not overtake the number of the hospital a CP can deliver to. (4) shows the sum of the hospitals served by the whole route should be the sum of the hospitals served by CP m. (5) shows every hospital served by a CP. (6) guarantee that every hospital must be visited only once by vehicle. (7)(8) is the value bound of the decision variables.

5. Clustering Algorithms with Multiple Health-care Facilities

5.1. Clustering based on distance and HBC

Clustering is organizing objects (data) into homogeneous groups based on specific characteristics (data points) so that similar objects are kept in groups, while dissimilar objects belong to different groups or in other cases ignored.

Clustering algorithms are used in many fields such as image segmentation to group similar regions, insurance fraud detection, document clustering based on topics keywords, identifying crime-prone areas, customer segmentation for grouping customers based on behavior and needs, and public transport data analysis.

With 12 administrative divisions as regions, Morocco is the most north-western country in Africa. The construction of centralized pharmacies is for serving the hospitals pharmacies in defined regions, several approaches have been proposed to cluster hospitals [1, 6, 7, 9, 10]. We propose in our work clustering by Performance indicators of regions instead of hospitals. The data provided are grouped by regions.

The performance indicators in our case are Hospital Bed Capacity (HBC) by region. Table 1 shows the HBC in Morocco by region with latitude and longitude for each region. Table 1 presents the old territorial division based on 16 regions besides the HBC

Table 1: The HBC by Region in Morocco for 2015

REGION	ID	HBC	Geo-Position
Oued Ed-Dahab-Lagouira	A	53	(22.733, -14,286)
Laâyoune-Boujdour-Sakia El Hamra	B	313	(25.443 -13,174)
Guelmim-EsSemara	C	322	(28.708 -9,545)
Souss-Massa-Draa	D	2199	(31.120, -6,067
Marrakesh-Tensift-El Haouz	E	2347	(31.562, -7,959)
Meknes-Tafilalet	F	663	(31.905, -4,727)
Tadla-Azilal	G	964	(32.004, -6,578)
Doukkala-Abda	H	2872	(32.599, -8,660)
Chaouia-Ouardigha	I	1343	(33.047, -7,265)
Fes-Boulemane	J	1474	(33.187, -4,233)
Oriental	K	1739	(33.419, -2,145)
Grand Casablanca	L	826	(33.520, -7,568)
Rabat-Sale-Zemmour-Zaer	M	2884	(33.817, -6,237)
Taza-Al Hoceima-Taounate	N	1506	(34.258, -4,233)
Gharb-Chrarda-BeniHssen	O	757	(34.543, -5,898)
Tangier-Tetouan	P	813	(35.262, - 5,561)

The above table contains different data about regions and bed capacity in each classified by region. The order in this table is random, there is no order by any indicator, and each region has a specific geographic and demographic diversity.

Morocco is bordering the Mediterranean Sea and the Atlantic Ocean each region in a vertical shape as shown in figure 4.

Each dot in the above figure shows the location of a region presented in the previous table with their id to simplify the graphical presentation of the clustering operations.

5.2. Clustering algorithms to determine construction regional pharmacies area

One of the major challenges is to determine the most appropriate place to install the central pharmacy to serve a region or a set of regions and ensure the connectivity between these latter. The cluster appears to be the remedy to similar issues in a larger area. We aim to find the location of CP by using a range of clustering algorithms as k-means, DBSCAN, and Weighted K-means. All clustering algorithms are implemented using Python with numpy to calculate distance and matplotlib to display results as shown in map figures. The simulation of the clustering algorithms is done in an Alienware 32Gb RAM and i7 CPU (the execution of the scripts was in CPU level and not the GPU).

Figure 4: Old Territorial Division Based On 16 Regions of Morocco.

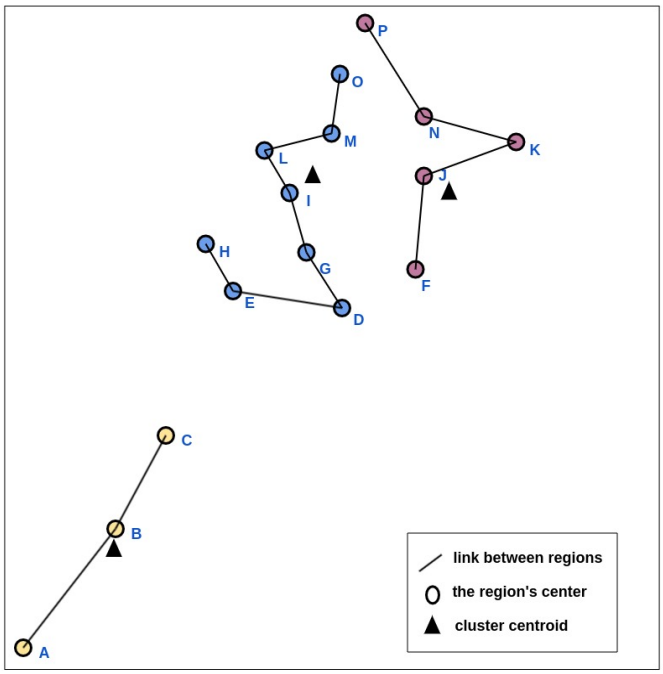


Figure 5: Geographic result as clustering with k = 3 using k-means.

5.2.1. Clustering with K-means

K-means is an unsupervised algorithm for non-hierarchical clustering, making it possible to group in K distinct clusters the observations of the data set. Thus, similar data will be in the same cluster. Moreover, an observation can only be found in one cluster at a time (exclusivity of belonging). The same observation, therefore, cannot belong to two different clusters.

We present each center of the region characterized by latitude and longitude as points in the cluster. Figures 5 and 6 present the result of clustering with k-means with 3 and 4 clusters in map shape.

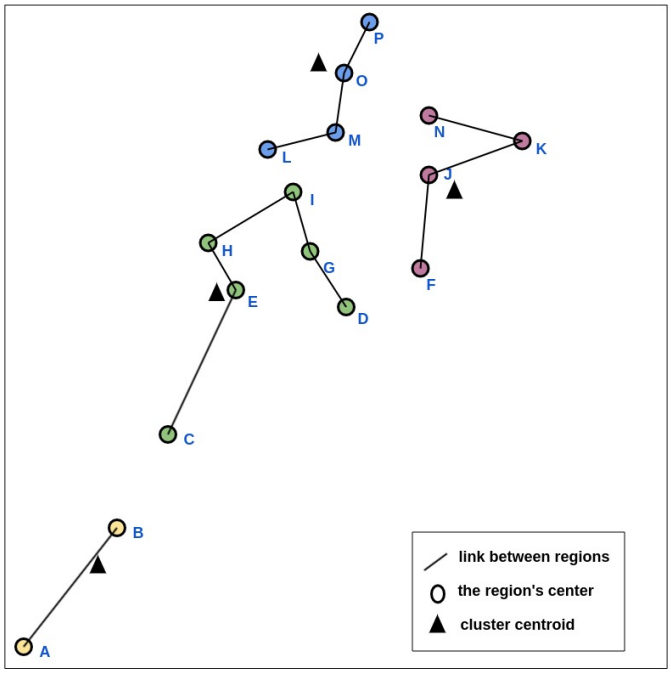


Figure 6: Geographic result as clustering with k = 4 using k-means.

The results of using k-means for 3 clusters show a favoring result for cluster 1 with a total of 15192 HBC and less than half for a cluster with 6195 HBC, for cluster 3 the HBC more minimal than the other clusters. Therefore, this method doesn't present equality or equilibrium between clusters, it shows a considerable imbalance between all three clusters. We can also notice the same disproportion for clustering with 4 clusters with different variations of total in every 4 clusters

5.2.2. Clustering using DBSCAN

DBSCAN is less popular than the k-means algorithm, based on the concept of density: cluster is an area of space where the density of observations is important. At the output, the algorithm generates as many clusters as areas of the high-density space. Isolated points are considered noise. The DBSCAN algorithm uses 2 parameters: the distance and the minimum number of MinPts (minimum number of points required to form a dense region) points to be in a radius for these points to be considered as clusters. The input parameters are therefore an estimate of the density of points of the clusters. The basic idea of the algorithm is then, for a given point, to retrieve its radius-neighborhood and check that it contains MinPts points or more. This point is then considered part of a cluster. We then go through the radius-neighborhood closer to find all the points of the cluster. Using the same data presented for k-means (without giving the number of clusters), the result was as following (Figure.7):

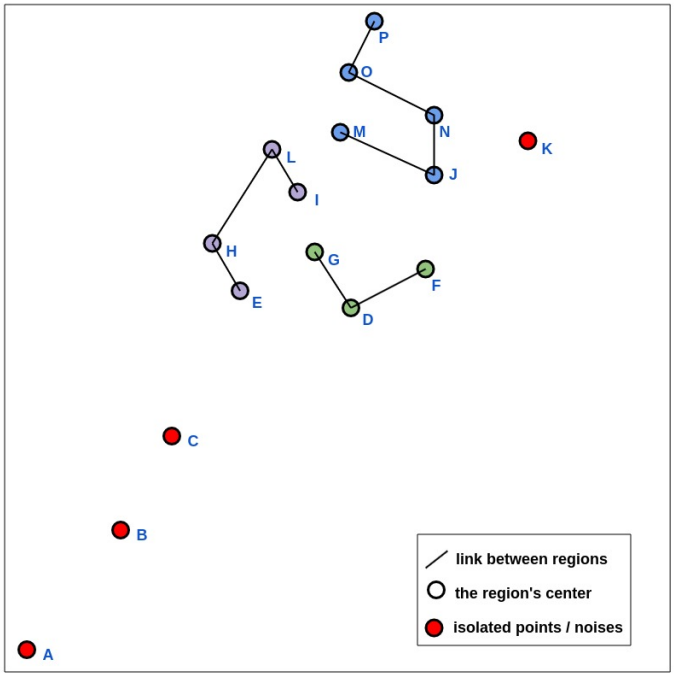


Figure 7: The result of clustering by using DBSCAN.

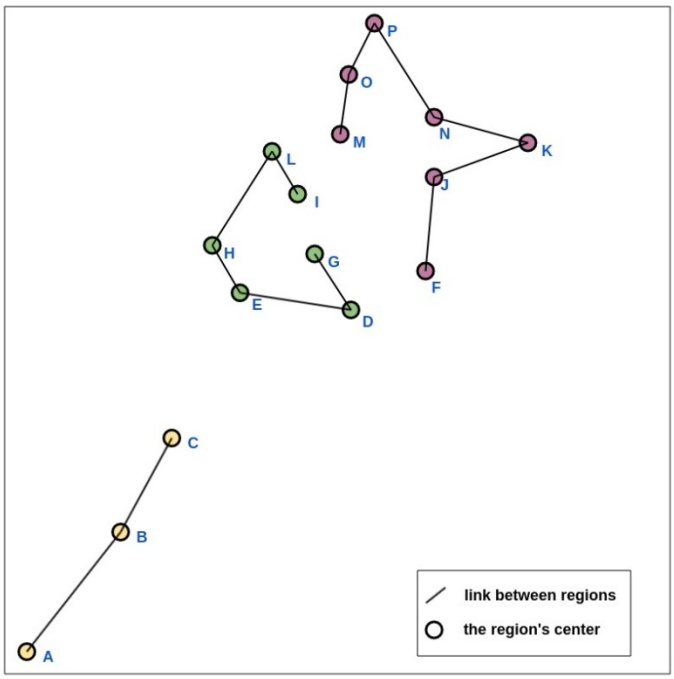


Figure 8: The result of weighted k-means for 3 clusters.

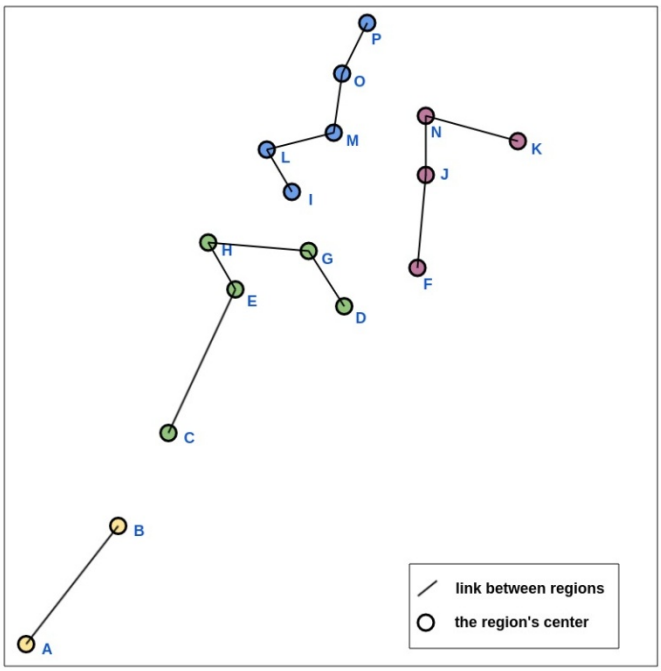


Figure 9: The result of weighted k-means fork = 4.

The result provided by this approach introduces the creation of 3 clusters where the density is outstanding. However, 4 regions are designed to be noise instead of points linked to a cluster, these are points that do not belong to any group/cluster. Noises or points that do not fit into any cluster are also significant at least in our case they present regions that need to be served, as they can represent a problem or rare event in data clustering, which we need to consider because it represents HBC in an isolated area, especially in epidemic cases, also this method cannot handle the number of clusters needed, the nature of DBSCAN consists on given number of clusters as a result which is not suitable in our case for decision making.

5.2.3. Clustering with weighted k-means

K-means algorithm groups data by assigning all points to the nearest clusters, and determining the cluster means. The algorithm repeats previous steps until points assemble. Weighted k-means automatically calculates the feature weights in a cluster, adopting the idea that features can have different degrees of relevance from other features. Figure 8 and 9 present the result of the current method for 3 and 4 clusters.

The result shows equilibrium between the two first clusters in the case of three clusters (first 11551, second 9836), and another equilibrium between the three first clusters in the case of four clusters (first 7623, second 5382, third 8704). Each cluster is unbiased in terms of HBC, south regions make an exception in this case because the points that represent the southern regions are far from the other points, which is tolerable since the integration of these regions with other regions will raise the costs of the transport.

Examination with k-means, DBSCAN, and weighted k-means shows progressively suitable results for weighted k-means. the method does not present an equilibrium between the clusters. However, clusters are unbiased, which the other clusters methods can't afford.

6. Genetic Algorithm

The Genetic Algorithm is a stochastic optimization technique based on a parallel search mechanism which makes it more efficient than other classical optimization techniques such as branch and bound, Tabu search method and simulated annealing. The main idea of GA is to avoid getting trapped in local optimum by genetic operators, crossover, and mutation. GA maintains a population of possible random solutions representing chromosomes that can evolve through a generation. The chromosomes are evaluated for fitness in each generation. The best-fitting chromosomes have a high probability to be selected for genetic operations such as crossover and mutation for the creation of the next generation.

GA is the most widely class of heuristics methods used as an application to the MDVRP problem, that why GA has received great attention from many authors, In [24] the author present a comparative study of different types of GA's applied to MDVRP, in [14] the author propose two solutions for the MDVRP. The first one is an implementation of the GA for solving the issues. The second one is a hybrid method with Clarke and Wright saving

method and nearest neighbor, and he proved that the last method performs better than the first one.

Initial population: or the first chromosome is formed based on the routes, CP and hospitals according to the clustering method we use in the first section (grouping, routing, and scheduling). Consider an MDVRP instance with 5 hospitals and 0 represent CP. If the path representation for this instance is (0 2 4 1 0 3 5 0), then two routes are required by the vehicles to serve all the hospitals. The first route: departs from the CP at 0 and travels to hospitals (2 4 1) as soon as the vehicle has returned to the CP. Second route: leaves the CP 0, serves hospitals 3, 5, and returns to the CP.

Evaluation: evaluate each chromosome by measuring the fitness value, the main function in MDVRP is to optimize the delivery time and decrease spending cost in n CP. The operation starts in every CP simultaneously, that's why it's normal that some vehicles in the network can achieve the objective (finish the delivery to the set of hospitals) while others can achieve the same objective later. In our context we consider fitness score calculated based on a total of cost route CR_{mk} of each vehicle (k) in every CP (m). The cost formula will be mathematically presented as the following:

$$\text{cost} = \sum^{m}\sum^{k} CR_{mk}$$

The cost for one vehicle of the i CP is calculated as the following:

$$CR_{i1} = D(d_i, c_1) + \sum D(c_j, d_i) + D(c_n, d_i) \text{ for all j } \in n_{jq}$$

Selection: for each generation, the best chromosome (based on fitness value) is selected for the mating pool and reproduction for the next generation, to generate new individuals in the population approach used in [14] use roulette wheel selection operation, while others use tournament selection [23]. The selection strategy is based on fitness evaluation, chromosomes will be selected with generated probability measured by the fitness function, the fittest individual we are selected for the next step.

Genetic operation: the mutation and crossover used in GA as genetic operation, both mutation and crossover affect the selected chromosomes with genetic operations rate.

Crossover: consists on combine genetic information from two genomes to a new generation. In [25] the author developed a technique for crossover based on BCRC - Best Cost Route Crossover - oriented to the vehicle routing problem with time windows (VRPTW). The classical order crossover is used by [14] to perform crossover into MDVRP. The aim is to combine the genetic information of exactly two parents to create new information

Mutation: consists of swapping two distinct customers i and j randomly selected, there are many propositions of mutation methods, and we can group them into two classes, heuristic mutation and inversion mutation. The effect is to flip a chromosome substring to create an offspring, mutation works on only one chromosome. The creation of the initial population will be projected as a chromosome with hospitals and regional pharmacies, hospitals presented as numbers while0_a, $0_b$are CP.

1	2	3	4	5	6	7	8	9	10	0_a	0_b

Grouping: We used the weighted-k means to cluster regions into the hospital's groups. Usually, in the same region, the geographical obstacle is not considerable. Also, the HBC for each hospital is not available. Therefore, in the grouping section, we use only Euclidean distance, however, if the space is large and can be affected by the shape of the globe, we can add haversine as distance instead of normal Euclidean one. The chromosome will be two sets of the group each group has its own pharmacy.

CP_A1, 2, 3, 4, 5. CP_B6, 7, 8, 9, 10.

The figure below (Figure 10) presents a flowchart to describe the initialization of parameters and the iteration on GA algorithm:

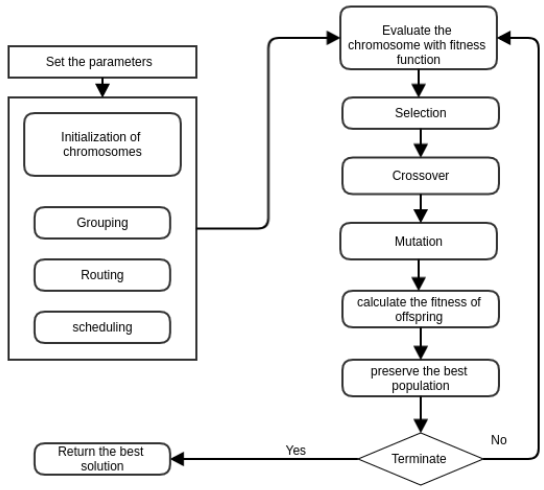


Figure 10: Flowchart of GA implementation.

7. Methods and Results

For the simulation, we used a list of hospitals by region from the Moroccan administration Open Data website, and we added demand for drugs to be served for the Casablanca region. Some of those health-care facilities use call for tender instead of begging to serve from CP based on category. For the purpose of this simulation, we consider all the facilities in table 2 are served by centralized pharmacies. Also, we consider three CP in the region only for the demonstration:

Table 2: The list of healthcare facilities in the grand Casablanca region.

ID	HOSPITAL	COORDINATE	DRUGS UNIT
1	Sekkat (Salam)	33.54631, -7.5884	30
2	Centre de Léprologie	33.55384, -7.59255	40
3	Mohamed V	33.58945, -7.55146	10
4	Med Bouafi	33.55874, -7.60972	15
5	Ben M'sick	33.55225, -7.5782	22
6	IbnRochd	33.58125, -7.61976	41

7	My Youssef	33.60274, -7.63226	13
8	20 Aout 1953	33.57474, -7.6199	8
9	Hôpitald'enfants	33.58088, -7.62111	21
10	Al Hassani	33.55849, -7.68736	26
11	Tit Mellil	33.54937, -7.48586	50
12	My Abdellah	33.69778, -7.38551	14
13	SidiOthmane	33.55804, -7.57312	7
14	Prince My Hassan	33.60457, -7.63394	13
15	Al Mansour	33.6078, -7.50362	30

Consider three CP with coordinate: CP1 {33.58212, -7.6594}, CP2 {33.57125, -7.5813}, CP3 {33.58200, -7.5289}

The capacity of each vehicle is 50 units, the population size is 10 chromosomes. In our demonstration we followed the standard GA for MDVRP mentioned by [23] after performing our clustering method mentioned above. The operation was simulated using C# as programming language on an Alienware aurora R2 CPU i7 and RAM 32 GB.

The grouping will give us three set CP1, CP2 and CP3:

CP1: [6-7-9-14-1-2-4] CP2: [5-8-10-11-13] CP3: [3-12-2015]

The bellow table (table 3) shows the simulation for weighted-k means by generations with their set of groups. The group here means the route that each vehicle needs to do to de starting from the source (CP in our case) which is presented as 0. Each 0 means a go back to CP to load drugs for the next route.

Table 3: Result of simulation for weighted-k means.

Generation	Fitness	Group	Path
Initial Generation	60000>	CP1	0-3-15-0-12-0
		CP2	0-1-13-0-11-0-10-5-0-8-2-0-4-0
		CP3	0-14-0-6-0-7-9-0
Generation 1	55986	CP1	0-3-15-0-12-0
		CP2	0-1-13-0-11-0-10-5-0-8-2-0-4-0
		CP3	0-6-0-14-7-9-0
Generation 20	51848	CP1	0-12-15-0-3-0
		CP2	0-8-10-0-1-0-11-0-4-5-13-0-2-0
		CP3	0-6-0-9-7-14-0
Generation 100	49948	CP1	0-12-15-0-3-0
		CP2	0-4-1-0-10-5-0-11-0-2-13-0-8-0
		CP3	0-14-7-9-0-6-0

Due to the mutation and crossover, each generation has its own fitness, a part of the above table presents an example of fitness by several generations (initial, 1, 20, 100). Figure 11 presents details of the evolution of fitness score by generation. The fitness is calculated based on several visits for the CP in each group, the figure also presents the comparison between k-means method used in [3] with the distance clustering and weighted k-means proposed in this paper. Both methods have the same initial data.

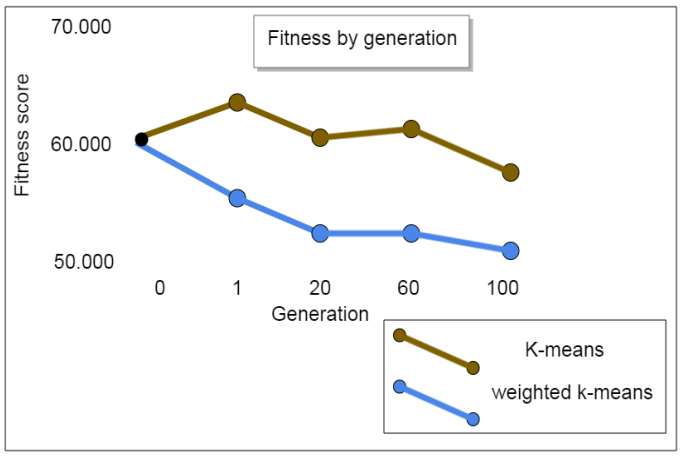


Figure 11: Fitness time for each generation.

Results show considerable regression in fitness score, which means acceptability of solution in newer generations in both clustering methods. Both methods have been improved over each generation, and the results shows that generation 100 is more fit in comparison with the others. However, the fitness score reached by weighted-k means is more acceptable than k-means.

The time simulation for each generation had shown a regression that can be presented in figure 12 as a chart for both methods. The time by minutes is calculated based on the distance crossed from each node of the set in each group with the visits of CP.

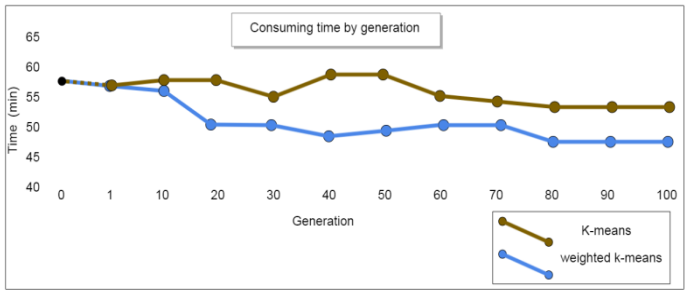


Figure 12. Consuming time for each generation.

In terms of time, the total consumption time from generation to generation is considerably changing in both methods. For the weighted k-means, the first generation shows better results in terms of time (less time-consuming means better results) in comparison with the initial random population (generation 0). At generation 40, the result is considered as the "better result", before time rises again. However, results show better results at generation 80 to 100 which means an escape from local minima to new global minima due to the use of crossover and mutation in GA and that shows less time in comparison with the normal k-means method.

Thus the result shows a less distance traveling between nodes (figure 11) and a faster route for each group using the weighted k-means in comparison of k-means (figure 12). Note that k-means has been improved over generations but not more fit then weighted k-means, which can impact the traveled cost by distance, the figure

13 shows the comparison between our proposed GA and the GA proposed by [3].

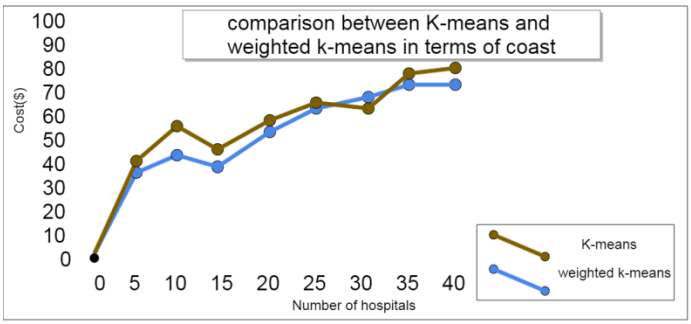


Figure 13: Comparison between K-means and weighted k-means in terms of cost.

The cost in weighted k-means is more efficient in comparison with traditional k-means, the figure 13 shows that the result using our proposed solution is competitive with k-means proposed by [3], however, the k-means solution can be more efficient if the hospitals' area in large number, which in our case not realistic due to the nature of the problem that we are resolving, the hospitals cannot exceed a large number in comparison with the CP and population in a specific region.

8. Conclusion and Future Work

Health-care and hospitals suffer from drug delivery delay and distribution due to the lack of infrastructure and distributed central/regional pharmacies. Our contribution is the adoption of clustering methods to locate the appropriate placement to set new central/regional pharmacy and group hospitals to favorable pharmacy based on hospital bed capacity, and using an algorithm based on genetic algorithms to set routing and scheduling plans for drug distribution at the level of one region or a set of regions. The MDVRP besides clustering is rarely adopted in the resolution of similar matters. The most relevant algorithm to adopt is weighted k-means one that showed favorable results, at least in our case. Also, even if this solution appears to be specific to the health-care sector, this work can concern further fields like agriculture and food-processing.

As an additional feature/improvement to the current model, it can be helpful considering applying parallelization and prediction for the GA algorithm and taking into consideration the conditions of transportation and providing an access to the result beside the transportation process with transparency.

References

[1] J.W. Thomas, J.R. Griffith, P. Durance, "Defining hospital clusters and associated service communities in metropolitan areas," Socio-Economic Planning Sciences, **15**(2), 45–51, 1981, doi:10.1016/0038-0121(81)90048-3.

[2] S. Zhang, W. Zhang, Y. Gajpal, S. Appadoo, Ant Colony Algorithm for Routing Alternate Fuel Vehicles in Multi-depot Vehicle Routing Problem: Performance and Safety Management, 251–260, 2019, doi:10.1007/978-981-13-0860-4_19.

[3] N. Mahmud, M.M. Haque, "Solving Multiple Depot Vehicle Routing Problem (MDVRP) using Genetic Algorithm," 2nd International Conference on Electrical, Computer and Communication Engineering, ECCE 2019, 1–6, 2019, doi:10.1109/ECACE.2019.8679429.

[4] K. Leungsubthawee, S. Saranwong, C. Likasiri, "Multiple depot vehicle routing problems on clustering algorithms," Thai Journal of Mathematics, 2018(Special Issue AnnualMeetinginMathematics), 205–216, 2018.

[5] T. Phiboonbanakit, T. Horanont, T. Supnithi, V.-N. Huynh, "Knowledge-Based Learning for Solving Vehicle Routing Problem," in Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, Association for Computing Machinery, New York, NY, USA: 1103–1111, 2018, doi:10.1145/3267305.3274166.

[6] J.R. Griffith, Quantitative techniques for hospital planning and control, Lexington, Mass., Lexington Books, Lexington, Mass., 1972.

[7] N. Albarakati, Z. Obradovic, "Disease-Based Clustering of Hospital Admission: Disease Network of Hospital Networks Approach," in 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), 636–641, 2017, doi:10.1109/CBMS.2017.87.

[8] A.K. Jain, "Data clustering: 50 years beyond K-means," Pattern Recognition Letters,**31**(8),651–666,2010, doi:https://doi.org/10.1016/j.patrec.2009.09.011.

[9] P. Delamater, A. Shortridge, J. Messina, "Regional health care planning: A methodology to cluster facilities using community utilization patterns," BMC Health Services Research, **13**, 333, 2013, doi:10.1186/1472-6963-13-333.

[10] M.R. Lavergne, "Identifying Distinct Geographic Health Service Environments in British Columbia, Canada: Cluster Analysis of Population-Based Administrative Data TT - Identifier les différents environnements de services de santé en Colombie-Britannique, au Canada: analyse typologique des données administratives en fonction de la population," Healthcare Policy = Politiques de Sante, **12**(1), 43–51, 2016.

[11] X. Chen, W. Yin, P. Tu, H. Zhang, "Weighted k-Means Algorithm Based Text Clustering," in 2009 International Symposium on Information Engineering and Electronic Commerce, 51–55, 2009, doi:10.1109/IEEC.2009.17.

[12] D. Serrou, A. Abouabdellah, An organizational strategy in order to optimize costs of medications return, 2016, doi:10.1109/GOL.2016.7731677.

[13] R.T. Sumichras, I.S. Markham, "A heuristic and lower bound for a multi-depot routing problem," Computers & Operations Research, **22**(10), 1047–1056, 1995, doi:https://doi.org/10.1016/0305-0548(94)00083-K.

[14] W. Ho, G.T.S. Ho, P. Ji, H.C.W. Lau, "A hybrid genetic algorithm for the multi-depot vehicle routing problem," Engineering Applications of Artificial Intelligence, **21**(4), 548–557, 2008, doi:https://doi.org/10.1016/j.engappai.2007.06.001.

[15] G. LAPORTE, Y. NOBERT, S. TAILLEFER, "Solving a Family of Multi-Depot Vehicle Routing and Location-Routing Problems," Transportation Science, **22**(3), 161–172, 1988.

[16] J.K. Lenstra, A.H.G.R. Kan, Some Simple Applications of the Travelling Salesman Problem, Operational Research Quarterly (1970-1977), **26**(4), 717, 1975, doi:10.2307/3008306.

[17] B. Crevier, J.-F. Cordeau, G. Laporte, "The multi-depot vehicle routing problem with inter-depot routes," European Journal of Operational Research, **176**(2), 756–773, 2007.

[18] N. Aras, D. Aksen, M. Tuğrul Tekin, "Selective multi-depot vehicle routing problem with pricing," Transportation Research Part C: Emerging Technologies,**19**(5),866–884,2011, doi:https://doi.org/10.1016/j.trc.2010.08.003.

[19] D.J. Gulczynski, "Integer Programming-Based Heuristics for Vehicle Routing Problems," 2010.

[20] A. Santoso, D.N. Prayogo, D.Y. Nugroho, "Model Fuzzy Multiobjective Vehicle Routing Problem untuk Produk Perishable dengan Pendekatan Algoritma Genetika," in Seminar Nasional Teknik Industri BKSTI 2014, IV–99, 2014.

[21] M. Santé Maroc/DPRF/DPE/S.E.I.S., "Santé en Chiffres 2015/ Edition 2016," 168, 2016.

[22] R. Hofman, "Why Linear Programming cannot solve large instances of NP-complete problems in polynomial time," 2006.

[23] S. Paneerselvam, S. Sai, "Solution to multi-depot vehicle routing problem using genetic algorithms," World Applied Programming, **1**, 118–131, 2011.

[24] S. Karakatič, V. Podgorelec, "A survey of genetic algorithms for solving multi depot vehicle routing problem," Applied Soft Computing, **27**, 519–532, 2015, doi:https://doi.org/10.1016/j.asoc.2014.11.005.

[25] B. Ombuki-Berman, A. Runka, F. Hanshar, Waste collection vehicle routing problem with time windows using multi-objective genetic algorithms, 2007.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1153-1162 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

The Design of a Hybrid Model-Based Journal Recommendation System

Adewale Opeoluwa Ogunde*, Mba Obasi Odim, Oluwabunmi Omobolanle Olaniyan, Theresa Omolayo Ojewumi, Abosede Oyenike Oguntunde, Michael Adebisi Fayemiwo, Toluwase Ayobami Olowookere, Temitope Hannah Bolanle

Department of Computer Science, Redeemer's University, Nigeria

ARTICLE INFO

Article history:
Received: 29 September, 2020
Accepted: 16 November, 2020
Online: 14 December, 2020

Keywords:
Academics
Collaborative filtering
Content-based filtering
Journal
Publications
Publication Authors
Recommender systems

ABSTRACT

There is currently an overload of information on the internet, and this makes information search a challenging task. Researchers spend a lot of man-hour searching for journals related to their areas of research interest that can publish their research output on time. In, this study, a recommender system that can assist researchers access relevant journals that can publish their research output on time based on their preferences is developed. This system uses the information provided by researchers and previous authors' research publications to recommend journals with similar preferences. Data were collected from 867 respondents through an online questionnaire and from existing publication sources and databases on the web. The scope of the research was narrowed down to computer science-related journals. A hybrid model-based recommendation approach that combined Content-Based and Collaborative filtering was employed for the study. The Naive Bayes and Random Forest algorithms were used to model the recommender. WEKA, a machine learning tool, was used to implement the system. The result of the study showed that the Naïve Bayes produced a shorter training time (0.01s) and testing time (0.02s) than the Random forest training time (0.41) and testing time (0.09). On the other hand, the classification accuracy of the Random forest algorithm outperformed the naïve Bayes with % correctly classified instance of 89.73 and 72.66; kappa of 0.893 and 0.714; True Positive of 0.897 and 0.727 and ROC area of 0.998 and 0.977, respectively, among other metrics. The model derived in this work was used as a knowledge-base for the development of a web-based application, named "Journal Recommender" which allowed academic authors to input their preferences and obtain prompt journal recommendations. The developed system would help researchers to efficiently choose suitable journals to help their publication quest.

1. Introduction

The overload of information available through the world wide web today has created a lot of challenges for information seekers. Researchers who are in need of journals related to their research area sometimes find it difficult to access and keep track of the relevant journals or research materials of their interest including searching for relevant journal to publish research outputs.

Conducting a very good research could be very stressful and searching for the right journal to publish it could even be more stressful. Academic authors are therefore faced with making their decision based on a number of competing preferences such as the scope of the journal, the access type, article processing charge, speed of the review process, impact factor, indexing and ranking just to mention a few. A simple approach used in searching for such journal publishers is to randomly surf the web by typing some specific keywords predetermined by the author. Results obtained from using this style may not be too beneficial as it is probability-based and highly dependent on the searching skills and maneuvering abilities of the author. Some authors, also, make their journal selection decision by checking out the reference lists form existing related journals available to them. This style is also limited in scope and it depends fully on related journals currently available to the author. Moreover, some potentially useful journals may be difficult to access by authors. It is also a fact that the article processing charge of some accessible journals may be too expensive especially for authors from low income countries where

*Corresponding Author: Adewale Opeoluwa Ogunde, Department of Computer Science, Redeemer's University, Ede, Osun State, Nigeria, +2347036090090, ogundea@run.edu.ng

access to research grants and funding are also a major challenge. There is, therefore, a need for a recommender system that can assist in solving the highlighted problems.

Recommender systems are described as information filtering systems that are designed to solve the challenge of information overload by filtering a large amount of information according to the user's observed behaviour, interest, preferences and producing very useful and vital information fragment as output [1]. Recommender systems have the ability to predict the choice of its user based on their past preferences. In this work, the use of a journal recommender systems to automatically suggest relevant journals to the researchers based on some initial information provided by the users and other information from journal publishing sites, was developed. To provide more relevant recommendations, the developed recommender system would incorporate the list of journals, the type of journal access whether it is an open access or restricted access, free or article processing fee, research topic, the field of study, indexing and rating. It is also specific to the field of study, which is Computer science-related journals.

This work resolves information overload problem to academic authors; whereby a researcher would want his/her authored works to be published, this might take a more extended period because the scope in the popular bibliographic database is vast. Also, in this age, when there is high proliferation of academic journals, getting the right journal to publish an article could turn out to be a huge task. A recommender system for guiding researchers when searching for journals to publish their articles promptly was developed in this work. This would make their publication quest easier and effective.

2. Literature review

In this section, some relevant literature tied to the research were reviewed and some closely related works were also presented.

2.1. Recommender Systems

The increasing use of internet services has resulted in a massive amount of data and information available on the internet. The nature of this data is so dynamic as well as diverse that it becomes difficult for one to look for the kind of data that matches the requirements. This challenge has aroused the development of new technology to assist Internet users to cope with information overload and indecisions on a product, service and items. They are called Recommender Systems. According to [2], recommender systems are applications that filters users' preferences, suggesting the most suitable items to specific users based on similar information about the users, the items, and their relationships. Solving the problem of information overload and filtering out the most important items, products and services are considered to be the primary purpose of recommender systems. Personalized recommendations could be generated by analyzing the behaviour of a user and other similar users to suggest user's interests and preferences [3]. Recommender systems are very useful in diverse domains such as education, commercial, entertainment, for example, movies, music and others. This work focused on deploying the recommender system in the academic domain in order to facilitate the search for useful educational web resources such as Journals, more comfortable to find, thereby helping researchers enhance their publishing process by getting access to very useful journal recommendations leading to faster publication time.

2.2. Techniques of Recommender Systems

The main recommendation techniques include traditional methods such as content-based, collaborative filtering-based, hybrid-based (a combination of content-based and collaborative filtering methods) and knowledge-based methods [4]. Other recently appearing variants in the literature are trust-based, group-based, fuzzy set-based, context awareness-based and social network-based methods. The most acceptable and widely used out of all are the traditional recommendation methods. Content-based is further divided into tag-based and descriptive-based. Collaborative filtering's subdivisions are Model-based and Memory-based, under model-based are Clustering, Association, Bayesian networks while for Memory-based are User-based and Item-based.

In content-based recommendation methods, items similar to items previously preferred by a specific user are usually recommended [5]. More on the basic principles, techniques, strengths and limitations of content-based recommender approaches can be found in [5]-[7]. On the other hand, collaborative filtering is a process of filtering information or pattern based on the opinion of other users, or the similarity between items [6]. This approach makes recommendation by finding similar users having a rating history similar to the new user. This implies that user preferences, in this case, are compared with the other users [7]. More information on the collaborative approach, its characteristics, techniques, strengths and drawbacks can be found in [6]-[10]. A hybrid recommendation system is an approach that deals with the combination of different recommendation models to evolve a more efficient and effective technique that would outperform the traditional stand-alone recommendation techniques. This is done to overcome the weaknesses of stand-alone recommendation techniques while combining their power and strengths to achieve better recommendation. There are seven main combination approaches for building hybrid recommendation system, and they are also grouped into three designs which are Monolithic, Parallelized and Pipelined. More details on hybrid recommendation systems, adapted in this work, could be obtained from the works of [4], [10] and [11].

2.3. Related Works

Apart from some of the previous works done by authors [12] and [13] on the exploration of recommender systems in the education sector, some other related works are also presented here. The authors considered only one algorithm in their implementations. A knowledge-base of product semantics recommender system was built by [14]. The recommender system was capable of determining semantic attributes of products, thereby helping the authors to understand the customer's taste. In order to learn the attributes of the product descriptions found on retailers' websites, supervised and semi-supervised learning techniques were applied to the system. The system was able to map products to an abstract layer of semantic features by recommending other items in the same class of products that

matches the user's model and understand the customer taste and also recommend items across categories. This approach is believed to enhance the user experience and build the user's confidences in the recommendations. The work of [14] focused on narrow classes of products from retailers' websites. A recommendation system based on collaborative filtering approach using Convolution Deep Learning Model was proposed by [15]. The convolution deep learning model used was based on the Label Weight Nearest Neighbor with three categorizations. The authors in [15] concluded that their proposed system outperformed the other traditional recommendation algorithms in terms of efficiency and effectiveness.

In [16], a product recommender system based on a hybrid recommendation technique was presented. The system focused on books as the product, and in order to recommend more accurately, a personalized book recommendation algorithm was used. It was based on the time-sequential collaborative filtering and content-based filtering. Essential factors such as time sequence of purchasing books of different users in the database and content of the user profile were considered in the system, and it is expected to satisfy users by providing best and efficient book recommendations. The accuracy of recommendation results in this work was low and this was due to the small volume of books used as the experimental dataset. The authors in [17] were able to predict fake information and fake Facebook accounts using a machine learning-based recommendation approach, which can also work on online social network. The limitation of the work was the low prediction accuracies recorded. In [18], a collaborative filtering job recommender system was developed, which was executed by exploring and analyzing Google employees' features to match the qualifications of job seekers and, thereafter, recommending Google to job seekers. The classification algorithms used were SVM, Naïve Bayes, Neural Networks and Decision Tree. The low F1 score and the few number of companies with few features was the limitation to the work.

3. Methodology

This section presents the research methodology, which includes the techniques used in the recommendation system and the design of the hybrid recommendation system.

3.1. Data collection and description

The primary aim of this work is to develop a system that recommends computer-related journals to academic researchers and authors. Two sets of data were used for the study. First primary data was collected through an online questionnaire prepared through google form and broadcasted across the web. Secondary data was also collected directly from major journal databases and indexing organizations available online. The sources of the data were obtained from bibliographic databases, journal publishers, and their journals, some of them are, Science Direct, Scopus, Researchgate, Elsevier, Google Scholar, DOAJ just to mention a few. The selected journals from these databases were specific to the field of computer science, and the topics selected covered every aspect of computer science including Computer programming, Big data analytics, Software engineering, Computer Architecture, Artificial Intelligence, Cybersecurity, Decision trees, Encryption, Data Mining, Recommendation systems, Human-Computer Interaction, Robotics and Cloud computing just to mention a few.

The questionnaire, used to collect the data was administered online using Google Forms, with eight hundred and sixty-seven (867) respondents, who are researchers that have published at least one article in a standard journal responded to the survey. There were three major sections in the online questionnaire. The first section contained the "Demographic data of the respondents". The second section contained "Questions on Articles Already Published by Respondents in Learned Journals" and "Questions on Publishing Preferences". The first section consisted of five questions. The second and third questions consisted of ten questions, each making a total twenty-five questions altogether. The demographic section contained five questions that required answers about the target audience which included their optional full names, institution, gender, highest academic qualification and work status of respondents. In summary, the majority (71%) of the respondents were males, and the remaining 29% were females. 76% of the respondents were PhD holders, and most (73%) of the respondents had published at least two articles in learned journals.

3.2. Attribute Representation and Categorization

Four major attributes were used for the recommendation in this work. They are Area of Research Interest (ARI), Journal Access Type (JAT), Article Processing Charge (APC), Indexing and Ranking (IAR). All the data were stored in Microsoft Excel and saved as a comma-separated (.CSV) files. The attributes used were assigned labels represented by three-character nominal values to make the data suitable for classification. These attributes represent preferences of future authors that will be matched in the journal recommendations model built from preferences selected by previous authors and secondary data obtained from publishing sites.

Eighty-one ARI was identified in Computer Science. They are Artificial Intelligence, Adaptive computing, Big Data Analytics, Bioinformatics, Biometric Technology, Classification, Cloud Computing, Clustering, Computational Biology, Computer Forensics, Computer Security, Computer Vision, Content Management Systems, Crowdsourcing, Cryptography, Cybernetics, Data Security, Data Management, Data Warehousing, Data Mining, Decision Support Systems, Decision trees and forests, Distributed systems, E-Commerce, E-Government, E-Learning Technologies, Embedded Systems, Evolutionary computing, Genetic Algorithms, Global Positioning System (GPS), Global System for Mobiles (GSM), Geographical Information Systems (GIS), Green Computing, Grid Computing, Human-Computer Interaction (HCI), Image processing technologies, Information systems, Internet Telephony, Intrusion Detection Systems, Knowledge based systems, Knowledge Management Systems, Machine Learning, Management Information Systems, Mobile agents, Mobile Computing, Multi-agent systems, Network Security, Artificial Neural Networks, Open Source Technology, Parallel Computing, Computer Architecture, Parasitic Computing, Pattern Recognition Techniques, Pervasive Computing, Proactive Computing, Real-time information systems, Recommender Systems, Robotics, Social Networks and Online Communities, Soft Computing, Support Vector Machines (SVM), Theoretical Computer Science, Internet of Things, Virtual Reality, Visualization, Voice Over Internet Protocol, Web-mining, Network Topology, Biosystems & Computational Biology, Cyber-Physical Systems and Design

Automation, Computer Engineering, Database Systems, Computer Education, Computer Graphics, Process Mining, Operating Systems, Programming Systems, Scientific Computing, Security, Theoretical Computing, Networking and Deep Learning. The JAT was divided into two categories, which are Open Access (OA) and Restricted Access (RA). Table 1 shows the distribution characteristics of the preferences of the respondents concerning JAT. The majority (62.51%) of the respondents preferred to publish in open access journals.

Table 1: Journal access type preferences

JAT	Frequency	Percentage (%)
OA	542	62.51
RA	325	37.49
TOTAL	867	100.00

APC was sub-divided into three other categories, which are low, medium and high with low having highest percentage (83.04%) preferences among the respondents, as shown in Table 2.

Table 2: Journal Article Processing Charge Preferences

Category	APC	Frequency	Percentage (%)
BELOW 500USD	LOW	720	83.04
500 - 1000 USD	MEDIUM	92	10.61
ABOVE 1000 USD	HIGH	55	6.34
	TOTAL	867	100.00

Furthermore, few (23.07%) of the respondents did not care about indexing and ranking of the journal while the majority (76.93%) considered ranking and indexing as the most crucial factor for publishing their articles in a journal (Table 3).

Table 3: Journal Indexing and Ranking Preferences

IAR	Frequency	Percentage (%)
YES	667	76.93
NO	200	23.07
TOTAL	867	100.00

3.3. The Design of the Recommender System

This section gives an accurate description of how the proposed system was designed. Unified Modeling Language (UML) was used to understand the flow of this system, and it shows the functions the users can do. The users of the system are represented as actors of the system, and the actions performed are called the use-case. There are two primary users of the system, the admin and the user. Figure 1 shows the use-case diagram for the Admin and User of the system. The Admin will normally log in as the system administrator. The System Admin can add journals and make it accessible to the system users with all the required details attached to it. The details are its Title, Topic category, Publisher, Its type of access, Its article processing fee, Issues per year and its link for easy access. System Admin can also add the category of a journal which also means the field of the journal, Publishers of Journals into the system and can View, Edit, Update and Delete Journals details. The User, on the other hand, can register and login to the system, search for Journals, Publishers, Categories with the use of keywords, view system recommendations regarding journals selected and generated by the system based on particular preferences.

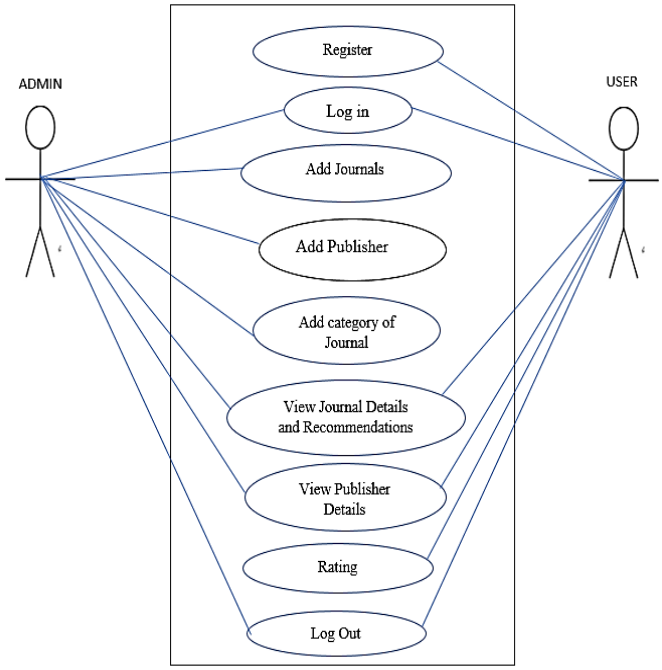


Figure 1: The Use-Case Diagram for the Recommendation System

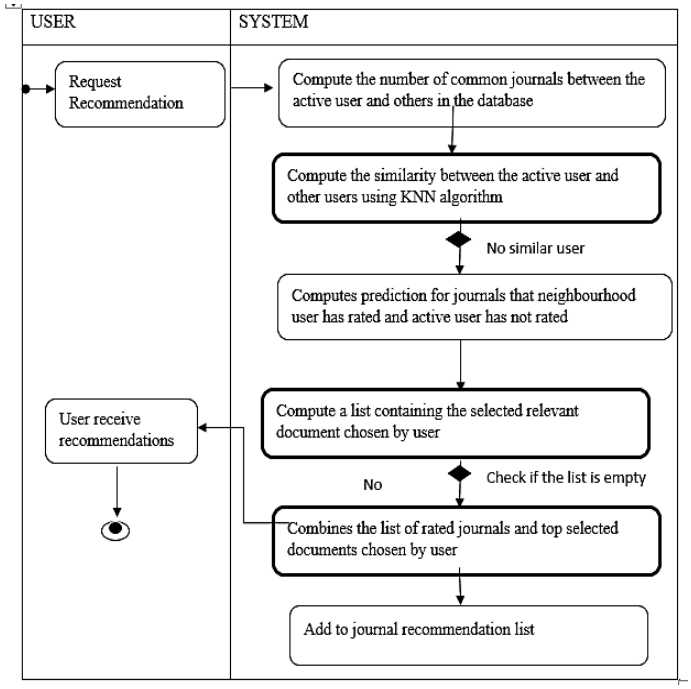


Figure 2: The Hybrid Filtering Recommendation Activity Diagram

3.3.1. Description of the hybrid filtering recommendation activity diagram

In a hybrid recommendation model shown in figure 2, recommendations were based on content-based filtering and collaborative recommendation combined using a monolithic feature augmentation approach. It also employed the combination of Naïve Bayes and Random forest algorithms for searching the desired journal. This model combined the results of collaborative and the result of content-based views the results to the users**.**

3.3.2. Description of the register process activity diagram

The activity diagram in Figure 3 shows that for a user to make use of the system, the user has to be registered with valid registration details. The user views the blank registration form and fills in personal details such as name, email address and phone number and other details the system requires. As the user begins to fill the form, the system simultaneously stores the information. The user is also expected to fill the preferences details such as the user's research topic, the type of access to journals, and field of study. Then the account form as part of the registration is the authentication using passwords. After the user has filled all the dialogue boxes for registration, the user's account is successfully created, and the account is active.

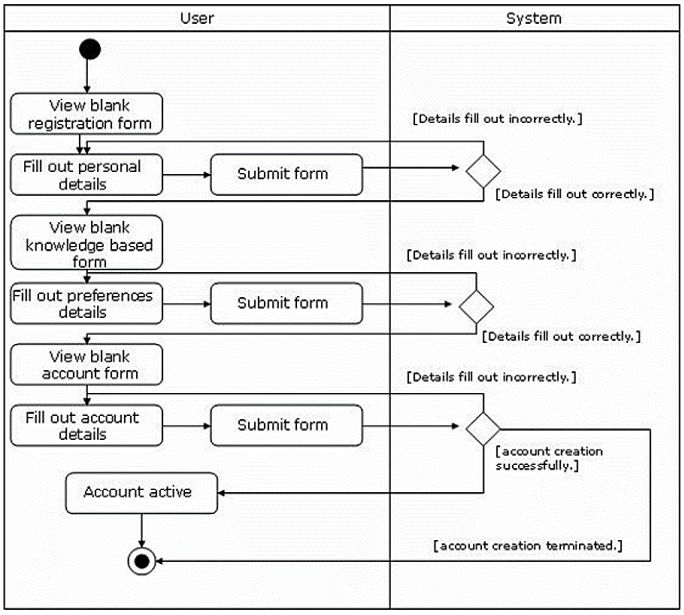


Figure 3: The Register Process Activity Diagram

3.3.3. Description of the search process activity diagram

The activity diagram in figure 4, shows the search process of when a user decides to search for a journal by using its name, publisher and the category, type of access, issue per year. A user can also make use of keywords as an alternative for the searching process, for example, searching for the word " Inter" for "International". The user selects the search form and fills it, the system checks for the search result, filters, sorts and views multiple lists of results to the user. The system can also inform the user when there are no matches of the searched journal. Every user can search as many as they require for journals of their interest.

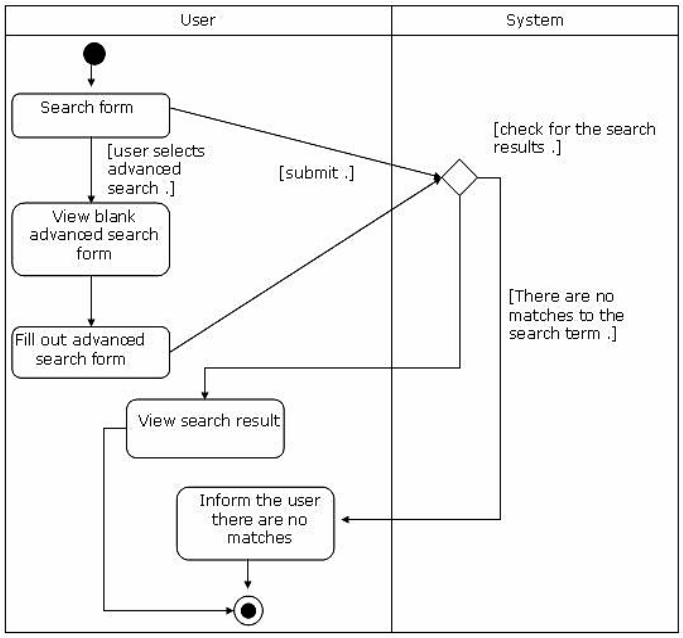


Figure 4: The Search Process Activity Diagram

4. Implementation and Results

This section provides the details of implementation followed by the discussion of the results.

4.1. Implementation

The WEKA Explorer version 3.9.4 was used for the training of data and implementation. This was done based on the data generated from the respondents' preferences discussed in section three. Naïve Bayes and Random Forest algorithms were used for the classification of the data and training of the model. The model generated was used as a knowledge-base for the development of the web-based journal recommender application. The data was saved in comma-separated values (CSV) format as required in the WEKA explorer and preprocessed by converting it into the *.arff* format, which is the format recognised by WEKA for a classification of this type. The preprocessed data were trained using the Naïve Bayes and Random Forests without any splitting. In classifying the data, accuracy was based on predicting the journal preferences class labels in the training data. Naive Bayes classifier is a major classifier that is very useful in Collaborative Filtering based Recommendation Systems. Naïve Bayes uses machine learning and data mining techniques to filter the information and use it to predict future occurrences. Naive Bayes algorithm can handle both continuous and discrete data. It requires less training data, can make probabilistic predictions, requires less training data, easy to implement and fast. Random Forest algorithm can be used for both classifications and regression task. It provides higher accuracy than other classification algorithms. Random forest classifier will handle the missing values and maintain the accuracy of a large proportion of data. If there are more trees, it will not allow overfitting trees in the model. Random Forests was chosen over decision trees and other classification methods because it is easy to interpret and make for straightforward visualizations, the internal workings are capable of being observed and thus make it possible to reproduce work, it can

handle both numerical and categorical data, and performs well on large datasets and are extremely fast.

4.2. Results and Discussion

Results of data training and classification carried out in WEKA explorer using the two ranking algorithms is presented in this section.

4.2.1. Results from Classification with Naïve Bayes Algorithm

Figure 5 shows the results of data training and model generation and with Naïve Bayes algorithm in WEKA explorer. Eight hundred and sixty-seven instances of the data and five attributes were trained using the 'Use training set' option.

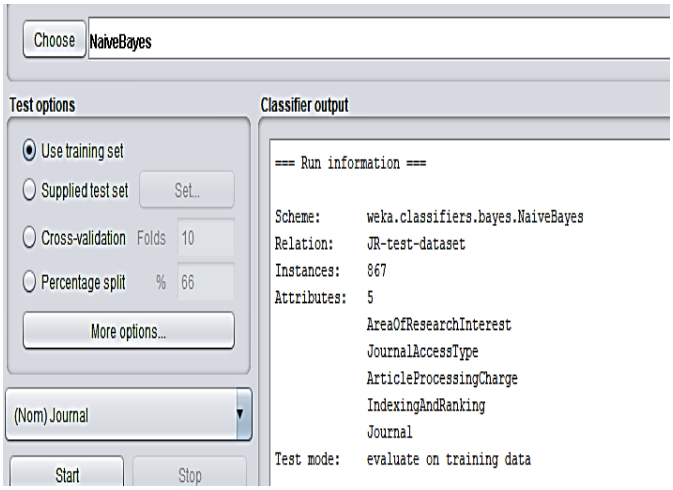


Figure 5: Details of Classification with Naïve Bayes Algorithm

Results from the classification using Naïve Bayes algorithm are shown in Figure 6. Time taken to build the model was 0.01 seconds. Time taken to test the model on training data was 0.02 seconds. There were six hundred and thirty (72.66%) correctly classified instances and two hundred and thirty-seven (27.34%) incorrect classified instances. Mean Absolute Error (MAE), which is a measure of the average magnitude of the errors in a dataset without considering their direction, was 0.0225. This result implied that the error associated with this classification was minimal and insignificant. Root Mean Squared Error (RMSE), which is a measure of the average magnitude of the error, was 0.0943. In this study, RMSE was larger than MAE, which implied that there was variance in the individual errors of the dataset. This is consonance with the standard as RMSE is always larger than or equal to the MAE. Kappa statistic, which measures the agreement of prediction with the actual class was 0.714. This showed a higher agreement between the predicted and the actual class. The evaluation measures that also showed the accuracy of the model are given as True Positive (TP) Rate, False Positive (FP) Rate, Precision, Recall, F-Measure, MCC, ROC Area and PRC Area. TP signifies the correct predictions. From the results obtained using Naïve Bayes classification, TP was an average of 0.727, which implied that more of the class labels were predicted correctly. False Positive (FP), which is the number of instances predicted positive that was negative, was an average of 0.014, which implied that fewer instances were not predicted correctly. Precision was an average of 0.789, which meant that many cases that were predicted positive are positive. The recall was an average of 0.727, which indicated that more of the positive instances were predicted positive. F-measure was an average of 0.760, which indicated that the precision and recall were evenly weighted. MCC was an average of 0.765. ROC Area was an average of 0.977. PRC Area was an average of 0.785. The high values obtained for MCC, ROC Area and PRC indicated that the model was predicting with a very high level of accuracy.

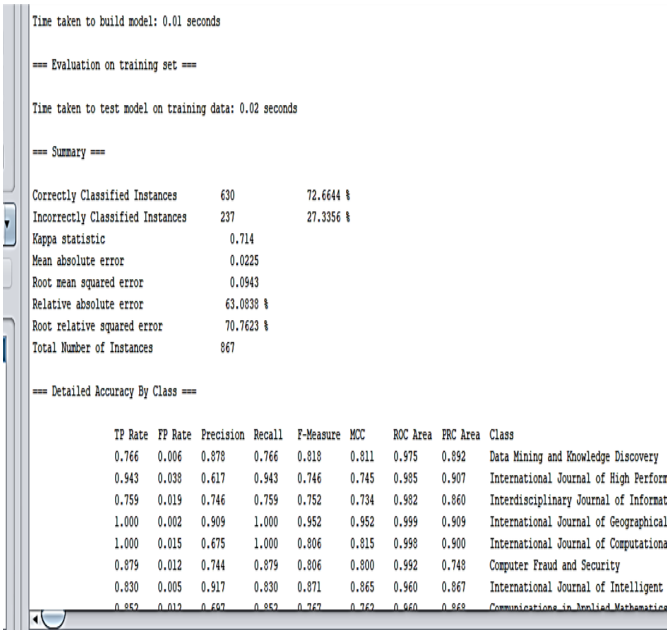


Figure 6: Results from Classification with Naïve Bayes Algorithm

4.2.2. Results from Classification with Random Forest Algorithm

Figure 7 shows the results of data training and model generation and with Naïve Bayes algorithm in WEKA explorer. Eight hundred and sixty-seven instances of the data and five attributes were trained using the 'Use training set' option.

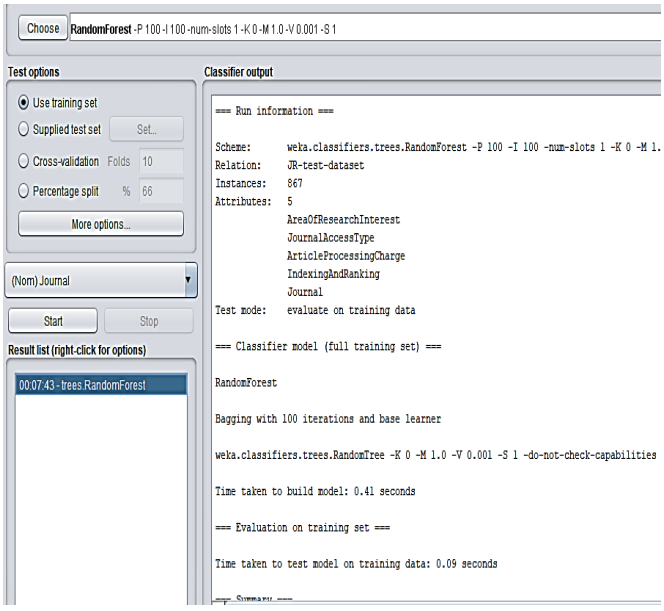


Figure 7: Details of Classification with Random Forest Algorithm

Results from the classification are shown in Figure 8. Time taken to build the model was 0.41 seconds. Time taken to test the model on training data was 0.09 seconds. There were seven hundred and seventy-eight (89.73%) correctly classified instances and eighty-nine (10.27%) incorrect classified instances. Kappa statistic was 0.893, which represents a higher agreement between the actual and the predicted class. MAE was 0.0061 and RMSE was 0.0526. which implied that there was variance in the individual errors of the dataset. Other evaluation measures that also showed the accuracy of the model are TP, which was an average of 0.897, which implied that more of the class labels were predicted correctly. Fewer instances were incorrectly predicted with FP as 0.004. The recall was an average of 0.897 and precision was an average of 0.900, which meant that many cases that were predicted positive are positive. F-measure was an average of 0.886, which indicated that the precision and recall were evenly weighted. ROC Area was an average of 0.998, PRC Area was an average of 0.912 and MCC was an average of 0.892. The high values obtained for ROC Area, PRC and MCC showed that the prediction accuracy of the model was very high.

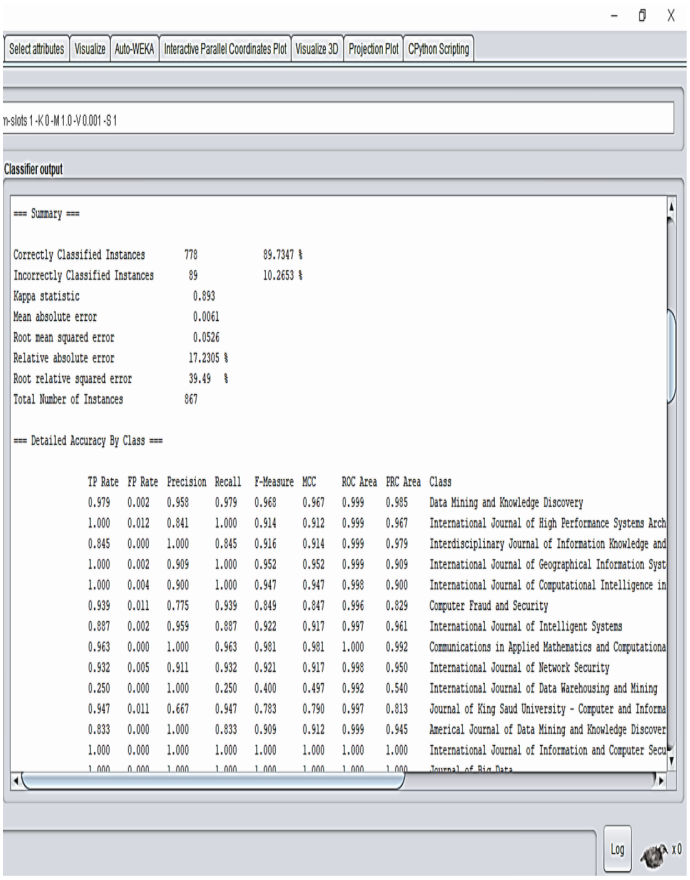


Figure 8: Results from Classification with Random Forest Algorithm

4.2.3. Comparing Results of Classification with Naïve Bayes and Random Forest Algorithms

The performance metrics for comparing the data training and model generation of the Naïve Bayes and Random Forest algorithms are shown in Table 4. The Naïve Bayes classification was faster than Random Forest in model building and testing. On the other hand, the Random Forest algorithm outperformed Naïve Bayes in terms of classification accuracy, Kappa statistics, true positives, ROC Area and all other performance metrics measured in this work.

Table 4: Results of Classification Using Naïve Bayes and Random Forest Algorithms Compared

PERFORMANCE METRICS	NAÏVE BAYES	RANDOM FOREST
Training time (s)	0.01	0.41
Testing time (s)	0.02	0.09
Instances correctly classified (%)	72.66	89.73
True positive	0.727	0.897
False positive	0.014	0.004
MAE	0.0225	0.0061
RMSE	0.0943	0.0526
Kappa Statistics	0.714	0.893
ROC Area	0.977	0.998
PRC Area	0.785	0.912
F-measure	0.760	0.886
Avg_Precision	0.789	0.900
MCC	0.765	0.892

4.2.4. Evaluation of Results

In this section, the results obtained from the study were compared with existing methods, which were already discussed in section two. After comparing the classification accuracies, the classification accuracy derived in this work (89.73%) outperformed four related works [13], [16]-[18].

Table 5: Results of Classification Using Naïve Bayes and Random Forest Algorithms Compared

Reference	Year of Publication	Algorithm Used	Classification Accuracy (%)
13	2019	C4.5	78.84
16	2017	Time sequence-based algorithm	60
17	2018	CNN	75
18	2015	Neural Network	63
This work	2020	Random Forest	89.73

4.3. System Interface

The section describes the features of the developed recommender system. The most important pages are described here and in the following subsections. The system's interface was built with HTML, CSS, PHP, Java scripts and AppServe. The interface for the application was thoroughly tested to ensure that it complies with the requirement of the recommendation system. The home page (Figure 9) is the first page a user is presented once the application loads. It shows the labels Home, Categories, Journals, Publishers, Recommendations, Register, Login and Search icons.

For example, users will click the 'Journals' menu to access the list of journals in the system and likewise click 'Recommendations' menu to make necessary recommendations after answering a few questions by selecting their options. The

'categories' menu shows the various fields of computer science available for searching.

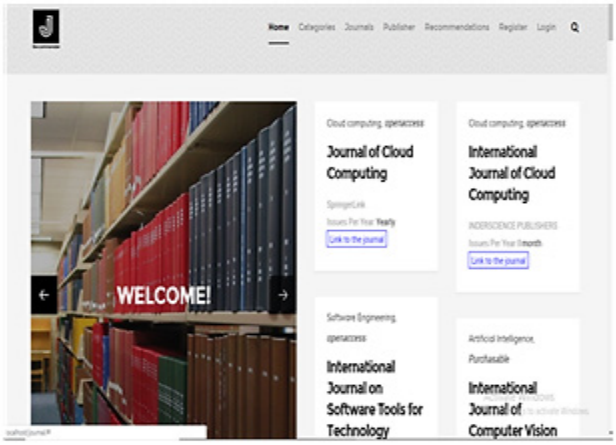


Figure 9: The home page of the 'Journal Recommender'

4.3.1. The Registration Page

The 'Registration page' is the page where a user creates an account. In this page the system requests for Name, Phone number, Email address, the type of user, Research topic, Selection of the field of study, the type of the Journal access he/she is interested in patronising, that is, open access or purchasable. The Registration Page (Figure 10) is significant because recommendations are generated from preferences any user inputs into the system, and the information supplied here helps to build the system's database with more robust information useful for future suggestions or recommendations.

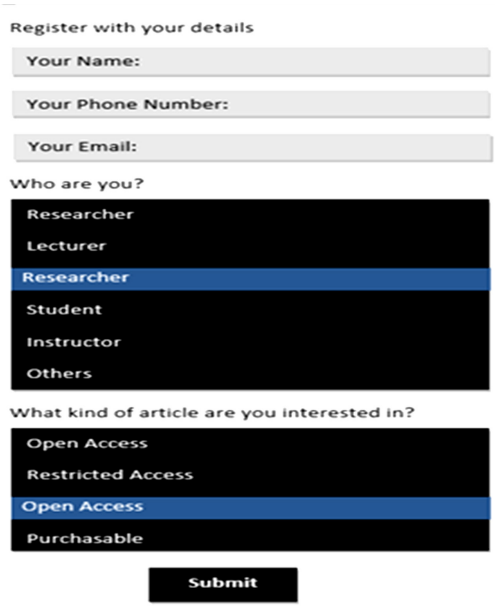


Figure 10: The Registration Page

4.3.2. The Journal Page

The 'Journal page' (Figure 11) is a direct link to all the journals available in the recommender system, and it gets updated from time to time.

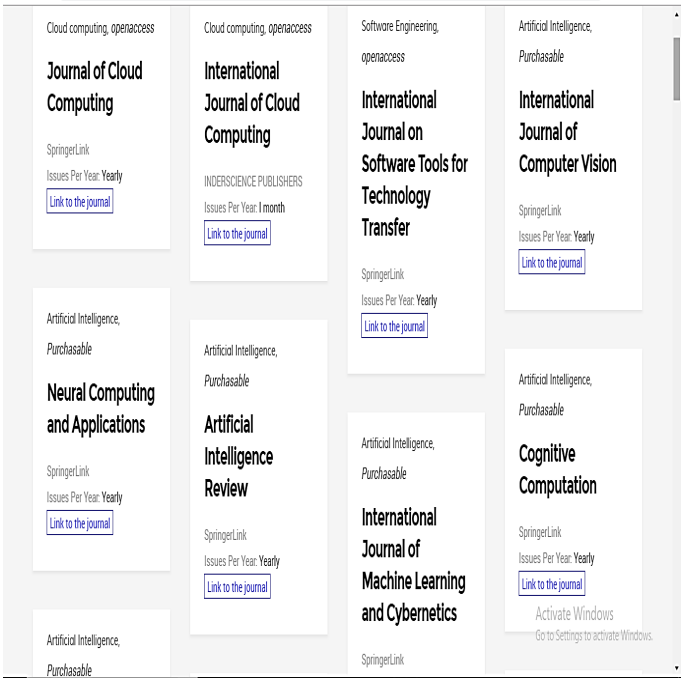


Figure 11: The Journal Page

4.3.3. The Publisher's Page

This 'publisher's page' (Figure 12) consists of the complete list of journal publishers and the journals each of them contain in the developed system.

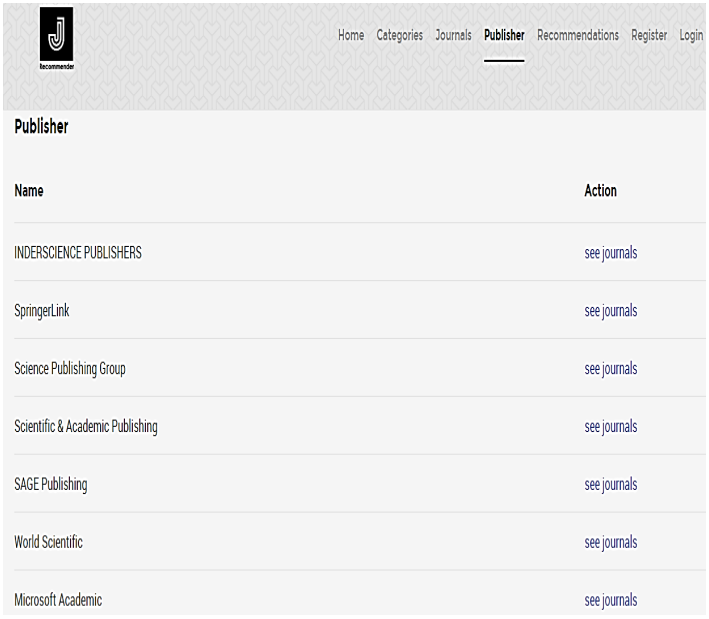


Figure 12: The Publisher's Page

4.3.4. The 'Get Journals Recommendation' Page

The 'Get Journals Recommendation' Page (Figure 13) is the main page displayed to the user when the "Recommendations" link is clicked. It requires the user to fill in the recommendation form, which is based on user's interests and selection of options. Some of the options that determine the final recommendation are the type of Journal Access, the kind of journals that serves user's interest,

the research topic, the amount of article processing fee that is charged and a few others. The user is expected to fill the form.

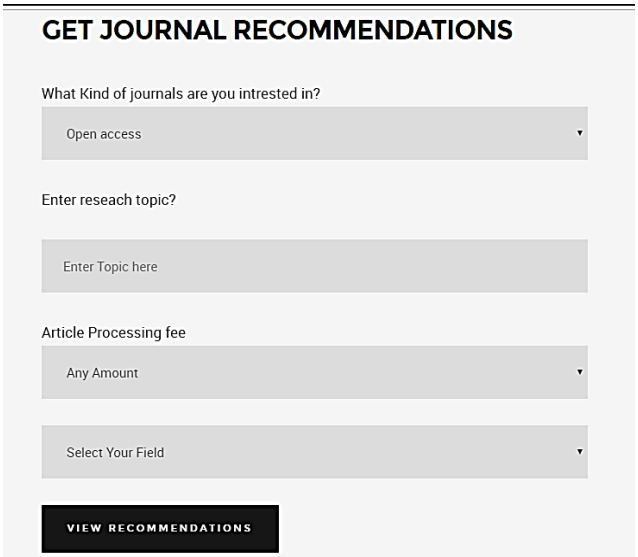


Figure 13: The Get Journals Recommendation Page

In Figure 13, the field "What Kind of Journals are you interested in?" refers to the type of journal access the user wants. Some journals are accessible to view, that is, open access, and others have to be purchased before viewing. The user is given a choice to choose the type of journal access he/she is interested in. Also, the field, Article Processing Fee', is also known as publication fee, which is charged to authors to make their work available in either the open access but free or restricted access and paid versions. This information is usually included in the Authors Instruction page of most journals. The range used for the purpose of the recommendation carried out in this work is from free, one to one-hundred USD, one hundred and one USD, two hundred and one to three hundred USD, three hundred and one to four hundred USD, four hundred and one to five hundred USD and five hundred USD & above. The category field defines what field of computer science the user is interested in. Examples of the fields in the drop-down list are Software Engineering, Information systems and Artificial Intelligence, just to mention a few. The rating field is used the system's feedback mechanism based on the user's acceptance of the recommendations. It has just two options, which are "Very Helpful" if the user is satisfied and "Not Helpful" if the user is dissatisfied with the recommendations from the system. This information is handy as the system makes use of it to self-improve itself from time to time.

4.3.5. System's extendability

The System Admin.'s page allows the flexibility and extendability of the system. The admin can add and update more publishers, more journals, more publishers, new users, review article processing charges and several other functions that extend the system's robustness and effectiveness.

5. Conclusion

An academic researcher can be an author who wants his or her article, papers or life's work to be published in journals through publishing groups. Over the years, the amount of information on the internet is enormous, and this causes information overload problem to researchers during their publication quest to finding the right journals to publish their works. This study proposed a solution that applied a monolithic hybrid recommender system with feature combination and augmentation along with two algorithms, which are Naïve Bayes and Random Forest. Data used in this work were computer science-related journals, collected from pre-existing data that are readily available on academic web sites and repositories. Five attributes were used as preferences for the recommendation of journals. They are the fields of study, type of journal access, research topic, article processing fee and rating of the journals. These attributes were trained and modelled in a machine learning tool, WEKA, and the model generated was used to develop a front-end interface application where prospective academic researchers and authors can enter preferences and view journal recommendations directly generated by the developed system. This system was tested and was found to make recommendations with high levels of accuracy. The recommender system, developed in this work, will help researchers and authors to quickly and dynamically determine which journal is best suitable for their ready-to-be published articles. It is highly recommended that universities, researchers, authors could start adopting the system to improve publishing speed and output. Future works will involve the use of ensemble machine learning algorithms. A comparative analysis would also be done to test the best method under different scenarios. Finally, future works would also include recommendations of other significant disciplines aside from computer science.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors appreciate the Department of Computer Science's laboratory, Redeemer's University, Ede, Osun State, Nigeria for providing a convenient environment for the analysis.

References

[1] J.A. Konstan, J. Riedl, "Recommender systems: From algorithms to user experience," User Modeling and User-Adapted Interaction, **22**(1–2), 101–123, 2012, doi:10.1007/s11257-011-9112-x.

[2] J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez, "Recommender systems survey," Knowledge-Based Systems, **46**, 109–132, 2013, doi:10.1016/j.knosys.2013.03.012.

[3] R. P, V. H.R., "Recommender Systems," Communications of the ACM, **40**(3), 77–78, 1997.

[4] R. Burke, "Hybrid recommender systems: Survey and experiments," User Modelling and User-Adapted Interaction, **12**(4), 331–370, 2002, doi:10.1023/A:1021240730564.

[5] M.. Pazzani, Billsus D., "Content-Based Recommendation Systems," The Adaptive Web, **43**(2), 325–341, 2007, doi:https://doi.org/10.1007/978-3540-72079-910.

[6] U. Shardanand, P. Maes, "Social information filtering: Algorithms for automating word of mouth," in Proceedings of the SIGCHI Conference on Human factors in Computing Systems, 210–217, 1995.

[7] G. Adomavicius, A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, **17**(6), 734–749, 2005, doi:10.1109/TKDE.2005.99.

[8] K. Madadipouya, Sivananthan Chelliah, "A Literature Review on Recommender Systems Algorithms, Techniques and Evaluations," BRAIN. Broad Research in Artificial Intelligence and Neuroscience, **8**(2), 109, 2017.

[9] C.C. Aggarwal, Recommender Systems: The Textbook, Srpinger International Publishing, Switzerland, 2016.
[10] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich, Recommender Systems, Cambridge University Press, Cambridge, 2010, doi:10.1017/CBO9780511763113.
[11] I. Kubinyi, Designing an Automatic Recommendation System for a Web Based Social e-Learning Application, Helsinki Metropolia University of Applied Sciences, 2013.
[12] A.O. Ogunde, E. Ajibade, "Adewale Opeoluwa Ogunde, Emmanuel Ajibade. A K-Nearest Neighbour Algorithm-Based Recommender System for the Dynamic Selection of Elective Undergraduate Courses," International Journal of Data Science and Analysis, **5**(6), 128–135, 2019, doi:10.11648/j.ijdsa.20190506.14.
[13] A.O. Ogunde, J.O. Idialu, "A recommender system for selecting potential industrial training organizations," Engineering Reports, **1**(3), 2019, doi:10.1002/eng2.12046.
[14] R. Ghani, A.E. Fano, Building a Recommender System using a Knowledge Base of Product Semantics, 2000.
[15] W.W. Zhang, F.G. Liu, L. Jiang, D.. Zu, "Recommendation Based on Collaborative Filtering by Convolution Deep Learning Model Based on Label Weight Nearest Neighbor," in 10th International Symposium on Computational Intelligence and Design, 504–507, 2017.
[16] S. Gore, J. Sonawane, L. Duggineni, "A Product Recommendation System Based on Hybrid Approach," **3**(5), 1–5, 2017.
[17] A. Kumar, T. Sairam, "Machine Learning Approach for User Accounts Identification with Unwanted Information and data," International Journal of Machine Learning and Networked Collaborative Engineering, **2**(3), 119–127, 2018, doi:10.30991/ijmlnce.2018v02i03.004.
[18] Y. Cai, R. Lin, Y. Kang, A Personalized Company Recommender System for Job Seekers, 2015.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 1163-1170 (2020)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

An Efficient Performance of OFDM-Shaped Symbol for 5G Green Communication Compared to FBMC

Mohamed Yasin Ibrahim Afifi*, El-Sayed Soliman Ahmed Said, Abd El-Hady Abd El-Azim Ammar

Electrical Engineering Department, Faculty of Engineering, Al-Azhar University, 11651, Egypt

ARTICLE INFO

Article history:
Received: 09 October, 2020
Accepted: 26 November, 2020
Online: 14 December, 2020

Keywords:
5G
FBMC
OFDM
STC
OOB
PAPR
ISI
ICI

ABSTRACT

The filter bank multi-carrier (FBMC) technique is being considered as the major candidates for the 5G schemes. The orthogonal frequency division multiplexing (OFDM) technique is used in 4G but it has drawbacks such as higher Out-of-Band (OOB) power reduction, and higher peak to average power ratio (PAPR). OFDM uses cyclic prefix (CP) extension to overcome on inter symbol interference (ISI) problem but reduces the spectral efficiency (SE). Wireless networks often suffer from extreme mobility that results to inter-carrier interference (ICI) and degradation of the performance system. In a previous work symbol time compression (STC) - Shaped system was proposed for the OOB power reduction, PAPR reduction, ISI reduction and removes the extension of cyclic prefix (CP). This paper will modify the STC- Shaped system for completely ICI cancellation due to high mobility with keeping of OOB power reduction, PAPR and ISI reduction as well as CP removal. Comparative study between this proposed scheme and FBMC will be also presented. As the windowing concept is applied to mitigate OOB, PAPR, ISI and ICI problems, it is implemented on overall symbol of the proposed scheme. A significantly transmitted power saving of this scheme would be achieved to be about -4.3 dB compared to the OFDM transmitted power. This recommends the system to be in satisfactory agreements with the 5G green communication needs. The system performance measurements have been fulfilled using cumulative distribution function (CDF), power spectral density (PSD) and bit error rate (BER). In this paper, the mathematical model of the proposed modified STC-Shaped technique will be introduced. Finally, the analysis of ICI power will be covered in this paper.

1 Introduction

The OFDM is the key technique for 4G networks, and it can effectively transform a frequency-selective (FS) channel into multiple frequency-flat channels at different sub-carriers, allowing easy one-tap channel equalization to alleviate the multi-path effect [1, 2]. Nonetheless, OFDM also has drawbacks, such as large side lobes in spectrum, large PAPR, higher adjacent channel interference and higher power radiation OOB. While OFDM uses the CP to mitigate the ISI problem but reduces the overall spectral efficiency [3]–[6].

OFDM main drawback [7] is its sensitivity to frequency offsets, also known as carrier frequency offset (CFO). This problem can be divided into two situations: (i) the frequency mismatch in the oscillators of the transceiver, and, (ii) the Doppler Effect as the transmitter and the receiver move relatively. The CFO is classified into two types: the integer offset frequency (IFO) and the fractional offset frequency (FFO). IFO causes the sub-carriers to shift on the receiver side, which causes a major deterioration of the bit error rate (BER) output but does not damage the Orthogonality between components of the sub-carrier frequency; therefore ICI does not appear. With FFO, its presence will demolish the orthogonality among the sub-carriers [8]. Therefore, the time varying channel deteriorates the BER performance of OFDM systems. ICI mitigation schemes for high mobility OFDM systems can be classified into two categories.

The first category is the receiver ICI cancellation schemes [9]. The second category is ICI self-cancellation schemes [10]. The receiver ICI cancellation schemes require estimation the characteristics of time varying channel to mitigate the interference of each sub-carrier to adjacent sub-carriers (ICI). This category handles normalized Doppler frequency up to $\epsilon_d = 0.3$ but suffers from less accuracy, Large computational complexity due to large number of iterations and needs extra BW due to pilots. The ICI self-cancellation scheme is based on the difference between the ICI coefficients of two consecutive sub-carriers is very small. Thus, the main idea is to

*Corresponding Author: Mohamed Yasin Ibrahim Afifi, mohamedyasin869@azhar.edu.eg

modulate the input data symbol onto a group of sub-carriers with predefined coefficients such that the generated ICI signals within that group cancel each other hence the name self-cancellation. The ICI self-cancellation schemes are not spectral efficient because redundant data in the transmitter. It suffers from reduction 50% of the system throughput, and it mitigates ICI up to $\epsilon_d = 0.2$.

Therefore, a half-time OFDM-STC is needed. The STC scheme raises the spacing of the sub-carrier twice, thus reducing the Doppler frequency by 50 percent and then improving the efficiency of the BER compared with OFDM. STC model is applied by distributing input information streams and combining them. STC has been significantly overcoming both of ICI and CFO [11, 12]. During an OFDM symbol duration expose time varying channel in high-speed mobile environments, thus being orthogonality disrupted resulting in ICI and therefore, the system efficiency is diminished [13].

Many other non-orthogonal filtering-base waveform have currently been introduced to satisfy the 5G specifications. With regard to the filtering's granularity, these waveform candidates can be categorized into two kinds: sub band filtering and sub-carrier filtering. FBMC [14, 15] is a sub-carrier filtering waveform. FBMC, theoretically, the frequency localization C/C's resulting from the long-length filters's prototype will make CFO more robust; it is hard to design a prototype filters with low OOB to mitigate ISI and ICI. In this paper, the FBMC scheme [16] will be implemented and compared with the proposed system.

This paper's key contribution is to address some of the problems in wireless communications networks that adopt 5G technology based on STC scheme. STC model compresses symbol duration of OFDM into the half. This paper would leverage the 50% symbol interval compression obtained by using the STC strategy by extending the duration of the symbol without degradation of the system's throughput. In this paper, the STC-Shaped system will be modified for completely ICI cancellation due to high mobility with keeping of OOB power reduction, PAPR reduction, ISI reduction and CP removal [17].

The proposed modified STC-Shaped system is achieved by using symbol extension, symbol shaper in the transmitter side and combiner process in the receiver side. The symbol extension will repeat the compressed OFDM symbol data in the transmitter side. Therefore, the entire OFDM symbol is composed of N data coefficients from $\frac{-N}{2}$ to $\frac{N}{2} - 1$. The first half of the entire OFDM symbol includes data from the STC accompanied by its repeater. Next, the shaper of symbol will be applied using the suggested raised cosine function (RCF). Then, data is transmitted via linear time varying (LTV) channel. The useful information and its repeater with using a symbol shaper will ensure proper retrieval of data in receiver. The combiner is attained at the receiver side.

This Paper contributes the following

- The proposed system decreases OOB power radiation for traditional OFDM system which is very near to the needs of 5G such as FBMC.
- And though, CP is removed, in multi-path environment the scheme is more rigorous against ISI.
- PAPR is also reduced in this method as opposed to traditional OFDM and FBMC.
- The performance of the proposed scheme is investigated for BPSK modulation technique in DVB-T2 system under COST 207, typical rural fading channel environment. The performance measurements are the ICI power and bit errors rate (BER). It is demonstrated that, under normalized Doppler frequency, the proposed modified STC-Shaped scheme significantly improves the BER efficiency.
- Also, the proposed scheme decreases the ICI power compared to OFDM and FBMC. A significant power saving of the proposed scheme is aimed during the system development and implementation.

The substance of this paper is structured as follows: The FBMC transceiver is presented in section two. The OFDM system in high mobility was presented in Section three. The proposed modified STC-Shaped model has been covered in Section four. Analysis of ICI power is demonstrated in Section five. The obtained simulation results are shown in section six. The complexity of the proposed system is introduced in section seven. Finally, Section eight indicated the conclusion of the proposed system.

2 Overview of Filter Bank Multi-Carriers

In the QAM modulator, the stream from a binary source is converted to parallel data streams. IFFT algorithm is applied for conversion to time domain. The output stream of IFFT is suitable to the FIR filter Poly-phase Network (PPN) at an overlap factor K=4. This transmitter-side PPN filters are known as synthesis filters. These filters output are again converted into serial pattern for transmission. The mathematical basis for the design of the PHYDYAS project synthesis filters has been obtained in Eq.1. Fig.1 displays the FBMC transmitter block diagram [16].

$$G(F) = \sum_{l=-(k-1)}^{k-1} \left(G_l \frac{\sin\left(\pi \left[\frac{f-l}{m_k}\right] N_k\right)}{N_k \sin\left(\pi \left[\frac{f-l}{m_k}\right]\right)} \right) \tag{1}$$

where:

Coefficient	Value
G_0	1
$G_{(\pm 1)}$	0.97196
$G_{(\pm 2)}$	$\frac{\sqrt{2}}{2}$
$G_{(\pm 3)}$	0.235147

At the receiver side of Fig.1, $y(n)$ denotes the received signal of Eq.2 where $t(n)$ is the transmitted signal and C is the channel impulse response, $w(n)$ symbolizes the AWGN added to the signal.

$$r(n) = C * t(n) + w(n) \tag{2}$$

The signal is transmitted through the RF down conversion, where it has been divided into parallel streams. The PPN is the analytic filter in the receiver-side, and then Fast Fourier Transform converts the signal into frequency domain. And finally OQAM is

converted to QAM which leads to recovery of initial bits of data. The block diagram of the overall FBMC system is shown in figure 1.

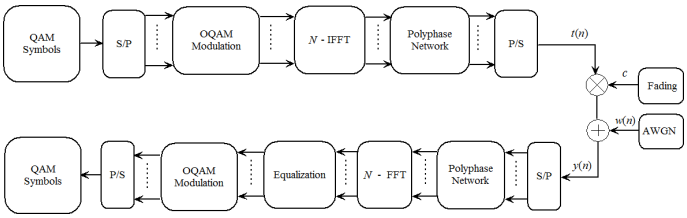


Figure 1: Block diagram of FBMC System[16].

3 Conventional OFDM Model Under Time Varying Channel

On the transmitter side, the following equation is described as a discrete OFDM symbol after IFFT [12, 13]:

$$x(n) = \frac{1}{N}\sum_{m=0}^{N-1}\left(X_m e^{\left(\frac{j2\pi nm}{N}\right)}\right) \quad -N_{cp} \le n \le N-1 \tag{3}$$

The nth received symbol of OFDM through multi-path time-varying channel after CP removal is as follows [12, 13]:

$$y(n) = \sum_{l=0}^{L-1} h_n^{(l)} X_{n-\tau^{(l)}} + W_n \quad 0 \le n \le N-1 \tag{4}$$

where:

Parameters	Symbolize
n	Index of time
x_n	Symbol of OFDM at the nth sample
N	Total number of useful sub-carriers in OFDM symbol
N_{cp}	The duration of CP
X_m	The complex transmitted data symbols on the mth sub-carriers ($m = 0, \ldots, N-1$)
L	The total number of channel paths
$h_n^{(l)}$	The lth path channel impulse response at the nth sample
$\tau^{(l)}$	Delay spread of the lth path
w_n	The additive white Gaussian noise (AWGN) with zero mean

The received OFDM symbol at the mth sub-carrier is written as follows [12, 13]:

$$\begin{aligned} Y_m &= \frac{1}{N}\sum_{l=0}^{L-1} y_n e^{\left(\frac{-j2\pi nm}{N}\right)} \\ &= \underbrace{H_{n.m}X_m}_{\text{desired signal}} + \underbrace{\sum_{m'=0}^{\frac{N}{2}-1}\sum_{m'\neq m} H_{m.m'}X_{m'}}_{\text{ICI terms}} + \underbrace{W_m}_{\text{AWGN}} \end{aligned} \tag{5}$$

where;

$H_{m.m'} = \frac{1}{N}\sum_{l=0}^{L-1} e^{-j2\pi\frac{m'}{N}\tau^{(l)}} \sum_{n=0}^{N-1} h_n^{(l)} e^{-j2\pi\frac{n}{N}(m-m')}$ does not vary during OFDM symbol duration, that is $0 \le n \le N-1$, thus there is no ICI terms. In contrast, if $H_{m,m'} \neq 0$ for $m' \neq m$ ICI occurs due to time variation of each complex path gain $h_n^{(l)}$, within one OFDM symbol $0 \le n \le N-1$. In the next section, the proposed modified STC Shaped scheme will be discussed and implemented.

4 The Proposed Modified STC-Shaped System

The proposed scheme is based on STC process, symbol extension and symbol shaping at the transmitter side, while only combining process at the receiver side. The desiring steps of the proposed method are followed;

In the transmitter side

- Input data streams are entered into STC stage.
- After STC process, the number of sub carriers decreased into half.
- Symbol extension process is proposed by repeating the compressed OFDM symbol in front of the useful part.
- The symbol shaper is suggested by using modified raised cosin function.
- Finally, the transmitted data has been sent to the LTV channel.

In the receiver side

- The received data are entered to the proposed combining stage for retrieving compressed OFDM symbol, which the extended part is combined to the corresponding information part.
- Finally, the STE stage is used to recovery the original input data streams.

The block diagram of the overall proposed system is shown in figure 2. After STC process, the number of sub-carriers becomes $N_c = \frac{N}{2}$ [17]. After IFFT, the nth compressed OFDM symbol is expressed by modifying Eq.4 as follows:

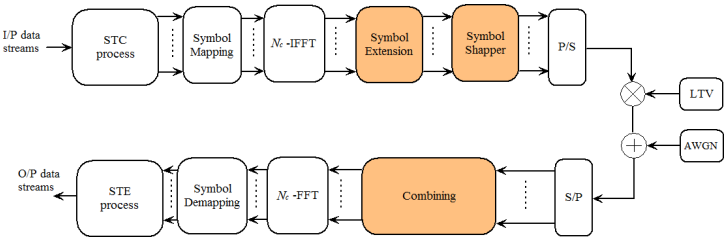


Figure 2: Block diagram of the proposed modified STC-Shaped System.

$$x_n^c = \frac{1}{N}\sum_{m=0}^{N_c-1} X_m^c e^{\left(\frac{j2\pi nm}{N}\right)} \quad 0 \le n \le N_c - 1 \tag{6}$$

Thus, after STC process, the transmitted signal as shown in figure 3

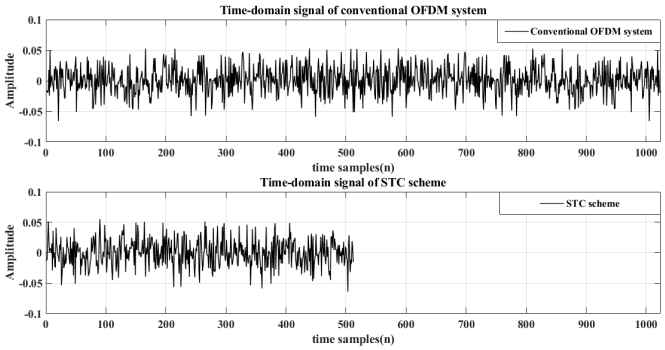


Figure 3: Time domain signal of the STC scheme and OFDM [17].

The symbol extension process is achieved by repeating the compressed OFDM symbol X_m^c in front of the useful part as shown in figure 4.

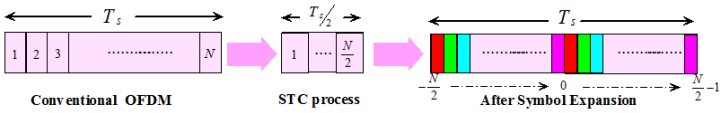


Figure 4: Symbol Expansion process in the transmitter for proposed system.

Thus, the time domain after symbol extension process becomes:

$$z_n = \frac{1}{N_c}\sum_{m=-N_c}^{N_c-1} X_m^c e^{\left(\frac{j2\pi nm}{N}\right)} \quad -N_c \le n \le N_c - 1 \tag{7}$$

The symbol shaper is achieved by using the raised cosine was proposed in [17], but with minor modification makes the raised cosine function (RCF) is symmetric around zero with keeping on the complementary concept for the negative part and positive part.

Thus, the modified symbol shaping is expressed as follows:

$$f_n = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi(n+N_c)}{N}\right)\right] \quad -N_c \le n \le N_c - 1 \tag{8}$$

The proposed modified RCF is showed in figure 5. After symbol shaper, the transmitted OFDM symbol becomes:

$$\begin{aligned} z'_n &= z_n * f_n \\ &= \frac{1}{2N_c}\sum_{m=-N_c}^{N_c-1} x_m^c - x_m^c \cos\left(\frac{2\pi(n+N_c)}{N}\right) e^{\left(\frac{j2\pi nm}{N}\right)} \\ &\quad -N_c \le n \le N_c - 1 \end{aligned} \tag{9}$$

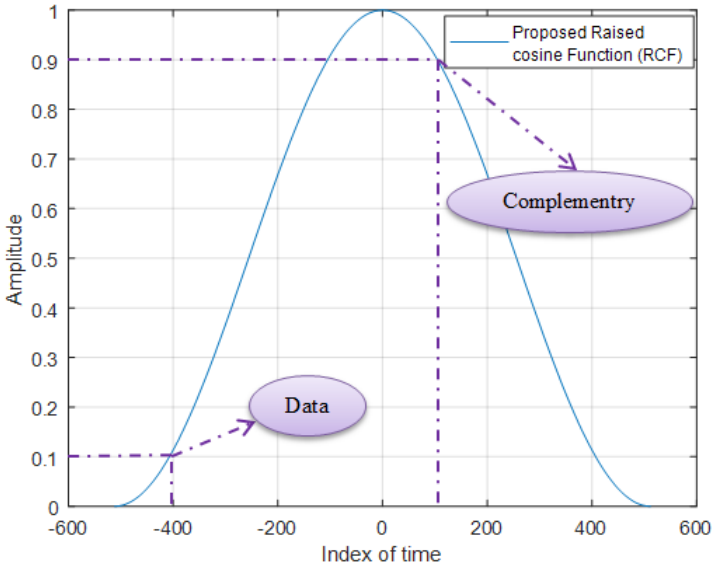


Figure 5: The proposed modified raised cosine function.

In this paper, the channel is assumed to be linear time varying channel (LTV) during OFDM symbol as follows:

$$h_n^{(l)} = 1 + j\left(\frac{\pi\epsilon_d n}{N}\right) \quad -N_c \le n \le N_c - 1 \tag{10}$$

$$\epsilon_d = \frac{d_f}{\Delta f} \tag{11}$$

Where:

Parameters	Symbolize
X_m^c	Compressed OFDM symbol at n^{th} sub-carriers
$h_n^{(l)}$	The lth path channel response at time instant n
ϵ_d	The normalized Doppler frequency
f_d	The maximum Doppler frequency
Δf	The sub-carrier spacing

Thus, the received signal over LTV multi-path channel is expressed as follows:

$$r_n = \sum_{l=0}^{L-1} h_n^{(l)} z'_{n-\tau^{(l)}} + W_n \quad -N_c \le n \le N_c - 1 \tag{12}$$

The proposed receiver for modified STC-Shaped system is achieved by only combining process. This stage is used to retrieve compressed symbol of OFDM in the receiver side. However, the extended part is combined to the corresponding information part. The combined time domain signal is derived as follows:

$$y_n = r_n + r_{n-N_c} \quad 0 \le n \le N_c - 1 \tag{13}$$

By substituting Eq.10 into Eq.11, the combined signal is expressed as follows:

$$\begin{aligned} r_n = &\sum_{l=0}^{L-1} h_n^{(l)} z'_{n-\tau^{(l)}} + \sum_{l=0}^{L-1} h_{n-N_c}^{(l)} z'_{n-N_c-\tau^{(l)}} \\ &+ W_n + W_{n-N_c} \quad -N_c \le n \le N_c - 1 \end{aligned} \tag{14}$$

Therefore, the equivalent channel time variation of the proposed modified STC-Shaped scheme is turned into flat with constant gain. As a result, the overall channel time variation is completely canceled without throughput loss.

5 ICI Power Analysis

In this section, the ICI power analysis of the proposed modified STC-Shaped scheme in high mobility will be discussed. The performance of the proposed modified STC-Shaped scheme is evaluated by computing the ICI power of the system. The ICI power of proposed modified STC-Shaped scheme is computed as follows:

$$\begin{aligned} p_{ICI} = &\frac{1}{N_c^2} \sum_{m'=0 m' \ne m}^{N_c-1} \sum_{l=0}^{L-1} p^{(l)} \sum_{n=0}^{N_c-1} \sum_{n'=0}^{N_c-1} E\left\{h_n', h_n'^*\right\} \\ &e^{\left(-j2\pi \frac{(n-n')(m-m')}{N_c}\right)} \end{aligned} \tag{15}$$

Where:

Parameters	Symbolize
$E\left\{h_n', h_n'^*\right\}$	The correlation function of Rayleigh fading channel of the lth path of the channel
$P^{(l)}$	The power of the lth channel path

The equivalent channel time variation after applying the proposed modified STC-Shaped scheme is expressed as follows:

$$h_n' = (f_n * h_n) + (f_{n-N_c} * h_{n-N_c}) \quad 0 \le n \le N_c - 1 \tag{16}$$

Thus, the time correlation function of the equivalent channel time variation in Eq.15 is expressed as follows [18] :

$$\begin{aligned} E\left\{h_n', h_n'^*\right\} = &\left(z_n' z_{n'}' + z_{n-N_c}' z_{n'-N_c}'\right) E\left\{h_n, h_{n'}^*\right\} \\ &+ z_n' z_{n'-N_c}' E\left\{h_n, h_{n'-N_c}^*\right\} \\ &+ z_{n-N_c}' z_{n'}' E\left\{h_{n-N_c}, h_{n'}^*\right\} \end{aligned} \tag{17}$$

where:

$$E\left\{h_n, h_{n'}^*\right\} = J_0\left(2\pi f_d T_s \frac{n-n'}{N_c}\right)$$

$$E\left\{h_n, h_{n'-N_c}^*\right\} = J_0\left(2\pi f_d T_s \frac{n-n'+N_c}{N_c}\right)$$

$$E\left\{h_{n-N_c}, h_{n'}^*\right\} = J_0\left(2\pi f_d T_s \frac{n-n'-N_c}{N_c}\right)$$

6 Simulation Results

In this paragraph, some numerical results for simulation are provided to analyze the performance of the proposed modified STC-Shaped model. The proposed modified STC-Shaped scheme is investigated in DVB-T2 OFDM system parameters in Table 1.

Table 1: Simulation system parameter [16, 17].

Parameters	Specification
OFDM sub-carriers N	1024
CP length N_{CP} in OFDM	256
CP length in proposed system	None
Modulation type	BPSK for (OFDM – STC)
	OQAM for (FBMC)
Overlapped factor K in FBMC	4
No of symbols	10000

The proposed modified STC-Shaped scheme is investigated under multi-path Rayleigh fading environment. The power delay profile of COST 207 rural area (RA) environment channel model which is summarized in Table 2.

Table 2: Power Delay Profile [10, 15].

	Delay ($\mu S\, ec$)	**Power**(dB)
tap 1	0	-3
tap 2	0.2	0
tap 3	0.5	-2
tap 4	1.6	-6
tap 5	2.3	-8
tap 6	5	-10

The proposed model is simulated by Matlab 2019 m-files software using intel Core i7-8550U CPU @1.99 GHz, 16.0 GB RAM, x64 Operating System for Windows 10 ultimate version. The results show the implemented model using Matlab M-file with respect to the theoretical results. The following paragraphs include demonstration ICI performance, BER, sidelobes suppression, and PAPR performance.

6.1 ICI Power Performance

Therefore, the time varying channel deteriorates the BER performance of OFDM systems. The proposed scheme will use proposed windowing technique for all symbol transmitted via channel, while OQAM-FBMC used a sub-carrier windowing waveform. Figure 6 is dedicated as illustrative graph for eight Consecutive OFDM sub-carriers in cases of without shaping and with shaping.

It is observed from figure 6 that, the symbol shaping reduces amplitude at sub-carrier index, so the ICI power will be reduced. figure

7 illustrates the ICI power of conventional OFDM, OQAM-FBMC and the proposed modified STC-Shaped schemes versus normalized Doppler frequency ϵ_d. figure 7 shows that the ICI power of the proposed modified STC-Shaped scheme is less than the OQAM-FBMC and conventional OFDM system for different normalized Doppler frequencies.

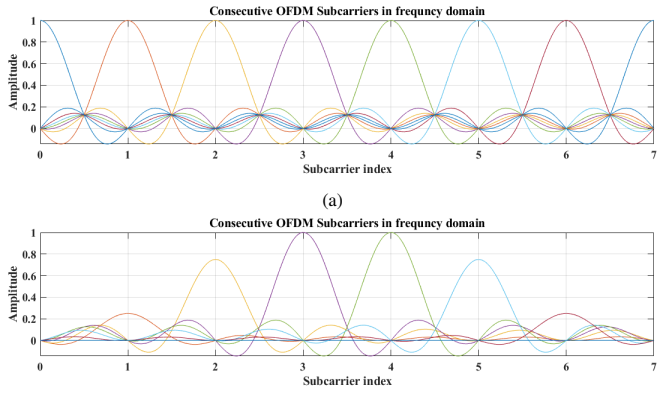


Figure 6: Consecutive OFDM subcarriers of (a) Without Shaping and (b) With Shaping.

At ϵ_d= 0.19 , the ICI power of the proposed modified STC-Shaped scheme is - 29.43 dB but the ICI power of OQAM-FBMC scheme is - 6.074 dB and the ICI power of conventional OFDM scheme is - 12.62 dB. Thus, the ICI power reduction using the proposed modified STC-Shaped scheme is 16.81 dB compared to OFDM and 23.36 compared to FBMC. As a result, the proposed scheme significantly reduces the ICI power compared to the OFDM and FBMC schemes. The simulated ICI power results of the proposed modified STC-Shaped scheme conform to the theoretical results.

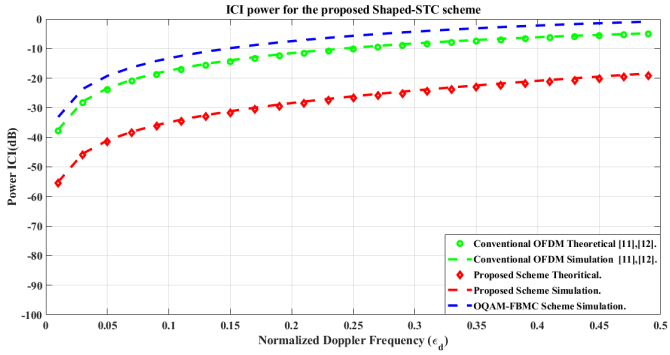


Figure 7: The ICI power of the proposed scheme versus ϵ_d.

6.2 Combined ISI & ICI Performance

This section evaluates the performance of OFDM system in time varying channel due to high mobility. The proposed modified STC-Shaped system is proposed for ICI reduction caused by high mobility. It is investigated in 2k mode DVB-T2 OFDM system parameters. It is evaluated under AWGN and multi-path Rayleigh fading channel for BPSK modulation scheme at 600 km/hr traveling mobile speed which corresponds to ϵ_d= 0.19. figure 8 and figure 9 show the BER performance of BPSK modulation scheme for the proposed modified STC-Shaped scheme compared to the conventional OFDM and OQAM-FBMC systems under AWGN and Rayleigh fading in high mobility respectively.

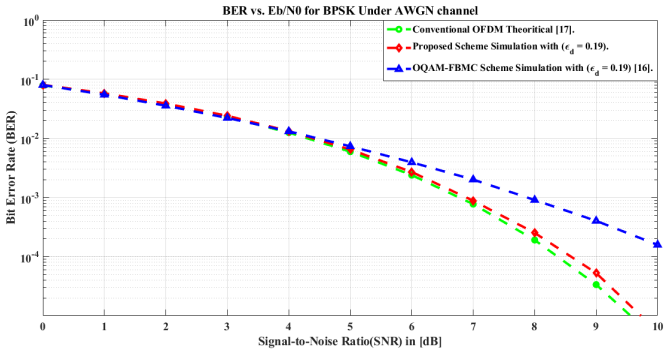


Figure 8: The BER under AWGN channel at ϵ_d=0.19.

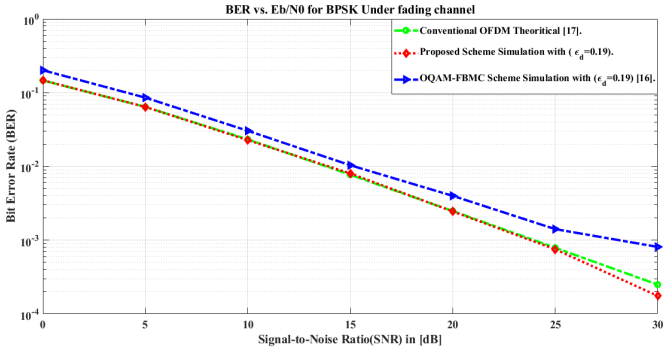


Figure 9: The BER performance under multipath fading channel at ϵ_d= 0.19 .

6.3 Side-lobe Suppression Performances

A comparison with the proposed system is discussed with the conventional OFDM, and FBM. figure 10a, figure 10b and figure 10c respectively depict the PSD for the traditional OFDM, the proposed modified STC-Shaped system and OQAM-FBMC.

6.4 PAPR Efficiency

The disadvantage of OFDM multi-carrier systems is a higher peaks of the coherently non-constant envelopes added [19]. The OFDM signal has addition of N sub-carriers. If N is large enough, a resulting signal is near a complex Gaussian model. This means that Gaussian is distributed for both its imaginary and real parts, and that its envelope and energy satisfy Rayleigh and exponential distributions respectively. PAPR is expressed as a discrete version of X[n] [20, 21].

$$PAPR(x[n]) = \max_{0 \le n \le N-1} \frac{|x[n]|^2}{E\left[|x[n]|^2\right]} \quad (18)$$

Figure 11 shows the calculated cumulative distribution function of proposed modified STC- Shaped system without CP. It has the

performance of PAPR which is referred to as 3dB better than the traditional OFDM and 4dB better than the OQAM-FBMC.

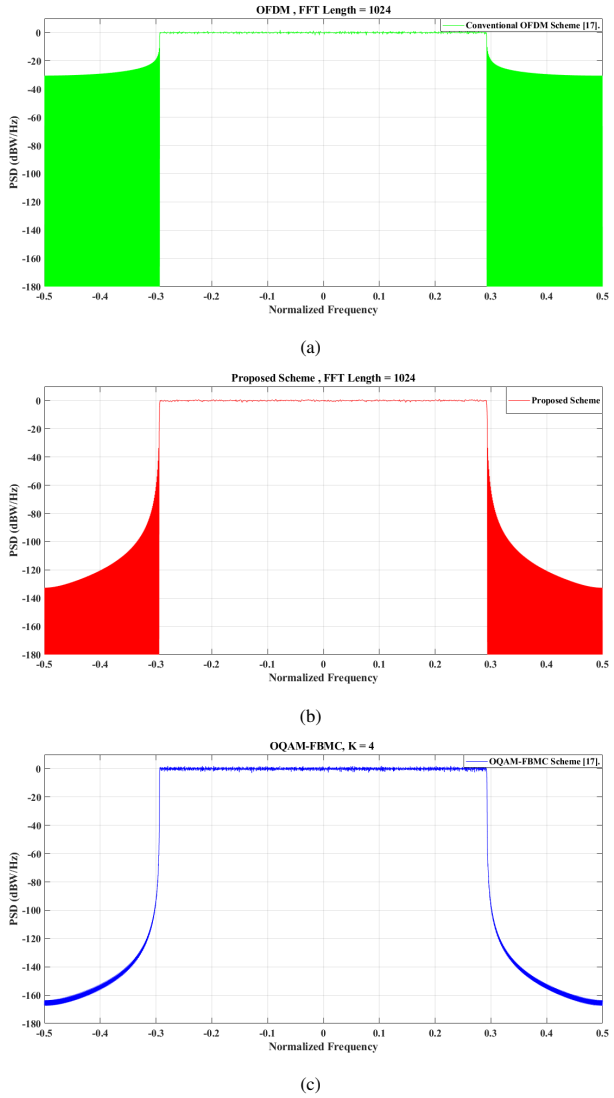


Figure 10: PSD of (a) traditional OFDM scheme[22], (b)proposed modified STC-Shaped system scheme and (c) FBMC scheme [23, 24].

6.5 *Power Saving*

The transmitted power for OFDM symbol after shaping can be estimated as follows:

$$\begin{aligned} P_T &= \frac{1}{N}\sum_{n=0}^{N}|h(n)|^2 .\hat{P} \\ &= \frac{\hat{P}}{N}\sum_{n=0}^{N}\left|\frac{1}{2}\left(1-\cos(\frac{2\pi n}{N})\right)\right|^2 \\ &= \frac{3}{8}\hat{P} \end{aligned} \tag{19}$$

where: P_T is the transmitted average power after shaping, $h(n)$ is the filter response and $\hat{P}$ is the power for symbol without shaping. The factor of $|h(n)|^2$ can be calculated from Eq.8 which is estimated by $\frac{3}{8}$. So, the transmitted after shaping is reduced by $\frac{3}{8}$ value of the power before. Thus, the proposed scheme would be saved to around -4.3 dB of the conventional OFDM transmitted power. Power amplifier rating could therefore be reduced by the estimated factor (-4.3 dB).

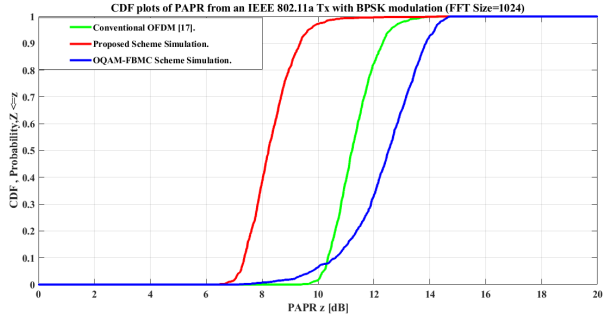


Figure 11: PAPR for OFDM, FBMC and proposed modified STC-Shaped systems.

7 Computational Complexity Analysis

For this paragraph, it is possible to obtain the computational complexity of the proposed modified STC-Shaped system in terms of number of multiplications as well as the number of additions compared to traditional OFDM, and OQAM-FBMC schemes as shown in Table 3.

Table 3: Comparative analysis of both the number of multiplication as well as addition processes.

Method	number of multiplications	number of additions
Traditional OFDM [18]	$2Nlog_2(N) - 2N$	$3Nlog_2(N) - N$
FBMC [25]	$2kN^2 + 3N$	$2N(NK-1)+3N+N-1$
Proposed modified STC-Shaped scheme	$2Nlog_2(N)$	$\frac{3}{2}Nlog_2(N) - \frac{N}{2}$

Finally, for the suggested scheme, the computational complexity is proportional to $Nlog_2(N)$, while FBMC has been used in 5G proportional to N^2. The proposed modified STC-Shaped model can thus fulfill very low computational complexity 5G requirements compared to FBMC requirements.

8 Conclusion

The optimal performance of mobile 5G green communications has been introduced and satisfactory implemented. An efficient performance STC based on time symbol modification has been suggested to increase the utilization factor of the band width as well as the reduction in band capacity. The shaped-STC will be adapted to an efficient STC-Shaped scheme for completely ICI cancellation due to high mobility with keeping of OOB power reduction in addition to PAPR, ISI reduction and CP elimination. The simulation

results compare the proposed scheme performance with OFDM and OQAM-FBMC.

The proposed modified STC-Shaped scheme is investigated under AWGN and multi-path Rayleigh fading channel for BPSK modulation scheme at ϵ_d=0.19. The proposed STC scheme has been achieved improvement on the BER performance compared to the conventional OFDM and OQAM-FBMC systems. It has been provided around -3 dB PAPR reductions about conventional OFDM, while for PAPR reduction -4 dB PAPR reductions compared to OQAM-FBMC. It had also offered a significant -80 dB lower than OFDM OOB radiation. The system results have been revealed a better performance throughout the computational complexity. The theoretical ICI power of the proposed STC scheme is computed. Finally this scheme has been saved the transmitted power to around -4.3 dB compared to the conventional OFDM transmitted power. The simulation results of the ICI power and BER performance were in a satisfactory agreement with the theoretical results.

The future work could cover latency, throughput, and higher order compression technique to meet the 5G requirements.

References

[1] A. Gupta, R. K. Jha, "A Survey of 5G Network: Architecture and Emerging Technologies," IEEE Access, **3**, 1206–1232, 2015, doi:10.1109/ACCESS.2015.2461602.

[2] Y. Tao, L. Liu, S. Liu, Z. Zhang, "A survey: Several technologies of non-orthogonal transmission for 5G," China communications, **12**(10), 1–15, 2015, doi:10.1109/CC.2015.7315054.

[3] K. Hussain, A. Lojo, R. López-Valcarce, "Flexible spectral precoding for sidelobe suppression in OFDM systems," in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4789–4793, IEEE, 2019, doi:10.1109/ICASSP.2019.8683162.

[4] Y. Mizutani, K. Mizutani, T. Matsumura, H. Harada, "A Low-pass Filtered Time-domain Window for DFTs-OFDM to Reduce Out-of-band Emission with Low Complexity," in 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC), 1–2, IEEE, 2019, doi:10.1109/CCNC.2019.8651816.

[5] S.-P. Lin, Y.-F. Chen, S.-M. Tseng, "Iterative smoothing filtering schemes by using clipping noise-assisted signals for PAPR reduction in OFDM-based carrier aggregation systems," IET Communications, **13**(6), 802–808, 2019, doi:10.1049/iet-com.2018.5421.

[6] S. Lu, D. Qu, Y. He, "Sliding window tone reservation technique for the peak-to-average power ratio reduction of FBMC-OQAM signals," IEEE Wireless Communications Letters, **1**(4), 268–271, 2012, doi:10.1109/WCL.2012.062512.120360.

[7] H. Nguyen, N. T. Le, T. L. Pham, Y. M. Jang, "A new method for mitigation of mobility effect of SUN MR-OFDM in Fast Fading channel," in 2019 International Conference on Information and Communication Technology Convergence (ICTC), 1422–1425, IEEE, 2019, doi:10.1109/ICTC46691.2019.8939879.

[8] M. Liu, J. Chen, B. Li, J. Li, "Fractional frequency offset estimation for OFDM systems in non-cooperative communication," China Communications, **13**(9), 65–71, 2016, doi:10.1109/CC.2016.7582298.

[9] L. Zhang, Z. Hong, L. Thibault, R. Boudreau, Y. Wu, "A low-complexity robust OFDM receiver for fast fading channels," IEEE transactions on broadcasting, **60**(2), 347–357, 2014, doi:10.1109/TBC.2014.2321681.

[10] K.-Y. Lin, H.-P. Lin, M.-C. Tseng, "An equivalent channel time variation mitigation schefme for ICI reduction in high-mobility OFDM systems," IEEE transactions on broadcasting, **58**(3), 472–479, 2012, doi:10.1109/TBC.2012.2197450.

[11] M. S. El-Bakry, H. A. El-Shenawy, A. E.-H. A. Ammar, "A symbol time compression for ICI reduction in high mobility OFDM systems," in 2017 29th International Conference on Microelectronics (ICM), 1–4, IEEE, 2017, doi:10.1109/ICM.2017.8268816.

[12] M. S. El-Bakry, H. A. El-Shenawy, A. E.-H. A. Ammar, "A time inversion and symbol time compression (TI-STC) scheme for ICI cancellation in high mobility OFDM systems," in 2017 Japan-Africa Conference on Electronics, Communications and Computers (JAC-ECC), 82–85, IEEE, 2017, doi:10.1109/JEC-ECC.2017.8305779.

[13] L. Yang, G. Ren, Z. Qiu, "A novel Doppler frequency offset estimation method for DVB-T system in HST environment," IEEE Transactions on Broadcasting, **58**(1), 139–143, 2011, doi:10.1109/TBC.2011.2170609.

[14] Q. Zheng, F. Wang, X. Chen, Y. Liu, D. Miao, Z. Zhao, "Comparison of 5G waveform candidates in high speed scenario," in 2017 XXXIInd General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS), 1–4, IEEE, 2017, doi:10.23919/URSIGASS.2017.8105377.

[15] A. Bazin, B. Jahan, M. Hélard, "Impact of the Doppler effect on the capacity of massive MIMO uplink systems: OFDM versus FBMC/OQAM," in 2017 24th International Conference on Telecommunications (ICT), 1–6, IEEE, 2017, doi:10.1109/ICT.2017.7998233.

[16] B. Ahmed, M. Zeeshan, "Effect of Carrier Frequency Offset on the Performance of FBMC and GFDM under Multipath Fading Channels," in 2018 21st International Symposium on Wireless Personal Multimedia Communications (WPMC), 240–245, IEEE, 2018, doi:10.1109/WPMC.2018.8712933.

[17] M. Y. I. Afifi, M. S. Elbakry, E.-S. S. A. Soliman, A. A. Ammar, "An efficient technique for out-of-band power reduction for the eliminated CP-STC-shaped system for 5G requirements," International Journal of Electrical and Computer Engineering, **10**(5), 5306, 2020, doi:10.11591/ijece.v10i5.pp5306-5315.

[18] F. Xiong, Digital Modulation Techniques, (Artech House Telecommunications Library), Artech House, Inc., 2006.

[19] P. Kundu, P. Kaur, "Comparison of peak to average power reduction techniques in OFDM," in 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 875–879, IEEE, 2014, doi:10.1109/ICACCI.2014.6968456.

[20] Y. Rahmatallah, S. Mohan, "Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy," IEEE communications surveys & tutorials, **15**(4), 1567–1592, 2013, doi:10.1109/SURV.2013.021313.00164.

[21] U. B. Mahadevaswamy, M. N. Geetha, "A comparative survey on PAPR reduction in OFDM signal," in 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 123–126, IEEE, 2016, doi:10.1109/ICEECCOT.2016.7955199.

[22] L. Pan, J. Ye, X. Yuan, "Spectral Precoding for Out-of-Band Power Reduction Under Condition Number Constraint in OFDM-Based Systems," Wireless Personal Communications, **95**(2), 1677–1691, 2017, doi:10.1007/s11277-016-3874-8.

[23] S. Nagul, "A review on 5G modulation schemes and their comparisons for future wireless communications," in 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES), 72–76, IEEE, 2018, doi:10.1109/SPACES.2018.8316319.

[24] S. Kundrapu, S. I. Dutt, N. K. Koilada, A. C. Raavi, "Characteristic Analysis of OFDM, FBMC and UFMC Modulation Schemes for Next Generation Wireless Communication Network Systems," in 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), 715–721, IEEE, 2019, doi:10.1109/ICECA.2019.8821991.

[25] A.-a. Husam, Z. Kollár, "Complexity Comparison of Filter Bank Multicarrier Transmitter Schemes," in 2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP), 1–4, IEEE, 2018, doi:10.1109/CSNDSP.2018.8471795.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1171-1177 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

PV Integrated Recursive Least Mean Square Estimation Based Shunt Active Power Filter

Ragam Rajagopal[*,1], K. Palanisamy[1], S. Paramasivam[2]

[1]*School of Electrical & Electronics Engineering, VIT Vellore, 632014, India*

[2]*Functional Head (R&D), Danfoss Drives, Chennai, 602105, India*

ARTICLE INFO

Article history:
Received: 27 August, 2020
Accepted: 19 November, 2020
Online: 14 December, 2020

Keywords:
Shunt Active Filter
Adaptive Linear Control
Harmonics
Power Quality

ABSTRACT

This paper proposes a photovoltaic (PV) integrated shunt active power filter (SAPF) for reactive power compensation, uneven loading and compensation of harmonic current. The control of SAPF is implemented by using adaptive based current estimation algorithm to maintain power factor near to unity and source current sinusoidal at source side. The weights are calculated online by recursive least mean square algorithm once the three-phase load current is detected. Secondly, the unit vector templates (UVT) are multiplied with these weights and thus the load current with fundamental active component is derived. PV system enhanced by Maximum Power Point Tracking (MPPT) plays the role of maintaining the dc link voltage constant. Finally, a current control scheme based on hysteresis is used for converter switching and thus the source current to follow reference current precisely. The superiority of the proposed method of compensation of harmonic current and reactive power using SAPF with Recursive Least Mean Square (RLS) algorithm is proved by the simulation results based on MATLAB/SIMULINK.

1. Introduction

The pervasive use of converters based on power electronics has given rise to increasing use of non-linear and inductive loads. The adjacent loads are negatively affected by the voltage variation at point of common coupling (PCC) caused by the harmonic current component of non-linear loads linked to the same network [1-4]. The increased demand of reactive power diminishes the flow capability of active power, and increases feeder losses, whereas the ordinary operation of transformer and generator are adversely affected by the unbalanced system. IEC-61000 and IEEE-519-2014 standards are put into practice to minimize such problems. The introduction of custom power device technology such as unified power quality controller (UPQC), dynamic voltage regulator (DVR), distributed static compensator (D-STATCOM) are effective for the electrical distribution network that faces these issues (unbalanced load, reactive power and harmonics). It is found that SAPF is suitable to efficiently compensate the afore mentioned issues [5].

The control method used to determine the efficacy of the SAPF presented. It mainly relies on the fundamental current extraction algorithm and converter switching generation using extracted reference currents.

The concept here to is to separate load current into three constituent factors. They are fundamental active (i_{Lfp}), reactive(i_{Lfq}) and harmonic(i_{Lh}) components of load current. References [6-9] have reported the low pass and band pass filter (LPF-BPF) based fundamental active current extraction scheme, per phase computation, scheme based on neural network, instantaneous reactive power (IRP) theory, rotating reference frame and instantaneous symmetrical components. Artificial neural network-based extraction scheme requires training of load current datasets for different load conditions which makes system complex and suitable for very specific loads. IRP theory works very well under sinusoidal supply voltage but fails to separate out working and wattless components of the load current under distorted supply condition.

Adaptive filtering techniques are proposed in latest literature because of its merits such as robustness, instinctive monitoring and tracing, less real-time calculation and a fast-dynamic response even under complex load conditions [10-11]. References [12-16] propose various adaptive filtering techniques such as adaptive notch filter, improved sinusoidal notch filter, fuzzy based adaptive control and synchronous adaptive filters. Reference [17, 18] presents least mean square (LMS) based fundamental active current extractor scheme. Least mean squares (LMS) algorithms has the least difficulty and most effectively applied versatile

[*]Corresponding Author: Ragam Rajagopal, ragamrajagopal@gmail.com

algorithms. The disadvantage of using LMS algorithm is its slow convergence rate. Excellent performance and prominent fidelity make the recursive least square algorithm (RLS) superior, however they accompany expanded intricacy and computational expense. In adaptive filtering applications, RLS approaches Kalman filter with reduced computational complexity in the signal processor. At the cost of requiring more computations, the RLS approach has quicker convergence characteristics compared to LMS algorithm and less error with respect to the unknown system. Reference [19] have reported PV integrated system-based voltage source converter (VSC) to keep a constant voltage at dc link. The development of PV technology has given motivation to the concept of combining clean and free solar energy for SAPF.

This paper proposes a PV integrated 3-phase shunt active filter with recursive least mean square based extractor scheme for fundamental active current extraction by recursively updating the weight of the input signal. The recursive approach finds the filter coefficients that minimize a weighted linear least squares cost function relating to the input signals. Minimization of the total weighted squared error between the desired signal and the output is the main objective of this approach.

It also uses variable forgetting factor to get fast merging rate and less steady state error. The usage of forgetting factor makes the de-emphasis of previous old data as compared to the new data. Moreover, PV system is integrated to keep the voltage of dc link constant. The performance analysis of RLS based SAPF reveals IEEE standard 519 compliant functionality under various operating conditions.

2. System Configuration

The Shunt active filter which act as voltage source converter is coupled to PCC through interfacing inductor L_f. The key functions of the SAPF include a) the elimination of harmonic component of load current thereby maintaining sinusoidal nature of source current with THD<5% b) maintaining source current in phase with pcc voltages achieving near unity pf and c) unbalanced load current compensation by maintaining source currents balanced irrespective of load conditions. The implementation of SAPF requires accurate connection of interlinking inductors, capacitor in the dc link, current and voltage sensors. The control algorithm must be fast enough for the successful functioning of the SAPF. The adaptive RLS algorithm used in this work has higher convergence rate as compared to other variants of least mean square algorithms. Figure 1 indicates the typical illustrative figure of 3-phase SAPF. Notations used in Figure 1 are described in section 2.1.

Three phase (i.e. a, b and c) load currents are determined by means of current sensors. Similarly, source currents, i_{sa}-i_{sb}-i_{sc} are measured. Three voltage sensors are used to measured three phase voltages v_a-v_b-v_c at PCC. PLL will determine θ, based on PCC voltages which in turn is used to generate UVTs for the three phase system. Fundamental active current extractor which makes use of RLS based adaptive algorithm as illustrated in figure 2 utilize this θ, to determine i_{Lfp}.

The product of UVT and i_{Lfp} results in the production of the reference source currents, i_{saR}-i_{sbR}-i_{scR}, [19]. These reference source currents along with the instantaneous values of the three phase source currents are fed into the hysteresis controller for the production of gate signals. These gate signals are in turn fed to voltage source inverter (VSI) for the generation of compensating currents which are injected into the system. DC link voltage is regulated by means of the PV array. Economic reliability is enhanced since the VA rating of photovoltaic panel used need not be very high as required for the injection of active power since the compensating currents injected need to provide only for harmonic currents and reactive power.

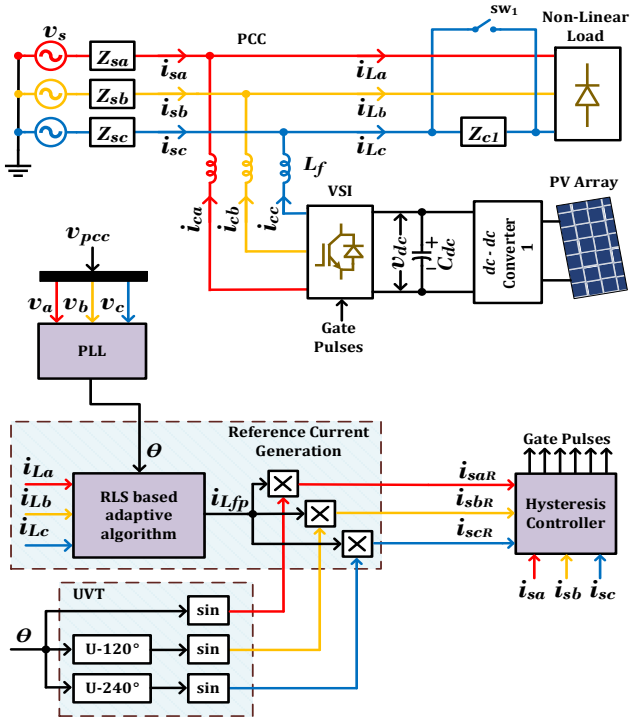


Figure 1: Block diagram of 3-phase SAPF

2.1. Nomenclature

a,b,c : Three Phases

v_{sa}-v_{sb}-v_{sc} : Supply voltage

v_a-v_b-v_c : PCC voltages

θ : instantaneous phase angle of fundamental component of supply voltage

Z_{sa}-Z_{sb}-Z_{sc} : Line impedances

i_{Lh},i_{Lfp},I_{Lfq}, : Harmonic load current component, Fundamental active and reactive components of load current

i_{La}-i_{Lb}-i_{Lc} : Load currents

u_{pa},u_{pb},u_{pc} : Unit vector template

i_{ca}-i_{cb}-i_{cc} : Compensating currents

i_{saR}-i_{sbR}-i_{scR} : Reference source currents

v_{dc} : DC link voltage

C_{dc} : Capacitor at dc link

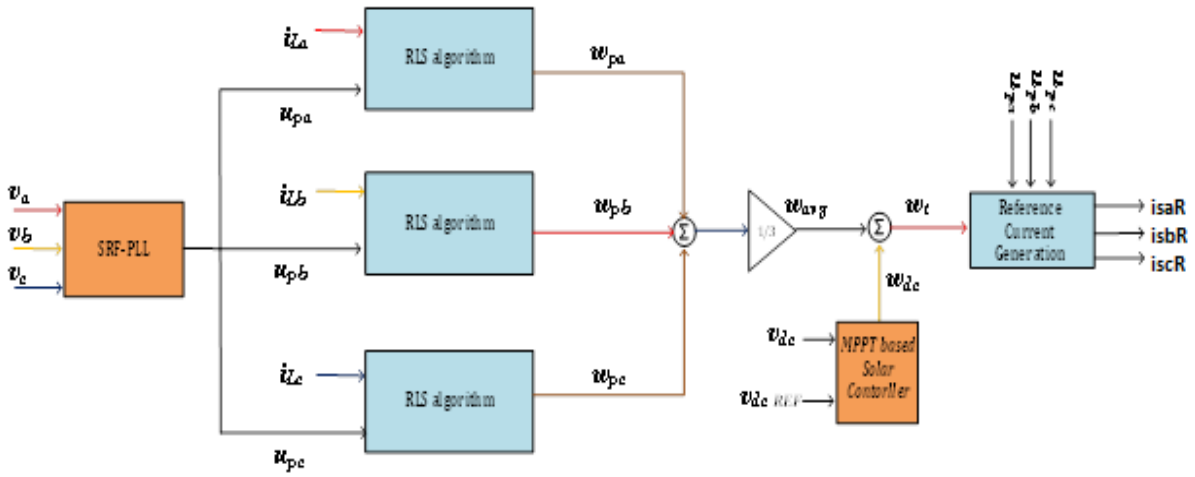


Figure 2: Control diagram for 3-phase SAPF

i_{Lp} : Fundamental active load current component's peak value

i_{Lq} : Fundamental reactive load current component's peak value

i_{Lm5}, i_{Lm7} : Peak value of 5th and 7th harmonic components of load current

i_{c1} : Fundamental compensating current component's peak value

i_{cm5}, i_{cm7} : Peak value of 5th and 7th harmonic components of compensating current

i_{sm1} : Fundamental source current component's peak value

i_{sm5}, i_{sm7} : Peak value of 5th and 7th harmonic components of source current

3. Control Scheme

The LS estimate of the filter coefficient w(n-1) at iteration n-1 is the basic technique behind RLS algorithm. This type of algorithm is known as recursive least mean square algorithm. This algorithm may be viewed as a special case of Kalman filter. Recursive filter is implemented by using the available starting conditions and then update the weight based on new arrived data. This leads to the minimization of cost function, where observed data's variable length is notated as n.

Cost function $J(n) = \sum_{k=1}^{n} w_n e^2(n)$ where w_n is weighting factor and e(n) is error.

This error e(n) can be expressed as,

$$e_j(n) = i_{Lj}(n) - i_{sj}(n) \quad (1)$$

where, j=a, b and c phase

$$e_j(n) = i_{Lj}(n) - w^T(n-1)x(n-1) \quad (2)$$

Weight updating equation for RLS control is given as,

$$g(n) = P(n-1)x(n)[\lambda + x^T(n)P(n-1)x(n)]^{-1} \quad (3)$$

$$P(n) = \lambda^{-1}P(n-1) - g(n)x^T(n)\lambda^{-1}P(n-1) \quad (4)$$

$$w(n) = w(n-1) + e(n)g(n) \quad (5)$$

where P(n) and g(n) represent the inverse input correlation matrix and Kalman gain vector. For algorithm learning, the term λ which is the forgetting factor plays a very major role, whose value is between 0 and 1.

The value of forgetting factor is closely related with convergence rate and tracking capability. In fixed forgetting factor based RLS algorithm, as λ increases and its value moves closer to 1, convergence rate increases and at the same time tracking speed will reduce. To improve the tracking capability of the algorithm λ value must be small which will diminish the convergence rate. These conflicting requirements are met by the usage of a variable forgetting factor RLS algorithm.

3.1. Generation of switching pulses for the operation of SAPF

Based on the three phase supply voltages, the control scheme shown in figure 2 will determine θ. This θ generates the three unit vector templates for the system. This UVT along with three phase load currents are the inputs to the control algorithm. The algorithm will generate weights corresponding to the three phase components of active power.

The mean weight magnitude equivalent to the components of active power (w_{pa}, w_{pb} and w_{pc}) from figure 2 is obtained by,

$$w_{avg} = \frac{w_{pa} + w_{pb} + w_{pc}}{3} \quad (6)$$

The dc bus voltage is sensed with the aid of a voltage sensor. This in turn is compared with the set point voltage of the dc bus (v_{dc}), and leads to the generation of the error signal (v_e). The MPPT based PI controller's tuned output (w_{dc}) ensures the steady maintenance of the dc bus voltage at its desired value.

The output of the PI controller at nth sampling instant is indicated below,

$$w_{dc}(n) = w_{dc}(n-1) + k_p(v_e(n) - v_e(n-1)) + k_i\, v_e(n) \quad (7)$$

where the gains of PI controller are notated as k_p and k_i

The sum of mean weight (w_{avg}) and PI controller's tuned output (w_{dc}) is utilized to find the active component of load current's weight.

$$w_t = w_{avg} + w_{dc} \quad (8)$$

UVT is used to calculate the reference source currents and denoted in (8) can be represented as,

$$i_{saR} = w_t u_{pa}, i_{sbR} = w_t u_{pb}, i_{scR} = w_t u_{pc} \quad (9)$$

The computation of VSC compensating current is carried out by subtracting the reference source currents (9) from load currents.

4. Performance Validation

MATLAB/SIMULINK platform is used to manifest the performance of the RLS based control algorithm. For various load conditions, validity of the proposed scheme has been checked as mentioned in the next section. The simulation parameters are detailed in Table 1. The SAPF comes into play at t=0.1 sec and the results of the three phase SAPF are as displayed in Figure 3.

Case 1: Steady state performance for Load-I

As shown in the Figure 3, from t=0.1s to 0.2s, undistorted v_{PCC}, supplies Load–I which has a Total Harmonic Distortion (THD) of 1.53% and RMS value of 230V. THD and max. value of load current is 27.16% and 1.84 A respectively.

Current harmonics is controlled by forcing peak value of fifth and seventh harmonic of compensating current to take a value as shown in Table 2. This will lead to a reduction in THD to 3.05%. Hence peak value of fifth and seventh harmonic of source current are found to be negligible. By ensuring source side pf at unity, enforcement of reactive power compensation is done by controlling the max. value of the compensating current's fundamental component to be at 0.17A. Allowing for a p-p ripple of 0.2 V, voltage at dc link, v_{dc} is maintained at a mean value of 220.2 V.

Table 1: System Parameters

System Parameters	Value
v_{pcc}(undistorted)/phase	230V, 50Hz
Load – I:	3-phase rectifier having load : (140+j31.42) Ω ;S_w closed; balanced load
Load – II:	3-phase inductive load having (90+j31.42) Ω ;S_w closed; balanced load
Load – III	3-phase rectifier having (90+j31.42)Ω ;S_w open; unbalanced load
Unbalanced load, Z_{c1}	(10+j3.14) Ω
Source impedance, Z_s	(0.02+j0.0314) Ω
DC link voltage, V_{dc}	220 V
Interfacing Inductor, L_f	4mH

Case 2: Dynamic performance under step change of load

Figure 3 manifests step load change from I to II at 0.2 sec, which demonstrates the SAPF performance under the condition of dynamic load change with the RLS control algorithm. From figure 3, peak value of source current under the transient condition is 2.02A. Values of all parameters in the new steady state condition with Load II is as indicated in Table 2. Compensation ensures that THD becomes well within the limit and also pf being improved to unity.

Under the sudden step load changing condition, voltage at dc link, v_{dc} droops to 215.8 V. The action of PV integration along with the MPPT control algorithm make sure that within 0.07 sec, voltage at dc link is restored to its set point value. The presented algorithm displays a satisfying performance by efficient tracking of the fundamental load current component, i_{Lfp} change from 1.28A to 1.90A.

Case 3: Steady state performance for unbalanced loading condition

Figure 3 shows the steady state unbalanced loading condition performance of 3-phase SAPF operating with RLS control scheme. Switch S_w is open at t=0.3sec to introduce the unbalanced loading scenario. The values related to unbalanced load currents and compensating currents for each phase are as indicated in Table 3. Unbalanced loading results in unbalanced compensating currents for the three phases. Each phase caters to the harmonic current and reactive power requirement set forth by the load. Once the load currents under unbalance are compensated, the source currents will automatically become balanced. The THD and max. magnitude for the source's currents i_{sa}, i_{sb} and i_{sc} are 3.21% and 1.61 A. The voltage at dc link, v_{dc} with p-p ripple of 0.21 V is sustained at a mean value of 221V. Also, the source power factor is kept at unity. This set up ensures unbalanced load compensation by the 3 phase SAPF with the presented control algorithm.

The results for the three cases are summarized in Table 2. From the table, it can be clearly understood that $i_{cm5} \cong i_{Lm5}$ and $i_{cm7} \cong i_{Lm7}$. This reveals harmonic current compensation with source current THD maintained well below the 5% limit. Unity power factor is well kept at supply end via reactive power compensation which is evident from the fact that $i_{cm1} \cong i_{Lq}$. Also unbalanced load current compensation is achieved by the algorithm which ensures that source current is balanced even under unbalanced load conditions. Thus the presented control scheme of 3 phase SAPF based on adaptive RLS algorithm is providing the necessary compensating currents, reactive power and unbalanced load compensation which leads to THD kept within 5% limit and unity power factor (UPF) at supply end.

From table 3, it is clear that convergence time, % steady state error and % THD is less for RLS based algorithm compared to LMS algorithms. Computational complexity is high for RLS algorithm since it considers the estimate of previous samples of output signal, error signal and filter weight. To include the effect of losses, distortion power should be taken into account because of the harmonic inducing nature non- linear loads connected to grid side. Thus the product of displacement power factor and distortion power factor gives the true power factor.

Table 2: Performance of the SAPF with the proposed control algorithm

Case	Category of load	v_{pcc}	i_L (A)				i_c (A)			i_s (A)			%THD	
			I_{Lp}	I_{Lq}	I_{Lm5}	I_{Lm7}	I_{cm1}	I_{cm5}	I_{cm7}	I_{sm1}	I_{sm5}	I_{sm7}	i_L	i_s
1	Load – I (balanced)	Undistorted (THD=1.53%)	1.22	0.14	0.25	0.13	0.16	0.24	0.13	1.28	0.01	0.01	27.16	3.05
2	Load – II (balanced)	Undistorted (THD=1.53%)	1.93	0.31	0.41	0.27	0.33	0.41	0.26	1.90	0.0	0.0	23.20	2.30
3	Load – III (unbalanced)	Undistorted (THD=26.5%)	1.78	0.21	0.28	0.34	0.25	0.27	0.34	1.61	0.01	0.01	31.6	3.21
			1.33	0.31	0.37	0.36	0.35	0.37	0.36	1.61	0.01	0.01	34.9	3.21
			0.45	0.2	0.13	0.31	0.40	0.12	0.31	1.61	0.01	0.01	34.8	3.21

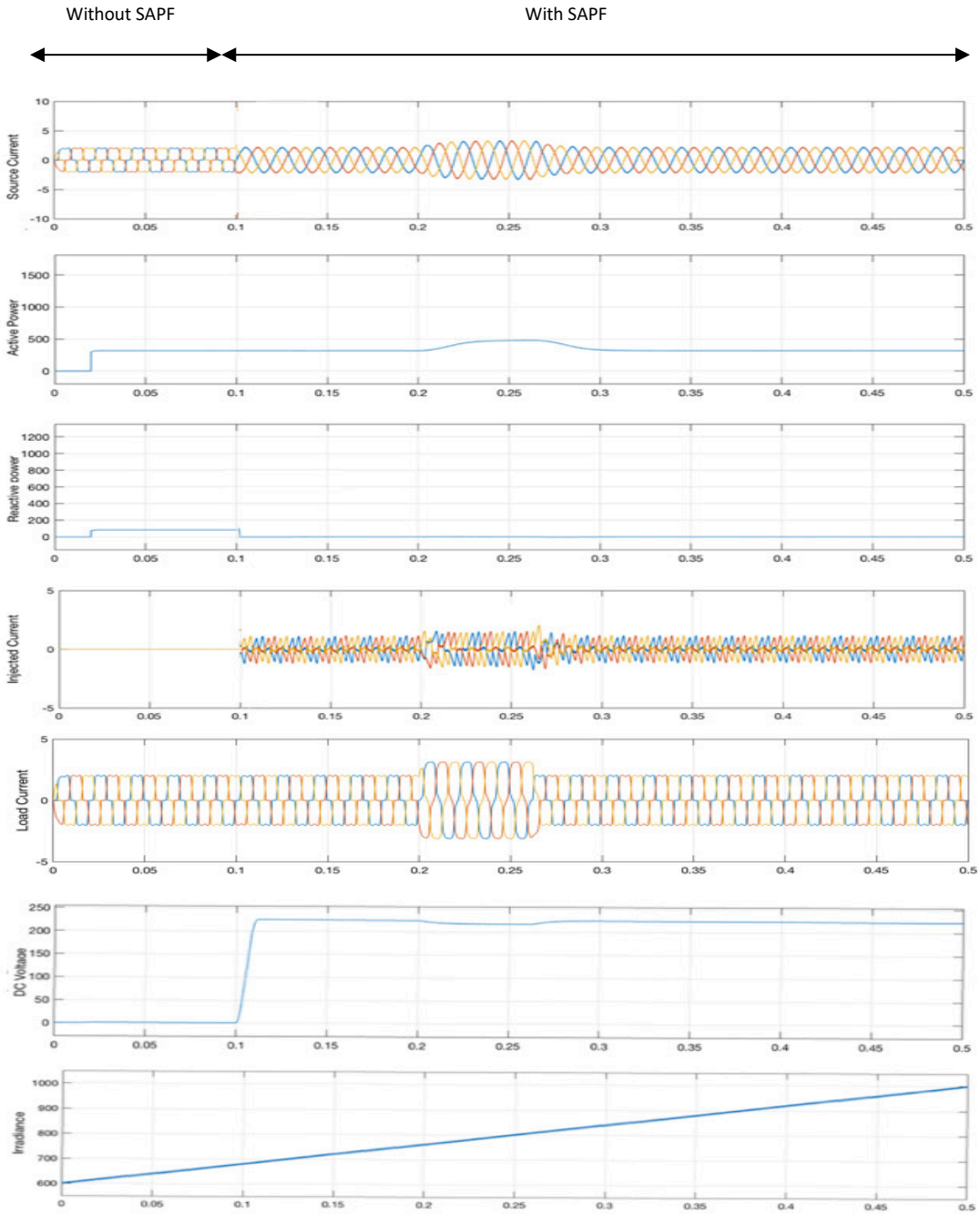


Figure 3: Simulation results for operation of the three phase SAPF with proposed algorithm

Table 3: Performance comparison of RLS algorithm with LMS algorithm [18]

Function	NLMS	VSSLMS	RLS
Convergence time	0.3 sec	0.12 sec	0.06 sec
Steady state error	2.11	0.85	0.56
THD	4.32	2.13	1.48
Complexity	Less	Less	High

Table 4: Power factors under varying load conditions

Case	Load	Distortion Factor	Displacement factor	True Power Factor
1	Load – I (balanced)	0.9995	0.9998	0.999
2	Load – II (balanced)	0.9997	0.9997	0.999
3	Load – III (unbalanced)	0.9995	0.9998	0.999

The equations (10-12) implies that the overall power factor gets diminished by the term cos γ. This leads to the conclusion that the average power transferred to the load gets diminished by cos γ, which depends on harmonic distortion induced in load current by the non-linear loads.

$$DF = \cos\gamma = \frac{I_{1(rms)}}{I_{(rms)}} \tag{10}$$

$$DPF = \cos\phi_1 = \frac{P_1}{S_1} \tag{11}$$

$$\begin{aligned} Power\ \ factor &= \frac{P}{S} = \frac{VI_1\cos\phi_1}{VI} \\ &= \left(\frac{I_1}{I}\right)\cos\phi_1 = \cos\gamma \times \cos\phi_1 \end{aligned} \tag{12}$$

The proposed SAPF configuration, maintains displacement power factor at 0.999 and the distortion power factor at near unity, for all the analysed load conditions. This leads to a magnitude of 0.999 for the true power factor. Table 3 manifests the true power factor under varying load conditions considered here.

5. Conclusion

This paper illustrates the performance aspects of PV powered RLS based shunt active power filter control utilizing the MATLAB/SIMULINK platform. Fundamental active components reference from the load currents are separated by the control law, based on the load currents, phase angle and reference currents of the previous sample. Load current's fundamental active component is only drawn from the source side and this is performed by the control system by ensuring that source currents track the reference currents with minimum steady state error. This makes the SAPF satisfy the reactive and harmonic power requisite of the considered non- linear load. Even under unbalanced loading condition, the source currents are in perfect balance. The put forward theory is validated by the results which shows that the source currents are drawn at UPF and they comply with the IEEE standard 519. Also, it is shown that dc voltage is maintained constant using PV integration into the network. With RLS algorithm, the 3-phase SAPF ensures sinusoidal and balanced supply currents which are maintained at UPF at various load conditions along with THD maintained under the 5% limit.

References

[1] B. Singh, A. Chandra and K. Al-Haddad, "Power Quality: Problems and Mitigation Techniques", John Wiley and Sons, 2015.

[2] R. Rajagopal, K. Palanisamy, S. Paramasivam, "A Technical review on control strategies of active power filters", IEEE 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR), TIST, Cochin. 1-6, 2018, doi: 10.1109/ICETIETR.2018.8529008.

[3] K.S. Rajesh, S.S. Dash, Ragam Rajagopal, R.Sreedhar "A review on control of ac microgrid", Renewable & Sustainable Energy Reviews, **71**, 814-819, 2017, doi: 10.1016/j.rser.2016.12.106.

[4] K.S. Rajesh, S.S. Dash, Ragam Rajagopal, "Hybrid improved firefly-pattern search optimized fuzzy aided PID controller for automatic generation control of power systems with multi-type generations", Swarm and Evolutionary Computation, **44**, 200-211, 2019, doi:10.1016/j.swevo.2018.03.005.

[5] S. R. Arya and B. Singh, "Performance of DSTATCOM using leaky lms control algorithm", IEEE J. Emerging Selected Topics in Power Electronics, **1**(2), 104-113, 2013, doi: 10.1109/JESTPE.2013.2266372.

[6] B. Singh, K. Kant, and S. R. Arya, "Notch filter-based fundamental frequency component extraction to control distribution static compensator for mitigating current-related power quality problems," IET Power Electronics, **8**(9), 1758-1766, 2015., doi: 10.1049/iet-pel.2014.0486.

[7] B. N. Singh, B. Singh, A. Chandra, and K. Al-Haddad, "Design and digital implementation of active filter with power balance theory," IEE Proceedings-Electric Power Applications, **52**(5), 1149–1160, 2005, doi: 10.1049/ip-epa:20050097.

[8] D. Chen, "A Novel Cost-Effective Two-Level Inverter with Combined Use of Thyristors and IGBTs", Journal of Electrical Engineering and Technology, **13**(1), 152-159, 2018, doi:10.5370/JEET.2018.13.1.152

[9] K. Karaarslan, B Arifoglu, E Beser, S Camur, "Level Number Effect on Performance of a Novel Series Active Power Filter Based on Multilevel Inverter", Journal of Electrical Engineering and Technology, **13**(2), 711-721,2018, doi: 10.5370/JEET.2018.13.2.711.

[10] B. Singh , J. Solanki, "An Implementation of an Adaptive Control Algorithm for a Three-Phase Shunt Active Filter", IEEE Transactions on Industrial Electronics, **56**(8), 2811-2820, 2009, doi: 10.1109/TIE.2009.2014367.

[11] M. Qasim , Parag Kanjiya , Vinod Khadkikar, "Optimal Current Harmonic Extractor Based on Unified ADALINEs for Shunt Active Power Filters", IEEE Transactions on Power Electronics, **29**(12), 6383-6393, 2014, doi: 10.1109/TPEL.2014.2302539.

[12] M. Cirrincione, M. Pucci, G. Vitale, "A single-phase DG generation unit with shunt active power filter capability by adaptive neural Filtering," IEEE Trans. Ind. Electron., **55**(5), 2093- 2110, 2008, doi: 10.1109/TIE.2008.918642.

[13] R. R. Pereira , C. H. da Silva , L. E. Borges da Silva , G. Lambert-Torres, "Application of adaptive filters in active power filters", 2009 Brazilian Power Electronics Conference, Bonito-Mato Grosso do Sul, 770-774, 2009, doi: 10.1109/COBEP.2009.5347654.

[14] K. Sebasthirani , M. Manimegalai, "Recent Advances in Shunt Active Power Filter with Recursive Least Square Algorithm and its Applications", 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, 1288-1293, 2018, doi: 10.1109/ICOEI.2018.8553806.

[15] R.K. Agarwal, I. Hussain, B. Singh, "Dual-function PV-ECS integrated to 3P4W distribution grid using 3M-PLL control for active power transfer and power quality improvement", IET Renewable Power Generation, **12**(8), 920-927 ,2018, doi: 10.1049/iet-rpg.2016.0723.

[16] M. Badoni , A. Singh , B. Singh, "Comparative Performance of Wiener Filter and Adaptive Least Mean Square-Based Control for Power Quality Improvement", IEEE Transactions on Industrial Electronics, **63**(5), 3028-3027,2016, doi: 10.1109/TIE.2016.2515558.

[17] R. Rajagopal, K Palanisamy, S.Paramasivam, "Comparative Performance of Normalized LMS and Variable Step Size LMS based Control for Shunt Active Filtering," 2019 9th International Conference on Advances in Computing and Communication (ICACC), Kochi, India , 136-141, 2019, doi: 10.1109/ICACC48162.2019.8986184.

[18] R. Belaidi, A. Haddouche, M. Fathi, M. M. Larafi and G. M. Kaci, "Performance of grid-connected PV system based on SAPF for power quality improvement," 2016 International Renewable and Sustainable Energy Conference (IRSEC), Marrakech, 542-545, 2016, doi: 10.1109/IRSEC.2016.7984050.

[19] R. Rajagopal, K. Palanisamy and S. Paramasivam, "Shunt Active Filter Based on 7-Level Cascaded Multilevel Inverter for Harmonic and Reactive Power Compensation" Journal of Electrical Engineering & Technology, Springer, **14**(4), 1543-1551, 2019, doi: 10.1007/s42835-019-00201-1.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1178-1189 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Defeating Anti-Debugging Techniques for Malware Analysis Using a Debugger

Jong-Wouk Kim[1], Jiwon Bang[2], Mi-Jung Choi[1,2,*]

[1]*Department of Computer Science, Kanwon National University, 24341, Republic of Korea*

[2]*Interdisciplinary Graduate Program in Medical Bigdata Convergence, Kangwon National University, 24341, Republic of Korea*

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 25 November, 2020
Online: 14 December, 2020

Keywords:
Anti-Debugging
Malware Analysis
Malware

ABSTRACT

Cyberattacks such as spear phishing and malspam pretending to be companies, institutes, and government officials are increasing and evolving. Malware has a variety of purposes, such as collecting personal information and illegal access to the system. New types of malware are increasing every day, and many malware programs spread all over the Internet, causing severe problems. To analyze such malware effectively, analysts first need to understand the inner structure of the malware. They can try to analyze malware manually and automatically. However, attackers who create malware use many different kinds of techniques, such as anti-reverse engineering, to hinder and delay analysis. They also extend malware life through a combination of different techniques, such as social engineering and anti-debugging. These techniques make the malware more sophisticated; thus, it is hard for an analyst to detect the malware. Anti-debugging, one way to protect malware, is a deadly poison to malware analysts because it makes the analysis more difficult by detecting a debugger or debugging environments. Therefore, this paper describes malware's anti-debugging techniques and how to defeat them through anti-anti-debugging mechanisms. It applies its findings to analyze a sample program, packed files, and actual malware with anti-debugging modules and performs various experiments to verify the proposed techniques. After the experiments, it confirms whether its countermeasure is useful for malware analysis.

1. Introduction

In the past, the development of malware was mainly used for malware programmers to show off. However, malicious software(*malware*) such as virus and keylogger in recent years illegally accesses server systems in various ways to steal personal and financial information and computing power. Fatal damage, such as the theft of essential data of institutes, corporations, public institutions, and personal user information, occurs. The malware damage can take many forms, from stealing personal information such as a user's financial information to destroying industrial facilities such as Stuxnet [1]. According to Symantect's "Internet Security Threat Report 2019", financial Trojans accounted for 16% of the many malware programs spread in 2018. In addition, Trend Micro's "Unseen Threats, Imminent Losses" said that a vulnerability was found in the Supervisory Control And Data Acquisition (SCADA) system, one of the industrial systems' components. Also, attackers can penetrate the system through malware that attacks this vulnerability and illegally controls the SCADA system.

New types of malware appear every day and spread over the Internet, causing serious problems frequently. The malware also uses anti-reverse engineering techniques such as Self-Modifying code, code compression, packing, and anti-debugging methods to hide internal code structures and interfere with the analysis. This paper is an extension of work initially presented in the *International Conference on ICT Convergence* [2] to solve these kinds of problems. It is challenging to analyze malware with these anti-reverse engineering techniques accurately and requires considerable time. Security experts reinforce security by using anti-virus software and various equipment types to minimize malware damage, while analysts scrutinize malware using various techniques, it is more likely that anti-virus software or an automated analysis tool fails to detect it. The workflow of malware analysis is as in Figure 1. Usually, when an analyst analyzes malware, an analyst gets information from automated and static tools [3]. For example, if malware has anti-reverse engineering analysis. After static analysis, the analyst analyzes malware behavior through dynamic analysis and investigates its behavior in more detail through advanced analysis. However, anti-debugging techniques are hinder all analysis steps especially advanced

*Corresponding Author: Mi-Jung Choi, 305 Hanbit-Gwan, 1 Gangwondaehakgil, Chuncheon-Si, Gangwon-Do, 24341 Repblic of Korea, mjchoi@kangwon.ac.kr

analysis. Most of the tools in the advanced analysis are insufficient to deal with anti-debugging techniques. The relevant APIs and structures such as PEB and HEAP could be modified to defeat anti-debugging techniques as soon as advanced analysis begins. This step makes it more convenient for analysts to analyze malware, and it could be analyzed more quickly because the malware interference is gone.

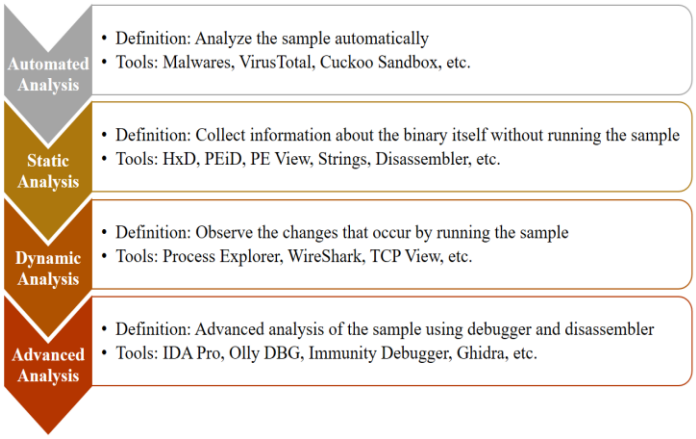


Figure 1: Workflow in safe and accurate analysis of malware

When analyzing malware in virtual environments such as QEMU, VMware, and VirtualBox, the virtual environments can be constructed similarly to the physical environments. The virtual environment analysis has advantages that are vastly similar to the actual execution, and does not have to worry about infecting a physical computer while performing malware analysis. However, there is a disadvantage in that analysis may be disrupted through the anti-VM (Virtual Machine) techniques. Anti-VM techniques disable malware analysis using CPU emulators or virtualization tools. The Dynamic Binary Instrumentation (DBI) tool can bypass anti-debugging by inserting or replacing code chunks into the operation command [4]. However, DBI has a disadvantage because it is not convenient, and it cannot execute huge or complex programs correctly.

The static analysis uses tools such as a decompiler and disassembler to analyze the binary or code of malware to infer the type of malware. However, this approach becomes more difficult and complicated by evasion techniques such as code obfuscation and packer [5]. A dynamic analysis examines the behavior of the target file. However, dynamic analysis tools cannot provide both accuracy and convenience when analyzing complex malware. The debugger, widely used to analyze malware dynamically, can examine changes in the machine's state while executing each command. However, the disadvantage of using a debugger is that there are anti-debugging techniques that can detect and interfere with the debugging process. Thus, if analysts can defeat anti-debugging techniques that interfere with the debugger, it becomes one of the most powerful methods for analyzing malware. Therefore, this paper introduces anti-debugging techniques that attack a debugger's vulnerabilities to interfere with analysis, and explains how to bypass analysis. In addition, it proposes anti-anti-debugging techniques that can defeat anti-debugging techniques, and verify the proposed method applying them to a sample program, packed files, and actual malware with anti-debugging modules. This paper aims to give direction for researchers who analyze malware or those who dream of being analysts to develop their methods or scripts that can defeat anti-debugging techniques when using a debugger.

This paper is organized as follows. Section 2 introduces research related to anti-debugging. Section 3 proposes anti-debugging techniques based on artifacts and time and describes correspondent anti-anti-debugging mechanisms. Section 4 presents the experiments performed and verifies anti-anti-debugging techniques with a sample program, packed files, and actual malware. Section 5 concludes the study.

2. Related Work

Analysts have continuously reviewed malware detection and binary code analysis. On the other hand, anti-reverse engineering research does not draw much attention from researchers except for some category topics [6], [7]. There are various topics in anti-reverse engineering such as anti-debugging, anti-VM, and code encryption. To understand anti-analysis, first, malware analysis must be briefly described. There are two different types of malware analyses, namely static analysis and dynamic analysis. Static analysis is a method of analyzing codes or binaries without executing them. Analysts can observe the file's operation code with a decompiler or disassembler and analyze the program flow and internal structure. Dynamic analysis probes its behavior by executing them using a sandbox, debugger, or emulator. Debugging refers to the process of detecting and eliminating errors or bugs when a process operates unexpectedly or crashes. A debugger is a tool or software that helps with this, therefore, it is a useful tool for understanding malware's inner workings. It likewise helps to observe changes in registry, memory, and stack closely while executing instructions in the target file (*debuggee*) line by line. Also, analysts use debuggers because they can manipulate elements such as code, environment, and memory. The sandbox case runs the file to record and analyze the process and network changes over a while [8].

However, malware developers use anti-analysis techniques against each analysis. Therefore, anti-debugging techniques delay malware detection and increase its life span by bypassing the anti-virus software. Anti-analysis is a series of techniques in which malware remains concealed, prevents detection, interferes with analysis, and makes analyzing it challenging. It is analyzed 6,222 samples of malware and studied how malicious behaviors differed in an environment with virtualization and a debugger [9]. As results, they found about 40 percent malware in an environment with a debugger performing less malicious behavior. Also, only two percent of the samples in a virtual environment showed malicious behavior when executed. This study shows malware with anti-VM and anti-debugging can bypass dynamic malware analysis tools such as a debugger, sandbox, and emulator.

Examples of anti-analysis techniques that interfere with static analysis include packer [10]-[12], code obfuscation [5], [13], [14], and opaque constants [15]. Packer is a technique that compresses the executable file and creates a new program entirely different from the original program. However, the new packed file has the same or similar behavior with various techniques such as Portable Executable (PE) relocation and anti-analysis. In the case of code obfuscation and opaque constants, these are the process of encrypting text, binary data, or code to make it difficult to understand. For these reasons, the malware detection method

based on signature became ineffective. Thus, dynamic analysis emerged to solve these kinds of problems. Debugging, which analyzes malware, carefully examines details such as behaviors, registers, and stacks while executing malware instructions by line. However, dynamic analysis using the debugger suffers from anti-analysis, such as anti-debugging, anti-VM, code compression, Self-Modifying, and blocking input devices. Among the anti-analysis, anti-debugging is an attempt by the debuggee to find or detect evidence of an analysis environment or debugger's presence. It is provided an overview of the analysis disruption techniques used by malware such as anti-debugging, anti-VM, and obfuscation [16]. Shang Gao and Qian Lin gave an overview of Windows's debugging mechanism, categorized according to the debugging methods used by the debugger, and explained the debugging functions and anti-debugging through exceptions [17]. It is classified general malware, targeted malware to attack specific systems and investigate how many anti-debugging techniques applied to it [18]. Most malware had anti-debugging techniques, however, it is found that the general malware had more anti-debugging techniques than targeted malware.

To defeat anti-analysis techniques, analysts can defeat most anti-debugging techniques using many CPU emulator and virtualization tools. However, these tools are not suitable for analyzing malware because they can analyze only a specific process or program. Dynamic Binary Instrumentation (DBI) framework can insert, modify, and delete code at the arbitrary address during the program's running [4], [19], [20], [21]. Therefore, the DBI can bypass the anti-debugging techniques by code insertion. However, analysis is inconvenient, and some DBIs cannot execute complex programs correctly [22]. Recently It is proposed the Apate framework to analyze malware [23]. Apate is a framework that hides the debugger from anti-debugging. The spectrum of powerful anti-debugging techniques is analyzed and classified into 79 attack vectors and six categories.

Apate is designed to defeat anti-debugging techniques by: (1) performing just-in-timed disassembling based on single-stepping and; (2) monitoring the execution flow of the debuggee and modifying the state of the debuggee to hide WinDbg. It is defined anti-debugging rule sets and proposed replacing the matching part of the binary with other instructions. The method proposed and modified the part that calls the anti-debugging functions in the debugger to other instructions [24], [25]. It is proposed a new framework for classifying, detecting and bypassing reverse engineering prevention techniques used by malware and protecting systems [9]. It is proposed UBER, a new system to neutralize anti-sandbox [26]. Sandbox has many artifacts of analysis environment. The anti-sandbox looks for various system artifacts expected to exist in the system for identifying the sandbox environment. UBER does not replicate artifacts or folders in use or directly simulate user behavior. Instead, it generalizes the user's computer usage pattern with an abstract behavioral profile, adopts the profile to simulate user and artifact creation operations, and then replicates the system with the created artifacts into a sandbox. Through this, UBER can defeat the anti-sandbox (i.e., usage artifacts analysis), while malware in the Windows or macOS malware is increasing. Since the tools for analyzing macOS based on malware are very limited. It is proposed Mac-A-Mal, a framework for analyzing Mac based malware [27]. It is developed a kernel extension to monitor malware behavior and bypass several evasion prevention techniques used in the wild, which uncovered 74 unknown malware programs.

3. Defeating anti-debugging techniques

This section first finds out how anti-debugging techniques work in Windows. Afterward, it explains how to defeat them [28].

3.1. Process Environment Block (PEB) Structure

The Process Environment Block (PEB) structure is a data structure in the Windows NT operating system. Each process has this structure and it contains each process information. Malware can detect the presence of debugger without using a specific function such as `IsDebuggerPresent`(). If implemented manually, there is no way to detect this in a static analysis or initial analysis. The anti-debugging based on PEB structure protects programs not only by malware but also many by many packers such as `PECompact`, `ASPack`, and `ASProtect` [29].

It confirms the `BeingDebugged`, `NtGlobalFlag`, and HEAP structure to perform anti-debugging among many PEB structure members. `BeingDebugged` is set to 0x0 if not debugged, and 0x1 if debugged. `NtGlobalFlag` has a value of 0x0 if not debugged, and is set to a value of 0x70 if debugged. Table 1 shows the flags of `NtGlobalFlag` while debugging. Not only the PEB structure, but HEAP structure can also detect the presence of a debugger.

Table 1: Flag configurations of `NtGlobalFlag` member in the PEB structure

Flag	Value
FLG_HEAP_ENABLE_TAIL_CHECK	0x10
FLG_HEAP_ENABLE_FREE_CHECK	0x20
FLG_HEAP_VALIDATE_PARAMETER	0x40

Table 2: Flag configurations of `Flags` member in the HEAP structure

Flag	Value
HEAP_GROWABLE	0x2
HEAP_TAIL_CHECKING_ENABLED	0x20
HEAP_FREE_CHECKING_ENABLED	0x40
HEAP_SKIP_VALIDATION_CHECKS (only in Windows XP or under)	0x10000000
HEAP_VALIDATE_PARAMETERS_ENABLED	0x40000000

Table 3: Flag configurations of `ForceFlags` member in the HEAP structure

Flag	Value
HEAP_TAIL_CHECKING_ENABLED	0x20
HEAP_FREE_CHECKING_ENABLED	0x40
HEAP_VALIDATE_PARAMETERS	0x40000000

Anti-debugging can detect debuggers using the `Flags` and `ForceFlags` members of the HEAP structure. `Flags` has a value of 0x2 if not debugged. It has a value of 0x50000062 in Windows XP or 0x40000062 in Windows 7 or later versions, if debugged. `ForceFlags` has a value of 0x0 if not debugged, and 0x40000060 if debugged. Table 2 shows the configuration flags of `Flags` and Table 3 shows the configuration flags of `ForceFlags`. To bypass anti-debugging based on PEB structure, analysts can defeat anti-debugging by modifying each member to an appropriate value, such as 0x0 in `BeingDebugged`.

3.2. IsDebuggerPresent()

IsDebuggerPresent() is the function in which software can easily detect the presence of a debugger. Malware and general software use this function to protect itself from reverse engineering. This function is also one of the anti-debugging techniques that most packers such as UPX and PECompact can apply to the original program. This function returns the BeingDebugged member of the PEB structure. To defeat this function, one must modify the code to return 0x0 or change the value of BeingDebugged member of the PEB structure. This paper uses PUSH and POP instructions to store 0x0 in the EAX register, shown in Figure 2. The original code of IsDebuggerPresent() is the left side of Figure 2, and the modification code of IsDebuggerPresent() is the right side of Figure 2.

```
MOV EAX,DWORD PTR FS:[18]          MOV EAX,DWORD PTR FS:[18]
MOV EAX,DWORD PTR DS:[EAX+30]      PUSH 0
MOVZX EAX,BYTE PTR DS:[EAX+2]      POP EAX
RETN                               RETN
```

Figure 2: The original code (left) and modification code (right) of IsDebuggerPresent()

```
MOV EDI,EDI                                  MOV EDI,EDI
PUSH EBP                                     PUSH EBP
MOV EBP,ESP                                  MOV EBP,ESP
CMP DWORD PTR SS:[EBP+8],0                   MOV EAX,DWORD PTR SS:[EBP+C]
PUSH ESI                                     PUSH 0
JE SHORT kernel32.77623FC1                   POP DWORD PTR DS:[EAX]
                                             XOR EAX,EAX
                                             POP EBP
                                             RETN 8
CALL DWORD PTR DS:[<&ntdll.NtQueryInformationProcess>]
TEST EAX,EAX
JL kernel32.77629176
XOR EAX,EAX
CMP DWORD PTR SS:[EBP+8],EAX
SETNE AL
MOV DWORD PTR DS:[ESI],EAX
XOR EAX,EAX
INC EAX
POP ESI
POP EBP
RETN 8
```

Figure 3: The original code (left) and modification code (right) of CheckRemoteDebuggerPresent()

3.3. CheckRemoteDebuggerPresent()

CheckRemoteDebuggerPrenset() function takes the Process Identifier (PID) of a particular process to detect whether it is debugged or not. It calls the NtQueryInformationProcess() function from inside to perform anti-debugging not only on malware but also on general software, and some packers use it to prevent debugging. The code on the left of Figure 3 is the original code of CheckRemoteDebuggerPrenset() and as shown on the left side of Figure 3, it calls NtQueryInformationProcess() from inside. There are two ways to defeat this function. Firstly, defeat the function called NtQueryInformationProcess() internally. Secondly, store 0x0 in the EAX register to make it always return 0x0.

This paper stores 0x0 in the stack using the PUSH instruction, as shown on the right side of Figure 3. It then stores 0x0 in the EAX register with POP instruction and terminates the function using RETN. There are three essential things to defeat this function. First, three existing instructions from the top of this code must be maintained. The second is to store 0x0 in the EAX register and return it. Last is to save the return address stored in the stack to the EBP register with POP instruction, then end with RETN. Otherwise, the flow of the debuggee gets tangled, and the analysis cannot proceed further.

3.4. ZwQueryInformationProcess()

The developer uses ZwQueryInformationProcess() functions for various reasons, such as finding a process path not only for anti-debugging. However, this section, only describes anti-debugging based on ZwQueryInformationProcess(). Figure 4 is a parameter for ZwQueryInformationProcess(). Among these parameters, ProcessInformationClass is the type of process information to be retrieved. It performs by set a specific value to the ProcessInformationClass parameter. If ProcessInformationClass is set to 0x7 which means ProcessDebugPort, it can determine whether it is debugged or not. In case of debugging, it returns 0xFFFFFFFF (-1). If not, it returns 0x0. Not only ProcessDebugPort, this function can executes anti-debugging bypassing 0x1E, which means ProcessDebugObjectHandle or 0x1F, which means ProcessDebugFlags to the parameter ProcessInformationClass. In the case of ProcessDebugObjectHandle, this function returns 0x000000B4 if debugged, and it returns 0x0 if not debugged. In the case of ProcessDebugFlags, this function returns 0x0 if debugged, and it returns 0x1 if not debugged. The code on the left of Figure 5 is the original code of ZwQueryInformationProcess().

```
NTSTATUS WINAPI ZwQueryInformationProcess(
    _In_      HANDLE              GetProcessHandle,
    _In_      PROCESSINFOCLASS    ProcessInformationClass,
    _Out_     PVOID               ProcessInformation,
    _In_      ULONG               ProcessInformationLength,
    _Out_opt_ PULONG              ReturnLength
);
```

Figure 4: Parameter of ZwQueryInformationProcess()

```
MOV EAX,0EA                        PUSH 100000
MOV EDX,7FFE0300                   RETN
CALL DWORD PTR DS:[EDX]
RETN 14                            MOV EAX,0EA
                                   MOV EDX,7FFE0300
                                   CMP DWORD PTR SS:[ESP+8],7
                                   JE SHORT 00100037
                                   CMP DWORD PTR SS:[ESP+8],1E
                                   JE SHORT 0010004A
                                   CMP DWORD PTR SS:[ESP+8],1F
                                   JE SHORT 0010005D
                                   PUSH 779A6052
                                   RETN
                                   MOV EAX,DWORD PTR SS:[ESP+C]
                                   PUSH 0
                                   POP DWORD PTR DS:[EAX]
                                   XOR EAX,EAX
                                   RETN 14
                                   MOV EAX,DWORD PTR SS:[ESP+C]
                                   PUSH 0
                                   POP DWORD PTR DS:[EAX]
                                   XOR EAX,EAX
                                   RETN 14
                                   MOV EAX,DWORD PTR SS:[ESP+C]
                                   PUSH 1
                                   POP DWORD PTR DS:[EAX]
                                   RETN 14
```

Figure 5: The original code (left) and modification code (right) of ZwQueryInformationProcess()

In order to bypass ZwQueryInformationProcess(), three things are crucial: (1) the first two instructions from the top of the original code must be executed; (2) it must return an appropriate

value; (3) it should not interfere with the flow of the debuggee. It needs a new virtual memory that must be allocated to handle the number of cases and consider non-anti-debugging cases. Also, `ZwQueryInformationProcess()` enters the kernel area through `KiFastSystemCall` and `SYSENTER` instructions. Since the debugger cannot enter the kernel, one can bypass these powerful instructions by writing new instructions in allocated virtual memory, as shown on the right of Figure 5. It determines whether the received `ProcessInformationClass` parameter is 0x7 (`ProcessDebugPort`), 0x1E (`ProcessDebugObjectHandle`), or 0x1F (`ProcessDebugFlags`). In each case, it jumps to the address with the specific instructions. Otherwise, it returns to the original instruction with `RETN`. The reason for returning to the original function is not to disturb the flow of the debuggee.

```
HWND FindWindowA(
  LPCSTR lpClassName,
  LPCSTR lpWindowName
);
```

```
HWND FindWindowW(
  LPCSTR lpClassName,
  LPCSTR lpWindowName
);
```

Figure 6: Parameters of `FindWindowA()` and `FindWindowW()`

```
MOV EDI,EDI
PUSH EBP
MOV EBP,ESP
XOR EAX,EAX
PUSH EAX
PUSH DWORD PTR SS:[EBP+C]
PUSH DWORD PTR SS:[EBP+8]
PUSH EAX
PUSH EAX
CALL USER32.76F8B865
POP EBP
RETN 8
```

```
MOV EDI,EDI
PUSH EBP
MOV EBP,ESP
XOR EAX,EAX
PUSH EAX
PUSH DWORD PTR SS:[EBP+C]
PUSH DWORD PTR SS:[EBP+8]
PUSH EAX
PUSH EAX
PUSH 220000
RETN
RETN 8

CALL USER32.76F8B865
CMP EAX,7018A
JE SHORT 000B000F
JMP SHORT 000B0011
XOR EAX,EAX
POP EBP
RETN 8
```

Figure 7: The original code (left) and modification code (right) of `FindWindowW()`

3.5. *FindWindow()*

The family of `FindWindow` functions can find the name of the specific window. It consists of `FindWindowW()`, `FindWindowA()`, `FindWindowExW()`, and `FindWindowExA()`. These are anti-debugging techniques that confirm the debugger's window name to see if the debuggers are running on the operating system. Figure 6 shows the parameters of `FindWindowA()` and `FindWindowW()`. The `lpClassName` gets a specific class name, and the `lpWindowName` gets a specific window name when calling these functions. If it finds the name of a specific window, it returns the handle of the window. If not, it returns 0x0. If malware calls these functions, it passes a particular debugger's window name to the `lpWindowName` parameter for anti-debugging. However, the analysts can defeat it always to returns 0x0.

The code on the left of Figure 7 is the original code of `FindWindowW()`, and the code on the right of Figure 7 is the code to defeat it. This study allocates a virtual memory and writes new instructions to disable anti-debugging techniques. After moving to the allocated address through the `PUSH` and `RETN` instructions, one must execute the existing `CALL` instruction. Next, it needs to get the debugger's window handle and compare it with the `EAX` register using `CMP` instruction. If it is the same, it jumps by `JE` instruction and stores 0x0 in the `EAX` register. Otherwise, it jumps by `JMP` instruction and restores the return address to the `EBP` register. The analysts can defeat this anti-debugging technique by writing new instructions in virtual memory. In Figure 7, the address of virtual memory is 0x00220000. However, this address must point to empty virtual memory.

3.6. *GetCurrentProcessId(), BlockInput()*

The commercial packer such as `Yoda's Protector` uses the two functions described in this section. The malware packed with `Yoda's Protector`, it calls `CreateToolHelp32Snapshot()` to obtain the PIDs and process all running processes' information. Next, it uses the `GetCurrentProcessId()` to get its process information and then compares the two results. If the parent process is a debugger, it interrupts analysis by terminating the process. Therefore, to defeat this anti-debugging technique, `GetCurrentProcessId()` always returns the PID of a debugger or analysis tool by modifying the PID of the debugger and debuggee to be the same. Analysts can defeat this technique. The code on the left side of Figure 8 is the original code of `GetCurrentProcessId()`, and the code is modified the same as on the right side of Figure 8 to defeat this function. To match the PID of the debugger and debuggee, it stores the PID of the debugger (0x0EA0 in Figure 8) in the `EAX` register and returns it. At this point, `NOP` instruction fills the remaining space of memory.

```
MOV EAX,DWORD PTR FS:[18]
MOV EAX,DWORD PTR DS:[EAX+20]
RETN
```

```
MOV EAX,0EA0
NOP
NOP
NOP
RETN
```

Figure 8: The original code (left) and modification code (right) of the GetCurrentProcessId()

```
MOV EAX,1141
MOV EDX,7FFE0300
CALL DWORD PTR DS:[EDX]
RETN 4
```

```
NOP
NOP
⋮
NOP
NOP
RETN 4
```

Figure 9: The original code (left) and modification code (right) of `BlockInput()`

The `BlockInput()` blocks the input events such as keyboard and mouse. When software calls this function, the computer blocks all events from input devices. By blocking all input devices' events, signals from devices do not affect the input queue's synchronous key state and asynchronous key state. If the software calls this function while running, the computer blocks the input devices' events, and there is no solution until the software calls again. The malware exploits this. If malware calls this function while debugging, the computer blocks all events from input devices, and

analysis stops. The only solution is to reboot the computer and analyze the malware again from the beginning. This function has a BOOL value as a parameter. If the parameter is TRUE, the computer blocks the events. If FALSE, the computer unblocks the events. This function returns to non-zero if it ends with no errors. The corresponding code is the left code of Figure 9. The way to defeat this technique is more straightforward than others. As shown on right side of Figure 9, this is defeated by modification with NOP instruction, which does nothing.

```
MOV ECX,DWORD PTR DS:[7FFE0324]
MOV EDX,DWORD PTR DS:[7FFE0320]
MOV EAX,DWORD PTR DS:[7FFE0328]
CMP ECX,EAX
JNZ SHORT KERNELBA.75D36770
MOV EAX,DWORD PTR DS:[7FFE0004]
MUL EDX
SHL ECX,8
IMUL ECX,DWORD PTR DS:[7FFE0004]
SHRD EAX,EDX,18
SHR EDX,18
ADD EAX,ECX
RETN
```

```
MOV EDX,7FFE0000
PUSH 5678ABCD
POP EAX
RETN
```

Figure 10: The original code (left) and modification code (right) of the GetTickCount()

```
CMP DWORD PTR DS:[73F78FD4],0
JNZ <JMP.&KERNEL32.GetTickCount>
CALL WINMM.73F526C0
SUB EAX,DWORD PTR DS:[73F78FD8]
PUSH 0
SBB EDX,DWORD PTR DS:[73F78FDC]
PUSH 2710
PUSH EDX
PUSH EAX
CALL <JMP.&ntdll._alldiv>
ADD EAX,DWORD PTR DS:[73F78FE0]
RETN
```

```
MOV EAX,0
RETN
```

Figure 11: The original code (left) and modification code (right) of timeGetTime ()

3.7. GetTickCount(), timeGetTime()

The anti-debugging techniques described so far are anti-debugging based on artifacts. Those are techniques in which malware detect analysis signs, analysis environment, or debugger. The anti-debugging based on artifacts is the most common and useful technique for malware to evade analysis. If a debugger runs the program, there are differences from normal execution. The anti-debugging based on artifacts confirms the debugger through these differences.

However, the anti-debugging described in this section is slightly different from the previous ones. The anti-debugging described in this section uses time elements. The anti-debugging based on time measures the time between a particular routine. It compares the execution in the real system and the execution analysis environment and indirectly detects the debugger. It measures the time stamp counter cycle before and after starting a particular routine in a program to confirm a debugger's presence because the analysis using a debugger takes a long time.

GetTickCount() starts counting after Windows boots. Through this function, malware or software can easily determine how long it has been since the computer booted up and got time value every time stamp counter cycle. The return value is the 32-bit form to keep the count for up to 49.7 days. Many developers use this function to measure elapsed time, and malware developers use it as well. The code on the left in Figure 10 is the original code of GetTickCount(). To defeat this, it must return the same value by hard coding. Hard coding is the way to develop software by embedding data directly into the source code or memory. In this way, it stores a meaningless value in the EAX register. Therefore, the return value of GetTickCount() is always meaningless value which hard coded. In other words, the measured time for a specific routine is always same.

timeGetTime() has less overhead than timeGetSystemTime(), because timeGetSystemTime() uses a MMTIME structure, which contains timing information for different multimedia data types. However, timeGetTime() uses a DWORD structure. Moreover, timeGetTime() reacts directly to timer interruptions and returns the time in milliseconds that have elapsed since the Windows booted, thus GetTickCount(). The debuggee can determine the presence of the debugger by using precise functions such as timeGetTime(). Disabling this is not much different with GetTickCount(). To defeat this function, one must to return the same value all the time. The code on the left of Figure 11 is the original code of timeGetTime(). However, if modified as the code on the right side of Figure 11, debuggee fails to detect the presence of debugger because it always returns the same value.

4. Experiments

This study implemented a sample program which includes anti-debugging techniques for the *(i) Sample-ex* experiment. The sample program has various anti-debugging techniques such as the PEB, and HEAP, among others. The sample program printed the return value to verify the results of the anti-anti-debugging intuitively. In the second experiment, *(ii) Packed-ex*, packed programs were used for packed malware or packed software. This study confirms the effectiveness of the proposed anti-anti-debugging mechanisms using these programs. Finally, real malware was employed in the in *(iii) Malware-ex* experiment. The anti-anti-debugging proposed in this paper works well with benign files, packed files, and actual malware through the step by step experiments.

Table 4: Anti-debugging techniques of each packer

Packer	Anti-Debugging Techniques
PECompact	IsDebuggerPresent()
Armadillo	IsDebuggerPresent()
ACProtect	IsDebuggerPresent()
	ZwQueryInformationProcess()
Themida	IsDebuggerPresent()
	ZwQueryInformationProcess()
Yoda's Protector	IsDebuggerPresent()
	GetCurrentProcessId()
	BlockInput()

Table 5: Information of real virus

File Type	Win.32.EXE
Virus Type	Worm/Win32.Abuse
MD5	068c3b80106b3300548bv6vv673a3db5
Anti-Debugging Techniques	`GetTickCount()`
	`FindWindowExA()`
	`FindWindowA()`

4.1. Experimental Data

This paper implemented a sample program including various anti-debugging techniques for *(i) Sample-ex* experiment, as shown in Figure 12. This program prints different results depending on the presence of a debugger. Next, sample files used in the *(ii) Packed-ex* experiment, packed the calculator program provided by Windows 7 (32bit) into five packers or protectors. The common feature of used packers or protectors is that they protect the original software by applying one or more anti-debugging techniques. Five packers called `Yoda's Protector`, `Themida`, `ACProtect`, `Armadillo`, and `PECompact` for *(ii) Packed-ex* experiment were used. These packers apply one or more anti-debugging techniques to protect original program against reverse engineering. Table 4 shows the anti-debugging techniques applied by each packer.

Among these, the most commonly used function is `IsDebuggerPresent()`, and the unusual cases are `GetCurrentProcessId()` and `BlockInput()` used by `Yoda's Protector`. Table 5 shows information about one of the malware samples collected from VirusShare[**], which is for the *(iii) Malware-ex* experiment. This malware sample uses `GetTickCount()`, `FindWindowExA()`, and `FindWindowA()` functions for anti-debugging. However, it is impossible to analyze some samples using different anti-debugging techniques such as `RDTSC` instruction and Self-Modifying technique.

`RDTSC` is an instruction that reads the current time-stamp counter variable. Unlike function such as `timeGetTime()`, `RDTSC` is executed as a single instruction. In order to bypass this instruction, it is necessary to modify the debuggee. However, modifying the debuggee requires careful attention. Self-Modifying is one of anti-analysis techniques. It is a technique to alter its own codes while the file is executing. Therefore, an analyst fails to identify the real instructions before executing Self-Modifying. Thus, `RDTSC` and Self-Modifying are out of our research scope and this paper does not handle these techniques.

4.2. Experimental Results

Figure 13 shows the results of analyzing sample data in the debugger in *(i) Sample-ex*. In the case of Figure 13 (a), all of the anti-debugging techniques used in the sample program detected a debugger. All of these anti-debugging techniques returned and printed artifact or evidence which indicates the presence of a debugger. However, an analysis was made again using anti-anti-debugging, and the results were confirmed, as shown in Figure 13 (b). It showed a completely different artifact or evidence and message against Figure 13 (a).

This research primarily confirmed that the anti-anti-debugging works successfully. Next, it conducted *(ii) Packed-ex* and analyzed using different data. The analysis from this point shows that the results of the analysis are limited. Therefore, the actual memory address and instructions were shown. Figure 14 (a) is the loaded file with `Yoda's Protector` in the debugger. The number series on the left correspond to each address, and instructions on the right correspond to the original code. Figure 14 (b) shows the new code to defeat `GetCurrentProcessId()`. It stored the PID of the debugger, which is 0x0B5C in the `EAX` register, to make it the same as shown in Figure 14 (b). This study also confirmed the results of `ZwQueryInformationProcess()` used by `ACProtect`, as shown in Figure 15. Figure 15 (a) is the original code and address of `ZwQueryInformationProcess()`, and Figure 15 (b) is the new instructions to defeat it. At this time, 0x003E0000 is the address of the virtual memory. Through this, whether the anti-anti-debugging is effective even in packed samples was secondarily ascertained.

```
void adbg_BeingDebuggedPEB(void)
{
    BOOL found = FALSE;
    _asm
    {
        xor eax, eax;           // clear eax
        mov eax, fs:[0x30];     // Reference start of the PEB
        movzx eax, [eax + 0x02];// PEB+2 points to BeingDebugged
        mov found, eax;         // Copy BeingDebugged into 'found'
    }

    if (found)
        printf("Caught by PEB.BeingDebugged check!\n");
    else
        printf("Not caught by PEB.BeingDebugged check!\n");

    printf("  Return value: 0x%08X\n", found);
}
```

(a) The code of `PEB.BeingDebugged` for sample program

```
void adbg_PEBNtGlobalFlag(void){
    BOOL found=FALSE;
    DWORD NtGlobalFlag;
    __asm
    {
        xor eax, eax // clear eax
        mov eax, fs:[0x30] // Address of PEB
        mov ecx, [eax+0x68] //Points to NtGlobalFlags
        mov NtGlobalFlag,ecx //FLG_HEAPENABLE_TAIL_CHECK(0x10)
                             //+ FLG_HEAP_ENABLE_FREE_CHECK(0x20)
                             //+ FLG_HEAP_VALIDATE_PARAMETERS(0x40)
    }
    if(NtGlobalFlag==0x70)
        printf("Caught by PEB.NtglobalFlag\n");
    else
        printf("Not caught by PEB.NtglobalFlag\n");
    printf("  Return value: 0x%08X\n", NtGlobalFlag);
}
```

(b) The code of `PEB.NtGlobalFlag` for sample program

```
void adbg_PEBHEAPFalgs(void){
    DWORD flags;
    __asm{
        mov eax, fs:[0x30]
        mov eax, dword ptr [eax+0x18]
        mov ecx, dword ptr [eax+0x40]
        mov flags, ecx
    }
    if(flags==0x40000062)
        printf("Caught by PEB.ProcessHeap.Flag\n");
    else
        printf("Not caught by PEB.ProcessHeap.Flag\n");
    printf("  Return Value: 0x%08X\n", flags);
}
```

(c) The code of `PEB.HEAP.Flags` for sample program

```
void adbg_PEBHEAPForceFalgs(void){
    DWORD forceflags;
    __asm{
        mov eax, fs:[0x30]
        mov eax, dword ptr [eax+0x18]
        mov ecx, dword ptr [eax+0x444]
        mov forceflags, ecx
    }
    if(forceflags==0x40000060)
        printf("Caught by PEB.ProcessHeap.ForceFlag\n");
    else
        printf("Not caught by PEB.ProcessHeap.ForceFlag\n");
    printf("  Return Value: 0x%08X\n", forceflags);
}
```

(d) The code of `PEB.HEAP.ForceFlags` for sample program

```
void adbg_IsDebuggerPresent(void)
{
    int found = 0;
    found = IsDebuggerPresent();
    if (found)
        printf("Caught by IsDebuggerPresent!\n");
    else
        printf("Not caught by IsDebuggerPresent\n");
    printf("  Return Value: 0x%08X\n",found);
}
```

(e) The code of `IsDebuggerPresent`()for sample program

```
void adbg_CheckRemoteDebuggerPresent(void)
{
    HANDLE hProcess = INVALID_HANDLE_VALUE;
    BOOL found = FALSE;
    hProcess = GetCurrentProcess();
    CheckRemoteDebuggerPresent(hProcess, &found);
    if (found)
        printf("Caught by CheckRemoteDebuggerPresent!\n");
    else
        printf("Not caught by CheckRemoteDebuggerPresent\n");
    printf("  Return Value: 0x%08X\n", found);
}
```

(f) The code of `CheckRemoteDebuggerPresent`()for sample program

```
void adbg_FindWindow(void){
    HWND result;
    result=FindWindow(NULL, L"Immunity Debugger - samplefile.exe");
    if(result==NULL)
        printf("Not caught by FindWindowW\n");
    else
        printf("Cautht by FindWindowW\n");
    printf("  Return value: 0x%08X\n", result);
}
```

(g) The code of `FindWindowW`() for sample program

```
void adbg_FindWindowA(void){
    HWND result;
    result=FindWindowA(NULL, "Immunity Debugger - samplefile.exe");
    if (result==NULL)
        printf("Not caught by FindwindowA\n");
    else
        printf("Cautht by FindWindowA\n");
    printf("  Return value: 0x%08X\n", result);
}
```

(h) The code of `FindWindowA`() for sample program

```
void adbg_gettickcount(void){
    DWORD beginTime=GetTickCount();
    DWORD endTime;
    __asm{PUSH ECX}
    __asm{MOV ECX, 3}
    __asm{PUSH EAX}
    __asm{XOR EAX, EAX}
    __asm{ADD EAX, 10}
    __asm{XCHG EAX, ECX}
    __asm{POP EAX}
    __asm{OR EAX, ECX}
    __asm{POP ECX}
    endTime=GetTickCount();
    if (endTime!=beginTime)
        printf("Caught by GetTickCount\n");
    else
        printf("Not caught by GetTickCount\n");
    printf("  Time difference: %d msec\n", endTime-beginTime);
}
```

(i) The code of `GetTickCount`() for sample program

```
void adbg_timegettime(void){
    DWORD beginTime=timeGetTime();
    DWORD endTime;
    __asm{PUSH ECX}
    __asm{MOV ECX, 3}
    __asm{PUSH EAX}
    __asm{XOR EAX, EAX}
    __asm{ADD EAX, 10}
    __asm{XCHG EAX, ECX}
    __asm{POP EAX}
    __asm{OR EAX, ECX}
    __asm{POP ECX}
    endTime=timeGetTime();
    if((endTime-beginTime)>0x10)
        printf("Caught by timeGetTime\n");
    else
        printf("Not caught by timeGetTime\n");
    printf("  Time difference: %d msec", endTime-beginTime);
}
```

(j) The code of `timeGetTime`() for sample program

```
void adbg_ZwQueryInformationProcess_ProcessDebugPort(void){
    DWORD ProcessDebugPort;
    PFZWQUERYINFORMATIONPROCESS pfZwQueryInformationProcess;
    HMODULE h_ntdll=GetModuleHandle(TEXT("ntdll.dll"));
    pfZwQueryInformationProcess=(PFZWQUERYINFORMATIONPROCESS)GetProcAddress(h_ntdll,"ZwQueryInformationProcess");
    pfZwQueryInformationProcess(GetCurrentProcess(),0x7,&ProcessDebugPort,4,0);

    if(ProcessDebugPort==-1)
        printf("Caught by ZwQueryInformationProcess(ProcessDebugPort)\n");

    else
        printf("Not caught by ZwQueryInformationProcess(ProcessDebugPort)\n");
    printf("  Return Value: 0x%08X\n", ProcessDebugPort);
}
```

(k) The code of `ZwQueryInformationProcess - ProcessDebugPort` for sample program

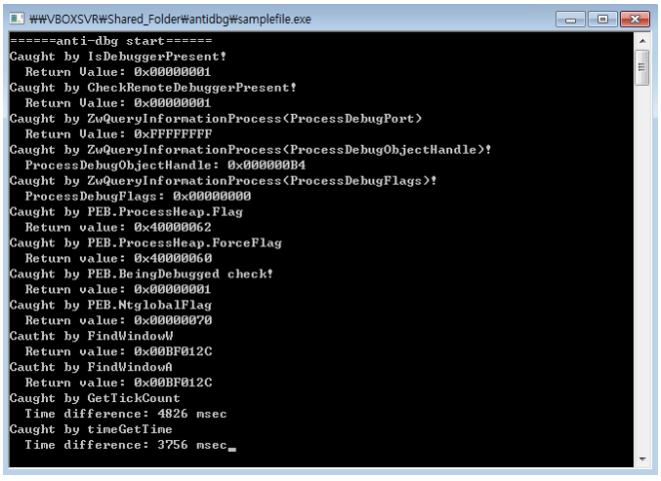
```
void adbg_ZwQueryInformationProcess_ProcessDebugObject(void){
    BOOL result=FALSE;
    DWORD fRet;
    DWORD ProcessDebugObjctHandle;
    PFZWQUERYINFORMATIONPROCESS pfZwQueryInformationProcess;
    HMODULE h_ntdll=GetModuleHandle(TEXT("ntdll.dll"));
    pfZwQueryInformationProcess = (PFZWQUERYINFORMATIONPROCESS)GetProcAddress(h_ntdll, "ZwQueryInformationProcess");
    fRet = pfZwQueryInformationProcess(GetCurrentProcess(), 0x1e, &ProcessDebugObjctHandle, 4, 0); // 0x1e : ProcessObjectHandle

    if(ProcessDebugObjctHandle)
        if(fRet==0) result=TRUE;
    if(result)
        printf("Caught by ZwQueryInformationProcess(ProcessDebugObjectHandle)!\n");
    else
        printf("Not Caught by ZwQueryInformationProcess(ProcessDebugObjectHandle)\n");

    printf("  ProcessDebugObjectHandle: 0x%08X\n", ProcessDebugObjctHandle);
}
```

(l) The code of `ZwQueryInformationProcess – ProcessDebugObjectHandle` for sample program

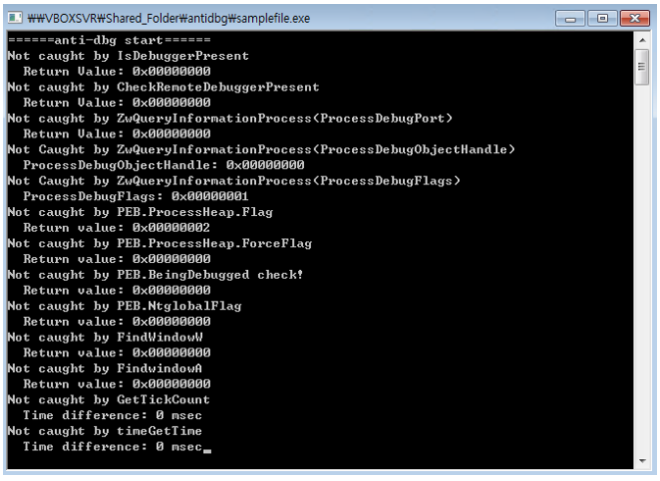
```
void adbg_ZwQueryInformationProcess_ProcessDebugFlags(void)
{
    BOOL result = FALSE;
    DWORD fRet;
    DWORD ProcessDebugFlags;
    PFZWQUERYINFORMATIONPROCESS pfZwQueryInformationProcess;
    HMODULE h_ntdll=GetModuleHandle(TEXT("ntdll.dll"));
    pfZwQueryInformationProcess = (PFZWQUERYINFORMATIONPROCESS)GetProcAddress(h_ntdll, "ZwQueryInformationProcess");
    fRet = pfZwQueryInformationProcess(GetCurrentProcess(), 0x1f, &ProcessDebugFlags, 4, 0); // 0x1f : ProcessDebugFlags

    if(!ProcessDebugFlags) result = TRUE;
    if(result)
        printf("Caught by ZwQueryInformationProcess(ProcessDebugFlags)!\n");
    else
        printf("Not Caught by ZwQueryInformationProcess(ProcessDebugFlags)\n");
    printf("  ProcessDebugFlags: 0x%08X\n", ProcessDebugFlags);
}
```

(m) The code of `ZwQueryInformationProcess – ProcessDebugFlags` for sample program

Figure 12: Anti-debugging functions for sample program



(a) Result without anti-anti-debugging



(b) Result with anti-anti-debugging

Figure 13: Results of (i) Sample-ex; (a) results without anti-anti-debugging; (b) results with anti-anti-debugging

```
75D368D0 MOV EAX,DWORD PTR FS:[18]
75D368D6 MOV EAX,DWORD PTR DS:[EAX+20]
75D368D9 RETN
```

(a) The original code and address of `GetCurrentProcessId()`

```
75D368D0 MOV EAX,0B5C
75D368D6 NOP
75D368D7 NOP
75D368D8 NOP
75D368D9 RETN
```

(b) The new code and address to defeat `GetCurrentProcessId()`

Figure 14: Results of (ii) Packed-ex in `Yoda's Protector`; (a) the original code and address of `GetCurrentProcessId()`; (b) the new code and address to defeat `GetCurrentProcessId()`

```
779A6048 MOV EAX,0EA
779A604D MOV EDX,7FFE0300
779A6052 CALL DWORD PTR DS:[EDX]
779A6054 RETN 14
```

(a) The original code and address of `ZwQueryInformationProcess()`

```
779A6048 PUSH 3E0000
779A604D RETN
003E0000 MOV EAX,0EA
003E0005 MOV EDX,7FFE0300
003E000A CMP DWORD PTR SS:[ESP+8],7
003E0015 JE SHORT 003E0037
003E0017 CMP DWORD PTR SS:[ESP+8],1E
003E0022 JE SHORT 003E004A
003E0024 CMP DWORD PTR SS:[ESP+8],1F
003E002F JE SHORT 003E005D
003E0031 PUSH 779A6052
003E0036 RETN
003E0037 MOV EAX,DWORD PTR SS:[ESP+C]
003E003E PUSH 0
003E0043 POP DWORD PTR DS:[EAX]
003E0045 XOR EAX,EAX
003E0047 RETN 14
003E004A MOV EAX,DWORD PTR SS:[ESP+C]
003E0051 PUSH 0
003E0056 POP DWORD PTR DS:[EAX]
003E0058 XOR EAX,EAX
003E005A RETN 14
003E005D MOV EAX,DWORD PTR SS:[ESP+C]
003E0064 PUSH 1
003E0069 POP DWORD PTR DS:[EAX]
003E006B RETN 14
```

(b) The new code and address to defeat `ZwQueryInformationProcess()`

Figure 15: Results of (ii) Packed-ex in ACProtect; (a) the original code and address of ZwQueryInformationProcess (); (b) the new code and address to defeat ZwQueryInformationProcess ()

```
76F86F69 MOV EDI,EDI
76F86F6B PUSH EBP
76F86F6C MOV EBP,ESP
76F86F6E PUSH 0
76F86F70 PUSH DWORD PTR SS:[EBP+14]
76F86F73 PUSH DWORD PTR SS:[EBP+10]
76F86F76 PUSH DWORD PTR SS:[EBP+C]
76F86F79 PUSH DWORD PTR SS:[EBP+8]
76F86F7C CALL USER32.76F88F4F
76F86F81 POP EBP
76F86F82 RETN 10
```

(a) The original code of FindWindowExA()

```
76F86F69 MOV EDI,EDI
76F86F6B PUSH EBP
76F86F6C MOV EBP,ESP
76F86F6E PUSH 0
76F86F70 PUSH DWORD PTR SS:[EBP+14]
76F86F73 PUSH DWORD PTR SS:[EBP+10]
76F86F76 PUSH DWORD PTR SS:[EBP+C]
76F86F79 PUSH DWORD PTR SS:[EBP+8]
76F86F7C PUSH 190000
76F86F81 RETN
76F86F82 RETN 10
00190000 CALL USER32.76F88F4F
00190005 CMP EAX,340722
0019000B JE SHORT 0019000F
0019000D JMP SHORT 00190011
0019000F XOR EAX,EAX
00190011 POP EBP
00190012 RETN 10
```

(b) The modification code of FindWindowExA()

```
76F88FF3 MOV EDI,EDI
76F88FF5 PUSH EBP
76F88FF6 MOV EBP,ESP
76F88FF8 XOR EAX,EAX
76F88FFA PUSH EAX
76F88FFB PUSH DWORD PTR SS:[EBP+C]
76F88FFE PUSH DWORD PTR SS:[EBP+8]
76F89001 PUSH EAX
76F89002 PUSH EAX
76F89003 CALL USER32.76F88F4F
76F89008 POP EBP
76F89009 RETN 8
```

(c) The original code of FindWindowA()

```
76F88FF3 MOV EDI,EDI
76F88FF5 PUSH EBP
76F88FF6 MOV EBP,ESP
76F88FF8 XOR EAX,EAX
76F88FFA PUSH EAX
76F88FFB PUSH DWORD PTR SS:[EBP+C]
76F88FFE PUSH DWORD PTR SS:[EBP+8]
76F89001 PUSH EAX
76F89002 PUSH EAX
76F89003 PUSH 170000
76F89008 RETN
76F89009 RETN 8
00170000 CALL USER32.76F88F4F
00170005 CMP EAX,340722
0017000B JE SHORT 0017000F
0017000D JMP SHORT 00170011
0017000F XOR EAX,EAX
00170011 POP EBP
00170012 RETN 8
```

(d) The modification code of FindWindowA()

```
75D36740 MOV ECX,DWORD PTR DS:[7FFE0324]
75D36746 MOV EDX,DWORD PTR DS:[7FFE0320]
75D3674C MOV EAX,DWORD PTR DS:[7FFE0328]
75D36751 CMP ECX,EAX
75D36753 JNZ SHORT KERNELBA.75D36770
75D36755 MOV EAX,DWORD PTR DS:[7FFE0004]
75D3675A MUL EDX
75D3675C SHL ECX,8
75D3675F IMUL ECX,DWORD PTR DS:[7FFE0004]
75D36766 SHRD EAX,EDX,18
75D3676A SHR EDX,18
75D3676D ADD EAX,ECX
75D3676F RETN
```

(e) The original code of GetTickCount()

```
75D36740 MOV EDX,7FFE0000
75D36746 XCHG EAX,EDI
75D36747 MOV EDI,5678ABCD
75D3674C XCHG EAX,EDI
75D3674D RETN
```

(f) The modification code of GetTickCount()

Figure 16: Results of *(iii) Malware-ex;* (a), (c) and (e) are the original code of each anti-debugging; (b), (d), and (f) are modification code to defeat each anti-debugging techniques

Table 6: Comparison results of anti-anti-debugging experiments with other works (O: successful, X: failed)

Technique	Proposed Work	Pin	Apate	Binary Substitution
`IsDebuggerPresent()`	O	O	O	O
`CheckRemoteDebuggerPresent()`	O	O	X	O
`FindWindow()`	O	O	X	O
`QueryInformationProcess()` (`ProcessDebugPort`)	O	O	X	O
`QueryInformationProcess()` (`ProcessDebugFlags`)	O	X	X	O
`QueryInformationProcess()` (`ProcessDebugObjectHandle`)	O	O	O	O
`GetTickCount()`	O	O	O	X
`timeGetTime()`	O	O	O	X
`GetCurrentProcessId()`	O	X	X	X
`BlockInput`	O	X	O	X
`PEB.BeingDebugged`	O	O	O	O
`PEB.NtGlobalFlag`	O	O	O	O
`HEAP.Flags`	O	O	O	X
`HEAP.ForceFlags`	O	O	O	X
`RDTSC` Instruction	X	X	O	O
`Self-Modifying`	X	X	X	X
Success Rate	87.5%	68.8%	62.5%	56.3%

Finally, this paper collected and experimented with malware, which has anti-debugging techniques for *(iii) Malware-ex*. As a result of the *(ii) Packed-ex* experiment, this paper listed the address and instructions in Figure 16, which shows the loaded malware sample results used in *(iii) Malware-ex* on the debugger and the results of defeating its anti-debugging techniques. Figure 16 (a) is the original code of `FindWindowExA()`, and Figure 16 (b) is the new code that defeats it. This study wrote a new code that was occupied in 0x00190000 virtual memory to defeat the function. Figure 16 (c) is the original code of `FindWindowA()`, and the code that defeats it is Figure 16 (d). This study occupied and wrote the new code on the memory address 0x00170000 to defeat this. As mentioned before, virtual memory addresses cannot always be the same and must occupy the memory with empty space. The last anti-debugging technique is `GetTickCount()`. Figure 16 (e) is the original code of this function, modified as in Figure 16 (f), in which malware failed to detect the presence of a debugger, and the malware analysis could proceed. Thus, this study experimented and confirmed step by step that the anti-anti-debugging proposed in this paper is effective.

Table 6 shows the experimental results for 16 anti-debugging techniques. The Binary Substitution [25], which shows the lowest success rate among each work, changes debuggee's binary. It requires special attention. Pin [4] is an excellent analysis tool, but it cannot defeat the anti-debugging techniques used by `Yoda's Protector`. Apate [23], which has a similar success rate to Pin, has also been unable to defeat the anti-debugging techniques used by `Yoda's Protector`, and it can defeat fewer anti-debugging techniques than Pin. As shown in Table 6, most anti-debugging techniques are based on APIs, PEB, and HEAP structures. However, there are far more anti-debugging techniques, and their research is needed for future works.

Among the many samples, there were some that this paper failed to analyze. Because this work can defeat some APIs and PEB structure-based anti-debugging techniques, the anti-debugging using different artifacts could not be defeated, such as `RDTSC` instruction, Memory Breakpoint, Self-Modifying, and Single-Step Detection [22].

5. Conclusion

Malware as well as commercial packers use various anti-debugging techniques to protect themselves from the analysis by reverse engineering. Anti-debugging techniques can bypass or neutralize debugging analysis. An analyst needs an advanced debugger and related knowledge in order to analyze malware with anti-debugging techniques. Thus, this study proposed an anti-anti-debugging against the anti-debugging techniques. In order to explain the anti-anti-debugging, we expounded the anti-debugging techniques used by malware and packers. Also, it set up a step by step experiment for verification. It implemented, experimented, and analyzed sample files, and packed files with anti-debugging techniques. Finally, it experimented with the actual malware with many anti-debugging techniques. As a result of a step by step experiment, it found that proposed anti-anti-debugging can defeat the anti-debugging techniques and deal with actual malware. It did not interfere with program execution flow as less as possible by minimizing the debuggee's direct modification. Of course, the method proposed in this paper cannot defeat all anti-debugging techniques. Nevertheless, the method serves an essential part in analyzing malware with a debugger. Furthermore, it improves the ability to handle sophisticated debugging evasion techniques.

Also, debuggees, such as `Yoda's Protector`, can call the anti-debugging techniques more than once to interfere with the analysis. However, the anti-anti-debugging proposed in this paper has the advantage of being able to defeat the anti-debugging

technique that is called multiple times by only modifying it once. There are several tools and methods to disable anti-debugging techniques, however, none elaborate on how to disable it or explain with assembly code. Thus, this paper aims to discover for analysts new ways to defeat anti-debugging techniques and help researchers develop their scripts for research and analyze malware.

Data Availability

The script file used in the experiments is available at https://github.com/goldbear564/antiantidebugging.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (No. NRF-2020R1A2C1012117, Development of machine learning based intrusion detection platform to secure end device in edge computing environment) and Institute of Information & communications Technology Planning & Evaluation (IITP) grant by the Korea government (MSIT) (No.2017-0-00158, Development of Cyber Threat Intelligence (CTI) analysis and information sharing technology for national cyber incident).

References

[1] R. Goyda, S. Sharma, and S. Bevinakoppa, "Obfuscation of stuxnet and flame malware," In. Proc. of the 3rd international conference on applied informatics and computing theory(AICT'12), 150--154, 2012.

[2] J. W. Kim, J. Bang, Y.S. Moon, and M. J. Choi, "Disabling anti-debugging techniques for unpacking system in user-level debugger," in 2019 International Conference on Information and Communication Technology Convergence(ICTC), Jeju Island, South Korea, 2019, doi:10.1109/ICTC46691.2019.8939719.

[3] M. N. Gagnon, S. Taylor, and A. K. Ghosh, "Software protection through anti-debugging," IEEE Security & Privacy, **5**(3), 82--84, 2007, doi:10.1109/MSP.2007.71.

[4] C.K. Luk, R. Chon, R. Muth, H. Patil, and A. Klauser, "Pin: building customized program analysis tools with dynamic instrumentation, " ACM SIGPLAN Notices, **40**(6), 190--200, 2005, doi:10.1145/1064978.1065034.

[5] B. B. Rad, M. Masrom, and S. Ibrahim, "Camouflage in malware: From encryption to metamorphism, " International Journal of Computer Science and Network Security, **12**(8), 74--83, 2012. doi:10.1007/s11416-017-0291-9

[6] S. Bardin, R. David, and J. Y. Marion, "Backward-bounded DSE: targeting infeasibility questions on obfuscated codes," 2017 IEEE Symposium on Security and Privacy, 633--651, 2017, doi:10.1109/SP.2017.36.

[7] T. Blazytko, M. Contag, and C. Aschermann, "Syntia: Synthesizing the Semantics of Obfuscated Code," In Proc. of USENIX Security Symposium, 643--659, 2017.

[8] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandram, and S. Venkatran man, "Robust intelligent malware detection using deep learning," IEEE Access, **7**, 46717--46738, 2019, doi:10.1109/ACCESS.2019.2906934.

[9] X. Chen, J. Andersen, Z. M. Mao, and M. Bailey, "Towards an understanding of anti-virtualization and anti-debugging behavior in modern malware," in 2008 IEEE international conference on dependable systems and networks with FTCS and DCC(DSN), 177--186, 2008, doi:10.1109/DSN.2008.4630086.

[10] M. J. Choi, J. Bang, J. W. Kim, H. Kim, and Y. S. Moon, "All-in-one framework for detection, unpacking, and verification for malware analysis," Security and Communication Network 2019, **2019**(5278137), 1--16, 2019, doi:10.1155/2019/5278137.

[11] S. Cesare, Y. Xiang, and W. Zhou, "Malwise – an effective and efficient classification system for packed and polymorphic malware," IEEE Transaction on Computers, **62**(6), 1193--1206, 2013, doi:10.1109/TC.2012.65.

[12] F. Guo, P. Ferrie, and T. C. Chiueh, "A study of the packer problem and its solutions," in International Workshop on Recent Advances in Intrusion Detection, 98--115, 2008, doi:10.1007/978-3-540-87403-4_6.

[13] J. M. Borello, and L. Mé, "Code obfuscation techniques for metamorphic viruses," Journal in Computer Virology, **4**(3), 211--200, 2008, doi:10.1007/s11416-008-0084-2.

[14] I. You, and K. Yim, "Malware obfuscation techniques: a brief survey," in 2010 International conference on broadband, wireless computing, communication and applications, 297--300, 2010, doi:10.1109/BWCCA.2010.85.

[15] A. Moser, C. Kruegel, and E. Kirda, "Limits of static analysis for malware detection," in 23rd Annual Computer Security Applications Conference(ACSAC'07), 421--430, 2007, doi:10.1109/ACSAC.2007.21.

[16] R. R. Branco, G. N. Barbosa, and P. D. Neo, "Scientific but not academical overview of malware anti-debugging, anti-disassembly and anti-vm technologies," in Blackhat, USA, 2012.

[17] S. Gao, and Q. Lin, "Debugging classification and anti-debugging strategies," in 4th International Conference on Machine Vision(ICMV): Computer Vision and Image Analysis; Pattern Recognition and Basic Technologies, 729--734, 2012, doi:10.1117/12.924835.

[18] P. Chen, C. Huygens, L.Desmet, and W. Joosen, "Advanced or not? a comparative study of the use of anti-debugging and anti-vm techniques in generic and targeted malware," in IFIP International Conference on ICT Systems Security and Privacy Protection, 323--336, 2016, doi:10.1007/978-3-319-33630-5_22.

[19] S. Bhansali, W. K. Chen, S. D. Jong, A. Edwards, "Framework for instruction-level tracing and analysis of program executions," In Proc. of the 2nd International Conference on Virtual Execution Environments, 154--163, 2013, doi:10.1145/1134760.1220164.

[20] P. Feiner, A. D. Brown, and A. Goel, "Comprehensive kernel instrumentation via dynamic binary translation," In Proc. of the 17th international conference on Architectural Support for Programming Languages and Operating Systems, 135--146, 2012, doi:10.1145/2150976.2150992.

[21] N. Nethercote, and J. Seward, "Valgrind: a framework for heavyweight dynamic binary instrumentation," ACM SIGPLAN Notices, **42**(6), 89--100, 2007, doi:10.1145/1273442.1250746.

[22] S. Choi, T. Chang, S. Yoon, and Y. Park, "Hybrid emulation for bypassing anti-reversing techniques and analyzing malware," The Journal of Supercomputing, 1--27, 2020, doi:10.1007/s11227-020-03270-6

[23] H. Shi, and J. Mirkovic, "Hiding debuggers from malware with apate," In Proc. of the Symposium on Applied Computing, 1703--1710, 2017, doi:10.1145/3019612.3019791.

[24] J. K. Lee, B. J. Kang, and E. G. Im, "Rule-based anti-anti-debugging system," In Proc. of the 2013 Research in Adaptive and Convergent Systems, 353--354, 2013, doi:10.1145/2513228.2513301.

[25] J. K. Lee, B. J. Kang, and E. G. Im, "Evading anti-debugging techniques with binary substitution," International Journal of Security and its Applications, **8**(1), 183--192, 2014, doi:10.14257/ijsia.2014.8.1.17.

[26] P. Feng, J. Sun, S. Liu, and K. Sun, "UBER: combating sandbox evasion via user behavior emulators," International Conference on Information and Communications Security, 34--50, 2020, doi:10.1007/978-3-030-41579-2_3

[27] D. P. Pham, D. L. Vu, and F. Massacci, "Mac-A-Mal: macOS malware analysis framework resistant to anti evasion techniques," Journal of Computer Virology and Hacking Techniques, **15**(4), 249--257, 2019, doi:10.1007/s11416-019-00335-w.

[28] O. Or-Meir, N. Nissim, Y. Elovici, and L. Rokach, "Dynamic malware analysis in the modern era- a state of the art survey," ACM Computing Surveys (CSUR), **52**(5), 1-48, 2019, doi: 10.1145/3329786.

[29] C. V. Liţă, D. Cosovan, and D. Gavriluţ. "Anti-emulation trends in modern packers: a survey on the evolution of anti-emulation techniques in UPA packers," Journal of Computer Virology and Hacking Techniques, **14**(2), 107--126, 2018, doi:10.1007/s11416-017-0291-9.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1190-1206 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Development of a Technology and Digital Transformation Adoption Framework of the Postal Industry in Southern Africa: From Critical Literature Review to a Theoretical Framework

Kgabo Mokgohloa*, Grace Kanakana-Katumba, Rendani Maladzhi

Department of Mechanical and Industrial Engineering, University of South Africa, Johannesburg, 0002, South Africa

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 23 November, 2020
Online: 14 December, 2020

Keywords:
Technology adoption
Systems approach
Digital transformation

ABSTRACT

The pressing and most urgent challenge for most of the Posts in Southern Africa is to transit from its historical reputation of being a snail-paced, inefficient, loss-making and ineffective service provider to an agile, innovative, solution-driven and highly competitive service provider that strive for excellence. In this context, adoption, and diffusion of technology and digital transformation by Posts should deliver the necessary traction to steer the digital transformation journey of the postal industry in Southern Africa with the necessary velocity while cognizant and mindful of the appropriateness of the technology in the African context in the light of a VUCA (Volatile, Uncertain, Complex and Ambiguous) world. The traditional technology adoption models are characterized by "linearity" which is the opposite of a dynamic setting characterized by "causality" which is the fundamental principle of a system thinking approach. The multi-stakeholders in the postal sector with often competing interests renders a linear approach defunct and requires a system approach to digital transformation and technology adoption. The conceptual framework developed from the critical review of literature offers considerable leverage to postal sector in developing countries and to Southern Africa in particular. The proposed conceptual framework integrates dimensions such as Postal Industry 4.0 Envisioning, Strategy, Institutional factors, Organizational factors, Individual factors, Industry 4.0 environment and outcomes.

1. Introduction

This paper is an extension of "Postal Development: Literature Review into Adoption Models" presented at IEEM 2019 Conference in Macau by authors in [1]. The rise of digitalization powered by a digital revolution of the 21st century ushered by the Fourth Industrial Revolution and personified by Industry 4.0, has prompted the postal sector across the world to expand their services well beyond the original service of postal operators (POs) which was delivery of physical mail in the form of letters and post cards.

In [2], the authors propose that although there are Posts in different nations that are struggling to keep afloat due to economic and political turbulence there are other Posts that are trailblazers in postal digital transformation which is underpinned by digital innovation which leads to digital disruption. It comes as no surprise that Swiss Post, The Netherland Post, Deutsche Post, La Poste, and Japan Post are trailblazers in digital innovation.

In [3], the authors postulates that societal anatomy is rapidly changing, and that the digital age shepherded changes in consumption patterns which has resulted in the evolution of the customer of the future with unique needs and expectations that industries should meet. A systems (holistic) approach is required to comprehend a new world that is driven by volatility, uncertainty, complexity, and ambiguity (VUCA). VUCA world is categorised by the traits of volatility, uncertainty, complexity and ambiguity [4]. In [5], the authors argue that Volatility, Uncertainty, Complexity and Ambiguity (VUCA) explains the competitive ecosystem of the digital economy in which corporations ought to acclimate business process to match up briskly shifting and more complex dynamic settings.

In [6], the authors propose that in the current Volatile, Uncertain, Complex and Ambiguous (VUCA) world, organizations and employees are faced with a plethora of challenges that overly dynamic and totally different from the challenges of the past. In [6], the authors further argue that there are several of drivers that propel these changes, and that

*Corresponding Author: Kgabo Mokgohloa, mokgok@unisa.ac.za

technology, accessibility to information, a fast shifting global economy, the establishment of a global consumer society and a very insistent customer with unique and diverse requirements are only some influences that have shaped an extremely complex and ambiguous world.

In [7], the authors advance that organizational and societal systems in the digital era encompass characteristics of volatility, uncertainty, complexity, and ambiguity; and as a result a systems approach to digital transformation and technology adoption is a key success factor for Posts in Southern Africa.

1.1. Problem statement

The crises that has befell on the Designated Postal Operators (DPOs) in Southern Africa has precipitated the near collapse of the postal logistics sector due poor performance on the four performance indicators of the 2IPD and consequently losses except for a few Designated Postal Operators making a profit, this is characterized by the scores respective Post in Southern Africa obtained on the 2IPD.

The underperformance by SADC DPOs is fuelled by a plethora of dynamics including inflexible or rather outdated business models that are not adapting to the technological changes that are disrupting current business models immensely, the archaic business models cannot compete in the 21st century and this result in majority of Postal Operators in Southern Africa posting losses year in year out. The unsustainability of the DPOs leave the State in respective countries with no option but to bail out these institutions. This performance is contrary to the first-tier countries who occupy pole positions on the integrated index on postal development (2IDP).

1.2. Research question

Can a technology and digital adoption framework in the context of postal industry 4.0 be developed for Posts in Southern Africa?

1.3. Concluding remarks to introduction

The paper is divided into (a) Introduction which introduces the phenomenon under study as well as articulating the problem statement and the research question which guided the research study. (b) Literature review which delved into the postal landscape, articulation of the Integrated Index on Postal Development (2IDP) which is a global measure of performance of the postal sector globally, and with specific reference to the top 5 performers on the 2IDP and digital readiness index. The literature review further delved into the phenomenon of industry 4.0, technology adoption, digital transformation, reductionist vs systems approach and capability maturity models. The report then detailed the research methods adopted, the discussion and findings that included the development of the theoretical framework, deductions and upcoming research concludes the study.

2. Literature review

2.1. The Postal landscape

In [2], the authors articulates an international view point of the Postal Industry, accentuating that the postal network encompasses a worldwide grid of over 677 000 retail offices, 5.3 million members and substantial infrastructure comprising 192 nations.

In [8], the authors hypothesizes that imperfect internet connectivity, inadequate investment and/or recapitalization of the network infrastructure, deprived value of service, restricted interoperability and connectivity among the systems of value chain participants, as well as uncertainty in postal industry characterisation in southern Africa are the foremost motives for the snail-paced advancement.

In [9], the authors argues that the regional scenery of the postal industry in southern Africa encompass 50 mail centres and a distribution infrastructure of 16, 064.394 deliverance points. Of these, South Africa has a lion share of 26 mail centres and 14, 106.896 delivery points. The postal industry in Southern Africa comprise of fourteen designated postal operators in South Africa, Botswana, Zambia, Eswatini, Seychelles, Democratic Republic of Congo, Malawi, Comoros, Madagascar, Mauritius, Tanzania, Lesotho, Mozambique, and Angola.

2.2. The Integrated Index on Postal Development

In [3], the authors note that most Posts are performing disappointingly on the Integrated Index for Postal Development (2IPD). In [3], the authors further proposes that measuring the multiple facets of postal development is no insignificant task, and hypothesize that in order to surmount this difficulty; the Universal Postal Union (UPU) has been tapping on abundance of postal big data to appraise the performance of Posts globally. One of the most important sequels of these endeavours was the development of the Integrated Index for Postal Development commonly known as the 2IPD. In [10], the author argues that 2IPD gauges the performance of Posts in the four vital aspects of postal advancement which are known as the 4Rs (reliability, reach, relevance and resilience).

In [11], the authors outlines the four aspects of postal advancement (development) as follows: (a) Reliability is a fusion of excellence of service delivery, including certainty across all categories of postal delivery service; it eventually gauges the level of postal operational efficiency (b) Reach is fusion of worldwide postal connectedness at intercontinental level across all categories of international postal delivery services, it ultimately gauges the level of globalization of the postal services (c) Relevance encompass the strength of demand for the full range of postal services in each postal section; it eventually gauges the level of attractiveness of the postal market (d) Resilience encompass ability to innovate, deliver wide-ranging postal services and assimilate sustainable development targets in postal business models; it ultimately gauges the level of adaptableness of postal business models.

In [3], the authors advances that a total of 172 countries were analysed for the development of the 2019 global ranking (based on complete data for 2018). Table 1 depicts the top 5 countries in alphabetical order with their corresponding scores on the 2IDP and are as follows: France, Germany, Japan, Netherlands, and Switzerland. In [12], the authors postulates that these top 5 DPO's are performing outstandingly on the 2IDP because they have steadily build dependable, well-connected, appropriate and agile postal services. The report further highlights that other DPOs

especially in developing countries are underperforming relative to the 2IDP indicators.

Table.1: Top 5 Designated Postal Operators performance on Integrated Index for Postal Development, adapted from [12]

Country	2IPD Score 2018	2IPD Score 2019
France	83.3	86.86
Germany	91.3	90.79
Japan	91.6	87.19
Netherlands	93.7	93.67
Switzerland	100	100

The analysis of performance on the 2IPD points to Posts in developing countries (Southern Africa included) as laggards' when compared to the top performing DPOs on the 2IPD. Table 2 lists the SADC (Southern African Development Community) Designated Postal Operators with their corresponding 2IPD scores.

Table.2: Southern African Development Community Designated Postal Operators performance on Integrated Index for Postal Development, adapted from [12]

Country	2IPD Scores 2018	2IPD Scores 2019
Angola	21.22	30.54
Botswana	23.72	21.00
Comoros	11.29	11.76
Democratic Republic of Congo	15.87	17.01
Eswatini	24.72	15.80
Lesotho	14.09	13.44
Madagascar	30.96	30.63
Malawi	24.78	13.74
Mauritius	49.17	40.53
Mozambique	4.95	6.28
Namibia	31.28	27.52
Seychelles	30.68	24.00
South Africa	33.34	33.34
Tanzania (Republic)	31.81	39.12
Zambia	7.73	8.75
Zimbabwe	19.53	12.74

In [13], the authors explains that the industrial sector has always been critical to the economic development of countries. Since the end of the 18th century, businesses have gone through enormous modifications that revolutionized a way how goods are produced and has yielded numerous benefits, predominantly related to efficiencies and productivity. In [14], the authors argues that presently, after three preceding industrial revolutions, the combination of advanced technologies and the internet age is completely transfiguring the industrial landscape and it is labelled the Fourth Industrial Revolution or its personification known as the Industry 4.0. The postal sector is not immune to these fast-paced technological revolutions that continue to chance the postal landscape daily.

The rise of digitalization brought upon the digital age has prompted the postal industry globally to enlarge their services further than the traditional mail delivery service of Posts. In [15], the authors hypothesise that with the global adoption of digitalization and digitization at an exponential rate, stakeholders are progressively demanding to interact directly with the postal service via digital networks.

2.3. Industry 4.0

In [13], the authors elucidates that the industrial sector has always been critical to the economic development of countries. Since the end of the 18th century, businesses have gone through enormous modifications that revolutionized a way how goods are produced and has yielded numerous benefits, predominantly related to efficiencies and productivity. In [14], the authors argue that today, after three preceding industrial revolutions, the combination of advanced technologies and the internet age is completely transfiguring the industrial landscape.

In [16], the authors argue that the aim of Industry 4.0 is to digitize and incorporate processes vertically across the entire value chain of the organisation, from conceptualization all the way to sales and after-sales support. Data throughout the value chain is available instantaneously, backed by augmented reality and optimized in a unified network Industry 4.0 is a farsighted initiative of the German State that traditionally supports advancement of the German industrial sector. In [17], the authors support this view and propose that in this sense, Industry 4.0 can be viewed as an action towards nurturing Germany's posture as one of the most leading technologically advanced nation.

In [18], the authors postulate that Industry 4.0 is envisioned at the fixated assemblage and application of prompt data and information by means of linking all individual rudiments of a system, with a view of reducing the intricacy of processes, while raising the efficacy and value of operations.

Figure 1 presents the Industry 4.0 framework and associated technologies at which the core of Industry 4.0 is based on the concept of the three underlying fundamentals which are:

- Digitalization and amalgamation of vertical and horizontal value chains
- Digitalization of offerings
- Digital industry models and client access

In [16], the author notes that the underpinnings to industry 4.0 could be elaborated as follows:

- Digitalization and amalgamation of horizontal & vertical value chains comprise of two features; The initial feature comprises amalgamation of processes vertically across the entire firm and entails all business processes from conceptual

design all the way to delivery of the product or service to the customer or end-user.

- Digitalization of offerings incorporates the extension of current offerings which includes the addition of smart sensors that could be utilized with data analysis tools as well as development of new digitized offerings. The ability to analyze big data could benefit firms to better respond to changing customer demands.
- Digital industry models and client access encompasses enlargement of offering by supporting disruptive digital solutions. Disruptive digital industry models have potential to create new revenue streams by responding with speed to the changing customer demand changes in a dynamic setting.

In [16], the authors argue that big data and data analysis tools are the principal competences of Industry 4.0 and is enabled by technologies such as smart sensors, advanced algorithms, Internet of Things, location technologies, and many other technologies that support the digital revolution.

Figure 1 presents the Industry 4.0 setting in which an argument is advanced that the crucial significance to Industry 4.0 is its ability to communicate with other smart "things" such as the digital capabilities of smart transportation, smart electrical grid, smart supply chain management, and other smart infrastructure. In [18], the authors argue that the linkage between social web and business web is of critical importance and will see further radical transformation in the future and will further reduce the digital divide between business and society at large as technology develops and becomes accessible to society.

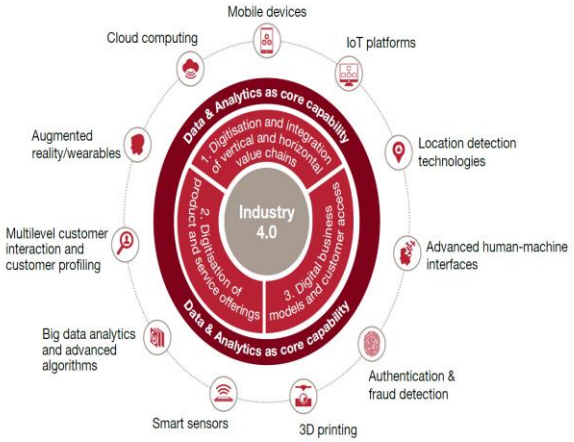


Figure 1: Industry 4.0 framework and contributing digital technologies [16]

In [19], the authors suggests that the important elements that outline the notion of Industry 4.0 are:

- Cyber-physical system (CPS), the term that describes the amalgamation of cyber with physical systems. CPS incorporates devices which can assemble and transmit data through the internet.
- Internet of things (IoT) facilitates the interaction with other systems and between systems and ultimately with the users.
- The ability to assemble data and analyze the data into useful information instantaneously is an important feature of a CPS and enables decision makers to make informed decisions.

In [20], the authors advocates that the concept of Industry 4.0 is underpinned by six design principles that are pertinent to its efficacious implementation in respective industries which are interoperability, virtualization, decentralization, real-time capability, service orientation and modularity depicted in Figure 2.

	Cyber-Physical Systems	Internet of Things	Internet of Services	Smart Factory
Interoperability	X	X	X	X
Virtualization	X	-	-	X
Decentralization	X	-	-	X
Real-Time Capability	-	-	-	X
Service Orientation	-	-	X	-
Modularity	-	-	X	-

Figure 2: Design principle for each of the main 4 Industry 4.0 components [20]

The Fourth Industrial Revolution and its application as espoused by Industry 4.0 offers both challenges and opportunities for the postal sector, the major challenge is that Posts can no longer operate using business models of the past because digitalization of the Post is something that is not optional but mandatory if the Post still must survive another 100 years.

In [21], the authors argue that today, the concept of the internet is often reflected as "given" due to its omnipresent manifestation and quickening effect on the ways in which humanity relates. The world is changing rapidly because of the digital revolution. In [22], the authors argue that one such manifestation of the digital revolution is the IOT (Internet of things) with which digitization is no longer a simulated reality. IOT links several "things" through the internet and allows them to communicate information amongst each other. In [23], the authors propose that the Internet of Things (IoT) is sensor technologies that enable physical objects to gather and transmit data through the Internet in instantaneously. It is one of the latest advances taking hold in many industries and most appropriate for the postal sector due to its possible applications in the postal value chain.

In [24], the authors propose that the IoT architype is envisioned at devising a sophisticated system with the combination sensors and associated technologies with ability to communicate in real-time. Contrarywise, assembling data and preserving the privacy of users, and then taking cognizance of confidentiality and internet security specification in IoT remains a significant concern amongst stakeholders in the IoT space.

In [23], the authors suggests that the current rise of IoT applications is fuelled by a conjunction of issues such as omnipresent connectivity and the diminishing expenditure and enhanced performance of measuring devices (sensors) and analytics. These technology developments relate to expanding consumer requirement for more data on the products and services they purchase. The flow of information is crucial to the "flow of things".

The United States Postal Service has embodied the Internet of Things (IoT) and coined the Internet of Postal Things (IoPT). In [23], the authors articulate a postal standpoint and reason that IoT can be better pronounced as "Things on the Internet" which are

unified webs of distinctively recognized physical "things" embedded with sensors that amass, converse, and act upon a widespread diversity of data such as locality pin-pointing, temperature, and movement. These "smart objects" amass and transmit data that can be scrutinized for improvement of business processes across organizations and for improved solutions to customer demands. Figure 3 depicts the chief elements of the IoT value-chain.

In [23], the authors argue that IoT unwraps new opportunities for various organizations and in particular entities like the postal service which has a geographic reach that surpass many industries. In [25], in the authors propose that the postal service world-wide has a widespread network of over 677,000 post offices, 5.32 million staff and physical network of buildings encompassing 192 countries and asserts that the postal sector performs a significant socio-economic role in society at large.

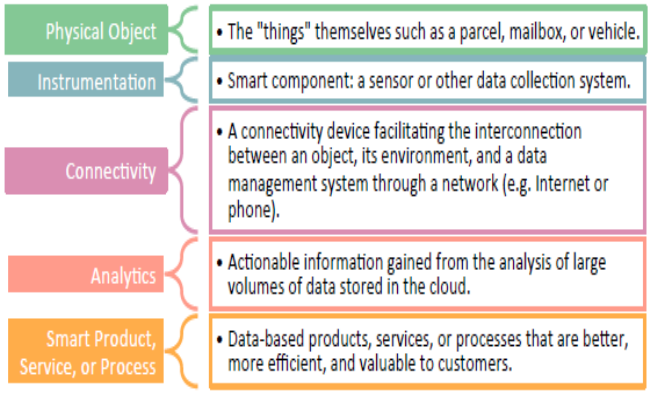


Figure 3: IoT value chain [23]

Posts could significantly profit in implementing Industry 4.0 fundamentals in the quest to adapt to challenges and opportunities presented by the digital era, and could greatly improve service delivery to new market demands and improve their sustainability (both economic, social and environmental). Applications of Industry 4.0 fundamentals can significantly improve sustainability of Posts in Southern Africa.

2.4. Innovation Diffusion Theory

In [26], the authors postulates that the most noticeable and dominant attempt to dissect the dynamics influencing the adoption and diffusion practices was pioneered by [27], however the authors argue that the framework is appropriate for conceptualization of adoption and diffusion from a qualitative point of view as opposed to quantitative prediction of adoption of innovative practices. In [28], the author argues that diffusion is the process by which an invention is conversed through networks over time amongst the participants of a network. In [29], the authors argue that technology adoption is a developed area of research in information systems (IS).

In [30], the authors articulates the difference between innovation and technology which is often confused and argues advance that technology is concerned in resolving real social problems to advance humanity. In [31], the authors argue that innovation on the other hand comprises of the creation of an innovative concepts and development of the concept into new offerings resulting in economic development. In [32], the authors argue that innovation is not merely creativity, research and development, invention, bright ideas or fancy gadgets. In [33], the authors argue that innovation is rather a development of new resources or new resource capacity that are able to create wealth. In [34], the authors argue that innovation is a complex, organization-wide effort, that demands a set of procedures and activities to shape, coordinate, and promote it.

In [28], the authors investigates drivers that define the speed of acceptance (adoption) of innovations and argues that the following factors regulate the degree of adoption. (a) Apparent characteristic of innovation affected by considerations such as relativity of the advantage to adopt, compatibility of adoption to current systems, complexity of the innovation, trialability of the innovation, and observability of the results. (b) Innovation choice, which is driven by issues such as is the innovation optional or mandatory (c) Type of societal structures including a component such as the degree of linkages of the societal structures; and (d) The effectiveness and efficiency of transformation (change) champions in advocating transformational change.

In [11], the authors suggests that each invention has distinctive attributes that affect the dispersion (diffusion) process. These attributes can be broken-down into five elements:

(a) Relative advantage is the extent to which an invention is understood as being superior to the concept it succeeds, typically stated as cost-effectiveness, societal stature, or other advantages.
(b) Compatibility as recognized by participants in society is constructively related to the level of acceptance (adoption) and influences the way customers act towards an invention.
(c) Complexity is the extent to which an invention is understood as comparatively complicated to comprehend and utilize.
(d) Trialability is the extent to which an invention may be tested on a restricted basis. The trialability of an invention is certainly related to the pace of acceptance (adoption).
(e) Observability is the extent to which outcomes are observable to the user and to others.

2.5. Technology Adoption Theory

In [35], the authors argues that Theory of Reasoned Action (TRA) is the early notion that attempted to explain user adoption of technology. TRA describes user behaviour from a social point, and with a specific aim of pin-pointing the causes of conscious behaviour. In [36], the authors propose that the model (Technology Acceptance Model- TAM) was developed from TRA as its basis but further refined.

In [37], the authors argues that the basis of TRA is that behaviour is predicted by intention (I) and further highlights that intentions are conjointly driven by two factors (a) Attitude (A) towards the act which is a function of attitudes about effects of this conduct and (b) subjective norms (SN) defined as a person's viewpoint of whether stakeholders key to the person consider the actions should be implemented. Attitude towards the behaviour is defined as the individual's negative or positive outlook about executing the behaviour.

In [38], the authors propose that TRA contends that views influence mindsets, which lead to intent and as a result engender behaviour. The TRA is graphically illustrated in Figure 4.

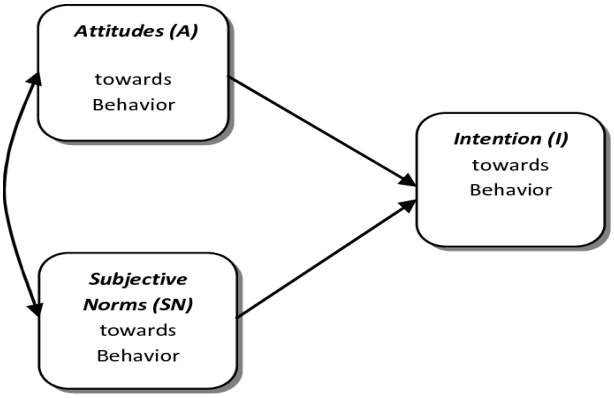


Figure 4: Theory of reasonable action. Source [37].

In [39], the authors state that consumer rationale to accept and adopt a technology could be explained by perceived usefulness (PU), perceived ease of use (PEU) and attitude towards the system. In [39], the authors hypothesize that attitude of a user towards the system was the core factor to determine if the user will accept or reject the innovation, and that attitude is directly influenced by (a) Perceived Usefulness (PU) and (b) Perceived Ease of Use (PEU); and concludes that apparent ease of use unswervingly influence perceived usefulness.

In [38], in the authors argue the concepts in the earliest TAM (see Figure 5) as follows: Perceived Usefulness (PU), Perceived Ease Of Use (PEOU), Attitude, and Behavioural Intention to Use. Among the constructs, PU and PEOU establish a user's confidence on a technology and consequently predict his or her stance on the technology, which in turn foresees its approval.

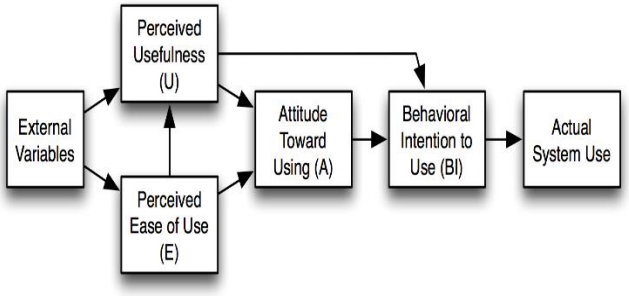


Figure 5: Technology Adoption Model. Source [38]

In [40], the authors maintains that while the TRA was intended to illuminate essentially any human behaviour, the TAM on the other hand was intended to offer an clarification of the factors of computer acceptance throughout a comprehensive range of end-user computing technologies and user populations. In [37], the authors subsequently contends that according to the TAM, actual system usage is directly defined by behavioural intention (BI) to use. BI is in turn swayed by the user's mindset towards utilizing the system and perceived usefulness. Attitude is propelled by perceived usefulness (PU) and perceived ease of use (PEOU) of the system. Both PU and PEOU are influenced by external variables, such as individual variations, situational constrictions, organizational traits, and system attributes.

As the TAM progressed past its initial positioning, several researchers established shortcomings in the model which required to be tackled. The outcome of research that was conducted by several researchers ensued in adjustments to TAM. In [37], the authors suggest that the adjustments entailed the establishment of additional variables and concepts resulting in TAM2 as depicted in Figure 6.

In [37], the authors argue that the major additions to the TAM that were introduced by TAM2 were the inclusion of societal influence activities (subjective norm, voluntariness, and image) and mental method procedures (job relevance, output quality and results demonstrability).

TAM2 theorizes that users' mental assessment of the match between considerable goals at work and the subsequent effects of executing tasks using the system provides a basis for establishing views concerning the usefulness of the system. The results of the research by uncovered that TAM 2 performed well in both voluntary and obligatory environments.

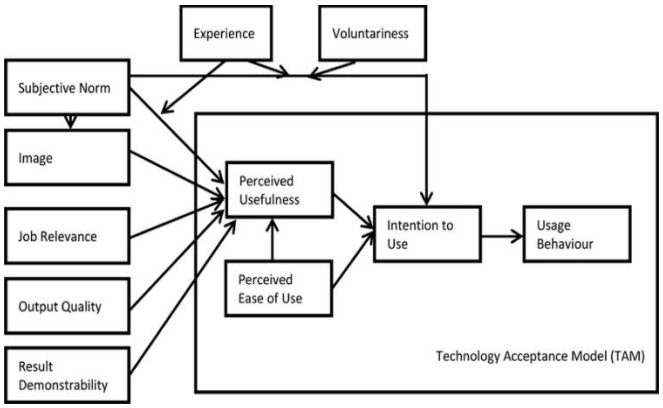


Figure 6: Technology Adoption Model 2. Source [42].

In [37], the authors argues that the succeeding foremost addition to TAM was the Unified Theory of Acceptance and Use of Technology (UTAUT), the author further proposes UTAUT was framed, with four fundamental determining factor of intention and usage and up to four moderators of crucial relations. The theory proposed that four important concepts perform a significant part as direct factors of user acceptance and usage behaviour: (a) performance expectancy (b) effort expectancy (c) social influence and (d) facilitating conditions. The UTAUT is presented in Figure 7.

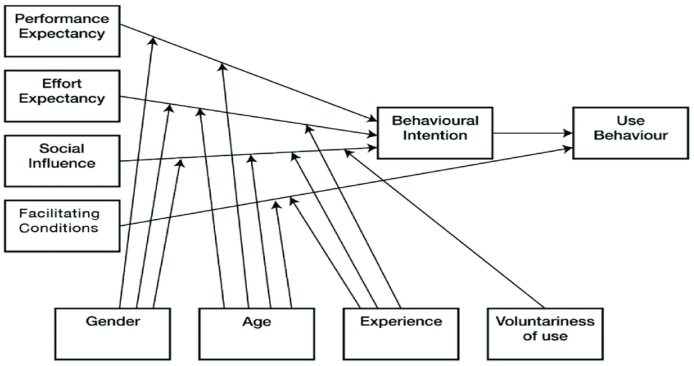


Figure 7: Unified Theory of Acceptance and Use of Technology (UTAUT. Source [43].

In [37], the authors propose that UTAUT advocates that facilitating conditions would unswervingly impact usage, while the remaining three concepts would influence Behavioural Intention (BI). In [44] the authors argue that the model describes facilitating conditions as the extent to which a person deems that organizational and technical infrastructure exists to support the utilization of a system they are accepting (adopting). In [45], the authors argue that the UTAUT further extended the concepts in the earliest TAM through the recognition of the vital role of endeavour in performing the new task and the role support structures determines adoption in organizations.

In [46], the authors argues that TOE (Technology-Organization-Environment) framework entails three categories of settings that may influence technological innovation adoption and implementation process. TOE is premised on the three context which are technological, organizational, and environmental. In [47], as represented in Figure 8, the authors explains that technological setting defines the internal and external technologies suitable to the organization, while organizational setting denotes descriptive actions about the organizational scope, size and structure; finally, environmental setting is the arena in which an organization conducts its business.

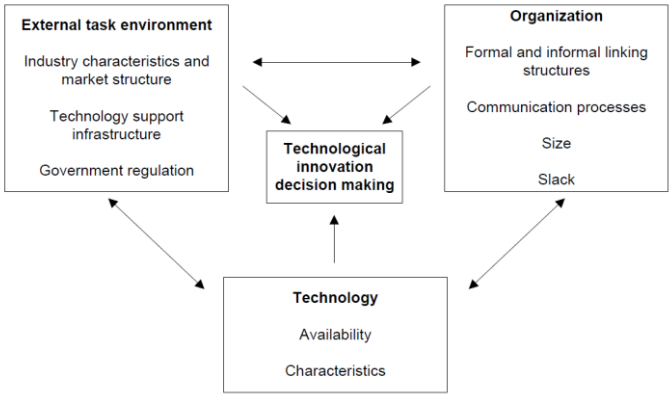


Figure 8: Technology, organization, and environment framework

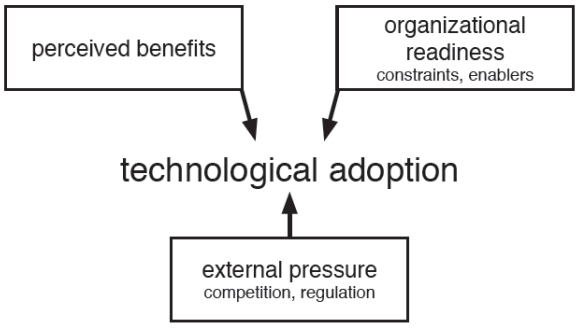


Figure 9: BOE Model [48]

In [48], the authors postulates that the TOE model defines the dynamics for technological adoption, focusing on the external stresses (e.g. market forces and governmental regulatory requirements), organizational structures, and technological availability; in addition the BOE model was originally developed to understand the adoption of EDI (electronic data interchange) technology, but has since been used as a general technology adoption model. The BOE model comprises three factors: external pressure, organizational readiness, and perceived benefits.

In [48], the authors advance that by distinction, the BOE model as illustrated in Figure 9 (Benefits, Organizational Readiness and External Pressures) integrates the organization and technology context of TOE into organizational readiness and augments a perceived benefit aspect for viewing at adoption of technology.

In [26], the authors argue that the Adoption and Diffusion Outcome Prediction Tool (ADOPT) was developed as a tool to encourage application in project planning and implementation of sound insight of the social and economic factors affecting acceptance (adoption) of agricultural innovations. In [49], the authors proposes that ADOPT focuses on factors that have demonstrated to impact the pace and/or highest point of adoption within a population. These are (a) attributes of the innovation, (b) attributes of the population, (c) tangible relative advantage of employing the innovation, and (d) learning of the actual relative advantage of the technology. In [26] the authors proposes that the variables of the theoretical context can be divided into two main categories: (a) those that relate to traits of the target population, and (b) those that relate to traits of the practice. These elements are presented in Figure 10.

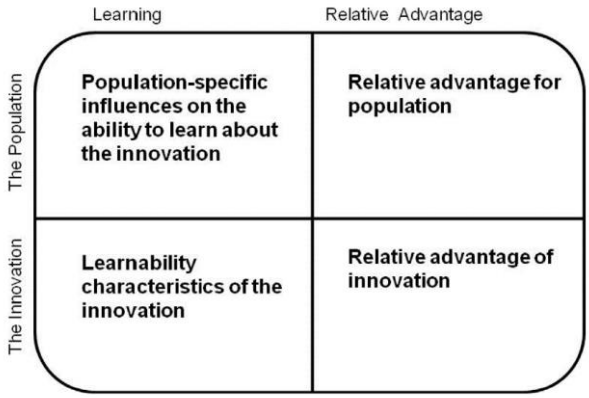


Figure 10: Conceptual characterization of factors included in ADOPT

2.6. Capability Maturity Models

In [50], the authors states that the original Capability Maturity Model (CMM) was developed in the software engineering space to support advances in the dependability of software development organizations, that is, in their aptitude to deliver software within time, performance, cost and quality. More specifically, it was "intended to assist developers to choose process-improvement strategies by defining their existing process maturity and finding the most critical issues to improving their software quality and process".

In [51], the authors describes the five maturity levels and further emphasize the basic process changes made at each level:

- Initial: The process is described as ad hoc, and sporadically even anarchic. Limited processes are defined, and accomplishment depends on individual effort.
- Repeatable: Fundamental business processes are established to track expenditure, plans, and functionality.

- Defined: The business processes are documented, standardized, and integrated.
- Managed: Detailed measures of the business processes are collected. Business processes are quantitatively understood and controlled.
- Optimizing: Continuous process improvement is enabled by quantifiable feedback from the business process.

The CMM provides a framework for shaping these incremental steps into five maturity levels that lie down consecutive grounds for continuous process improvement. These five maturity levels characterize an ordinal scale for evaluating the maturity of an organization's processes and for appraising its process capability. The levels also help an organization highlight its improvement endeavours. The steps above are graphically presented in Figure 11.

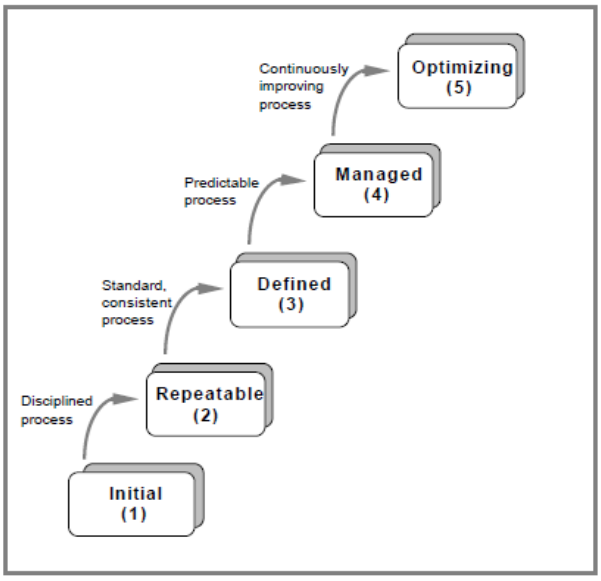


Figure 11: The five levels of software maturity [51]

In [52], the authors argue that capability maturity models are used in three ways; (a) Descriptive use which focuses on (1) To assess the 'as-is' state of specific organizational capabilities (2) As an analytical diagnostic tool (3) To communicate maturity levels to internal and external stakeholders (b) Prescriptive use which focus on (1) To ascertain desired maturity level and receive guidance on how to continuously improve from current to desired level (2) To follow specific and detailed sequences of action (3) Comparative use which emphasizes on comparing the performance of the specific organisation through internal or external benchmarking (best practices).

In [53], the authors affirms that one of the barriers to industry's wider headway in digital transformation is the absence of a vivid industry focused blueprint. The report proposes that the Digital Maturity Model presented in Figure 12 is a valuable instrument to offer guidelines for a vivid pathway throughout the transformation journey.

The five core dimensions underpinning the model are:

- Customer, this facet is focused on offering an understanding where consumers perceive the organization as their digital collaborator utilizing their desired networks of interaction to control their connected future on and offline
- Strategy, this facet stresses on how organizations transform or operate to enhance its competitive advantage through digital projects; it is entrenched within the whole organizational strategy
- Technology, this facet emphasizes the success of digital strategy by enabling the creation, processing, storage, securing and transmission of data to meet the needs of customers effectively and efficiently.
- Operations, this facet is premised on executing and evolving processes and tasks by leveraging on digital technologies to drive strategic management and enhance business efficiency and effectiveness
- Organization and culture, this facet focuses on identifying and cultivating an organizational culture with governance and talent management processes to boost advancement along the digital maturity curve, and the ability to realize expansion and invention (innovation) goals.

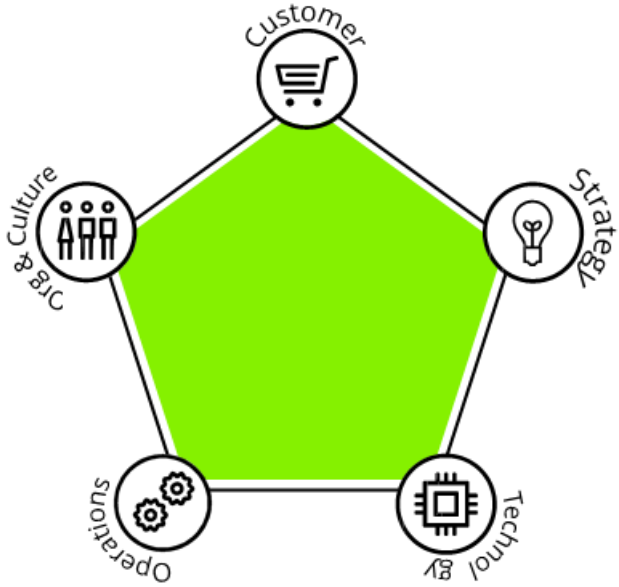


Figure 12: Deloitte Digital Maturity Model [53]

Figure 13 crystalizes the sub-dimensions that are associated with the five core dimensions presented in Figure 12. The sub-dimensions elucidate and unpacks each of the five core dimensions that are critical to digital transformation, these dynamics are such as customer engagement, customer insight and behaviour, brand management, finance & investment, connected things, data & analytics, real-time insights & analytics, agile change management, organizational culture, leadership & governance, organizational design & talent management, workforce enablement, automated resource management, stakeholder management, ecosystem management, strategic management, customer trust & perception, market & customer, network, security technology architecture and innovation management are key towards a successful digital transformation journey.

In [52], the authors propose a Digital Maturity Model which consists of four stages. The stages proposed in their model are (a) Initiate (b) Competent (c) Purposeful and lastly (d)

Transformative. The key attributes of businesses related to each of these maturity stages are presented in Figure 14.

Customer	Strategy	Technology	Operations	Organisation & Culture
Customer Engagement	Brand Management	Applications	Agile Change Management	Culture
Customer Experience	Ecosystem Management	Connected Things	Automated Resource Management	Leadership and Governance
Customer Insights & Behaviour	Finance & Investment	Data & Analytics	Integrated Service Management	Organizational Design & Talent Management
Customer Trust & Perception	Market & Customer	Delivery Governance	Real-time Insights & Analytics	Workforce Enablement
	Portfolio Ideation & Innovation	Network	Smart and Adaptive Process Management	
	Stakeholder Management	Security	Standards & Governance Automation	
	Strategic Management	Technology Architecture		

Figure 13: The 5 core dimensions divided into 28 sub-dimensions [53]

In [54], the authors suggests that the use of maturity models to the Industry 4.0 phenomenon could potentially aid corporations to assimilate this practice into their ethos and argue that the Industry 4.0 vision portrays a completely new approach to business process and a completely new business model. They propose a three-step systematic process to guide organizations in their respective Industry 4.0 journey to find footing and determine their vision which is precursor for strategy.

In [54], the authors proposes that the three-stage process comprise of (a) Envision 4.0 which relates to defining a tailor-made Industry 4.0 vision and it requires companies to develop their comprehension of the generic Industry 4.0 concepts with company-specific proficiencies and resources (b) Enable 4.0 starts with the significant Industry 4.0 vision developed in the previous phase, and based on this vision, the organization defines the digital portfolio and proficiencies required to provide support to the new products or services identified. (c) The enact phase is characterized by a well-timed, directed and multi viewpoint plan of the overall strategy towards the Industry 4.0 vision that fosters the strategic setting for tangible measures and outcomes.

In [54], the authors further argue that in order to employ maturity models to execute the three-staged process described earlier in the context of Industry 4.0, the maturity gauges (scales) for the three-stage model are described as follows: (a) Level 1 represents "initial" level where there is no industry 4.0 vision in place, (b) Level 2 represents "managed" level whereby a Roadmap of industry 4.0 strategy is in place, (c) Level 3 represents "defined" level whereby customer sections (segments), value intention (proposition), and significant (key) resources and competencies are clearly defined; (d) Level 4 which represents "transform" level whereby strategic imperatives are transformed into tangible projects, and lastly (e) Level 5 which represents "Detailed Business Model" level whereby the business model is completely transmuted in line with Industry 4.0 objectives.

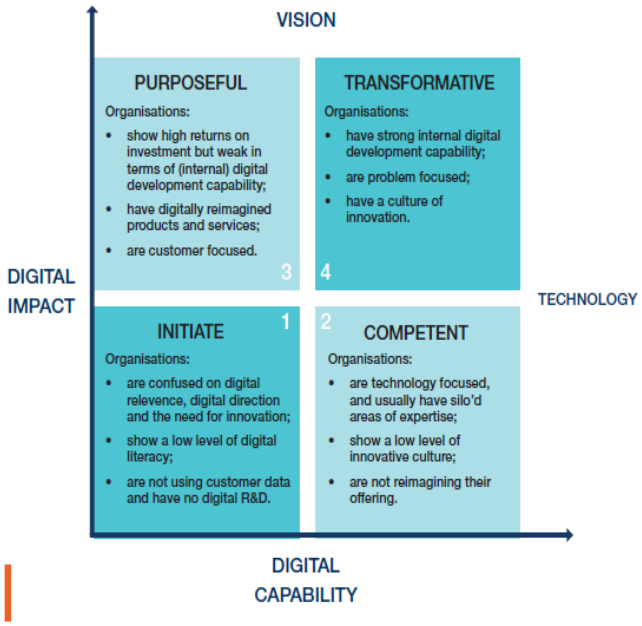


Figure 14: Stages of Digital Maturity and Key Attributes of Businesses in Each Maturity Levels [52]

2.7. Barriers towards movement to digital Post

In [15], the authors articulates the top five barriers (obstacles) towards movement into digital Posts by Region and for purposes of this study only the top performers (Industrialized countries) and Africa will be considered. These barriers are tabled in Table 3.

Table 3: Barriers for digital adoption and diffusion in the postal sector [15]

Geographic area	**Barriers**
Globally	Resource constraints
	Transition towards a digital culture
	Limitations of IT infrastructure
	Lack of adequate inhouse expertise required for development of digital services
	On-time custom clearance capability and capacity
Industrialized countries	The time taken for the corporation to shift in the direction of a digital culture
	Slow overall consumer embracing of digital postal services
Africa	Limited resources
	Lack of IT infrastructure
	Poor digital culture
	Lack the internal resources to develop digital services
	Slow overall consumer embracing of digital postal services

Clearly, Africa is faced with a plethora of challenges as depicted in Table 3, and Southern Africa as a Region in Africa suffers from the same challenges. In [8], the authors hypothesises that inadequate internet connectivity, inadequate capital outlay and

or recapitalization of the network infrastructure, meagre excellence of service, inadequate interoperability and connectivity among the business value chain and vagueness in postal sector definition in Southern Africa are proposed as foremost motives for the low advancement of the Region in embracing technologies associated with the digital revolution.

In [15], the authors affirms that the best ranking designated operator when it comes to digital readiness achieves a normalized score of 1, while the lowest possible performer obtains a normalized minimum score of 0. All normalized scores can be understood as the gap of any given designated operator with respect to the best (score 1), the midway (score above 0.5) or the worst (zero score) global performer. Table 4 represents the postal digital readiness of the industrialized countries with Switzerland leading with a full score of 1 while Table 5 portrays Southern Africa as a region and South Africa leads with a score of 0.46 which is just below the intermediate score of 0.5 while the worst performing countries are Madagascar and Democratic Republic of Congo which are both at 0.05. The distance between the worst performer and the best performer stands at a mammoth gap of 0.95. Clearly Southern Africa is faced with enormous challenges to realize a digital Post.

Table 4: Digital readiness score of Industrialized countries [15]

Country	Normalized score 2019
France	0.81
Germany	0.81
Japan	-
Netherlands	0.62
Switzerland	1.00

Table 5: Digital readiness score of Southern Africa countries [15]

Country	Normalized score 2019
Angola	-
Botswana	0.32
Comoros	0.14
Democratic Republic of Congo	0.05
Eswatini	-
Lesotho	0.19
Madagascar	0.05
Malawi	0.08
Mauritius	-
Mozambique	0.11
Namibia	0.27
Seychelles	-
South Africa	0.46
Tanzania (Republic)	0.22
Zambia	-
Zimbabwe	0.14

The literature has pointed out that adoption of technology of organizations in general or digitalization (Adoption of digital technologies) of the Post is a complex non-linear phenomenon with a variety of drivers and barriers. Adoption and diffusion barriers and drivers reveal a systemic character, which demands an all-inclusive approach to managing the complex nature of adoption and diffusion barriers and drivers that involve exploration of interactions between adoption and diffusion barriers and drivers as well as management of the casual relationship between the drivers and barriers of adoption and diffusion of technology/digitalization.

2.8. From reductionist to system thinking approach

In [55], the authors explains that reductionism is understood as the conventional means of tackling problem solving and is widespread throughout numerous fields. Reductionism also goes under many names including 'stepwise enhancement', 'disaggregation' and merely 'breaking problem into sub-problems'. Reductionism inclines to refer to comprehension instead of problem-solving but the latter seems more appropriate in this case. In [56], the authors argue that reductionism as the opposite of holism and systems thinking agrees with the view that all objects or systems are breakable to lower levels in the hierarchy of their composition.

In [57], the authors contends that in accordance with reductionist thinking "The whole can be broken down into its parts and put back together from its parts. Parts are related through a simple cause-effect relationship; Thus, its defining characteristics exist in its parts." However, there are limitations to a reductionist thinking approach because systems most often defy the reductionist analysis because by focusing on the system's parts, researchers miss the whole, which frequently took on a form that was not distinguishable from the reconstructed parts. This view can be interpreted that it assumes that for a system to perform effectively and efficiently, its components (parts) should perform effectively and efficiently.

Conversely, systems are complex and as result the interaction of the components (parts) in a dynamic setting invariably affect performance of the entire (whole) system. Therefore, effective, and efficient technology adoption must focus on the dynamic setting (interaction) of drivers and barriers rather than on their separate actions. In [58], the authors argue that system is a unified set of components that are logically organized in a way that achieves a particular objective. In [59], the authors proposes that a system is a sequence of intermingling components organised to achieve stated goals.

In [60], the authors suggests that there are key dissimilarities in behaviours of reductionist thinkers when contrasted with systems thinkers and the crux of the differences is that while a reductionist approach simplifies the problem to its smallest essence to enable resolution of the problem through analytical methods; a system thinking approach on the other hand seek to understand the bigger picture and how elements of the problem interact in a dynamic setting. Because organizations are complex due to many interacting dynamics and forces, it is probable to propose that postal sector in Southern Africa lags in adoption of innovative technologies due to its reductionist thinking approach rather than a system thinking approach.

In [61], the authors propose that an event-oriented viewpoint is logical, action-oriented, irresistibly uncomplicated and frequently narrow-minded. Figure 15 portrays this thinking as a concept. In [62], the authors further argues that this event-oriented mind-set

signifies a conviction that challenges are irregular, emanating from overwhelming events in the outside world and that life is unpredictable. He argues that this mind-set has a conviction that events pop out from nowhere or at least there is no time to worry about their causes, what is important he argues, is to fix the problem as soon as possible. The typical thinking is characterized by linearity whereby problems are events and solutions as fixes.

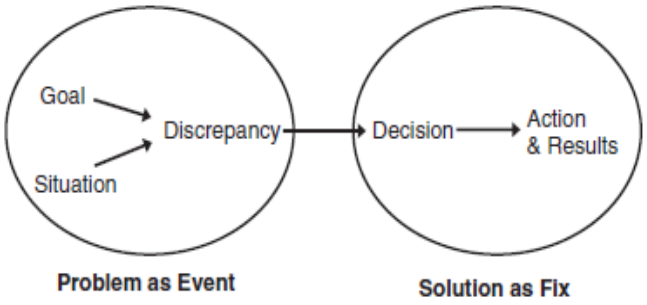


Figure 15: Event oriented worldview [62]

In [62], the author contends that feedback systems thinking is paradigm shift, a new approach of understanding the world, and a remedy to narrow thinking; and further advances that problems and solutions are viewed as intertwined as depicted in Figure 16.

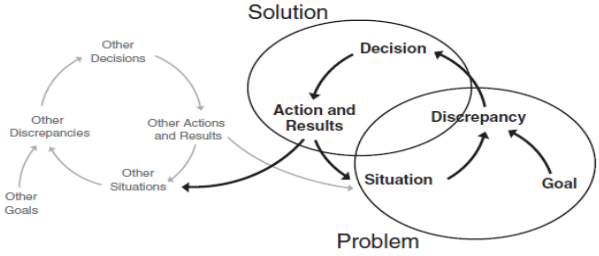


Figure 16: A shift of mind: A feedback perspective [62]

In [61], the authors contends that the change is the ultimate constant, and that change confronts conventional traditions, methods, and values. Important, most of the changes we battle to grasp occur as after-effects, deliberate, and not deliberate. In many cases, well-intentioned exertions to resolve unrelenting problems often lead to policy resistance, where policies are overdue, weakened, or thrashed by the unanticipated reactions of other people or of nature. This is a frequent challenge in the postal sector globally, in Africa and for purposes of this study in Southern Africa where for instance the digital age has brought unintentional consequences of the archaic postal business models that are putting in peril the very survival of postal operators.

In [63], the authors affirm that a system thinking approach is different from a reductionist approach, the reductionist approach focuses on analysis which is about dissection of complexity into manageable components or an event-oriented approach which has the conviction that problems emanate from nowhere. This approach clearly fits into the linear and mechanistic worldview, where the world and problems are fragmented into parts. Diametrically opposed to a reductionist approach of "analysis" is systems thinking approach which relies on synthesis rather than analysis. Fusion (synthesis) relates to comprehending the whole and the fragments (parts) simultaneously, alongside the relationships and links that constitutes the dynamic setting of the whole.

In [57], the authors supports this notion and argues that the whole emanates because of the interfaces between its components or fragments and proposes that fragments are related through intricate multiple effects. Thus, its defining attributes do not live in its fragments but in the whole. In [60] the authors suggest that systems thinking means the aptitude to distinguish and analyse the inter-connections within and between systems.

In [64], the authors contend that system thinking is a strong approach to deal with the system of innovation. In [64], the authors defines systems thinking as a field for observing aggregates and a context for spotting interrelatedness rather than things, for spotting patterns of change instead of stagnant snapshots. In [65], the authors define system thinking as the capability to epitomize and measure dynamic intricacy (e.g., behaviour that develops from the interface of a system's agents over a period), of systems.

In [66], the authors maintain that systems thinking is a set of harmonious diagnostic competencies used to enhance the ability to discover and comprehend systems, forecasting their conducts, and developing adjustments to them to create the anticipated effects. These competencies act as a system. In [63], the authors suggest that systems thinking is a process of examining and improving mental models and its core is based on four fundamental concepts (a) connectedness (b) Synthesis (c) Feedback loops, and lastly (d) Causality

In [60], the authors suggest that Systems Thinking requires recognition that systems designed by humans, recurrent occurrences or relations originate from systemic constructs which emanate from mental models. In [60], authors further propose that the Iceberg Model (Depicted in Figure 17), is a fundamental aspect of systems thinking. The Iceberg Model contends that actions (events) and patterns (which can be observed) are triggered by universal (systemic) structures and mental models, which are usually not observable.

In [60], the authors further propose that systemic constructs as in structures are the structural pyramid; societal order; interrelatedness; instructions and techniques; establishments and endorsement levels; process flows and methods; enticements (incentive programme), reward (compensation), aims, and system of measurement (metrics); manners (attitudes); responses and the incentives and qualms that cause them; organizational culture; feedbacks and interruptions (delays) in a dynamic setting; and fundamental forces that occur in organizations. Behaviours develop from these complex interactions, which exist because of mental models.

In [67], the authors argue that system dynamics, a subset of systems thinking signifies variations in a complicated system, and is beneficial for mimicking the stock-flow-feedback processes that usually trigger the behaviour of intricate systems. In [68] the authors argues that the modelling approach depend on the following crucial elements. "Stocks" exemplify the rise of something; "flows" exemplify the rate of flow into or out of a stock.

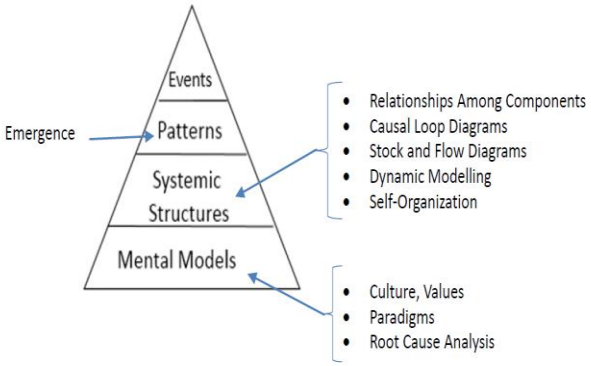


Figure 17: Integrated Model of Systems Thinking [60]

Stocks and flows, and variables characterises a system of linked non-linear differential equations, which can be modelled over a period to ascertain how stocks shift over time. In [69] the authors argues that one benefit of application of system dynamics approach over more conventional models is the addition of feedback consequences. Feedbacks ensure that a model's output is precise and uncovers intricacy that could be unobserved in other modelling approaches. Therefore, decision-makers can more definitely distinguish the interrelatedness of the model in a dynamic environment.

In [69], the authors argue that learning is a response (feedback) process. In [67], the author argues that as decisions are made these action modify the real world, and information feedback about the real world is received; and applying new insights, interpretation of the world is adjusted and becomes nearer to the goals.

This observation is principally valuable in the context of postal sector in Southern Africa which is overwhelmed by multiple stakeholders with multiple interests including policy directions internationally (Universal Postal Union), continentally (Pan African Postal Union), regionally (Southern African Postal Operators Association as implementing Agency of Southern Africa Development Community) and lastly nationally where individual postal operators operate. In this research the enablers and impediments that play in the adoption and diffusion of technology are identified including feedback loops that associate the several factors involved in the adoption and diffusion of technology process.

3. Methodology

In [70], the authors advances that research design can be classified in three categories which are (a) Exploratory research studies; (b) Descriptive and diagnostic research studies, and (c) Hypothesis-testing research studies. This study adopted the qualitative research design approach which is exploratory in nature.

In [71], the authors suggest that the idea behind qualitative research is to tenaciously hand-pick partakers or locations (or records or graphic materials) that enables the researchers to comprehend the conundrum (problem) and the research question. This fundamental feature of qualitative research design resulted in the in collection of specific secondary data that provides insight to the researcher to comprehend the phenomenon under study.

In [72], the author argues there are diverse forms of review methodologies; systematic, semi-systematic and integrative approaches and argues that each approach is appropriate depending on the purpose of analysis. This study adopted an integrative approach to literature reviews. An integrative approach is characterized by tenets which are articulated in Table 6.

Table 6: Tenets of an integrative approach to literature reviews

Tenet	Integrative Approach
Typical rationale	Critique and integrate
Research questions	Narrow or broad
Exploration approach	Usually not methodical
Sample attributes	Peer reviewed articles
Evaluation	Qualitative
Examples of contribution	Classification (Theoretical models or frameworks)

The data was collected from archival data from literature and industry reports, this data focused on a variety of drivers and barriers common in adoption of technology in organizations as well as characteristics of factors that drive or impede technology adoption in organizations and as well as in the postal sector.

This exploratory, qualitative research design using integrative approach was aimed at answering at great length the first research question which is, "Can a generic framework for technology and digital transformation adoption be developed for Posts in Southern Africa?

In [71], the author avows that "The data collection steps include setting the boundaries for the study, collecting information through unstructured or semi structured observations and interviews, documents, and visual materials, as well as establishing the protocol for recording information". In [71], the author further argues that the objective of qualitative research is to tenaciously choose members or sites (or documents or visual material) that will assist the investigator to fully grasp the challenge and the research question. The study purposefully selected relevant literature from Universal Postal Union, the different postal operators, specific literature that deals with technology adoption, ICT adoption and digital transformation with the view of comprehending the problem and the associated research question.

4. Discussion and findings

The literature review accentuates the significance of not only dealing with barriers and drivers to adoption of technology and Industry 4.0 technologies but highlights the fact that these barriers and drivers are not linear but intermingles with each other in a dynamic setting. Literature divulges that adoption and technology diffusion processes are very complex and entails multiple interdependent drivers and barriers which are very dynamic involving numerous feedback processes that demonstrate nonlinearity characteristics.

Literature details the postal sector landscape, and reflects on the digital panorama globally, continentally and in the region

(Southern Africa) and points to serious challenges faced by developing and less developed countries from a digital ecosystem point of view. These challenges are reflected in several barriers that need to be overcome and drivers that needs to be leveraged on in a dynamic setting.

Literature points to vast variances in digital adoption readiness between developed countries and developing countries, with specific reference to Southern Africa as depicted in Table 4 and Table 5 respectively and graphically depicted Figure 18 and Figure 19, respectively.

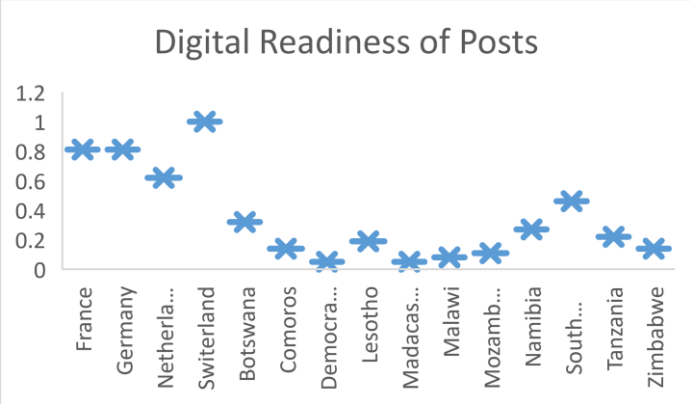


Figure 18: Digital readiness score of Posts

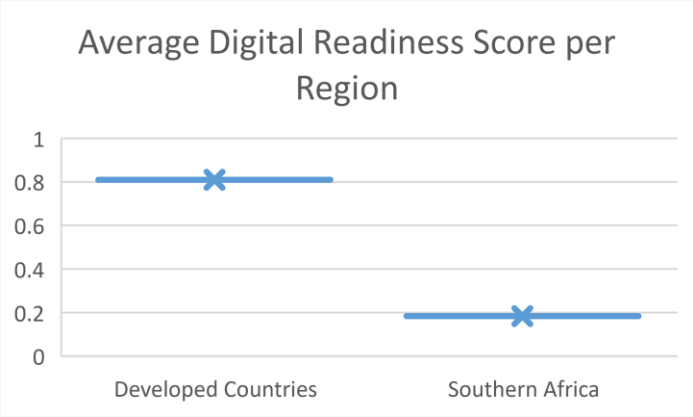


Figure 19: Average Digital readiness score per Region

The digital readiness score illustrates that Switzerland scores 100% and this is further supported by the performance of Switzerland on the Integrated Index on Postal Development in which it scores 100%. Switzerland is followed by both Germany and France scoring 81% and The Netherlands scoring 62%. The average of the four leading countries stands at 81%.

Southern Africa ranges from maximum of 46% which represent the performance of South Africa to a minimum of 5%, the variation within Southern Africa cumulatively results in an average of 18%. The performance of Southern Africa should be seen in the context of a plethora of challenges facing the continent and the Southern Africa in the context of this study. These challenges require a concerted effort by Posts and respective national governments and regional bodies to address the legacy of colonialism but also the problems that occurred post-independence that continue to halt prospects of development.

Literature also points to Technology Acceptance Models TAM, TAM2, TRA, UTAUT, TOE, BOE, and ADOPT are among the most popular which highlight factors related to system users; however, there is a literature knowledge gap in these models handling other stakeholders who are directly or indirectly affected by the adoption process and the interaction of drivers and barriers in a dynamic setting. These technology acceptance models lack the important aspect of feedback systems which are at the core of systems thinking approach.

TRA, TAM, TAM2, and UTAUT are significant pioneers in the technology adoption space by elucidating the aspects that perform a prominent role in the process of technology adoption. The evolution from TRA to UTAUT is exemplified in Figure 20.

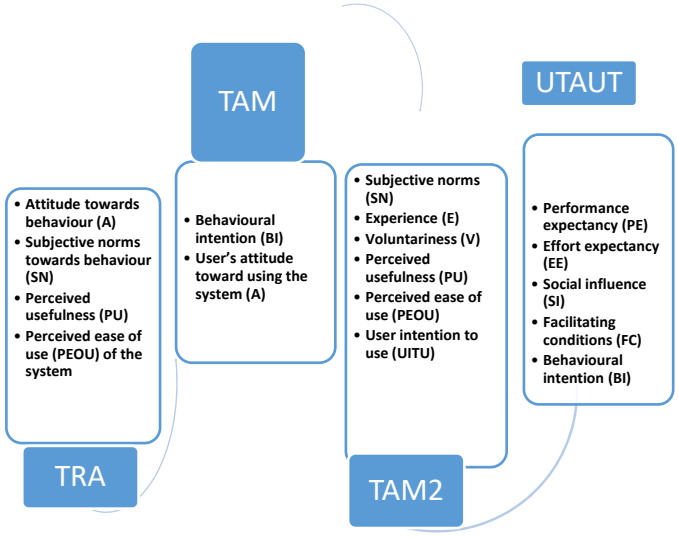


Figure 20: Evolution of technology adoption models

The factors elucidated in Figure 20 directly determine the user adoptability potential of technology, however they are applicable to individuals rather than organizations. The BOE and TOE attempts to address the limitations of TRA, TAM, TAM2 and UTAUT which are only applicable to individual's adoption of technology through articulation of constructs that are applicable to organization-wide adoption of technology. The constructs of BOE (Perceived Benefits, Organization Readiness and External Pressures) are depicted in Figure 21 and TOE constructs (Technology, Organization and External Environment) are depicted in Figure 22.

These factors associated with both TOE and BOE are an attempt to integrate organization-wide factors and external factors that are at play in the context of organization-wide adoption of technology adoption as opposed to the traditional technology adoption models in the form of TRA, TAM, TAM2 and UTAUT which are relevant to individual adoption of technology or new innovation. While TOE and BOE attempts to integrate factors associated with organization-wide adoption of technology and new innovation, they both lacked "causality" which is at play in a dynamic setting resulting in new emergence of patterns and mental models due to interacting factors that inhibit or enable adoption

and diffusion of technology in a dynamic setting which result in even greater complexity. BOE and TOE are therefore not sufficient to explain nor explore the complex nature of interacting factors in the dynamic setting of the postal sector in Southern Africa.

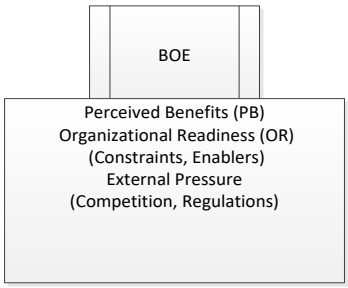


Figure 21: Constructs of the BOE model

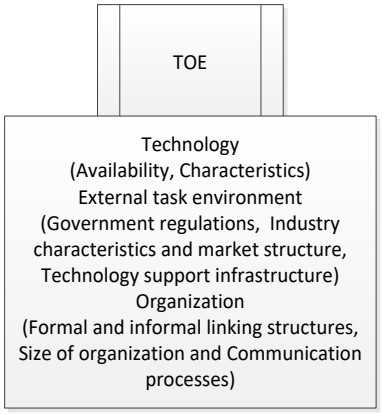


Figure 22: Constructs of the TOE model

The purpose of this research is to explore how best to close the limitations of the "linear" technology adoption models through the application of a holistic or systems thinking approach to identify systemic issues in technology adoption within Southern African context by considering the complete stakeholder system a "system of stakeholders" with their corresponding interests giving rise to various dynamic settings that could impede or enable technology adoption and digital transformation in the postal sector in Southern Africa.

The System Thinking approach which inherently integrates feedback systems is suitable for complex phenomenon like adoption of technology and digital transformation in the postal sector which is underpinned by multiple stakeholders with often opposing interests. A systems approach in comparison to traditional "linear" technology adoption models (TRA, TAM, TAM 2, UTUAT, TOE, BOE and ADOPT) was found to be the most appropriate and suitable for managing the interaction of barriers and drivers that are at play in a dynamic setting in the context of the postal sector in Southern Africa.

The literature suggests that Industry 4.0 is centred on digital transformation of business models. At its core is digitalization and integration of vertical and horizontal value chains, digitalization of offerings (products and services) and digitalization of business models and customer access. Of which, data and data analytics are core competencies interacting with digital technologies such as wearables and augmented reality, IoT platforms, mobile devices, big data and advanced algorithms, location detection technologies, 3D printing and other related Industry 4.0 technologies.

The critical review of literature explicates that the foundation to digital transformation is a systematic approach that is directed from a digital vision envisaged by the organization, the model proposed in [52] where the authors argues that the transformation requires organizations to move from level 1 (Initiate), to level 2 (Competent) to level 3 (Purposeful) and finally to level 4 (Transformative). The levels start from chaotic and no ambition or culture for innovation (level 1) to structured, systematic and innovative (level 4).

In [54], the authors articulates the three-stage model of envisioning of Industry 4.0, enablement of Industry 4.0 and lastly enacting of Industry 4.0. The three-staged model proposes 5 levels of digital maturity ranging from the lowest level (Level 1) in which the Industry 4.0 vision is not defined to the highest level (Level 5) in which there is complete business model transformation. The insights and inferences that could be drawn that without a solid vision and an underpinning strategy which acts as a roadmap for organizations with demonstrable projects; the adoption of technological innovation and digital transformation is set to fail.

In [53], the authors proposes the five dimensions that underpins digital transformation journey and are (a) Customers, (b) Strategy, (c) Technology, (d) Operations and (e) Organization and culture are key towards digital transformation journey that is structured to ensure success.

The proposed technology and digital adoption conceptual framework as presented in Figure 23 espouses the insights drawn from the critical review of selected literature.

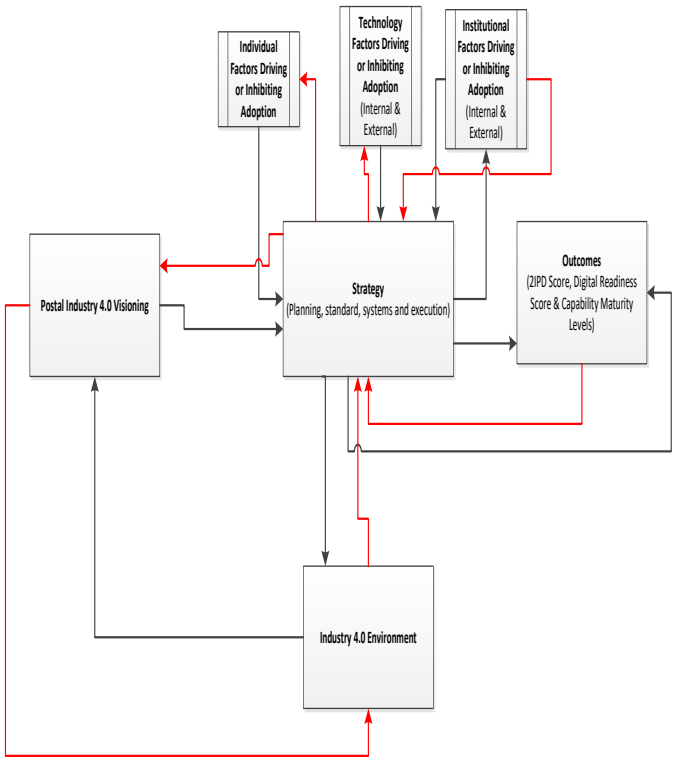


Figure 23: Proposed postal industry 4.0 adoption framework

The framework proposed commences with the visioning dimension which is a key critical success factor in adoption of digital technology and digital transformation. It sets the agenda for digital transformation of the Postal Sector in Southern Africa and without a vision for postal industry 4.0, the digital transformation process is bound to fail. The visioning dimension is supported by the strategy which integrates aspects of planning, systems relating to Industry 4.0, standards relating to Industry 4.0 and execution of strategic imperatives into concrete programmes and projects.

This dimension of strategy interacts with factors that inhibit or enable digital technology adoption and technology adoption broadly. These factors that either inhibit or enable adoption of digital technology and technology broadly are technological, institutional, and individual focused. These factors should be factored in during the strategy development phase and take cognizance of Industry 4.0 standards and systems prevailing. The strategy dimension also feedforwards to the Industry 4.0 environment which is meant to facilitate the core capabilities of Industry 4.0 which are data and analytics, aimed at supporting the core fundamental goal of Industry 4.0 which are: Digitalization and integration of horizontal & vertical value chains, digitalization of products and service offerings includes the expansion of existing products; and lastly, digital business models and customer access.

The execution of the vision for postal industry 4.0 through execution of strategy is envisaged to improve outcomes in the form of an improved 2IPD score, improved digital readiness score and improved capability maturity for the postal industry in Southern Africa.

The technology and digital objectives are underpinned by 5 strategic objectives which are:

1. Adoption of Industry 4.0 in the postal sector (Postal Industry 4.0)
2. Digital transformation readiness
3. Improvement of digital readiness by Posts in Southern Africa through capability maturity modelling
4. Adoption and diffusion of technological innovation
5. Improvement on the 2IPD score

A system approach is represented by feedback systems that are represented as red lines and arrows. Feedback systems are the foundation of the proposed approach as they mimic a dynamic setting due to interaction of a plethora of factors and outcomes.

The critical findings of this study can be summarized as:

(a) The traditional technology adoption models are particularly relevant to individual level rather than organizational level

(b) The TOE (Technology, Organization and Environment) model is the first attempt to incorporate organizational-wide adoptability of technology. However, the approach is linear in nature rather than systemic as feedback systems which are at play in a dynamic setting are lacking.

(c) That there is a big variation in performance on the 2IPD and on digital readiness index between the first-tier countries and Southern Africa. This variation is driven by digital culture which is lower in Southern Africa as compared to top-tier countries (Switzerland, The Netherlands, Germany, France, and Japan).

(d) Visioning and strategy drive digital transformation, without a robust visioning and strategy (including digital culture), the outcomes in terms of performance on both 2IDP and digital readiness index will always be below expectation.

(e) A linear approach to digital transformation and technology adoption fails to capture the complexities associated with interaction of factors in a dynamic setting resulting with even greater complexity.

5. Conclusions

Universally, the postal sector is faced with a plethora of challenges regarding the ever-changing world that is driven by volatility, uncertainty, complexity, and ambiguity (VUCA). This research presented the postal landscape with specific reference to performance of Southern Africa as a Region against the top 5 postal operators on the Universal Postal Union's Integrated Index on Postal Development.

The limitations of the technology adoption models were presented and discussed in context of their "linearity" or rather, a reductionist approaches to technology adoption as opposed to a systems approach which was presented and discussed in detail together with its respective tenets of which feedback systems are fundamental cornerstone to a system approach where the paradigm is "causality" rather than "linearity".

The conceptual framework that incorporates a system approach was presented and discussed in detail, with its respective tenets derived through insights obtained from critical review of literature. In [1], the authors presented a theoretical model at IEEM 2019. The proposed conceptual model as depicted in Figure 23; it is an enrichment of the earlier theoretical framework presented. It is a build-up towards the final goal of developing a dynamic model towards technology and digital adoption in the postal sector in Southern Africa. The conceptual framework answers the research question which was the guiding ingot to the critical literature review.

The research gap identified in this study is the failure of traditional technology adoption models in driving digital transformation agenda. The linearity fails to capture adoption dynamics in a dynamic setting as opposed to the proposed adoption model with build-in feedback loops to better understand interaction of drivers and barriers in an evolving dynamic setting as in the real world.

Future research will be the development of a conceptual framework using system dynamics where the drivers and barriers identified in the research are modelled in a dynamic setting to understand patterns that emerge from their interaction. The developed system dynamic conceptual model will be rigorously tested by selected participating Posts to further refine the model until it closely represents reality. The validated model could be used by Posts to model different drivers and barriers with a view of mitigating against barriers that would enable Posts in Southern Africa to fully adopt a digital transformation journey in order to be sustainable now and in the future.

6. Conflict of Interest

The authors declare no conflict of interest.

References

[1] K. Mokgohloa, M.G. Kanakana-Katumba, R.W. Maladzhi, J.A. Trimble, "Postal Development: Literature Review into Adoption Models," IEEE International Conference on Industrial Engineering and Engineering Management, **2**, 764–768, 2019, doi:10.1109/IEEM44572.2019.8978531.

[2] UPU, Annual Strategic Review 2017, Bern, 2017.

[3] UPU, "Postal Development Report 2018 - Benchmarking a critical infrastructure for sustainable development," (April), 1–32, 2018.

[4] C. Barman, A., Potsangbam, "Shifts of Strategic Paradigms in the VUCA World- Does 'outside the box thinking' a meaningful cliché for the business world?," (July 2018), 2017.

[5] B. Cousins, "Design thinking: Organizational learning in VUCA environments," Academy of Strategic Management Journal, **17**(2), 1–18, 2018.

[6] N. Balasubramanian, "Innovative Operation Strategies in VUCA world," 2319–7668, 2018.

[7] P. Hallo, Leonie; Nguyen, Tiep; Gorod, Alex; Tran, "Effectiveness of Leadership Decision-Making in Complex Systems," Systems, **8**(5), 2–21, 2020.

[8] SADC, SADC postal strategy 2017-2020, 2017.

[9] SAPOA, SAPOA Strategy Implementation Plan 2017 - 2020 Version 4, SAPOA, Maseru, 2016.

[10] UPU, postal Development Report 2018, Universal Postal Union, Bern, 2018.

[11] Universal Postal Union, The digital economy and digital postal activities – a global panorama, Universal Postal Union, Bern, 2019.

[12] UPU, Postal Development Report 2019, Berne, 2019.

[13] T. Santos, Beatrice; Chaura-Santos, F; Lima, "Industry 4.0: an overview," Research G, (July), 2018.

[14] H. Lasi, P. Fettke, H.G. Kemper, T. Feld, M. Hoffmann, "Industry 4.0," Business and Information Systems Engineering, **6**(4), 239–242, 2014, doi:10.1007/s12599-014-0334-4.

[15] Universal Postal Union, The digital economy and digital postal activities – a global panorama, 2019.

[16] G. Reinhard, V. Jesper, S. Stefan, Industry 4.0: Building the digital enterprise, 2016, doi:10.1080/01969722.2015.1007734.

[17] A. Rojko, "Industry 4.0 Concept: Background and Overview," International Journal of Interactive Mobile Technologies (IJIM), **11**(5), 2017, doi:10.3991/ijim.v11i5.7072.

[18] Deloitte, Industry 4.0 Challenges and solutions for the digital transformation and use of exponential technologies, 2015, doi:10.1057/9780230514027_2.

[19] L.D. Galindo, The Challenges of Logistics 4.0 for the Supply Chain Management and the Information Technology, 2016, doi:10.1038/npp.2015.289.

[20] B. Hermann, Mario; Pentek, Tobias; Otto, Ultrafast synthesis of bifunctional Er 3+ /Yb 3+ -codoped NaBiF 4 upconverting nanoparticles for nanothermometer and optical heater, Gallen, 2015, doi:10.1016/j.jcis.2017.12.027.

[21] J. Macaulay, L. Buckalew, G. Chung, Internet of Things in Logistics, 2015.

[22] F. Jindal, R. Jamar, P. Churi, "Future and Challenges of Internet of Things," International Journal of Computer Science and Information Technology, **10**(2), 13–25, 2018, doi:10.5121/ijcsit.2018.10202.

[23] USPS, The Internet of Postal Things Report Number RARC-WP-15-013, Virginia, 2015.

[24] M.A. Rahman, A.T. Asyhari, "The emergence of internet of things (Iot): Connecting anything, anywhere," Computers, **8**(2), 8–11, 2019, doi:10.3390/computers8020040.

[25] UPU, Postal development report, Berne, 2018.

[26] G. Kuehne, R. Llewellyn, D.J. Pannell, R. Wilkinson, P. Dolling, J. Ouzman, M. Ewing, "Predicting farmer uptake of new agricultural practices: A tool for research, extension and policy," Agricultural Systems, **156**(August 2016), 115–125, 2017, doi:10.1016/j.agsy.2017.06.007.

[27] E.M. Rogers, Diffusions of Innovations, 1995.

[28] E.M. Rogers, M. Everett, Diffusion of Innovatons, 1983.

[29] R. Sharma, R. Mishra, "A Review of Evolution of Theories and Models of Technology Adoption," Indore Management Journal (IMJ), **6**(2), 17–29, 2014.

[30] C. Dube, V. Gumbo, "Diffusion of Innovation and the Technology Adoption Curve: Where Are We? The Zimbabwean Experience," Business and Management Studies, **3**(3), 34, 2017, doi:10.11114/bms.v3i3.2500.

[31] A. Kogabayev, Timur; Maziliauskas, "The definition and classification of innovation," Holistica, **8**(1), 59–72, 2017, doi:10.1515/hjbpa-2017-0005.

[32] S. Xaba, Evaluation of South African Post Office SOC Ltd Innovation Efficacy, University of Pretoria, 2012.

[33] P.F. Drucker, Innovation and Entrepreneurship: Practice and Principles, HarperCollins Publishers, London, 1985.

[34] M. De Jong, N. Marston, E. Roth, "The eight essentials of innovation," McKinsey Quarterly, (2), 36–47, 2015.

[35] J.P. Wentzel, K.S. Diatha, V.S.S. Yadavalli, "An application of the extended Technology Acceptance Model in understanding technology-enabled financial service adoption in South Africa," Development Southern Africa, **30**(4–5), 659–673, 2013, doi:10.1080/0376835X.2013.830963.

[36] D. Mugo, K. Njagi, B. Chemwei, J. Motanya, "The Technology Acceptance Model (TAM) and its Application to the Utilization of Mobile Learning Technologies," British Journal of Mathematics & Computer Science, **20**(4), 1–8, 2017, doi:10.9734/BJMCS/2017/29015.

[37] J.P. Wentzel, Modelling Financial Services Adoption Through an Intermediary in South Africa: TAM and SEM Approach, Pretoria, 2012.

[38] Q. Ma, L. Liu, "The Technology Acceptance Model," Advanced Topics in End User Computing, **4**, (January), 2011, doi:10.4018/9781591404743.ch006.ch000.

[39] M. Chuttur, "Technology Acceptance, Information System Deployment, TAM, Information System Theory," Sprouts, **9**(2009), 2009.

[40] P.. Davids, F.D; Bagozzi, R.P; Warshaw, "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," Management Science, **35**(8), 982–1003, 1989.

[41] P. Lai, "the Literature Review of Technology Adoption Models and Theories for the Novelty Technology," Journal of Information Systems and Technology Management, **14**(1), 21–38, 2017, doi:10.4301/s1807-17752017000100002.

[42] V. Venkatesh, F.D. Davis, "Theoretical extension of the Technology Acceptance Model: Four longitudinal field studies," Management Science, **46**(2), 186–204, 2000, doi:10.1287/mnsc.46.2.186.11926.

[43] F. Venkatesh, Viswanath; Morris, Michael; Davis, Gordon; Davids, "User Acceptace of Information Technology: Towards a Unified View," MIS Quarterly, **27**(3), 95–98, 2003.

[44] C.M. Chao, "Factors determining the behavioral intention to use mobile learning: An application and extension of the UTAUT model," Frontiers in Psychology, **10**(JULY), 1–14, 2019, doi:10.3389/fpsyg.2019.01652.

[45] Y.K. Dwivedi, N.P. Rana, A. Jeyaraj, M. Clement, M.D. Williams, "Re-examining the Unified Theory of Acceptance and Use of Technology (UTAUT): Towards a Revised Theoretical Model," Information Systems Frontiers, **21**(3), 719–734, 2019, doi:10.1007/s10796-017-9774-y.

[46] H. Gangwar, H. Date, R. Ramaswamy, "Understanding determinants of cloud computing adoption using an integrated TAM-TOE model," Journal of Enterprise Information Management, **28**(1), 107–130, 2015, doi:10.1108/JEIM-08-2013-0065.

[47] T. Oliveira, M. Martins, "Literature Review of Information Technology Adoption Models at Firm Level.," Electronic Journal of Information …, 2011.

[48] A. Dasgupta, S. Wendler, "AI Adoption Strategies," (9), 1–13, 2019.

[49] G. Kuehne, C. Nicholson, M. Robertson, R. Llewellyn, C. McDonald, "Engaging project proponents in R&D evaluation using bio-economic and socio-economic tools," Agricultural Systems, **108**, 94–103, 2012, doi:10.1016/j.agsy.2012.01.009.

[50] K. Crowston, J. Qin, "A capability maturity model for scientific data management: Evidence from the literature," Proceedings of the ASIST Annual Meeting, **48**(February), 2011, doi:10.1002/meet.2011.14504801036.

[51] M.C. Paulk, B. Curtis, M.B. Chrissis, C. V Weber, "for Software , Version 1 . 1 Software Engineering Institute," Office, **40**(February), 1–26, 1993.

[52] M. Shahiduzzaman, M. Kowalkiewicz, R. Barrett, M. McNaughto, Digital Business: Towards a Value Centric Maturity Model, 2017.

[53] Deloitte, Deloitte Digital Maturity Model, 2018.

[54] J. Ganzarain, N. Errasti, "Three stage maturity model in SME's towards industry 4.0," Journal of Industrial Engineering and Management, **9**(5), 2016, doi:10.3926/jiem.2073.

[55] M. Rafferty, "Reductionism, Holism and System Dynamics," Proceedings of the 2007 International Conference of the System Dynamics Society, 2007.

[56] M. Dafermos, "Encyclopedia of Critical Psychology," Encyclopedia of Critical Psychology, (April), 2014, doi:10.1007/978-1-4614-5583-7.

[57] P. Mella, "Systems Thinking: The art of understanding the dynamics of systems," Journal of Learning, **15**(January), 2015.

[58] D.H. Meadows, Thinking in Syatems, Earthscan, London, 2009.

[59] H. Sillitto, P. Godfrey, "Defining " System " : A Comprehensive Approach," in 27th Annual INCOSE International Symposium, INCOSE, Adelaide,

2017.
[60] M. Jamshidi, "Systems thinking," IEEE Transactions on Neural Systems and Rehabilitation Engineering, **16**(4), 1–4, 2008, doi:10.1109/TNSRE.2008.929253.
[61] J.D. Sterman, Business Dynamics: Systems thinking and modeling for a complex world, McGraw-Hill, London, 2000, doi:10.4324/9781351152723-7.
[62] J.D.W. Morecroft, Strategic Modelling and Business Dynamics, 2015, doi:10.1002/9781119176831.
[63] E.. Schumacher, Introduction to Systems Thinking Principles and Analytical Tools, 2019.
[64] S. Peter, The Fifth Discipline:The art and practice of learning organisation, Doubleday, New York, 1990.
[65] J.D. Sterman, L.B. Sweeney, "Bathtub dynamics: initial results of a systems thinking inventory," System Dynamics Review, **16**(4), 249–286, 2000.
[66] R.D. Arnold, J.P. Wade, "A definition of systems thinking: A systems approach," Procedia Computer Science, **44**(C), 669–678, 2015, doi:10.1016/j.procs.2015.03.050.
[67] J.D. Sterman, "System dynamics modeling: Tools for learning in a complex world," IEEE Engineering Management Review, **30**(1), 42–52, 2002, doi:10.1109/EMR.2002.1022404.
[68] M. Reinker, E. Gralla, "A System Dynamics Model of the Adoption of Improved Agricultural Inputs in Uganda, with Insights for Systems Approaches to Development," Systems, **6**(3), 31, 2018, doi:10.3390/systems6030031.
[69] D.K. Fisher, J. Norvell, S. Sonka, M.J. Nelson, "Understanding technology adoption through system dynamics modeling: Implications for agribusiness management," International Food and Agribusiness Management Review, **3**(3), 281–296, 2000, doi:10.1016/s1096-7508(01)00048-9.
[70] P.D. Leedy, J.E. Ormrod, John W. Creswell, S. Greener, W. Askarzai, B. Unhelkar, D.J. Currie, C. Smith, P. Jagals, D. Williams, H. Snyder, H.K. Mohajan, H.K. Mohajan, J.M. Brayer, O. Skovsmose, M. Borba, M.A. Ragab, A. Arisha, S. Rajasekar, P. Philominathan, S. Macdonald, N. Headlam, U. Sekaran, C. Kothari, Research Design, Eleventh E, SAGE, Essex, 2004, doi:10.1186/s12889-018-5318-8.
[71] J.W. Creswell, Research Design, Fourth Edi, SAGE, London, 2014.
[72] H. Snyder, "Literature review as a research methodology: An overview and guidelines," Journal of Business Research, **104**(August), 333–339, 2019, doi:10.1016/j.jbusres.2019.07.039.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1207-1216 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Evolution of Teaching Approaches for Science, Engineering and Technology within an Online Environment: A Review

Rendani Wilson Maladzhi*, Grace Mukondeleli Kanakana-Katumba

University of South Africa, Department of Mechanical and Industrial Engineering, Johannesburg, South Africa

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 25 November, 2020
Online: 16 December, 2020

Keywords:
Augmented reality
Online teaching and learning
Science
Engineering and technology
Teaching Approaches
Virtual reality
Virtual laboratory

ABSTRACT

The emergence of the COVID-19 pandemic earlier this year destabilised the operations in the education sector, particularly in institutions of higher learning. Most of these institutions are now expected to teach online, assess their students using non-venue examinations and offer remote laboratory practical and experiments. However, many of these institutions were not prepared for such dynamic change in a such short space of time. Consequently, most of the institutions moved their mid-year examinations to the October/November period. In addition, they are finding methods to conduct laboratory practical and experiments. Prominent researchers agree that institutions of higher learning are challenged by the current dispensation where academics are expected to implement new pedagogical approaches and take advantage of information and communications technology (ICT) for teaching and learning. Every module requires a specific teaching method, as a result, academics need to know about various methods available. Generally, the sciences and engineering modules are taught within the classroom environment and require experiential platforms to solidify the theoretical knowledge gained. The current study is aimed at assessing the evolution of teaching approaches and technologies in distance education regarding online delivery particularly in science, engineering and technology. Various search engines such as Google Scholar, Scopus, Sabinet, ProQuest and EBSCO were used to obtain relevant literature to depict popular teaching approaches and the relevant technologies. In order to access relevant literature, various key strings were used. The findings of the study revealed problem-based, apprenticeship and experimental, and competency-based learning as the most popular teaching approaches, particularly between 2010 and 2020. Project-based learning, case-based learning and communities of practice equally share the third position followed by integrative learning. Students and academics show that instant feedback provided by virtual labs yielded encouraging results. This shows that many students and their academics prefer the introduction of the virtual labs in their learning environment, especially from 2012 until recent. The literature further confirms the current findings that institutions of higher learning need to equip academics to be able to integrate technology within their teaching methods for students to continue to learn at ease.

1. Introduction

This paper is an extension of "Online Learning Approaches for Science, Engineering and Technology in Distance Education" presented at the IEEM 2019 Conference in Macau [1]. Higher education institutions are currently endorsing online courses as a business reality since their budgets are now linked to their enrolments [2]. The introduction of an online learning environment has become a long-term business strategy [3] for many institutions worldwide, to the extent that the proportion of students taking at least one online course has increased from 1 in 10 to nearly one-third in 2010 [3]. The accessibility of the internet of things (IOT) in universities worldwide caused them to invest in e-learning systems to enhance their teaching activities to improve their students' learning experiences and performance [4]. Online teaching and learning environments have brought some mixed feelings to students who believe that video lecturers slightly outperform face-

*Corresponding Author: Rendani Wilson Maladzhi, maladrw@unisa.ac.za

to-face lecturers [5] and some students feel interactive online videos are even better [6]. In addition, others feel that online homework is as effective as paper and pencil [7], [8] and intelligent tutoring is seen to be as effective as human tutors [9]. The dynamics above describe how teaching and learning have evolved over the years with reference to five generations of distance education.

The transition through the five generations of distance education has, over the years, witnessed a change in the use the delivery media of teaching and learning that ranged from text and images, sound and video, live television broadcasting to Web 2.0 interactive online technologies such as podcasts, video conferences, virtual spaces etc. [10]-[12]. Based on the above stated transition, the exchange of information between academics and students remained key, but the mode of delivery and the technological tools made all the difference. Hence, various disciplines require different teaching methods accompanied by various technological tools, as is the case with the current research that focusses on the science, engineering and technology sector. Teaching science, engineering and technology in an online environment has been considered an impossible mission for several years [13]. It is a fact that the sciences and engineering disciplines customarily require laboratory work where students perform their practicals to demonstrate their understanding of the theoretical knowledge.

In another instance, an Engineering Graphics Online Course was first introduced in Tianjin University, where 20 colleges and universities participated to cater for almost 1500 students. As a result, this online environment enabled educational resources to be shared in different areas while quality teaching of engineering graphics was made possible [15].

Laboratory classes are an essential component of most science and engineering courses with the potential to achieve several practical and theoretical objectives [16]. Arguably, students must not only learn manipulative techniques, but also link theory to practice, solve different kind of problems, interpret data, interact with staff and other students, and successfully navigate the lab [16]. Learning in this situation can be greatly assisted by academics who are able to guide students through this complex process. However, the effectiveness of laboratory classes is often not achieved to their full potential [16].

The emergence of the COVID-19 pandemic brough along uncertainties for the institutions of higher learning regarding teaching activities and practical and laboratory experiments. On the other hand, it presented distance education institutions with an opportunity to introduce non-venue-based assessments and remote practical and laboratory experiments. The traditional teaching and learning platform greatly benefited by the development of internet and ICT capabilities [17]. Virtual labs afforded students an opportunity to revive their curiosity and creativity while learning from their own respective spaces [18]. The mode of teaching such as synchronous and asynchronous modes made it possible for teaching and learning to take place at a distance. As such, students taught through the asynchronous mode displayed higher cognitive achievement while those taught through the synchronous e-learning mode displayed improved skills acquisition [19]. Hence, the importance of academics understanding the engineering, science and technology (SET) environment to better equip the students appropriately. A synchronous mode of delivery facilitates teaching and learning in real time, whereas an asynchronous mode requires self-paced teaching and learning activities that make use of email, blogs, wikis, discussion boards, web-supported textbooks, hypertext documents, audio or video courses, and social networking using Web 2.0. [19]. The emergence of Massive Open Online Courses (MOOCs) has made it possible for students all over the world to access free credible multimedia resources utilised by academics from recognised universities across the world [20].

Theories about teaching methods do not exist, particularly in post-secondary level, but apprenticeships that resemble learning by teaching, do exist [21]. However, other researchers viewed apprenticeship or experiential learning or learning by doing as an opportunity for SET learners to engage in active learning while reflecting and putting into practice what they learned in the classroom or outside of the classroom [22]. Traditionally, SET experiential learning methods such as laboratory work, workshops; apprenticeships; problem-based learning; case-based learning; project-based learning; inquiry-based learning; and cooperative learning [21], were mainly found within the SET environment. Additive manufacturing [23], augmented reality, virtual labs, online teaching modules and educational aids (Kit) [24]; modelling activities by 3D printers [25]; Engineering Graphics Online Course [15]; online tutorial system [30]; flipped classroom [23]; virtual simulation teaching [26]; virtual reality; the Digital Engineering module were all transformed it into an online teaching environment.

The purpose of this paper is to determine teaching approaches in SET that are gaining popularity due to the evolution of technology within the online environment to improve students' performance. The technological advances over the years contributed positively in turning traditional teaching approaches in science, engineering and technology into suitable online learning.

2. Literature Review

Teaching and learning have evolved over the years due to technological advancements and knowledge transfer between academics and students. It is a fact that teaching modalities also depend on various disciplines offered in various institutions of higher learning.

Experiential learning took a form of learning by doing and as such, is preferable in the SET environment [21]. However, its successful implementation particularly in an online environment requires virtual technologies. As a result, literature that covers teaching methods and respective technologies are viewed as follows: integrative, apprenticeship and experiential learning, problem-based learning, project-based learning, case-based learning, inquiry-based learning, competency-based learning and communities of practice. It has now become apparent that technologies that enhance teaching and learning be exposed to the world for the benefit of the SET sector.

2.1. Integrative approach

The integrative approach is regarded as an innovative practice due to its ability to introduce a concerted approach to mathematics

and science academics in the teaching and learning environment [27]. It permits students to learn about life's origin, then life's evolution, and thereafter about the future of life and humankind in the Earth and Life Sciences, Mathematical modelling concepts, such as the modelling cycle, solves problems regarding real-life situations [27]. Some of the characteristics of the interactive approach are like those in seminars and tutorial [27]. The emergence of distance learning through online environments calls for teaching activities to be conducted differently. In other instances, the teaching of mathematics education in engineering spaces is enriched through modelling activities by 3D printers [25]. As a result, academics should be provided with technological skills to function better and teach effectively.

Virtual learning systems (VLSs) in the academic fraternity play a similar role as enterprise resource planning systems (ERPSs) in the administration arena [28]. Virtual learning systems are known for their role in the teaching and learning avenues that coordinate activities such as learning objectives, class exercises, quizzes and tests in both synchronous and asynchronous communication [29]. The introduction of online tutorial systems within the Australian universities has seen a tremendous increase in the number of students entering the universities [30]. It is a fact that the new developed online teaching approaches introduced a diverse background of skills, motivations and prior knowledge. In addition, the online tutorial system was found to be instrumental in modules such as engineering physics and mathematics, and as such, students did not only pass their examinations, but it also improved the retention rate of first-year engineering students with the financial benefit to the institution [30].

2.2. Apprenticeship and experiential learning

The teaching approaches such as apprenticeship, experiential learning, cooperative learning and adventure learning are commonly used in the fields of SET. They are preferred for their capability to provide students with prospects to experimentally practice their learned experiences [21]. Learning by doing was traditionally instrumental in the mechanical, civil, electrical and electronics vocational fields [21], [27]. However, its transition towards online experimentation and remote/virtual labs presented learners with vast opportunities to control real or virtual equipment at a distance in order to perform scientific experiments [20]. It appears that different researchers refer to virtual laboratories differently; some view it as simulation-based learning, remote experimentation and remote inquiry-based learning [18], [31], [32].

Lack of reputable laboratory experiments contributed negatively to engineering education and resulted in students utilising outdated experiments [18]. The introduction of virtual laboratories gave students an opportunity to make use of good laboratory facilities operated by trained academics as they made use of remote experimentation [18]. It should be noted that virtual laboratories allow students to take ownership of the experiments and the quality control thereof [18]. Virtual laboratories become a multi-institution and yet multidiscipline initiative aimed at benefiting the students within the engineering fields [18]. In its nature, it addresses the issue of lack of good laboratories facilities, as well as trained teachers, by making remote experimentation possible [18]. It provides students with an opportunity to take ownership of teaching and learning space while academics provide minimal guidance. According to [18] and [31], Virtual Laboratory reached over 100 000 students, and are accessible after hours [18,31].

It has been proven that virtual experiments are proven to be more beneficial for teaching and learning [33].

Factually, learning is currently shifting from synchronous in-class settings to asynchronous online platforms so that education can be accessible to a wider audience [34]. It is believed that engineering education always requires laboratory work and hands-on practical components that are not possible to conduct online. However, it was only made possible by the introduction of simulations in fluid mechanics laboratories [34].

Virtual reality (VR) puts the user into a computer-created world that seems real [35]. In addition, VR creates an environment in which the user feels and seems to be moving inside a computer-created virtual world in the same way people move in and around a natural environment; while immersed in the virtual world, the user cannot perceive the real one which still surrounds him [36]. The findings on the study conducted by Luo depicted that students enjoyed taking measurements in the laboratory. However, other students on the other hand were sleeping around with nothing to show. As a result, virtual practicals are preferred as opposed to the traditional way of conducting practical work [26] for mechanical, electrical, civil engineering and other qualifications.

Augmented reality (AR) utilises smart devices and applications such as smart phones, tablets, wearable headsets and immersive technologies that enhance information in real-time to become intelligent [37]. It contemplates real-world events with computer-generated digital information by allowing the user to experience the interaction of the two worlds [38]. AR is capable of making required information available, presented and experienced by users 'augmenting the real world with digital information [39]. The practical application of augmented reality proved itself to be economically and technically viable for societal usage [39]. Hence, it is preferred within the mechanical, industrial, electrical, civil and other engineering disciplines.

2.3. Problem-based learning

Problem-based learning (PBL) enables the students to gain knowledge about and apply it to different problems they investigate [40]. As they seek solutions for the problems at hand, students have an opportunity to argue and debate based on their own understanding and the acquired facts [40]. PBL teaches students not to depend on their lecturers for information but to rather resort to other sources to make informed decisions. It is important to note that PBL is more suited for student-centred and community-based learning environments [41].

Additive manufacturing (AM) is the preferred technology of teaching used by academics to introduce topics that are industry and academic experts, laboratory and final engineering projects. AM is a PBL in nature due to its innovative features dealing with devices for people with disabilities. Other PBL-related projects include the development of an opener for medicine containers, a device for pouring liquids for people who have Parkinson's disease, and a 3D puzzle for people who have visual impairments.

[23]. It has the capability to encourage students' creativity, learning motivation and engagement [23].

The flipped classroom (FCM) has emerged as a learning model that increases students' active learning, collaboration and scaffolding during the learning process through better allocation of teaching time [43]. As such, the FCM method recognises that teaching activities within face-to-face environment should not be about academics teaching but should rather focus on creating an environment for students to effectively collaborate with each other while guided by the academics through scaffolding methods [44], [45]. It is believed that digital technologies have made it possible for teaching and learning to take place without students being confined to a classroom, and that learning can take place from anywhere as long as study material, videos, online quizzes etc. are available.

The FCM method is preferred by both practitioners and researchers in various subjects such as Mathematics, ICT, Social Studies, Humanities etc. [46]-[50], [89].

The results from various studies further confirmed the FCM method is an effective technology that raises the performance of students who face difficulties in face-to-face educational contexts, and equally assists academics and institutions to improve learning conditions [51]. The FCM's nature affords academics an opportunity to scaffolding and create an enabling environment for engagement with hands-on practices [51].

Traditionally, the Digital Engineering lectures were limited to the face-to-face mode of delivery where students only listened to the lecturer delivering the lecture [24]. It was indeed a tremendous challenge since there was no point of demonstrating the application of electronic circuits, digital electronics and high logic circuits for students to apply in industry [24]. The introduction of online teaching methods utilising the Proteus software to enhance skills of learners began to make a difference [24]. Both problem-based learning and competency-based learning suited the Digital Engineering module. This is due to the application of the software that demonstrates teaching and learning activities to students, which also requires both online and off-line learning with computer-based learning assistance [24].

Arguably, teaching practical science at a distance has been a challenge for a while in most institutions across the world. However, in July 2013, the Open University (OU) launched the Wolfson Open Science Laboratory (OSL), which offered a variety of practical science experiences to their distance learning students [13]. With this initiative, students are provided an opportunity to observe, investigate, gather and analyse data. In fact, numerous studies confirmed the effectiveness of PBL because it fostered active learning.

2.4. Case-based learning

Case-based learning (CBL) is an interactive group learning pedagogy preferable in various disciplines including business, law, health [14], [52] and in sciences [53]. Case-based learning is defined as an educational approach that allows students to work in small, collaborative groups to solve problems [54]. It is gaining popularity at a faster pace within the online fraternity [14]. (2018: 283-300). Furthermore, it stimulates students' thinking and discussions amongst students [14]. It is paramount to mention that the case presentation teaching method is frequently utilised in medicine [52].

Case-based is predominantly used where complex and interdisciplinary themes openly discussed [27]. Application oriented teaching platform became a reality in various university through online course construction [17].

According [55], civil engineering requires modern teaching methods such as hybrid, project, case and discussion into the classroom while utilising engineering examples to enhance teaching and learning.

2.5. Project-based learning

Project-based learning (PJBL) is a twenty-first century innovative learning approach that teaches various strategies to ensure students' success [56], particularly in higher learning environments. With this approach, students are afforded an opportunity to experience the reality of the real-world while boosting their communication skills and interdisciplinary learning [57]. However, PBL places more emphasis on learning through internal rather than external projects. Industrial collaboration projects build partnerships and provide opportunities to work on real problems, often seeking multidisciplinary approaches. This study explores and suggests the benefits that students derived from their participation in an external research and consultancy project at a University in Oman [57].

In 2016, Ralph conducted a research study looking into the post-secondary project-based learning in science, technology, engineering and mathematics [58]. The findings indicated that students benefitted from the project-based method regarding content knowledge, and interdisciplinary skills. Their learned skills would be instrumental in their future classes and career settings. However, most students did not enjoy the teamwork that PJBL brings but some learned from other group members [58]. All in all, the findings recommended that PJBL should be implemented institutionally rather than per discipline.

Typically, students using the project-based learning method work together as a team in order to solve the given problem by developing the product and development process [59]. PJBL is found to be an efficient teaching method in the twenty-first century as it promotes critical thinking, problem-solving skills, interpersonal communication, information and media literacy, collaboration, teamwork and leadership, creativity and innovation [60].

Other researchers conducted a study on Reform Practice of Engineering Drawing Courses in Chinese Colleges in 2019 [61]. The findings highlighted amongst others the emergence of the MOOC platform of outcomes-based education for engineering education models. All mechanical drawing courses implemented project-based teaching methods that were suited for the online environment. The MOOC platform catered from various online learning engagement models for a variety of questions, discussions sessions and projects [61].

2.6. Inquiry-based learning

The inquiry-based instructional approach is learner-centred and requires the learner to observe, generate questions, discover existing gaps and find appropriate resources to overcome these

gaps [62]. These students take charge of their own learning by planning, monitoring and evaluating their progress made [63]. In 2017, researchers conducted a research study about preparing teacher-students for the 21st century. An inquiry-based teaching method was utilised when 110 physics and chemistry teachers were involved. In their tasks, they managed to successfully explore and analyse the problems and formulated hypotheses and conducted their experiments. ICT applications were integrated by means of smart phones with cameras, laptop computers and software applications. As a result, the experiments were fruitful as they worked together with one common purpose [63].

It is a fact that in an inquiry-based phenomenon, students become their own teachers and take lead in the teaching and learning processes. The emergence of technology makes it easy for students to access their laboratory through the internet of things without them being in the physical laboratory [64-66]. It is imperative to note that virtual laboratory makes things bearable and cost-effective as compared to traditional laboratories as they are easy to maintain and safe to use. The science, engineering, and technology students find virtual laboratories user-friendly as they interact in inquiry-based classes where they can implement and analyse their own experiments and learn by using virtual objects and apparatus [67]. This teaching method also equips students with critical thinking, and innovative and team skills, which are valued for the twenty-first century job market [67]. A previous study initiated the STEM build website to coordinate online teaching and learning where instructors, markers, researchers, and Universal Design for Learning experts can easily interact [68].

Previously, engineering drawing was hand-based, now, online tutors are currently assisting these students. The utilisation of artificial intelligence makes it possible to provide real-time human-like feedback to students [69].

The emergence of the COVID-19 epidemic has seen the closure of campuses for schools and universities to ensure the safety of their students. As a result, online teaching and learning has become the reality in the 2020 academic year [70]. Online teaching and learning were found to be conducive for teaching students about circuits in electronic engineering [70] (Hu, 2020).

2.7. Competency-based learning

Over the past years, competency-based learning has rapidly gained momentum within the higher education space [71] and has been supported by ICT [74]. It is noticeable that over 896 institutions (universities and colleges) across the world prefer the competency-based learning method in their teaching and learning activities [71, 72]. Within competency-based learning, students depend more on their competency levels than on other people [21, 73]. Competency-based learning enables students to return to one or more competencies that are not yet mastered to avoid having to repeat courses as a result of failing [74]. Competency-based learning equips students with explicit and transferable learning objectives that provide them with personalised support and formative feedback so they can foster critical-thinking, problem-solving, communication and collaboration skills [74]. Competency-based models are dependent on measurable assessments linked to the related learning outcomes [75]. In most cases, competency-based learning is demonstrated through both formative and summative assessments [75]. However, at the Minnesota State University, both assessments methods (summative and formative) were successfully utilised for the Honors Program to assess the competency level of the students over time through electronic portfolios. As such, this assessment method replaces the standard research thesis required in numerous traditional Honors Programs [76].

In most instances, the competency-based method is used by professional bodies to measure the level of acquired competencies of graduates regarding their required training expected in industries [77]. Employers like to state to the institutions of higher learning what competencies they expect from the graduates. Hence, the importance of institutions of higher learning to always consult the companies based on what they expect their graduates to possess is imperative [78].

2.8. Communities of practice

The communities of practice approach combine experiential learning, social constructivism and connectivism, which allows people to work together and give each other moral support [79] and compassion [80] as they study. This method permits students to put their skills into practice and largely share their experiences [10, 81]. Communities of practice (COP) allow the combination of experiential learning, social constructivism and connectivism can while demonstrating the limitations of classifying learning theories. In addition, COP allows group of engineers to work together in order to solve general challenges [81-84]. It is argued that for teaching and learning to effectively take place, precedence should be on cognitive, social and teacher presence [85].

3. Research Methodology

As mentioned before, the current study aims to determine teaching approaches in SET, which are gaining popularity due to the technological evolution over the years. It also aims to ascertain the state of distance education evolution with reference to technology adoption within the Science, Engineering and Technology qualifications in the institutions of higher learning. As mentioned earlier, engineering students were traditionally afforded opportunities to experientially practice learned knowledge in a laboratory setup. It is therefore of vital importance to assess the literature if science, engineering and technology students in higher education are to continue receiving laboratory experiments while studying in an online environment. The study targeted literature from 2010 to 2020 to determine teaching approaches that are gaining popularity in SET within an online environment. The reviewed literature was obtained from various search engines such as Google Scholar, Scopus, Sabinet, ProQuest and EBSCO.

The following search strings were used to search the open-journal sites:

- Teaching approaches for science, engineering and technology
- Teaching approaches for distance education
- Teaching technologies for SET in distance education
- Integrative approach for distance education in the SET
- discipline
- Apprenticeship and experiential learning for distance

- education in SET
- Problem-based learning for distance education in SET
- Project-based learning for distance education in SET
- Competency-based learning for distance education in
- SET
- Inquiry-based learning for distance education in SET
- Case-based learning for distance education in SET
- Community of practice-based learning for distance
- education in SET
- AR in SET
- VR in SET
- Virtual laboratory in SET

The research study concentrated only on journals that focussed on institutions of higher learning within the SET discipline. All other journal papers focussing on primary schools and industry were excluded from the study. The results of the reviewed literature were presented in both table and graph formats for appropriate analysis. Pareto analysis was conducted to determine teaching methods and their respective technologies preferred in the SET qualifications in an online environment.

4. Research Findings and Discussion

Theories such as cognitive behaviour, social constructivism and connectivism formed the basis of the different approaches discussed in the current study. It is clearly indicated that teaching approaches for SET always pay attention to the practical implications on what students theoretically learn in a class setup. The evolution of distance education has seen traditional teaching approaches such as competency-based learning, project-based learning, problem-based learning, inquiry-based learning, case-based learning, and apprenticeship and experiential learning as becoming relevant in the twenty-first century through technologies such as VR, AR, virtual laboratories, AM, FCM, Proteus software, Wolfson Open Science Laboratory, Virtual Microscope etc. It should be noted that the findings of this study are entirely based on the reviewed literature sources. The results in Table 1, depict how different approaches transited from traditional face-to-face to online virtual practical so that students could continue putting their knowledge into practice.

The results from both Figure 1 and Table 1 show the popularity of these teaching approaches in an online environment emerging from 2010 until recently. The results also display that problem-based, apprenticeship and experiential learning, competency-based learning and inquiry-based learning are more popular in an online environment, particularly in SET. It suffices to mention that all highlighted teaching approaches are preferred in SET within online space but differ due to their popularity over the years.

Figure 1 elaborates on the popularity of teaching approaches suitable for online teaching and learning as per the author's contributions. The study results show that problem-based learning has gained popularity and exceeded all the other approaches, followed by apprenticeship and experiential approaches. The third important approach was found to be competency-based. However, case-based, project-based and community of practice-based approaches shared the fourth position for popularity. Last, but not the least, integrative approach proved to have less popularity. It is important to emphasise that all these approaches are applicable in an online environment within SET disciplines. It is evidence that from 2009,2010, 2011, 2012 these various teaching approaches were utilised mainly in face-to-face institutions. However, the emergence of the internet of things made it possible for utilising them for open and distance education, particularly 2017, 2019 and recently. In recent years as from 2015 until recent Virtual labs, virtual reality gave problem-based, apprenticeship and experiential learning as well as competency-based approaches impetus within the science and engineering space as seen in the table below. It is evidence from Figure 1 and Table 1 that other technologies such as LMS, FCM in education, 3D printers and AM in the manufacturing spaces continued to gain recognition as from 2016 and prominently in 2019. The study rightfully, depicted the gap in open and distance learning environment where similar teaching approaches in science and engineering where applicable.

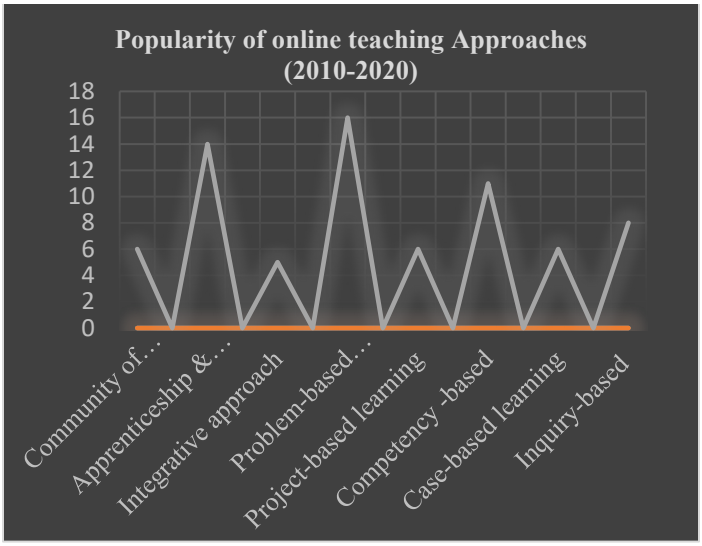


Figure 1: Popularity of teaching approaches

Table 1: Authors who endorsed teaching approaches

Authors endorsed these approaches	Learning Approaches & Preferred technologies	Used in SET	Online Learning delivery
[2011],[2015] [2017],[2018] [2019],[2019]	Community of Practice learning **Technologies:** Learning management system (LMS)	(6)	Yes
[2011][2012] [2013][2013] [2014][2015] [2017][2018] [2018][2018] [2018][2019] [2019][2020]	Apprenticeship & experiential **Technologies:** Virtual labs, AR, VR, LMS,	(14)	Yes

[2016][2017] [2018][2018] [2019]	Integrative approach **Technologies:** Virtual learning Systems; 3D printing	(5)	Yes
[2010][2017] [2017] [2017][2018] [2019][2020] [2020]	Inquiry-based learning **Technologies:** LMS	(8)	Yes
[2012][2014] [2014][2015] [2015][2017] [2015][2016] [2016][2017] [2017][2018] [2018][2018] [2019][2019]	Problem-based learning **Technologies:** AM; FCM; Proteus software; Wolfson Open Science Laboratory; Virtual Microscope	(16)	Yes
[2010][2016] [2016][2017] [2019][2020]	Project-based learning **Technologies:** FCM and LMS	(6)	Yes
[2015][2015] [2016][2016] [2016][2016] [2015][2016] [2017][2019] [2020]	Competency - based **Technologies:** FCM and LMS	(11)	Yes
[2011][2016] [2018][2018] [2019][2019]	Case-based learning **Technologies:** FCM and LMS	(6)	Yes

5. Conclusion

The emergence of the COVID-19 pandemic has provided an opportunity to many institutions to introduce various technologies to enhance teaching and learning activities for students and academics. Much more importantly, many institutions were not ready to offer non-venue-based examinations and remote laboratory practical. As a result, some institutions forged ahead and offered mid-year non-venue-based examinations while other institutions postponed all their examinations to October/November 2020. However, the issue of credible examination remained a concern to these institutions. These institutions want their students to write non-venue-based examinations, which are remotely invigilated by means of proctoring system to maintain their quality standards.

Equally important, institutions which offered campus-based practical experiments opted for alternative practical such as video-based and virtual practical to their students. Many of the institutions are only considering virtual practical now, as the COVID-19 situation does not show any sign of coming to an end soon. Hence, the importance of the current study so that institutions across the world can see various teaching approaches and technologies suitable for the online environment.

Traditionally, all practical experiments in science, engineering and technology institutions were conducted in physical laboratories. As such, most distance education institutions across the world collaborated with other campus-based institutions for their students to conduct laboratory work at those institutions, see Figure 1 and Table 1. However, such institutions always complained of the quality of practical work done in other institutions. In most cases, the Engineering Council of South Africa (ECSA) was concerned about the teaching and learning activities pertaining to engineering courses offered by distance education and those that were conducted online. ECSA, is a professional body that accredits all engineering qualifications in South Africa. They were more concerned about the practical work and the credibility of the assessments would ensue. Hence, the importance of the current study to publicise what other institutions across the world are doing regarding the virtual spaces utilising AR and VR technologies.

The findings of the current study indicate that virtual laboratories are the solution for most of the SET institutions operating in an online environment. The traditional teaching methods for civil engineering courses were no longer appropriate due to technological advancements [86]. Hence, the adoption of online courses for efficient learning. The introduction of micro-videos and final assessment methods made it possible for students in other areas to be involved [86]. Notably, most of the distance education institutions are already employing teaching approaches such as problem-based (16), apprenticeship and experiential learning(14), competency-based learning (11), inquiry-based learning (8), community of practice learning (6), case-based learning (6) and integrative-based learning (5).

Education technology has made it possible for students to learn outside of the classroom when they are at home. Both synchronous and asynchronous learning allows students to access teaching materials in their own comfortable space [87]. The study confirmed that many students in the engineering discipline developed an interest in teaching and learning where they were exposed to various assessment methods, linking theory for practice and case discussion [86]. This statement shows that the introduction of new technologies within the institutions of learning are enabling students to perform better, see Figure 1 and Table 1. The usage of different teaching methods plays an important role towards students' performances [88].

Both students and academics show that instant feedback provided by virtual laboratories yielded encouraging results [67]. Virtual laboratories began to gain popularity as early as 2012 and increasingly so in this digital era. This shows that many students and their academics prefer the introduction of the virtual laboratories in their learning environment. The institutions of higher learning are challenged by the current dispensation where

academics are expected to implement new pedagogical approaches and take advantage of ICT for teaching and learning, as can be seen in both Figure 1 and Table 1. As a result, these institutions need to equip the academics to be able to integrate technology within their teaching methods so that students continue to learn at ease. It is affirmed that the world-wide education system should incorporate the twenty-first century skills into curricula [63]. The current study acknowledges the challenges brought by the COVID-19 pandemic into the online institutions of learning, but the introduction of various teaching methods along with various technologies might bring relief if implemented correctly. Institutions are confronted with an opportunity to assess their current resources and equip their academics appropriately so that they can structure their assessments and practicals accordingly.

The current study proposes that a future study assess the implementation of various teaching approaches along their suitable technologies within the SET institutions in the southern Africa region.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research was made possible by the funding received from the National Research Foundation of South Africa and the research unit of the University of South Africa.

References

[1] M. G. Kanakana-Katumba, R. Maladzhi, "Online Learning Approaches for Science, Engineering and Technology in Distance Education," In 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 930-934. IEEE, 2019, doi.org/10.1109/IEEM44572.2019.8978892.

[2] C. Hall, W. Swart, S. Duncan, "Balancing customer needs and standards in higher education," Quality Approaches in Higher Education, **3**(1), 2–7, 2012.

[3] E.I. Allen, J. Seaman, "Changing course: Ten years of tracking online education in the United States," Sloan Consortium. PO Box 1238, Newburyport, MA 01950, 2013.

[4] T. Kattoua, M. Al-Lozi, M., A.A. Alrowwad, "A review of literature on E-learning systems in higher education," International Journal of Business Management & Economic Research, **7**(5), 754-762, 2016.

[5] P.A. Cohen, B.J. Ebeling, J.A. Kulik, A "meta-analysis of outcome studies of visual-based instruction," Educational Technology Research and Development, **29**(1):26–36, 1981.

[6] J. Barbara, A. McNeil, "Meta-analysis of interactive video instruction: A 10 year review of achievement effects," PhD thesis, University of Idaho, 1989.

[7] S.W. Bonham, D.L. Deardorff, R.J. Beichner, "Comparison of student performance using web and paperbased homework in college-level physics," Journal of Research in Science Teaching, **40**(10):1050–1071, 2003. https://doi.org/10.1002/tea.10120

[8] H. Fynewever, "A comparison of the effectiveness of web-based and paper-based homework for general chemistry," The Chemical Educator, **13**(4):264–269, 2008.

[9] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems," Educational Psychologist, **46**(4), 197-221, 2011, doi.org/10.1080/00461520.2011.611369.

[10] T. Bates, "Second Wave of DE and History of the Open University," 2011.

[11] J.F. Heydenrych and P. Prinsloo, "Revisiting the five generations of distance education: Quo vadis ?," Progressio. **32**(1):5–26, 2010.

[12] L.S. Mbati, "Exploring media blends for constructivist learning in open and distance e-learning (ODeL) environments," Diss, 2013.

[13] T. W. Argles, M. Brodeur, N. St J. Braithwaite, "Delivering authentic online practical science teaching–geoscience perspectives from the OpenScience Laboratory," In International Geological Congress, Abstracts, 35, 2017.

[14] N. Wengrowicz, W. Swart, R. Paul, K. Macleod, D. Dori, Y.L. Dori, "Students' collaborative learning attitudes and their satisfaction with online collaborative case-based courses, " American Journal of Distance Education, **32**(4), 283-300, 2018, doi.org/10.1080/08923647.2018.1511509.

[15] S. Jiang, J. Xu, "Construction and Commissioning of Engineering Graphics Online Course," In Education Teaching Forum, **12**, 102-103, 2018.

[16] N. Brouwer, G. Fleerackers, N. H. Majcen, I. Maciejowska, C. McDonnell, and M. Mocerino, "Online course to improve university laboratory teaching practice," VIRT&L-COMM, 10, L-COMM, 2016.

[17] J. Wang, "Research on the classroom teaching reform of financial management" under the background of online course construction," Think Tank Era, **34**, 66-67, 2019.

[18] R. Bose, "Virtual Labs Project: A Paradigm Shift in Internet-Based Remote Experimentation," Access IEEE, **1**, 14, 2013, doi.org/10.1109/ACCESS.2013.2286202.

[19] E.G. Dada, A.H. Alkali, D. O. Oyewola, "An Investigation into the Effectiveness of Asynchronous and Synchronous E-learning Mode on Students' Academic Performance in National Open University (NOUN)," Maiduguri Centre. International Journal of Modern Education & Computer Science, **11**(5), 2019. doi: 10.1002/tea.10120

[20] A. Mikroyannidis, A. Gómez-Goiri, A. Smith and J. Domingue, "Online experimentation and interactive learning resources for teaching network engineering," IEEE. Global Engineering Education Conference (EDUCON), Athens, 181-188, 2017, doi.org/10.1109/EDUCON.2017.7942845.

[21] T. Bates, "Teaching in a Digital Age: Guidelines for Designing and Learning," Tony Bates Associates Limited, 2019.

[22] A. Y. Kolb, D.A. Kolb, "The learning way: Meta-cognitive aspects of experiential learning," Simulation & gaming, **40**(3), 297-327, 2009, doi.org/10.1177/1046878108325713.

[23] A. Stern, Y. Rosenthal, N. Dresler, Ashkenazi, D, "Additive manufacturing: An education strategy for engineering students," Additive Manufacturing, 27, 503-514, 2019, doi.org/10.1016/j.addma.2019.04.001.

[24] N. Kholis, M. S. Zuhrie, R. Rahmadian, "Innovation Online Teaching Module Plus Digital Engineering Kit with Proteus Software through Hybrid Learning Method to Improve Student Skills," Materials Science and Engineering, 336, 012036, doi.org/10.1088/1757-899X/336/1/012036.

[25] G. Guler, C. Sen, Z.S. Ay, A. Ciltas, "Engineering Skills that Emerge During Model-Eliciting Activities (MEAs) Based on 3d Modeling Done with Mathematics Pre-Service Teachers," International Journal of Education in Mathematics, Science and Technology, **7**(3), 251-270, 2019.

[26] L. Luo, "Analysis on Online and Offline Blended Teaching Design Taking Engineering Survey," Course of Non-surveying and Mapping Engineering Specialty as an Example. In 3rd International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2019), Atlantis Press, 2019.

[27] N. Mestrinho, B. Cavadas, "Innovation in Teacher Education: An Integrative Approach to Teaching and Learning Science and Mathematics," in Multidisciplinary Digital Publishing Institute Proceedings, **2**(21), 1343, 2018, doi.org/10.3390/proceedings2211343.

[28] A.O. Akande, J.P. Van Belle, "The use of Software as a Service to Facilitate Quality Learning Experience for Students in Higher Education Institutions," Simulation (ISSN 2518-0959 Online), **2**, 2017, doi.org/10.24178/ijsms.2017.2.2.08.

[29] P. Padayachee, S. Wagner-Welsh, H. Johannes, "Online assessment in Moodle: A framework for supporting our students," South African Journal of Higher Education **32**(5): 211-235, 2018, doi.org/10.1080/03043797.2015.1095162.

[30] R. Bhathal, "An appraisal of an online tutorial system for the teaching and learning of engineering physics in conjunction with contextual physics and mathematics, and relevant mathematics," European Journal of Engineering Education, **41**(5), 504-511, 2016.

[31] S. Frerich, D. Kruse, M. Petermann, A. Kilzer, "Virtual Labs and Remote Labs: Practical experience for everyone," Proc. IEEE Global Engineering Education Conference (ED UCON), 312-314, 2014, doi.org/10.1109/EDUCON.2014.6826109.

[32] J. M. Marquez-Barja, N. Kaminski, F. Paisana, C. Tranoris and L. A. DaSilva, "Virtualizing testbed resources to enable remote experimentation in online telecommunications education," Proc. IEEE Global Engineering Education Conference (EDUCON), 836-843, 2015, doi.org/10.1109/EDUCON.2015.7096069.

[33] L. Li, Y. Chen, Z. Li, D. Li, F. Li, H. Huang, "Online Virtual Experiment Teaching Platform for Database Technology and Application," 13th International Conference on Computer Science & Education (ICCSE), Colombo, 1-5, 2018. http://dx.doi.org/10.1109/ICCSE.2018.8468849

[34] G. Altuger-Genc, Y. Han, Y. Genc, "Towards Simulation Aided Online Teaching: Material Design for Applied Fluid Mechanics," International

Journal of Online and Biomedical Engineering (iJOE), **14**(12), 112-125. 2018, doi.org/10.3991/ijoe.v14i12.8630.
[35] M. Hincapié, A. Caponio, H. Rios, E.G. Mendívil, "An introduction to Augmented Reality with applications in aeronautical maintenance," In 2011 13th International Conference on Transparent Optical Networks, 1-4, IEEE, 2011, doi.org/10.1109/ICTON.2011.5970856.
[36] I. Sanchez-Rola, M. Dell'Amico, P. Kotzias, D. Balzarotti, L. Bilge, P.A. Vervier, I. Santos. "Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control," In Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security, 340-351, 2019.
[37] H.K. Wu, S.W.Y. Lee, H.Y. Chang, J.C. Liang, "Current status, opportunities and challenges of augmented reality in education," Computers & education, **62**, 2013, 41-49, doi.org/10.1016/j.compedu.2012.10.024.
[38] R. Azuma, R. n.d, "A Survey of Augmented Reality. iitk.ac.in". Retrieved August 2020.
[39] T. Olsson, T. Kärkkäinen, E. Lagerstam, L.Ventä-Olkkonen, "User evaluated of mobile augmented reality scenarios" Journal of Ambient Intelligence and Smart Environments, **4**, 2012, 29-47.
[40] N.S. Ismail, J. Harun, M.A.Z.M. Zakaria, and S.M. Salleh, "The effect of Mobile problem-based learning application DicScience PBL on students' critical thinking," Thinking Skills and Creativity, **28**, 177-195. 2018, doi.org/10.1016/j.tsc.2018.04.002.
[41] S. Han, R. Capraro, M.M. Capraro, "How science, technology, engineering, and mathematics (STEM) project-based learning (PBL) affects high, middle, and low achievers differently: The impact of student factors on achievement," International Journal of Science and Mathematics Education, **13**(5), 1089-1113, 2015, doi.org/10.1007/s10763-014-9526-0.
[42] S. Ford, T. Minshall, "Invited review article: Where and how 3D printing is used in teaching and education," Additive Manufacturing, **25**, 131-150. 2019, doi.org/10.1016/j.addma.2018.10.028.
[43] J. Bergmann, A. Sams, "Flip your classroom: Reach every student in every class every day International Society for Technology in Education," Washington, DC, 2012.
[44] S.J. DeLozier, M.G. Rhodes, "Flipped classrooms: A review of key ideas and recommendations for practice," Educational Psychology Review (2016), 1-11, doi.org/10.1007/s10648-015-9356-9.
[45] Y. Chen, Y. Wang, Kinshuk, N.S. Chen, "Is FLIP enough? Or should we use the FLIPPED model instead?" Computers & Education, **79**, 16-27, 2014, doi.org/10.1016/j.compedu.2014.07.004.
[46] M. Katsa, S. Sergis, D.G, Sampson. "Investigating the potential of the flipped classroom model in K-12 mathematics teaching and learning Proceedings of the 13th international conference on cognition and exploratory learning in digital age", 2016.
[47] T. Muir, V. Geiger, "The affordances of using a flipped classroom approach in the teaching of mathematics: A case study of a grade 10 mathematics class," Mathematics Education Research Journal, **28**(1), 149-171, 2016 doi.org/10.1007/s13394-015-0165-8.
[48] M.N. Giannakos, J. Krogstie, N. Chrisochoides. "Reviewing the flipped classroom research: Reflections for computer science education," Proceedings of the computer science education research conference, 23-29, 2014, doi.org/10.1145/2691352.2691354.
[49] S. Reza, M.I. Baig, "A study of inverted classroom pedagogy in computer science teaching," International Journal of Research Studies in Educational Technology, **4**(2), 19-30, 2015, doi.org/10.5861/ijrset.2015.1091.
[50] V. Aidinopoulou, D.G. Sampson, "An action research study from implementing the flipped classroom model in primary school history teaching and learning," Journal of Educational Technology & Society, **20**(1), 237-247, 2017.
[51] S., Sergis, D.G. Sampson, L. Pelliccione, "Investigating the impact of Flipped Classroom on students' learning experiences: A Self-Determination Theory approach," Computers in Human Behavior, **78**, 368-378, 2018, doi.org/10.1016/j.chb.2017.08.011.
[52] B. Turk, S. Ertl, G. Wong, P.P. Wadowski, H. Löffler-Stastka, "Does case-based blended-learning expedite the transfer of declarative knowledge to procedural knowledge in practice?" BMC Medical Education, **19**(1), 1-10, 2019, doi.org/10.1186/s12909-019-1884-4.
[53] G. Marcus, R. Taylor, A. Ellis, "Implications for the design of online case based learning activities based on the student blended learning experience," in Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference, 557-586, 2004.
[54] P. Nicklen, J.L. Keating, S. Paynter, M. Storr, S. Maloney, "Remote-online case-based learning: a comparison of remote-online and face-to-face, case-based learning-a randomized controlled trial," Education for Health, **29**(3), 195, 2016.
[55] Newson, T.A, N.J. Delatte, "Case methods in civil engineering teaching," Canadian Journal of Civil Engineering, **38**(9), 1016-1030, 2011.
[56] S. Bell, "Project-based learning for the 21st century: Skills for the future," The clearing house, **83**(2), 39-43, 2010, doi.org/10.1080/00098650903505415.
[57] R. Belwal, S. Belwal, A.B. Sufian, A. Al Badi, "Project-based learning (PBL): outcomes of students' engagement in an external consultancy project in Oman," Education and Training, 2020.
[58] R. A. Ralph, "Post-secondary project-based learning in science, technology, engineering and mathematics," Journal of Technology and Science Education, **6**(1), 26-35, 2016, doi.org/10.3926/jotse.155.
[59] D. Kokotsaki, V. Menzies, A. Wiggins, "Project-based learning: A review of the literature Improving Schools," **19**(3) (2016), 267-277.
[60] S. K.W. Chu, R.B. Reynolds, N.J. Tavares, M. Notari, C.W.Y. Lee, "21st Century skills development through inquiry-based learning," Singapore: Springer Singapore. 2017, doi.org/10.1007/978-981-10-2481-8.
[61] Y. Dong, J. Song, S. Yi, X. Li, "Reform Practice of Engineering Drawing Courses in Chinese Colleges by the Blended Teaching Method with Different Teaching Modes and Resources Based on the XuetangX MOOC Platform," In 2019 2nd International Conference on Education, Economics and Social Science (ICEESS 2019). Atlantis Press, 2019, doi.org/10.2991/iceess-19.2019.70.
[62] C.E. Hmelo-Silver, C.A. Chinn, Chan, A.M. O'Donnell, "International handbook of collaborative learning," New York, NY: Routledge, 2013, doi.org/10.4324/9780203837290.
[63] P. Häkkinen, S. Järvelä, K. Mäkitalo-Siegl, A. Ahonen, P. Näykki, T. Valtonen, "Preparing teacher-students for twenty-first-century learning practices (PREP 21): A framework for enhancing collaborative problem-solving and strategic learning skills Teachers and Teaching," **23**(1), 25-41, 2017, doi.org/10.1080/13540602.2016.1203772.
[64] C. Soudien, "Transformation in higher education : A briefing paper," 2010.
[65] K. Kaen, "A Contextual O nline Game based on Inquiry Learning Approach for Improving Students' Learning Performance in a Chemistry Course," 601-608, 2017.
[66] V. Fruiht, "Project-Based Curriculum as a Guide for the Re-Sequencing of Discipline-Specific Statistics Courses," 2018.
[67] T. Lynch, I. Ghergulescu, "Review of virtual labs as the emerging technologies for teaching STEM subjects," In INTED2017 Proc. 11th Int. Technol. Educ. Dev. Conf. 6-8 March Valencia Spain, 6082-6091, 2017, doi.org/10.21125/inted.2017.1422.
[68] M. V. Ramirez, C.L. Gordy, "STEM build: an online community to decrease barriers to implementation of inclusive tactile teaching tools," Journal of Microbiology & Biology Education, **21**(1), 2020, doi.org/10.1128/jmbe.v21i1.1963.
[69] E. Hilton, B. Williford, W. Li, T. Hammond, J. Linsey, "Teaching Engineering Students Freehand Sketching with an Intelligent Tutoring. Inspiring Students with Digital Ink: Impact of Pen and Touch on Education", 135, 2019.
[70] H. Hu, S., Song, Y. Chai, Y. Gong, "Study on Online Teaching of High Frequency Circuit under COVID-19 Epidemic", Open Journal of Social Sciences, **8**(3), 493-496, 2020, doi.org/10.4236/jss.2020.83041.
[71] T.R. Nodine, "How did we get here? A brief history of competency-based higher education in the United States," The Journal of Competency-Based Education, **1**(1), 5-11, 2016, doi.org/10.1002/cbe2.1004.
[72] R, Kelchin, 2015, "The landscape of competency-based education: Enrollments, demographics, and affordability," Washington, DC: American Enterprise Institute.
[73] S. Directors, C. Technical, E. Consortium, "Building a strong competency-based pathways and career," 2015.
[74] P. Williams, "Does competency-based education with blockchain signal a new mission for universities?" Journal of higher education policy and management, **41**(1), 104-117, 2019, doi.org/10.1080/1360080X.2018.1520491.
[75] A.A. Kytmanov, M.V. Noskov, K.V. Safonov, M.V., Savelyeva, V.A. Shershneva, "Competency-based learning in higher mathematics education as a cluster of efficient approaches," Bolema: Boletim de Educação Matemática, **30**(56), 1113-1126, 2016, doi.org/10.1590/1980-4415v30n56a14.
[76] G. Walters, "Developing competency-based advising practices in response to paradigm shifts in higher education," NACADA Journal, **36**(1), 66-79, 2016, doi.org/10.12930/NACADA-15-033.
[77] S. Perera, S.O. Babatunde, J. Pearson, D, Ekundayo, "Professional competency-based analysis of continuing tensions between education and training in higher education," Higher Education, Skills and Work-Based Learning, 2017, doi.org/10.1108/HESWBL-04-2016-0022.
[78] J. Henrich, "Competency-based education: The employers' perspective of

higher education," The Journal of Competency-Based Education, **1**(3), 122-129, 2016, doi.org/10.1002/cbe2.1023.
[79] D, Brittany, J.E. Coryell, "Relationships, Participation, and Characteristics of a Community of Practice in S hort-Term Adult Study Abroad," Adult Learning, 2019.
[80] A. ettit, A. McVicar, P. Knight-Davidson, A. Shaw-Flach. "Releasing Latent Compassion through an Innovative Compassion Curriculum for Specialist Community Public Health Nurses," Journal of Advanced Nursing **75**(5), 53-62, 2019.
[81] B.W. Trayner, "Communities of practice a brief introduction," 1-8, 2015.
[82] A. Cox, "What are communities of practice? A critical review of four seminal works. Paper presented at the Fifth European Conference on Organizational Knowledge, Learning, and Capabilities," in Fifth European Conference on Organizational Knowledge, Learning, and Capabilities, 2004.
[83] B.W. King, "Communities of Practice," no. August, 2017.
[84] L. Leonard, C. Kies, N. Braaf, "Reflections: eLearning Colloquia enhance a Community of Practice," 2018, doi.org/10.4324/9780429283499-2.
[85] D. R. Garrison, M. C. Leveland-Innes, "Facilitating Cognitive Presence in Online Learning : Interaction Is Not Enough in Online Learning : Interaction Is Not Enough," Am. J. Distance Educ., **19**(3), 133-148, 2005.
[86] C.H.E.N. Tao, L.I. Ju-xuan, "Teaching Reform in the Course of Civil Engineering Introduction Under the Background of Online Courses," In International Conference on Modern Educational Technology and Innovation and Entrepreneurship (ICMETIE 2020), 76-80, Atlantis Press, 2020.
[87] M.M. Sahabadi, M. Uplane, "Synchronous and Asynchronous e-learning Styles and Academic Performance of e-students. Procedia - Social and Behavioral Sciences", **176**, 129-138, 2015.
[88] P. Ahmadi, K.I, Yousaf, "An Empirical Study of Teaching Methodologies and Learning Outcomes for Online and in-class Networking Course Sections," arXiv preprint arXiv:1711.11411, 2017.
[89] Y, Grossman, "Introduction to flavour physics," In LHC Phenomenology, 35-80. Springer, Cham, 2015, doi.org/10.1007/978-3-319-05362-2_2
[90] P. Fain, "Moving ahead with competency," Inside Higher Education, 2015, https://www.insidehighered.com/news/2015/02/17/two-year-colleges-washington-state-expand-competency-based-project-business-degree
[91] T. Mitchell, "Guidance to competency-based education experimental site released," Home Room, the official blog of the U.S. Department of Education, 2015.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1217-1224 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Development and Performance Analysis of HRPL Using 6LoWPAN CC2538 Module for IoT Ecosystem

Nin Hayati Mohd Yusoff*, Nurul Azma Zakaria

Center for Advanced Computing Technology, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia, Durian Tunggal, Melaka, 76100, Malaysia

ARTICLE INFO

Article history:
Received: 09 September, 2020
Accepted: 01 December, 2020
Online: 16 December, 2020

Keywords:
6LoWPAN
Internet of Things
Smart Home
Performance Analysis

ABSTRACT

The Internet of Things (IoT) application has been experiencing increasingly progressive demand, especially for embedded devices (ED). However, the ED has limited capabilities, low power consumption resources, and low bandwidth in connecting to the Internet by using Wireless Sensor Networks (WSNs). Therefore, the WSNs form creates a necessity for new technologies and protocols for the IoT implementation. Thus, IPv6 Low Power Area Network (6LoWPAN) was designed by the Internet Engineering Task Force (IETF) to overcome the Internet Protocol (IP) based communication that allows direct communication between each ED. Nevertheless, the communication between ED using 6LoWPAN becomes challenging in designing routing protocols to achieve the efficient performance Quality of Service (QoS). Among the existing protocols for the 6LoWPAN network, RPL is considered effective for the 6LoWPAN system. However, the Power Consumption (PC) and routing overhead of RPL is high when it was implemented in a real scenario. Therefore, HRPL was proposed to enhance the RPL by introducing the rebroadcast technique in order to minimize the routing overhead at the same time reduced the PC usage. Thus, this paper is an extended version, in which 6LoWPAN Smart Home Testbed (6LoSH) was developed to investigate the impact of the number of nodes on PC usage for both protocols (HRPL and RPL) in a real scenario. The result shows for this instance, HRPL has succeeded in reducing the use of PC for both experiments (simulation and 6LoSH). On the other hand, the number of nodes had given an impact on PC usage. For further work, we plan to use multiple topologies and the different number of nodes to explore the HRPL for some performance metrics such as convergent time and latency.

1. Introduction

This paper is an extended version of work published in [1]. We are making this version available in order to have more results and discussions in comparison to its short version.

The 6LoWPAN was introduced by IETF to support a huge number of IP-based embedded devices that are connected to the Internet with a low power network capability. Moreover, the rapid development of the Internet of Things (IoT) ecosystem in multiple areas [2] such as healthcare, smart building automation [3], vehicle [4], etc. had made the embedded device to be in high demand [5]. The embedded devices are categorized as limited power usage, limited cost, limited bandwidth, and small memory usage [6]. Meanwhile, the development of IoT devices become a serious problem to enable interoperability and scalability among devices and systems [7,8]. Thus, the 6LoWPAN protocol was designed to complete and complement the technological constrain as an essential part of the IoT development application and ecosystem. 6LoWPAN that enables the ED which are mainly sensor nodes to connect to the Internet over IEEE 802.15.4 in Physical (PHY) layer and Medium Access Control (MAC) sublayer. Figure 1 presents the 6LoWPAN stack where the adaptation layer was added between the data link physical layer and network layer. 6LoWPAN is a wireless embedded internet [9]

*Corresponding Author: Nin Hayati Mohd Yusoff, nin6699@gmail.com

that integrates IPv6 with WSN [10] that allows the IP-based ED to connect and exchange information directly with other devices on the Internet [9]. Due to this matter, the energy used for connectivity will save, and for this condition, 6LoWPAN network becomes an essential part of the IoT deployment [11].

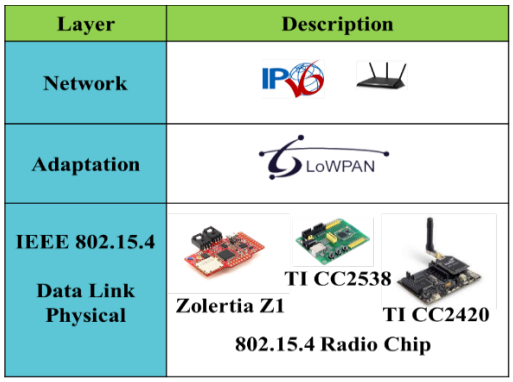


Figure 1: 6LoWPAN Stack

However, the infrastructure of IPv6/6LoWPAN and how IPv6 can integrate heterogeneous technology (system, tool, application, etc.) towards a new embedded application protocol (6LoWPAN) for IoT ecosystem become a challenging issue. This because the current standard of 6LoWPAN is limited to 250 kbps, and 127 bytes of frame length compared to IEEE 802.15.4 packet size is 1280 bytes. As a result, it is critically affected for the routing algorithm to optimize the path formation. Indirectly, it will degrade the interoperability of 6LoWPAN devices/applications. Besides, the implementation, integration and evaluation for large scale deployments of sensors networks in buildings, the potential heterogeneous combination of wireless and wired need to optimize the energy consumed by the system itself.

In addition, the 6LoWPAN protocol develops to optimize the routing scheme for converging cast traffic patterns for low power nodes [12]. One of the 6LoWPAN routing protocols is Protocol Over Low-Power and Lossy Networks (RPL) as defined in RFC6550 [13]. Based on previous research, the RPL protocol in 6LoWPAN is effective in terms of delay and Packet Delivery Rate (PDR). But, it can be further improved in terms of overhead as the original value is very high [14] [15]. According to Feschel, Huhn, & Schwetlick in 2012, the fragmentation of the frames should minimize the routing overhead and power consumption in order to save network lifetime that is a critical parameter to 6LoWPAN nodes. On the other hand, to optimize network performance, the routing overhead must be reduced. At the same time, the 6LoWPAN system offers the scalability and as well as interoperability between nodes. Therefore, this research project addresses algorithmic, networking, and application development using the 6LoWPAN CC2538 module in improving the 6LoWPAN protocol to achieve efficient performance QoS of the network in terms of PC. The proposed algorithm is implemented and tested using a simulation study and validated with the experimental approach by developing the 6LoWPAN Smart Home testbed (6LoSH). The initial phase of this study focused on a literature review that has been done in [17] to address the two (2) following challenges: i) Comparative analysis of existing systems and solutions and ii) Comparative analysis of existing standards and protocols. As a reminder of this paper, we present our work as follows. We describe a methodology used in Section II follows by Section III present the experiment setup. Then, Section IV describes the evaluation results. Finally, we emphasize the contribution of this paper and future works.

2. Methodology

The methodology is divided into three parts, as presented in Figure 2. The first part is the development process, followed by implementation, and the last part is testing and validity.

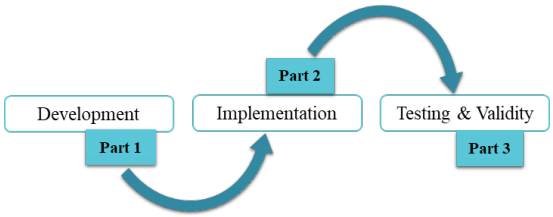


Figure 2: The Methodology Design

2.1. Development

The development of 6LoSH involves three (3) important aspects. First, the integration between the 6LoWPAN system requirement and building infrastructure with different technologies like sensors, actuators, and mobile devices. Figure 3 presents the proposed 6LoSH network design, which implements a simple node architecture [11], [9], [18]. 6LoWPAN nodes communicated with each other through border router, and share a backbone link to connect to the Internet. This allows all the 6LoWPAN nodes remain the same as the IPv6 prefix. Besides, 6LoWPAN connects to another IP network via a border router. At the same time, the border router is handling the 6LoWPAN compression and Neighbor Discovery (ND) for 6LoWPAN nodes. According to [11] the power consumption usage using simple architecture remain low when the intelligent objects are idle, and increases the data rate in both host and router mode. The end-users will be able to interact with the system by running several applications from their PCs, mobile phones, notebooks, and voice. These applications will make use of the services provided by the building system. They will allow end-user to obtain information about the state of the building and switched off/on the application over the Internet.

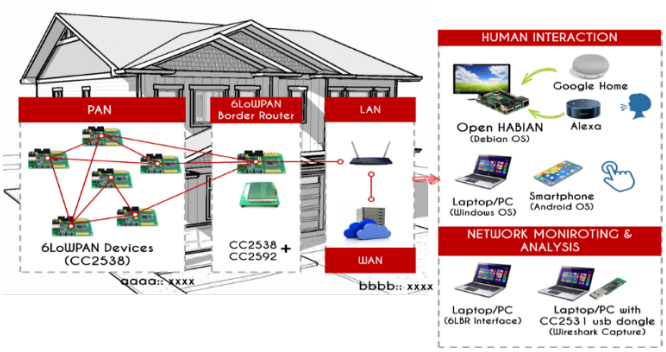


Figure 3: Simple 6LoWPAN Nodes (CC2538 Module) Architecture for 6LoSH

Second, the 6LoWPAN node, other hardware, and software requirements towards a new embedded application in developing 6LoSH. Table 1 shows the list of the equipment required for the implementation of the 6LoSH physical deployment. We used the TI CC2538 module [19] as a 6LoWPAN node and border router because it is a very low power wireless module for IoT applications [20], Figure 4. In contrast, Raspberry Pi 3 Model B is used for running the 6LoWPAN Border Router (6LBR). We observed 6LoWPAN packets by using a CC2531 traffic sniffer and the Wireshark packet capture tool with baud rate 115200. Besides, the Contiki OS [21] [22] by performed COOJA simulator is used to set up the CC2538 program and running simulation of the 6LoWPAN network.

Table 1: The equipment required for the implementation of the 6LoSH physical deployment

Hardware and Software Requirement		
Equipment	Count	Types, and Description
Node	10	• The CC2538 Texas Instrument
Border Router	1	
Sniffer	1	• CC2531 USB Dongle
Raspberry Pi 3	1	• Running 6LoWPAN Border Router (6LBR) • Model B (Debian OS)
Contiki OS (Cooja Simulator)		• Perform simulation testing • Writing a program for CC2538 module
Wireshark		• Packet capturing and analyzing the OTA packet

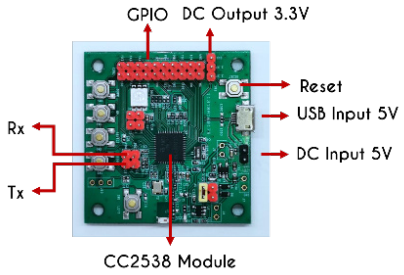


Figure 4: CC2538 Board

Finally, a novel algorithmic models and scalable solutions for the 6LoWPAN system. The on/off an application for embedded devices has been designed and flashed onto the CC2538 board, namely 6Node. The real-time arrangement of the 6Node setup is shown in Figure 5. Table 2 presents the proposed 6Node protocol parameter setup. The default trickle algorithm [15] is used to control the amount of DIO message in routing traffic by setting the parameter of DIO Interval Min as 12ms and DIO interval doublings as 16ms. To achieve the best Packet Delivery Ratio (PDR), the Minimum Rank with Hysteresis Objective Function (MRHOF, RFC6719) with Expected Transmission Count (ETX, RFC 6551) metrics are used as path selection mechanism [23]-[24].

Figure 5: Real-Time Arrangement of CC2538 setup

Table 2: System Parameter for CC2538 Protocol

Parameter	Algorithm/Coding

Trickle Algorithm	case RPL_TRICKLE_T_TIMEOUT:
	rpl_trickle_handle_t_timeout();
	break;
	case RPL_TRICKLE_I_TIMEOUT:
	rpl_trickle_handle_i_timeout();
	break;

Nodes rank	#define RPL_OCP_OF0 0 In #define RPL_OCP_MRHOF 1
RPL_Dio Interval Min	#ifdef RPL_CONF_DIOINTERVALMIN #define RPL_DIOINTERVALMIN RPL_CONF_DIOINTERVALMIN #else #define RPL_DIOINTERVALMIN 12 #endif
DIO interval doublings	#ifdef RPL_CONF_DIOINTERVALDOUBLINGS #define RPL_DIOINTERVALDOUBLINGS RPL_CONF_DIOINTERVALDOUBLINGS

2.2. Implementation

The implementation process involves the deployed hardware infrastructure, physical network, and integrated with software services for managing and controlling the hardware. Dealing with devices and technologies, it has explored existing platforms and solutions that could be relevant for 6LoSH to achieve effective communication (interoperability) between the nodes and systems [7], [8]. The home appliances are connected to nodes through a relay which in turn is connected to another digital pin of the board as illustrated in Figure 6. Figure 7 present the 6Node prototypes and Figure 8 show the development of 6Node prototype activities.

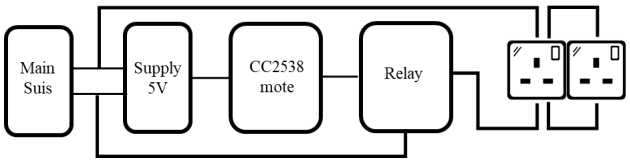


Figure 6: The Block Diagram of CC2538 Application (6Node)

2.3. Testing and validity

The performance of the 6LoSH platform has been evaluated and validated through simulation and calibrate the simulation result using an experiment to ensure the 6Node reach the efficiency of QoS. The HRPL protocol introduced in [1] and RPL

standard was used to assess the effectiveness of 6Node in reducing routing overhead toward new embedded applications for the 6LoWPAN environment.

Figure 7: The 6Node Prototypes

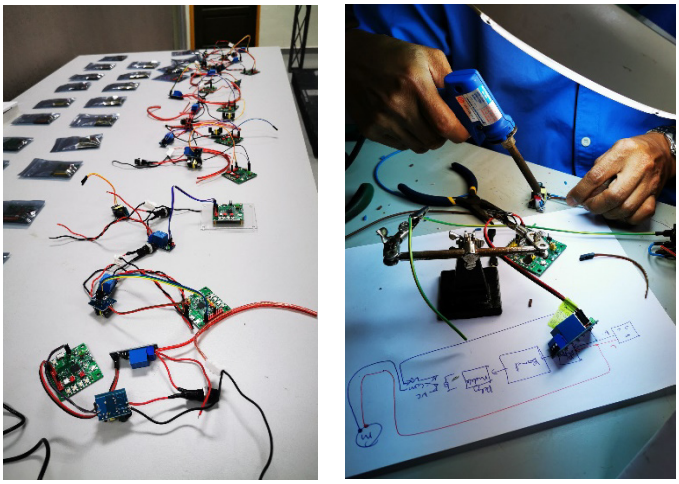


Figure 8: The development of 6Node Prototype Activities

3. Experiment Setup

Figure 6 shows the testing design and Table II describes a summary of the evaluation parameter setting for both experiments that was adapted from the existing model in [25] [26,27]. The comparative study in terms of power consumption has been done by using standard HRPL and RPL protocol to investigate the scalability and interoperability communication between devices (6LoWPAN appliances) and network performance. To be fair, the open-source implementations run on the same operating system (OS) which is Contiki OS. The HRPL and RPL protocol was simulated using the Cooja simulator [28,29] (Testing 1). Then, we compared the simulation results with the evaluation of HRPL and RPL through testbed experiments (testing 2). We created similar scenarios for both experiments. Lastly, we evaluated the performance of the HRPL protocol based on defined metrics. Figure 10 present the real-time arrangement of 6LoSH testing.

3.1. Simulation (Testing 1)

Tmote Sky CC2420 was used as a platform. The set of simulation scenarios (4, 6, 8 and 10 nodes) used in this study as shown in Figure 11. Firstly, the implementation of the RPL protocol from the standard examples of RPL collects file that contains source code of sink and sender nodes are simulated. We set a sink node in the middle of the network and deploy the sender node randomly. We set a transmission range (Tx) to 20m and interference range (INT) to 20m at Unit Disk Graph Model (UDGM) [30]. The radio messages were enabled to capture the traffic and saved as a pcap file. The saved pcap file would be used to analyze the network using Wireshark network analyzer. Then, the standard RPL collect file is modified by changing RPL library to HRPL library, in both source codes of sink and sender nodes. Next, the HPRL protocol is simulated in the same scenario. Lastly, the result between RPL and HRPL is compared with four types of power consumption: CPU Power, Transmit Power, Listen Power, and Low Power Mode (LPM). The simulation results will be evaluated for the different number of nodes (4, 6, 8, and 10 nodes).

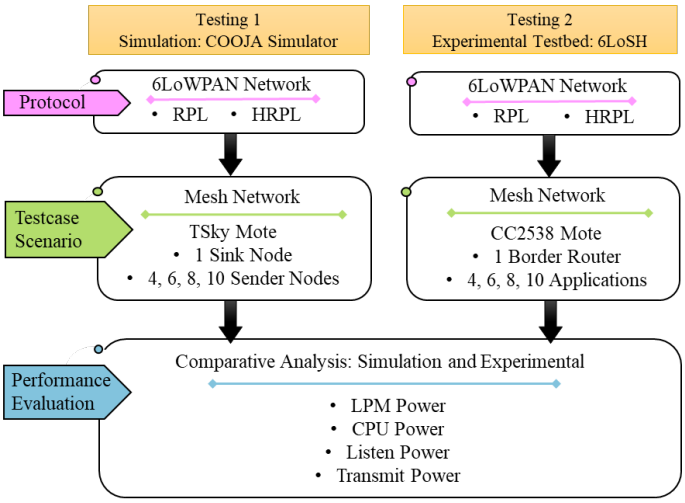


Figure 9: The 6LoSH Testing Design

Table 3: The Summary Parameter Seting for 6LoSH Testing

Method	Testbed (6LoSH)	Simulation
Protocol	HRPL	HRPL
Transport	UDP/IPv6	UDP/IPv6
Mote Type	CC2538 Module	Tmote Sky
Reception Ratio (Rx) and Transmission ratio (Tx)	-	100% & 100%
Packet Size	160bytes	160bytes
Queue	150 packets	250 packets
Transmission Power	-17dBm	0dBm
Transmission range	20m	20m
Network Scale	4, 6, 8, 10 nodes	4, 6, 8, 10 nodes
Length	60 minutes	60 minutes
Packet Capturing	Pcap (Wireshark)	Pcap (Wireshark)

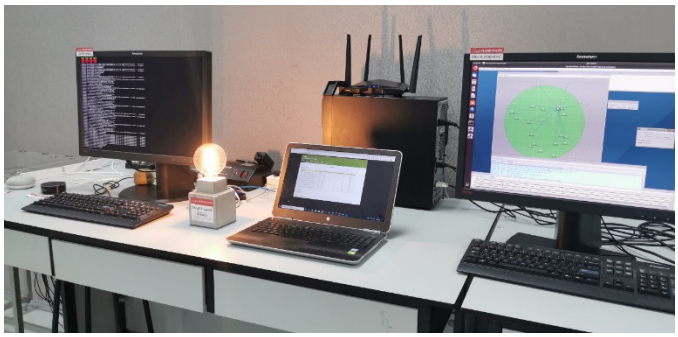


Figure 10: The Real-Time Arrangement of 6LoSH Testing

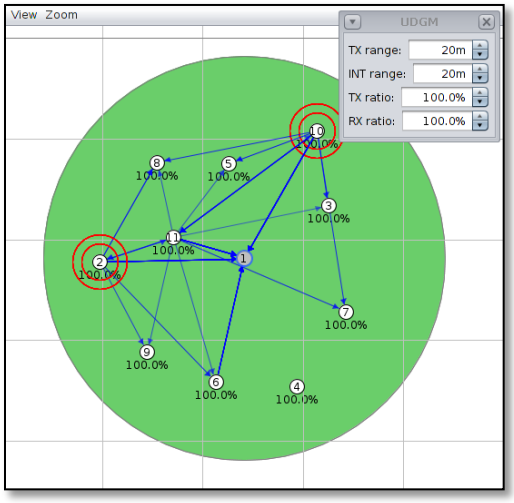


Figure 11: The Set of Simulation Scenario

3.2. Testbed (Testing 2)

The testbed results are used to calibrate the simulation results and also to validate the performance of HRPL in a real-scenario. In this scenario, one border router (6LBR) is located in the center and the different number of applications (4, 6, 8, and 10 nodes) were deployed randomly on the first floor in one of the houses in Malaysia, as presented in Figure 12.

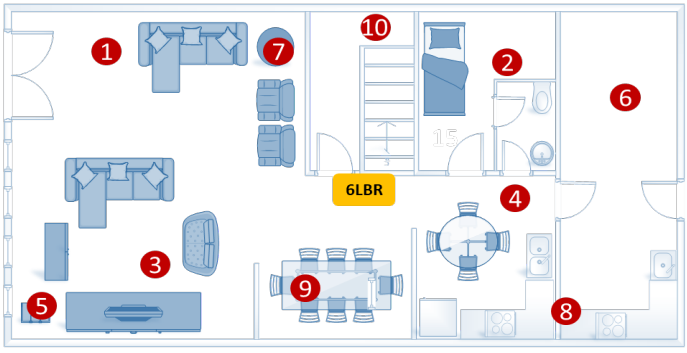


Figure 12: The 6Node Deployment

3.3. Performance Metric

The power consumptions are monitor in both experiments through the development of the Powertrace application in the Contiki operating system. There are four types of power are measured:

- CPU Power,
- Low Power Mode (LPM) Power
- Transmit Power and
- Listen Power.

CPU power refers to the power of computation required by the node in active mode while LPM power refers to the power used by the nodes in CPU sleep mode. Meanwhile, the energy used by each node to transmit the data package to the neighbors is called transmit power, and listening power refers to the energy required by each node in ready mode to receive the data package.

4. Results and Discussion

We assumed our experiment a random topology where the nodes were deployed randomly in an area of 20m x 20m (simulation) and 20m x 20m (testbed). We put the sink node in the center of the network. In order to enable interconnection between all the sender nodes, we set up each node with multi-hop transmission distances from the border router. For both experiments, we considered the DAG root at the border router. Then we set up the nodes to generate packets 1 packet/min randomly in a fixed interval time. We derived this analysis using a script to determine when all nodes had printed to the simulator output (simulator) and web interface (testbed) that they joined the DAG.

Table 4: Analysis of Total Power Consumption

Number of Nodes	Testing 1 (Simulation)			Testing 2 (6LoSH)		
	RPL (mW)	HRPL (mW)	%	RPL (mW)	HRPL (mW)	%
4	0.995	0.629	36.8	1.576	1.0405	34.0
6	1.015	0.641	36.9	1.607	1.0515	34.6
8	1.021	0.650	36.3	1.623	1.07225	33.9
10	7.239	3.175	56.1	7.906	4.1099	48.0

Figure 13 shows the results obtained from the analysis of HRPL and RPL power consumption usage in different nodes (4, 6, 8, and 10) for testing 1 and testing 2. By comparing the four different nodes, it is evident that the power consumption usage for Number of Nodes (NoN) = 4 is the lowest followed by NoN = 6 then NoN = 8 while NoN = 10 is the highest power consumption usage. The result shows that HRPL is better in all NoN and managed to reduce the overall power consumption by 36.8% (NoN=4), 36.9% (NoN=6), 36.3% (NoN=8) and 56.1% (NoN=10) lesser than RPL in testing 1. While in testing 2 HRPL has obtained a better result that can reduce power consumption by 34.0% (NoN=4), 34.6% (NoN=6), 33.9% (NoN=8) and 48.0% (NoN=10) lesser than RPL. In this instance, according to the observation, the HRPL had generated a smaller power consumption, and due to the number of nodes, however it had managed to obtain the increases of power consumption in real time scenario (6LoSH) compared in simulation. It is figured out which the power consumption value would have effect under different platform even used the same scenario and system setting. The cooja simulator enable the node which make a simple network and doesn't requirement gateway to connect to the internet. But

different condition in 6LoSH, raspberry pi (powered by battery) used as a 6LBR to deploy the node for physical network and this condition increase the radio transmission power received by each node. Besides, the retransmission activities detected leading to higher power consumption in real scenario.

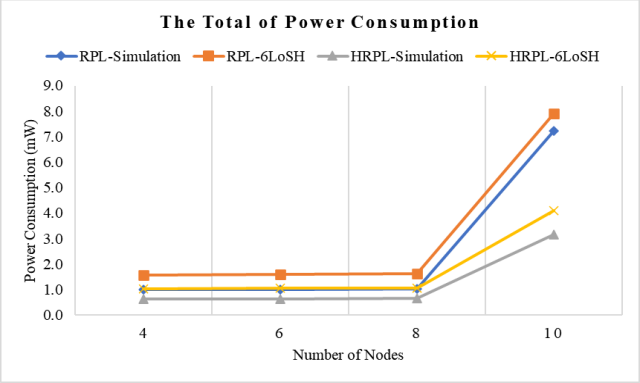


Figure 13: The Total Power Consumption for Both Experiment.

Table 5 presents the analysis of each type of power consumption for testing 1 and testing 2. The CPU Power, LPM Power, Listen Power, and transmit Power values obtained for each of the experiments were presented in Figure 14-17. It is observed that for NoN = 4 gave a minimum value of LPM Power, CPU Power, and Transmit Power. However, the value of Listen Power is the highest in comparison with others. Based on observation, the node turns on its radio during all its idle slot when it's set as a parent. This condition not only becomes the listening power increase but also wastes their energy for the whole network. Interestingly, the computation of each node is increased due to the way the node determines the best route from the source node to the destination node[31]. Meanwhile, the T-Sky CC2420 module that is used in simulation has less power compared to the CC2358 module that is used in 6LoSH.

Table 5: Analyses of Power Consumption for Testing 1 (Simulation) and Testing 2 (6LosH)

Types of Power	Number of Nodes	Testing 1 (Simulation)		Testing 2 (6LoSH)	
		RPL (mW)	HRPL (mW)	RPL (mW)	HRPL (mW)
CPU Power	4	0.332	0.209	0.453	0.285
	6	0.349	0.213	0.487	0.309
	8	0.357	0.218	0.514	0.328
	10	0.376	0.228	0.586	0.362
LPM Power	4	0.153	0.101	0.224	0.156
	6	0.153	0.104	0.218	0.143
	8	0.153	0.105	0.224	0.145
	10	1.519	0.988	0.551	0.315
Transmit Power	4	0.092	0.054	0.211	0.132
	6	0.086	0.056	0.214	0.134
	8	0.08	0.056	0.214	0.134
	10	0.88	0.504	2.101	1.211
Listen Power	4	0.418	0.265	0.688	0.468
	6	0.427	0.268	0.688	0.466
	8	0.431	0.271	0.671	0.465
	10	4.464	1.455	4.668	2.222

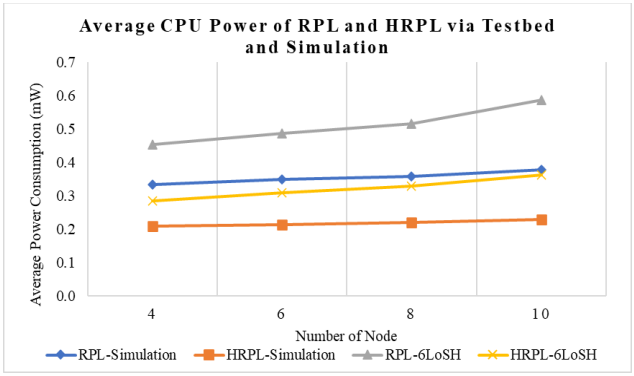


Figure 14: The CPU Power of HRPL and RPL

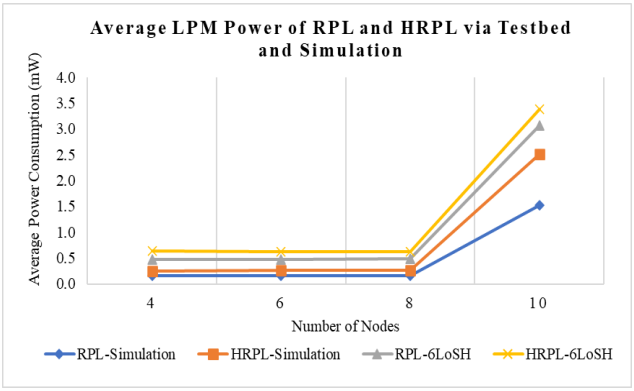


Figure 15: The LPM Power of HRPL and RPL

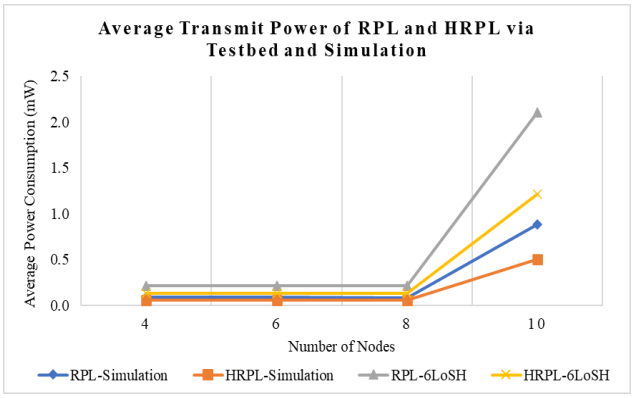


Figure 16: The Transmit Power Of HRPL and RPL

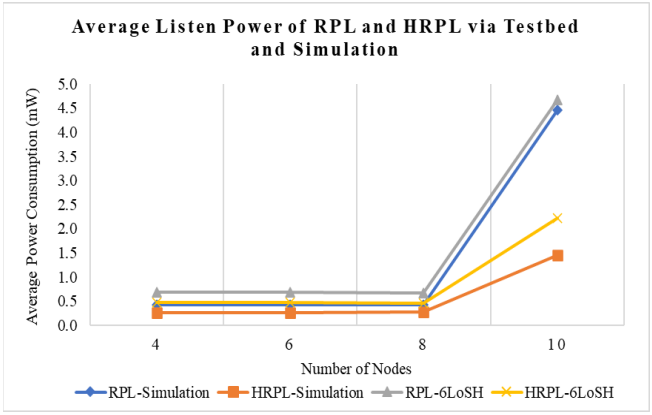


Figure 17: The Listen Power of HRPL and HRPL

5. Conclusion

6LoWPAN system enables the ED to communicate with each other using the IEEE 802.15.4 wireless network. Due to this matter, 6LoWPAN had become a priority network for IP-based communication for ED. However, the design of the 6LoWPAN system in the real-world scenario has become a challenging issue, such as interoperability, scalability, and integration among the nodes, and network. Thus this study aims to develop the 6LoWPAN smart home system (6LoSH) to evaluate the performance of HRPL and compared the result with RPL. The result of this study shows the PC of nodes reached its most minimum for NoN =4 scenarios, while low link quality is linked to NoN = 10 scenarios respectively for both experiments. These results confirm that the 6Node (6LoSH) reaches an efficiency of QoS and HRPL reliability to reduce the PC. The routing overhead for the number of nodes increases significantly with the increase of the PC. The main contribution of this study, the present result shows that HRPL outperforms the RPL standard of PC in both experiments. However, these results were derived from a limited number of nodes, the Transmission Range (Tx) is 5m, and it was tested for random topology. Further research should be conducted to investigate large-scale networks in multiple topologies for different distances and metrics.

Acknowledgment

The authors would like to thank the Ministry of Higher Education Malaysia and Universiti Teknikal Malaysia, in support the Fundamental Research Grant Scheme (FRGS) with reference number (FRGS/2018/FTMK-CACT/F00392).

References

[1] N.H. Mohd Yusoff, N.A. Zakaria, A. Sikora, J.S. Sebastian, "6LoWPAN Protocol in Fixed Environment: A Performance Assessment Analysis," Proceedings of the 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2019, **2**, 1142–1147, 2019, doi:10.1109/IDAACS.2019.8924283.

[2] H. Yadav, C. Agrawal, "A Review of Enhancing Energy Efficiency of Wireless Sensor Network," International Journal of Electrical, Electronics and Computer Engineering (IJEECE), **8**(1), 46–49, 2019.

[3] V.C. Thang, N. Van Tao, "A performance evaluation of improved IPv6 routing protocol for wireless sensor networks," International Journal of Intelligent Systems and Applications, **8**(12), 18–25, 2016, doi:10.5815/ijisa.2016.12.03.

[4] T. Abdelkader, K. Naik, A. Nayak, N. Goel, V. Srivastava, "A performance comparison of delay-tolerant network routing protocols," IEEE Network, **30**(2), 46–53, 2016, doi:10.1109/MNET.2016.7437024.

[5] J.A. Alzubi, R. Manikandan, O.A. Alzubi, N. Gayathri, R. Patan, "A Survey of Specific IoT Applications," International Journal on Emerging Technologies, **10**(1), 47–53, 2019.

[6] S.N.S. Vishalakshi R, Chaithra P, "6LoWPAN Routing Issues," International Journal Of Engineering And Computer Science, **3**(7), 2014.

[7] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, W. Zhao, "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications," IEEE Internet of Things Journal, **4**(5), 1125–1142, 2017, doi:10.1109/JIOT.2017.2683200.

[8] S. Li, L. Da Xu, S. Zhao, "The Internet of Things: A Survey," Information Systems Frontiers, **17**(2), 243–259, 2015, doi:10.1007/s10796-014-9492-7.

[9] Z. Shelby, C. Bormann, 6LoWPAN : The Wireless Embedded Internet, 2009.

[10] P. Pongle, G. Chavan, "A survey: Attacks on RPL and 6LoWPAN in IoT," in 2015 International Conference on Pervasive Computing: Advance Communication Technology and Application for Society, ICPC 2015, 0–5, 2015, doi:10.1109/PERVASIVE.2015.7087034.

[11] S.N. Han, Q.H. Cao, B. Alinia, N. Crespi, "Design, implementation, and evaluation of 6LoWPAN for home and building automation in the Internet of Things," Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA, **2016**–**July**, 2016, doi:10.1109/AICCSA.2015.7507264.

[12] A.J. Albarakati, J. Qayyum, K. a. Fakeeh, "A Survey on 6LowPAN & its Future Research Challenges," International Journal of Computer Science and Mobile Computing, **3**(10), 558–570, 2014.

[13] T. Winter, P. Thubert, A. Brandt, J. Hui, R. Kelsey, P. Levis, K. Pister, R. Struik, J. Vasseur, R. Alexander, "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks Abstract Low-Power," IETF RFC6550, doi:10.1017/CBO9781107415324.004.

[14] H. Xie, G. Zhang, D. Su, P. Wang, F. Zeng, "Performance evaluation of RPL routing protocol in 6lowpan," in Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS, 2014, doi:10.1109/ICSESS.2014.6933646.

[15] P. Levis, T.H. Clausen, J.W. Hui, O. Gnawali, J. Ko, "The Trickle Algorithm," IETF RFC6206, 2011, doi:http://dx.doi.org/10.17487/RFC6206.

[16] M. Feschel, A. Huhn, H. Schwetlick, "Routing protocols for 6LoWPAN," International Conference on IT Revolutions, **82 LNICST**, 71–83, 2012, doi:10.1007/978-3-642-32304-1_7.

[17] N.H. Mohd Yusof, N.A. Zakaria, N. Harum, "Design of Smart Building Automation System for 6LoWPAN Network : The Platform Structure," in Procedings National Symposium on Human Computer Interaction, 2019, 35–37, 2019.

[18] E. Toscano, L. Lo Bello, "Comparative assessments of IEEE 802.15.4/ZigBee and 6LoWPAN for low-power industrial WSNs in realistic scenarios," 9th IEEE International Workshop on, IEEE, 2012.

[19] CC2538 Powerful Wireless Microcontroller System-On-Chip for 2.4-GHz IEEE 802.15.4, 6LoWPAN, and ZigBee ® Applications 1 Device Overview.

[20] G.A. Naidu, J. Kumar, V. Garudachedu, P.R. Ramesh, "6LoWPAN Border Router Implementation for IoT Devices on RaspberryPi," SSRN Electronic Journal, 316–323, 2018, doi:10.2139/ssrn.3166729.

[21] The Official Contiki OS Blog: Contiki 3.0 Released, New Hardware from Texas Instruments, Zolertia.

[22] A. Velinov, A. Mileva, "Running and Testing Applications for Contiki OS Using Cooja Simulator," International Conference on Information Technology and Development of Education, 279–285, 2016.

[23] N. Pradeska, Widyawan, W. Najib, S.S. Kusumawardani, "Performance analysis of objective function MRHOF and OF0 in routing protocol RPL IPV6 over low power wireless personal area networks (6LoWPAN)," Proceedings of 2016 8th International Conference on Information Technology and Electrical Engineering: Empowering Technology for Better Future, ICITEE 2016, (7863270), 7863270, 2017, doi:10.1109/ICITEED.2016.7863270.

[24] J. Vasseur, M. Kim, K. Pister, N. Dejean, D. Barthel, "Routing Metrics Used for Path Calculation in Low-Power and Lossy Networks," IETF RFC6551, doi:10.1017/CBO9781107415324.004.

[25] Y. Tahir, S. Yang, J. McCann, "BRPL: Backpressure RPL for High-throughput and Mobile IoTs," IEEE Transactions on Mobile Computing, **2**(c), 1–1, 2017, doi:10.1109/TMC.2017.2705680.

[26] C.-A. La, M. Heusse, A. Duda, "Link reversal and reactive routing in Low Power and Lossy Networks," IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC, 3386–3390, 2013, doi:10.1109/PIMRC.2013.6666733.

[27] E. Ancillotti, R. Bruno, M. Conti, E. Mingozzi, C. Vallati, "Trickle-L2: Lightweight link quality estimation through Trickle in RPL networks," Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014, WoWMoM 2014, 2014, doi:10.1109/WoWMoM.2014.6918951.

[28] A. Dunkels, B. Gronvall, T. Voigt, "Contiki - a Lightweight and Flexible Operating System for Tiny Networked Sensors," Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks (LCN'04), (January), 2010, doi:10.1109/LCN.2004.38.

[29] Y. Bin Zikria, M.K. Afzal, F. Ishmanov, S.W. Kim, H. Yu, "A survey on routing protocols supported by the Contiki Internet of things operating system," Future Generation Computer Systems, **82**, 200–219, 2018, doi:10.1016/j.future.2017.12.045.

[30] L. Sitanayah, C.J. Sreenan, S. Fedor, "A Cooja-based tool for coverage and lifetime evaluation in an in-building sensor network," Journal of Sensor and Actuator Networks, **5**(1), 1–22, 2016, doi:10.3390/jsan5010004.

[31] L. Lassouaoui, S. Rovedakis, F. Sailhan, A. Wei, "Evaluation of energy

aware routing metrics for RPL," 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 1–8, 2016, doi:10.1109/WiMOB.2016.7763212.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1225-1236 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Priority-based Scheduling Algorithm for NOMA-integrated V2X

Ala Din Trabelsi*, Hend Marouane, Faouzi Zarai

National School of Electronics and Telecommunications (ENET'COM), NTS'COM Laboratory, Sfax University, 3029, Tunisia

ARTICLE INFO

Article history:
Received: 30 August, 2020
Accepted: 26 November, 2020
Online: 16 December, 2020

Keywords:
5G
V2X
Scheduling Algorithm
Resource allocation
PSA

ABSTRACT

The fifth-generation (5G) wireless system was implemented due to the massive connectivity, especially Internet of Things (IoT) including industrial automation, Vehicle to Everything (V2X) communications, which is shelter an extensive range of applications services and use cases and Intelligent Transportation Systems (ITS). A cellular Non-Orthogonal Multiple Access (NOMA) based on radio resource allocation schemes need to be carefully designed to serve and support the diversity of applications and services, through increased safety and traffic efficiency usage of connected vehicle and ranging from infotainment services. In this paper, we develop a novel scheduling and resource allocation algorithm for vehicular communications named Priority -based Scheduling Algorithm for NOMA integrated V2X (PSA). The novel scheme could make the limited radio Resource Blocks (RB) to be assigned in a more efficiently manner to the different traffic classes (safety and non-safety traffic) imposed by V2X exigency, in order to improve the performances of the system in terms of average throughput, Quality of Service (QoS), Average Blocking Rate (ABR) and to ensure fairness criteria between Vehicular User Equipments (VUEs). The performance of PSA is shown for urban macro-cell and rural macro-cell scenarios and we compared it with the conventional OMA and some recent works for NOMA based allocation.

1. Introduction

New 5G technologies supporting vehicular communications considered as a promising and vital role in the blueprint of Intelligent Transportation Systems (ITS), assisting to solve the main problems of transport engineering, improves road safety, traffic efficiency and comfort and infotainment services, try to prevent any fostering and dangerous situations investments in the automotive market [1-3].

V2X connectivity has been developed as a potential concept for 5G which strive to give communication models that can be supported by vehicles in several application contexts including vehicle-to-vehicle (V2V), vehicle-to-infrastructure/network (V2I/N) and vehicle-to-pedestrian (V2P) [4].

V2X will include two complementary sets of specifications for direct transmission which are 802.11p and cellular V2X (C-V2X) standards. The first one is proposed as the lower layers of Wireless Access in the Vehicular Environment (WAVE), it is standardized for V2V defined in USA and recognized as Dedicated Short-Range Communications (DSRC). 802.11p is also designed as the lower layers of the ITS-G5 standard in Europe. It is based on OFDM and which is known as the case of fourth generation (4G) technology LTE (Long Term Evolution). However, the adoption of DSRC has been delayed due to its Medium Access Control (MAC) and physical layer designs that have been planned for Wireless Local Area Networks (WLAN) and its poor scalability and the high-mobility environments that impose communication challenges, and lack of QoS guarantee. In addition, 802.11p indicates that it's restricted range, reliability and undefined performance especially in urban propagation scenarios when scaling up the number of transmitting VUEs [4].

Cellular-V2X is proposed to innovate and deliver an evolution path to 5G and it comes to handle the challenges of 802.11p. The main advantage of C-V2X that is based on a technology originally optimized for high-speed mobile services, which has been enhanced specifically for safety and automotive applications. C-V2X is outperformed 802.11p due to modulation and coding enhancement, to support a better communication range, greater capacity, better congestion control in denser environments with lower packet error rate, higher reliability and superior Non Line-of-Sight (NLoS) performance [4].

*Corresponding Author: Ala Din Trabelsi, trabelsi.ala93@gmail.com

Accordingly, 27 use cases for V2X communications are designed by third-Generation Partnership Project (3GPP) in Release 14 that including two categories of service requirements as safety related applications, like car platooning, autonomous driving, vulnerable road user warning, etc and non-safety related uses cases exemple, dynamic digital map update, mobile hot spot, mobile high data rate entertainment, office, home, etc [5].

More recently, 5G networks anticipate an explosive demand for massive connectivity driven by the ever-increasing number of connected vehicles, Internet of Things (IoT) and the explosive proliferation of mobile Internet, over limited radio resources [6].

The limitation of OMA-based scheme that it has not been efficiently and fully assigning the limited radio resource blocks causing to serious problems as lack of QoS guarantee for NLoS environment. This defiance can be exemplified with the following scenario, consider a single User Equipment (UE) with mediocre channel condition needs to be scheduled for fairness purposes, example this UE has superior priority information or has not been scheduled for a long time. This causes that one of the limited bandwidth resources is only allocated by this UE. In addition, the limitation of the conventional OMA is the number of users that can be served simultaneously. Towards these challenges, 3GPP is motivated to develop new technologies in order to maximizing the achievable throughput. Among various solutions, non-orthogonal multiple access (NOMA) is a very promising access technology. It has been envisioned and recognized as an innovation for 5G mobile networks to increase the number of links, enlarge the spectral efficiency and to ensure the high data rate necessity. It provides enhancements in terms of high reliability, low latency and communication's range for vehicular environments [7].

Using non-orthogonal radio resource allocation, make NOMA to multiplex more than a few users with different power levels into a same frequency resource and time to transfer information. Not only does NOMA support users with good channel conditions but also it serves both others with poor channel conditions and those operate in the same bandwidth resources [8][9].

To enable V2X communications, radio resource management is a vital specification of the MAC layer of the eNodeB, that aims to optimize the use of the limited availability of the spectrum while meeting the requirements of QoS and insure efficient resources scheduling, resource allocation and interference management. Even in 5G networks, a resource allocation technique requires to be carefully developed to satisfy different QoS requirements to achieve target throughput and to maximize spectrum efficiency for several V2X traffics and services.

The paper contributions are as following:

- Taking into account traffic classification, such as Safety Traffic (ST) and Non Safety one (NST) are considered
- A resource allocation algorithm called Priority -based Scheduling Algorithm for NOMA integrated V2X (PSA) is proposed.
- Adding a scheduling priority which is determined according to the channel condition for each VUE, users with higher priority will be selected for NOMA sharing.
- Due to the diversity and multitude of the propagation scenarios, various simulations are presented for performance evaluation for rural macro-cell and urban macro- cell, to demonstrate the efficiency and the applicability of the proposed PSA algorithm for any propagation scenarios.
- Validate the proposed PSA algorithm with several Matlab simulations showing the aggregate throughput, the fairness index and the ABR among different VUE traffic classes.
- Compare the proposed algorithm with the conventional OMA
- Evaluate the PSA algorithm with some recent works for NOMA based allocation.

This paper is presented as follows we study same proposed method as related works in section 2. We detail our approach in section 3. In the section 4 and 5 we portray its different steps. We evaluate and analyze the performances of the developed scheduler in section 6. Finally, in Section 7, we determine our work by conclusion and perspective as future work.

2. Related Work

The radio resources management task to enable V2X communications has been investigated in several research works such as:

In [10], the authors developed a novel resource allocation algorithm to maximize the number of scheduled V2N communications while satisfying the required latency for the several priority classes of the V2V links (two class of priority are defined, where the messages in the class 1 are more critical and aware to the latency comparing to messages in the second one). They perform the power allocation for potential resource reusing. Then, they selected the finest reusing pattern by employing the Hungarian scheme. However, this method does not take NOMA systems in consideration and the performance evaluations focused only to present the average latency.

In [11], the authors studied the resource allocation process when the periodic messages are not all at the same size in cellular V2X networks. They addressed an optimization problem for maximizing number of vehicles that can be allocated simultaneously without interfering each other. Then, they provided the same algorithm to solve optimization problem then evaluate its performance by increasing number of vehicles when a trade-off between range and density of vehicle is guaranteed. However, NOMA scenarios issue was usually ignored in design of the resource allocation algorithm.

In [12], the authors developed a novel resource allocation scheme for Power domain based a cooperative multi-cell network by using the cellular mode. This algorithm aims to improve the spectral efficiency of the system and achieves the desired QoS. They assume a communication scenario for two pairs of users sharing the same sub-carrier to transmit their information, where the primary users characterized by delay-sensitive communication that requires stringent QoS exigency as well as the minimum transmission rate, share the same radio RB with the secondary users which support non-delay sensitive or elastic traffic and exploit the physical resources of the network in a best-effort manner. Successive convex approximation (SCA) method is the

best solution to solve general optimization issues, they adopt the successive convex approximation (SCA). Furthermore, they investigate the computational complexity of the designed approach and they compare our work with existing transmission methods based on Orthogonal Frequency Multiple Access (OFDMA). However, this approach does not hit the V2X links requirements.

Some recent research has begun to explore efficient energy NOMA design were given in [13-16]. They presented a general overview of NOMA techniques and seek to optimize the power allocation problem in order to improve Energy Efficiency (EE). Then, they proposed a low complexity power allocation approach. However, those algorithms do not cover V2X links, traffic classifications and heterogeneous QoS requirements for each VUE link.

In [17], the authors focus on the problem of clustering users in groups of different size, and assigning RB to each cluster based on Power-Domain (PD-NOMA) scheme. Clustering and RB allocation are performed as stated by Proportional Fairness (PF) technique with the main objective that obtaining high cell throughput, low complexity and guaranteeing NOMA relevance. The performance evaluation of the clustering algorithm proposed Relevant Clustering (RC) compared with two other algorithms named Ordered Clustering (OC) and Disparity Clustering (DC). However, traffic classifications such safety information for the V2X links is not considered in this algorithm.

In [18], the authors improved proportional fair scheduling scheme for NOMA system, called a spectrum resource and power allocation algorithm with Adaptive Proportional Fair (APF) user pairing with low complexity. This algorithm aims to maximize system throughput and considers the fairness among UE by selecting the user pairing with higher priority expressed by the channel gain. For the optimizing the power allocation, they explore different power allocation schemes which focusing on the proportional rate constraint so as to ensure the best throughput performance. The limitation of this approach that does not considering V2X traffic classifications.

In [19], the authors have addressed and investigated the power allocation problem in downlink Multiple Carriers Non-Orthogonal Multiple Access (MC-NOMA) systems for maximizing Weighted Sum Rate (WSR) by considering imperfect successive interference cancellation, in order to present an iterative power allocation with low-complexity algorithm. Moreover, they have presented a joint user scheduling and power allocation scheme for the MC-NOMA systems. The proposed methods can achieve an optimal WSR performance and user fairness improvement compared to OMA and other existing schemes. However, the system model only considered each sub- channel can be multiplexed by two users no user priority and service classification are treated.

In [20], the authors investigated the benefits of 5G technologies based on NOMA to support massive connections services to group and assign the required sub-channel of the internet of things equipment. To show how the quality of experience requirements of IoT links affect the needed number of physical RBs in real time, they developed a novel method to improve spectrum distribution and to control IoT links mobility by using Non-Orthogonal Multiple Access technique for autonomous vehicles. The designed solution aims to maximize throughput, to contribute QoS improvement and exploiting k-means grouping to reply to the dynamics of the system and minimize the complexity of re-arranging NOMA clusters. The regrouping of IoT connections according to distance while considering the maximum supported devices by group. Therefore, they determined the requisite set of RB for each group of UEs based on the QoS requirement. However, IoT equipments are clustered without considering their type of traffic.

In [21], the authors is introduced an over view on optimization problems of scheduling and NOMA approaches. They planned a novel scheduling and resource allocation scheme based on NOMA-enhanced cellular network for device-to-device (D2D) communications called Tree-based Search power allocation Algorithm (TSA). This scheme is compared and evaluated with several statistic models for user scheduling algorithms, users selection problems for the D2D communications underlying cellular network based on NOMA technology and power allocation strategies such as FRPA (FRactional Power Allocation) and FIPA (FIxed Power Allocation algorithm) where their mean objectives are mitigating computational complexity of the algorithm, increase network capacity, guarantying the QoS and maintaining a better system throughput. The simulation results clearly demonstrate the superiority of TSA compared with traditional algorithms. However, the proposed algorithm does not consider V2X links and traffic classification (ST and NST).

In [22], the authors designed a new method entitled proportional-fairness based user pairing and power allocation algorithm for NOMA uplink transmission. This method is introduced to serve two scenarios: the first one is when user equipments are only distributed in the coverage area of the cell and the second one which is the more complex that is when the interference users are distributed outside of the cell. For scenario one, to optimize the problem they decoupled the User Pairing (UP) part and Power Allocation (PA) part. In the power allocation part, they classify the pairs of users by considering their channel condition and SINR constraint and they developed different power allocation techniques. While in the user pairing part, they designed a probability-based tabu search user-pairing algorithm (PTS). For scenario two, they carried out a prediction-based particle swarm optimization (PBPSO) power allocation algorithm with a low computational complexity to obtain a good proportional fairness performance under the constraints of outage probability. Then, they combine PBPSO and PTS to achieve high system performance in terms of average throughput and outage probability. However, this algorithm assigns the RB for the pairs of users independently of their type of traffic class.

In [23], the authors investigated the global throughput maximization for two UE with timeliness transmission over fading channels for partial and full Channel State Information at the Transmitter (CSIT). This approach supports certain level of fairness and the minimum average throughput constraint. In the full case of CSIT case, the dynamic resource allocation method was optimally determined by using lagrangian dual decay, while in the partial CSIT one, the power allocation technique was proposed focused on investigative results. However, the limitation of this algorithm that does not takes into account the demanded QoS requirements for users and the traffics classifications. The simulation results are investigated with limited parameters setting.

In addition, the required bandwidth to enable ST V2X services in NOMA has not been studied.

In [24], the authors are introduced an over view on the optimization scheduler problem joined with NOMA techniques. They developed and simulated some statistic models for user scheduling strategies and power allocation approaches which aim to achieve the different objectives as minimizing the computational complexity of the algorithm, satisfy the required QoS and obtaining a better system performance in terms of average sum rate and equity criteria. Also, they select Proportional Fairness (PF) and Round-Robin (RR) as a shared model of scheduling and NOMA techniques, to evaluate the optimization problem of power allocation for combined NOMA and round-robin scheduling scheme and proposing an optimization strategy for joint NOMA algorithms and Proportional Fairness scheduling which assure the required QoS. However, the proposed scheme doesn't reflect on the V2X link environments and traffic classifications.

The table 1 illustrates all the mentioned algorithms above. All related work mentioned above doesn't take charge the traffic classifications while seeking the throughput performance gain. Nevertheless, the benefits of V2X ST class at the road cannot be exploited successfully by these algorithms. Despite the existing literature, there is a lack of priority constraint-based scheduling scheme that marked the transmission ranking to each connection but it restricted just by the Channel Quality Indicator (CQI). In addition, diverse algorithms presented above didn't balance cell throughput, equity index and take the system-level QoS for V2X links, new solutions must be provided.

3. System model

The used input and variable parameters are given below.

3.1. System Model

In order to evaluate our proposal work, we consider NOMA scheme consists of a single base station and N connected VUEs denoted by I = {1,....j,…,N}. We suppose that the number of available RBs in the cell is M where the resource blocks set U = {1,…i,…,M}. To assure the required capacity of the N vehicular connections, we suppose that PD-NOMA (Power-Domain NOMA) technique linking by the base station. NOMA is a scheme used to appreciate multiple accesses by utilizing power domain in a spectrally resourceful way and can schedule multiple vehicular user equipments in the same radio resource. This can be maintained by assigning dissimilar power levels to VUEs. Moreover, NOMA scheme is expected to allocate more power level to the VUEs with poor channel qualities. Denote the power of VUE_j on subcarrier k by $P_{j,k}$, satisfying $\sum_{j \in N} P_{j,k} \leq P_j^{max}$, where $P_j^{max} = P$ is the maximum power constraint allocated for VUE_j. Therefore, we assume that $N \geq 2M$ for dense environment and each vehicular user assign to one RB in the same Transmission Time Interval (TTI) while the entire other not scheduled vehicular users are not allocated. In addition, we assume that the distribution and the position of links are supposed random in the coverage area.

To maintain the improvements of V2X link, radio resource allocation and Transmitter -Receiver selection are managed by the eNodeB in each TTI.

Table 1: Comparison between different algorithms studied above

Ref	NOMA systems	Throughput optimization		Qos satisfaction	Services classifications ST / NST	Fairness
		Average	Instantaneous			
[10]	√	√				√
[11]	√	√				
[12]	√	√		√		
[13]	√	√				
[14]	√	√				
[15]	√	√				
[16]	√	√				√
[17]	√	√				
[18]		√				
[19]	√	√		√		
[20]	√	√				
[21]	√	√	√	√		
[22]	√		√			√
[23]	√	√		√		√
[24]			√		**Only ST**	

Table 2: Notations

Symbol	Description
N	Connected VUEs.
I	Set of vehicular uses equipments.
M	Available RBs.
U	Set of resources blocks

P_{ST}, P_{MT}, P_{IT}	Percentage of safety, management and infotainment class of traffic.
N_{ST}, N_{MT}, N_{IT}	Number of active VUEs for each class of traffic.
φ, ψ	Set of safety and non safety VUEs.
$N_{NST}= N_{MT}+N_{IT}$	Number of NST VUEs.
Ć, Ŕ	Set of ST and NST VUEs candidates that can be served to different resource.
P_{factor}	The power level for VUE.
H_ST	Channel coefficient of safety VUE.
H_NST	Channel coefficient of non-safety VUE.
thr_min_safety	Minimum throughput for safety traffic requirement.
H_safety	Matrix contains the channel coefficient of safety VUEs.
H_nonsafety	Matrix contains the channel coefficient of non- safety VUEs.

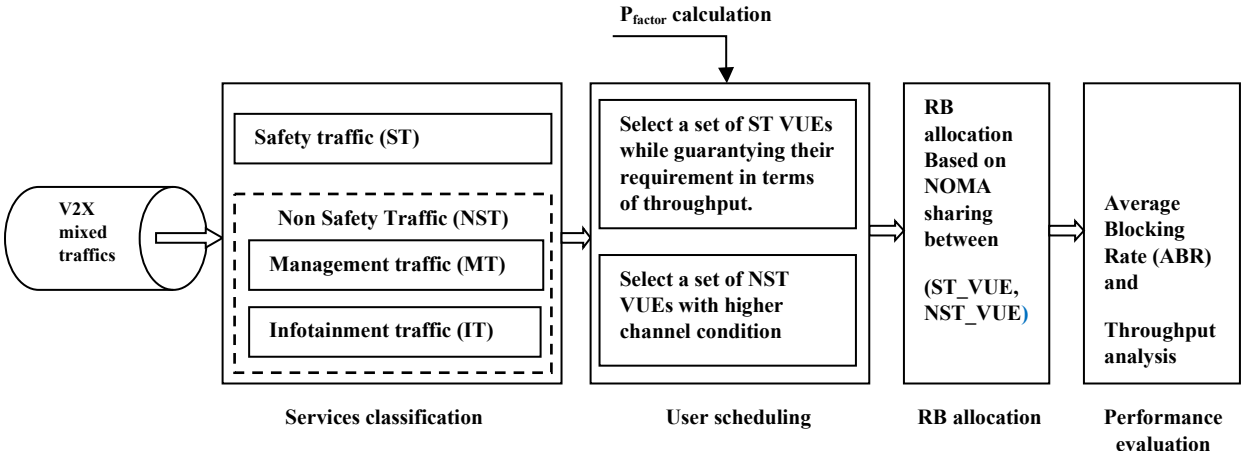


Figure 1: PSA system model

Using NOMA system, the eNodeB superimposes the message waveforms with different assigned power level for its selected users. At the Rx, successive interference cancellation (SIC) schemes are employed to cancel multi-user interferences and each user detects its own signal. The receivers with higher channel gain apply SIC for signal decoding. Therefore, those the modulated symbols are subtracted and removed from the united data until the associated UE obtains the desired message [6].E.g., consider two Tx: VUE_{j1} and VUE_{j2} occupying the RB_i in the current TTI transmitting to receiver VUE_m satisfying $P_{j1,k}^{(i)} \left|H_{j1,m,k)}^{(i)}\right|^2 < P_{j2,k}^{(i)} \left|H_{j2,m,k}^{(i)}\right|^2$, where $P_{jn,k}^{(i)}$ representing the transmit power of the transmitter VUE_{jn}, n $\in\{1, 2\}$. $H_{j2,m,k}^{(i)}$ presentes the channel coefficient of the subchannel k which allocated for the vehicular user VUE_{jn} in the RB_i. The receiver VUE_m deals the signal of the great channel condition VUE_{j2} as a noise, first, detects and remodels the signal of the inferior channel gain of VUE_{j1}, then, exploits the SIC technique to cancel or withdraw the signal of VUE_{j1} which has the mediocre channel condition to get signal of VUE_{j2}.

Generally, the signal x_k, transmitted by the eNodeB at a certain sub-band k is a simple power multiplexing of the symbols generated by the scheduled users.

$$x_k = \sum_{j\in N_m} x_{j,m,k} = \sum_{j\in N_m} \sqrt{P_{j,k}^{(i)}}\, s_j^{(i)} \quad (1)$$

Denotes $N_m=\{1 \le j \le N \mid \max (SINR_{j,m}^{(i)})\}$: presents the set of VUEs inside receiver vehicular UE m's communication range of interest. $P_{j,k}^{(i)}$ is the power allocated to user j at sub-band k and $S_j^{(i)}$ indicates the transmitted signal of Tx VUE_j.

The received signal vector corresponding to the time slot i can be defined by:

$$Y_{m,k}^{(i)} = H_{j,m,k}^{(i)} x_{j,m,k} + n_m^{(i)}$$
$$= \sum_{j\in N_m} \gamma_j^{(i)} \cdot \mu_{j,k}^{(i)} \sqrt{P_{j,k}^{(i)}} H_{j,m,k}^{(i)} S_j^{(i)} + n_m^{(i)} \quad (2)$$

where:

$\gamma_j^{(i)}$: Binary variable denotes that RB_i is assigned by the BS to VUE_j.

$\mu_{j,k}^{(i)}$: Binary variable indicates that subchannel k is assigned by the eNodeB to VUE_j.

$H_{j,m,k}^{(i)}$: presents the channel coefficient of subchannel k between Tx_j and Rx_m in RB_i.

$n_m^{(i)}$: CN~$(0,\sigma^2)$ is the additive white Gaussian noise (AWGN) for reciever VUE_m, and σ^2 is the noise variance. Furthermore, the channel gain can be defined as:

$$H_{j,m}^{(i)} = G \cdot h_{j,m,k}^{(i)} \cdot \beta \cdot \left(d_{j,m}^{(i)}\right)^{-\alpha} \quad (3)$$

Denotes:

G: Constant power gain factor implied by the amplifier and antenna.

$h_{j,m,k}^{(i)}$: Complex Gaussian variable CN (0,1) indicates the rayleigh fading.

β: Shadowing fading.

$d_{j,m}^{(i)}$: Distance between VUE_j and VUE_m.

α: Path loss exponent.

3.2. Propagation Scenario

Many covered propagation scenarios are detailed in [25] and [26]. It is indispensable to extend such modeling and measurement efforts to various environments that are also very important for V2X communications, mainly when safety applications are supported. Moreover, modeling a radio channel has been one of the most challenges in the mobile radio system design, and it is based on measurements specifically made for an intended communication system. Propagation models are the base for channel modeling, as they try to describe the way a radio signal changes during its travel distances from the Tx to the Rx and to mitigate interference. Also, radio propagation has a significant impact on the performance of cellular mobile radio systems due to increased data rate, bandwidth, mobility and QoS.

Then more, the trajectory between the Tx and the R can differ from a simple Line Of Sight (LOS) to a line that is obstructed by mountains, buildings, or foliage, named Non line of sight (NLOS).

The main characteristics of the radio channel - such as path loss, fading and time-delay spread and the parameterization vary from each scenario to an other one similar environments.

The macro-cell propagation can be divided into two classes as urban macro-cell propagation and rural macro-cell one. As the surrounding building heights, the density of obstacles and the number of users is very high in urban area, the cellular network is more robust and complicated where one macrocell antenna serves many microcells and picocells to ensure a better service compared to rural area because of the presence of less buildings and obstacles in the area [26].

For our work, we focus on rural macro-cell and urban macro-cell senarios with vehicles travelling on multiple lanes and on random position in the cell, including other pedestrians. Since we are considering the network-controlled allocation under independent Rayleigh fading channel and the line of sight (LOS) propagation properties of the vehicular networks, without loss of generality, we only incorporate the effect of the distance dependent path loss between the eNodeB and each VUE into analysis. Therfore, Path-loss models play an important role in the design of cellular systems to specify key system parameters such as antenna heights, frequency, transmission power and so on.

According to [26], The path-loss models are parameterized and controlled using results from a wide-ranging set of measurement campaigns, and they are extensively used and accepted.

Path loss (PL) is a measure of the average radio frequency attenuation suffered by a transmitted signal when it reaches the receiver, after traversing a path of several wavelengths [25]. It is defined by:

$$PL=\alpha*\log10(d[m])+B+C*\log10(fc[Ghz]/5) \quad (4)$$

Were, the fitting parameter α includes the path-loss exponent, B is the intercept, C describes the path loss frequency dependence and fc is the center frequency.

However, the path loss frequency dependencies model used in our scenarios are defined below:

$$PL_{Rural}[dB] = 40*\log10(d[m])+10.5-18.5*\log10(h_{BS}[m])-18.5*\log10(h_{MS}[m])+1.5*\log10(fc*1e-9/5) \quad (5)$$

$$PL_{Urban}[dB] = 40*\log10(d[m])+13.47-14*\log10(h_{BS}[m])-14*\log10(h_{MS}[m])+6*\log10(fc*1e-9/5) \quad (6)$$

where, h_{BS} and h_{MS} are the antenna heights at the base station and the mobile station respectively. d denotes the distance between the Tx and the Rx.

4. Proposed Algorithm

This proposed technique aims achieve a higher cell throughput, consider the heterogeneity of VUEs channel gains and the equity criteria. This algorithm trys to increase the number of served safety V2X links in the cell while a minimum requirement must be guaranteed on terms of throughput and to maximize the average data rate of the NST VUEs.

Our proposed algorithm named Priority -based Scheduling Algorithm for NOMA integrated V2X (PSA), can be considered as an assignment solution that aims to obtain an efficient and optimal spectrum allocation for VUE. Firstly, we begin by services classification, also, the incoming links of the VUE are classifyed according to their types of services as: safety traffic and non safety traffic . The NST is divited into two type of services as Management Traffic (MT) and Infotainment Traffic (IT).

Then, we assign the physical resources blocks to the pairs VUEs which can occupy the same spectrum by taking into account the channel condition of the transmitters on each RB. The designed approach enhances the equity index in the processus of resource allocation among diverse VUEs class of services. In fact, if the selected VUE has already assigned to an RB during the current TTI, this one will have't another radio resource block in the same TTI while all the selected users are not served. This, at the same time, raises the spectral efficiency and the capacity of the system inside the coverage area, and ensures a better compromise between the average throughput and the fairness index.

Once, the RB allocation is performed, a power control process is used to treat the interference and minimize the binary error rate by calculating per each resource block the P_{factor} that allows the safety user to obtain the needed throughput as well as maximizing throughput of the non-safety user, also, each VUE can attain on the allocated RB without minimizing its achieved throughput. The non-scheduled safety VUEs in the present TTI will be selected with the superior priority to be served for the subsequently TTI. The scheduling algorithm of our system is shown in Figure 1.

The mean objective of our proposed algorithm PSA is maximizing: the cell throughput, the number of served VUEs, the number of served safety VUEs, the sum rate of non-safety communications. We try to minimize the interference inter VUEs using the same RB by imposing a power constraint for different vehicular UEs.

5. Description of the Proposed Algorithm

In this part, we present more details for all steps of our proposed strategy.

5.1. Step 1: Service Classification

NOMA resource allocation schemes need to be designed to serve the ever-increasing number of vehicles that are connected for different services, such as safety features and passengers multimedia amenities.

V2X services are a real time traffics and arrive with strict exigency on throughput, reliability and latency. Also, we consider two types of service classes as Safety Traffic (ST) that imposes timeliness requirements, network scalability, and stringent reliability and refers to the case when the transmit signals need high QoS and requires to be selected almost momentarily in order to minimize road accidents, such as traffic condition warning, curve speed warning, vulnerable road user warning and emergency vehicle warning [27].

The NST (Non Safety traffic) included two categories of services: MT and IT. The Management traffic required to be scheduled with most importance thanks to their significance which aims to enhance traffic flow by minimizing traffic congestion and travel time, examples traffic information and recommended, intersection management, regulatory speed limit and automated parking system [28].

Infotainment Traffics are characterized by the delay tolerance and provides convenience applications and comfort for VUEs to improve their driving experience, exemples media downloading, vehicle software update and data provisioning [28].

The idea of the first step is to sort the arriving vehicular links according to their type of service that are: ST, MT and IT.

5.2. Step 2: User Scheduling

Scheduler functionality is to allocate the RBs to the UEs according to their requested requirements and their radio conditions at every TTI. So, the available resources blocs are very limited and it has to be allocated with an optimal manner that highest cell throughput and equity index are ensured among all VUEs, applications or service classes.

The system comprises two categories of V2X communications. Let $\varphi=\{1\ldots j,\ldots N_{ST}\}$ the set of Vehicular users with Safety Traffic (VST) and $\psi=\{1\ldots j',\ldots N_{NST}\}$ the set of VUEs with Non Safety Traffics (VNST), where $\{\varphi, \psi\} \in \{I\}$.

Moreover, the purpose of this step is: firstly, to select a set of VST by $\acute{C} \in \{\varphi\}$ while guarantying their requirement in terms of throughput. The idea is to search the combination (VST_j, RB_i) where $1 \leq j \leq N_{ST}$ and $1 \leq i \leq M$, with the highest metric on the RB_i. This metric expresses the channel gain between the VST_j, $j \in N_{ST}$ and the receiver m on subcarrier k, denotes by $\left|H_{j,m,k)}^{(i)}\right|^2$, which includes both small-scale Rayleigh fading and large-scale path loss.

Secondly, to select a set of VNST by $\acute{R} \in \{\psi\}$ that will be shared the same $RB_i$with VST. The selection of VNST is based on the higher channel condition expressed by the estimated SINR which corresponds to the CQI (Channel Quality Indicator) index sent by the VNST to the eNodeB.

Also, priority users scheduling considered as an important tasks to enhance the network performances in terms of aggregate sum rate and equity index before the resource allocation.

5.3. Step 3: RBs Allocation

The third step is designed based on NOMA to maximize the aggregate throughput of the system by assigned the limited physical RB to pair of VUEs (VSTj, VNSTj') selected in the above step. Exploiting the knowledge on vehicle positions and propagation characteristics at the eNodeB, this latter allocate the resources to guarantee the required quality of service and maximize the network capacity. Moreover, the proposed RB allocation method is divided on two processes. Starting with the RB allocation for ST communications which aims to allocate the candidate $\acute{C} \in \{\varphi\}$ safety users to different resource blocs and set a minimum requirement must be guaranteed for those. Then, the RB allocation for VNST candidate $\acute{R} \in \{\psi\}$

The proposed scheme offered superior priority to the safety links than others type of traffics because they can't tolerate delay and must be scheduled and served immediately without forgetting to taking into considiration the equity index. Moreover, in this step the scheduler checks if each VST has the minimum requirement in terms of throughput. In fact, if the VUE is't satisfied, for the scheduler, this vehicular user is reselected to transmit for the next TTI. This allows us to maximize the served VUEs number by trying to give each one his requested requirement. The algorithm for RB allocation is summarized as following.

Algorithm 1: RB allocation

```
{Initialization};
// select the set B safety VUE candidates can be served to
different resource blocs;
Check each line of the matrix H_safety and calculate the
matrix rb_acceptable_safety : if a single allocation is
acceptable that is used (where an allocation can be used is '1'
or can't be used is '0');
at_least_one_allocated = true;

While i=1 to N at_least_one_allocated do
    at_least_one_allocated = false;

    For i=1 to all the vehicle of safety traffic do
        if sum(rb_acceptable_safety(i,:))==1 then
            Find the best VUE candidate to be served at the
            rb_to_allocate;
            Select the best VUE candidate to be served at the
            rb_to_allocate;
```

```
        Store the number of selected VUE corresponds to
        the index the rb_to_allocate which granting the
        minimum requirement in terms of throughput;
        at_least_one_allocated = true;
      end
    end
  End
  // Allocate the different RBs for the selected Ć safety
  VUEs.
  While sum(sum(rb_acceptable_safety(:,:))) > 0 do
    Calculate H_safety matrix of the channel coefficient per
    each selected VUE at the rb_to_allocate;
    Find (ue_to_allocate,rb_to_allocate) = max h_matrix
    (ue_to_allocate,rb_to_allocate), with
    (ue_to_allocate,rb_to_allocate) ∈ Ć X U;
    Selecting the best VUE with highest channel gain;
    Remove the ue_to_allocate row from h_matrix;
    Assign rb_to_allocate to ue_to_allocate ;
    Update U: U = U \ { rb_to_allocate };
    Update Ć: Ć = Ć \ { ue_to_allocate };
  End
```

We define an indicator defined by a variable s_j^i , (i∈M, j∈N) which indicates if the ith RB is assigned to j^{th} VUE or not.

$$S_j^i = \begin{cases} 1 & \text{if } RB_i \text{ is used by } VUE_j \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

This step is designed to maximize non safety vehicular users throughput (for both management and infotainment VUEs) by prioritizing the users with good channel condition and not to starve others with bad channels condition.

Specifically, the SINR at the receiver Rx with the highest channel gain which can cancel the interference is computed as:

$$SINR_{j,k} = \frac{s_j^{(i)} P_{factor}^{(i)} P \rho_{j,m,k}^{(i)}}{1 + \sum_{l \in L_n, l \neq j} (1 - P_{Factor}^{(i)}) P \rho_{l,m,k}^{(i)}} \quad (8)$$

where Ln represents the subset of the interfered VUEs when the receiver m exploiting the resource block for the transmission.

5.4. Step 4: Performance Evaluation

To collect the benefits of the output of our proposed scheme, certain performance evaluations are need to be addressed such as cell throughput and average blocking rate.

Algorithm 2 briefly describes this process.

Due to the unavailability of radio resources, ABR is a typical constraint for VUE with safety traffic (VST) in PSA which aims to maximize served VST thanks to their high priority, great QoS requirement and due to better road utilization and surroundings. The average blocking rate is calculated as following:

$$ABR = \frac{\text{Number of blocked transmitters}}{\text{Total number of VST}} \quad (9)$$

Algorithm 2: Throughput calculation

```
For i=1 to N do
  RBi is used by safety VUE then;
  Determinate the channel coefficient H_ST of the selected
  safety VUE;
  Determinate the channel coefficient H_NST of the selected
  non safety VUE;

  If H_ST > H_NST then
    % safety is the primary VUE that cancels the
    interference
    Calculete P_factor;
    P_factor= (N0/(P*H_ST))*(2^(thr_min_safety/Bw_rb)-
    1);
    % N0: Additive White Gaussian Noise (AWGN);
    Calculate the throughput of safety VUEi;
    thr_user_safety=(Bw_rb)*log2(1+(P*(P_factor)*H_ST/(P
    *(1-P_factor)* H_ST + N0)));
    Calculate the throughput of non safetyVUEi;
    thr_user_nonsafety= (Bw_rb)*log2(1+(P*(1-P_factor) *
    H_NST / (P * P_factor * H_NST+N0)));
  else
    % safety is the secondary VUE that cannot cancel the
    interference;
    P_factor= (2^(thr_min_safety/Bw_rb)-1)*(P*H_ST+N0) /
    ( 2^( thr_min_safety / Bw_rb)*(P*H_ST));
    Calculate the throughput of safety VUEi:
    thr_user_safety=(Bw_rb)*log2(1+(P*(P_factor)*H_ST/(
    P*(1-P_factor)*H_ST+N0)));
    Calculate the throughput of non- safety VUEi;
    thr_user_nonsafety = (Bw_rb)*log2(1+(P*(1-P_factor)*
    H_NST / N0 ));

  end
end
```

In this step we calculate per each resource block the Pfactor that allows the safety VUE to obtain the needed throughput and to maximize the sum rate of non-safety VUE at the same time. Indeed, the objective is to take into account power split between the two vehicular users which share the same RB to minimize the interference.

The achievable data rate obtained from VUE_j over sub-channel k at the RB_i can be mathematically formulated as:

$$R_{j,m,k}^{(i)} = B\_W \log_2 \left(1 + \frac{s_j^{(i)} P_{Factor}^{(i)} P \rho_{j,m,k}^{(i)}}{1 + \sum_{l \in L_n, l \neq j} (1 - P_{Factor}^{(i)}) P \rho_{l,m,k}^{(i)}} \right) \quad (10)$$

6. Simulation Results

In this section, simulation results based on Matlab are presented and performed to demonstrate and validate the efficiency of the developed PSA schemes. We assume that the position and the distribution of vehicular users are randomly in the

coverage area for the current TTI. Simulation results provide a better insight of various issues i.e. average throughput, BER and fairness analysis. In addition, we vary the percentage of ST and NST services, we take as an example three scenarios that are mentioned in table 3where scenario 1 is oriented Non Safety Traffic (NST), scenario 2 is marked by equity of percentage of traffic and the last one is oriented Safety Traffic (ST), in order to evaluate the system performance and show that our algorithm is efficient and applicable for any scenarios. So the objective of these scenarios is to highlight the effect of the priority on the average blocking rate, the cell throughput and the equity index. Moreover, the performance results of the developed algorithm are compared with the conventional OMA based allocation which is detailed in [29] and some recent works for NOMA based allocation to demonstrate the benefits of our proposed approach.

Table 3: Parameter setting

Parameter	Value		
Bandwidth	10 MHz		
Available resource blocks	50		
connected VUEs in the cell	[100, 150]		
eNodeB in the coverage area	1		
Transmitted power	15 dBm		
Number of VUEs allocated the same spectrum	2		
TTI	1 ms		
	Scenario 1	Scenario 2	Scenario 3
Percentage of ST services	30%	50%	70%
Percentage of MT services	40%	30%	20%
Percentage of IT services	30%	20%	10%

The used parameters for simulation results are announced below in Table 3.

6.1. ABR Analysis

ABR is an important Key Performance Indicators (KPIs) [30]. It is presenting the ratio between the number of blocked transmitter VUEs due to the limited spectrum and the total number of vehicular users defined for each scenario. Moreover, the average blocking rate achieved by the PSA algorithm was calculated by their average values for all scenarios and traced as illustrated in Figure 2.

While increasing the number of connected vehicular UEs in the coverage area, the average blocking rate raises a little or will be perpetually starved in scenario 1 due to the limitation of radio resource blocks. We mark that the raise of average blocking rate is less important for safety traffic in scenario 1 than the others one, those explain that the RBs are assigned based on the type of traffic connections. Scenario 1 is oriented NST traffics, so the number of VST is less than the number of VNST. In addition, the main objective of the proposed approach is to maximize the number of vehiclar users served by assigning a physical RB to two VUEs and by giving more priority to the safety links which must be served immediately due to their timeliness requirement. In scenario 1 and 2, the user with good channel condition can guarantee the minimum required QoS, has a superior priority to be scheduled while VUE with a mediocre channel gain waiting their access, which conducts to a equity problem and high latency. PSA aims to allocate multiple VUEs with diverse channel qualities simultaneously. Moreover, it can supply an enhancement by ensuring massive connectivity as indicated in Figure 2.

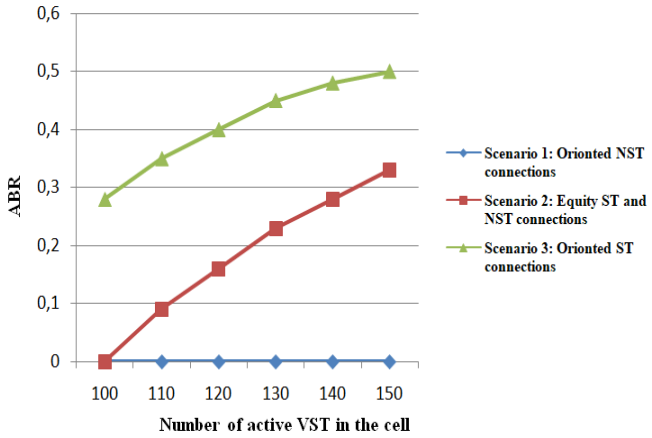


Figure 2: ABR analysis for ST communications

6.2. Cell Throughput

Maximizing average throughput is the principal goal of our algorithm. Figure 3 illustrates the performance of PSA scheme in terms of overal throughput for the three already mentioned scenarios versus the number of connected links in the cell.

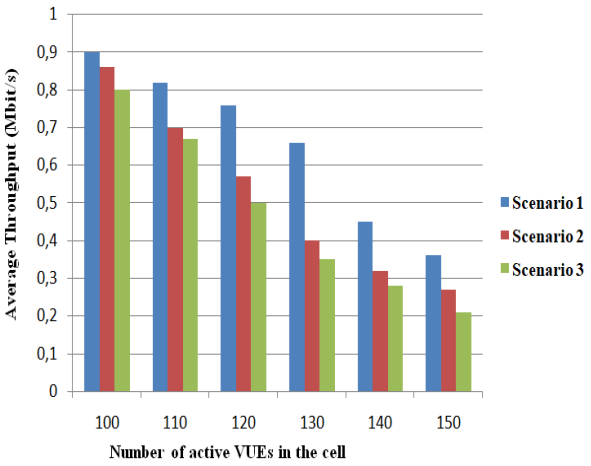


Figure 3: Average throughput analysis for NOMA (Rural Macro-cell)

We show that the raise in cell throughput is more better for scenario 1 than for scenario 2 and scenario 3 which is less efficient. From the Figure 3, it is observed that the total throughput of the proposed algorithm gradually decreases when the number of VUEs increases. This is due to the large demand on resources blocs. Since the proposed algorithm cancels the waste of the RBs by updating the metric before each new allocation and by maximizing the use of available RBs. Indeed, the proposed algorithm achieve 0,9 Mbps (scenario 1) when the targeted active VUEs in the cell is

equal to 100. This scenario oriented to be NST, while this type of traffics needs a high throughput to provide desired QoS. On the other hand, scenario 3 is oriented safety traffic. Taking into account the type of communications, this increase affects the throughput because in real world scenarios not all V2X connected links are safety communication and not all VUEs are close to the base station. ST connections often requires stringent exigency as will as low latency and high reliability. Indeed, safety traffics representing small data packets compared to non safety traffics. For ST connection like control loss warning, pre-crash sensing warning, forward collision warning and queue warning, the short transmit packets have a size of 50-300 Bytes [31] [32]. Figure 3 shows that the acheived throughput for ST can support several packets sizes. So, PSA allows exchanging packets size along to 120 Bytes when the average sum rate is equal to 0.9 Mbps. Moreover, the results satisfy the recommendation which is imposed by 3GPP organization when the lowest message size is 50 Bytes.

Indeed, Figure 4 shows how the cell throughput varies according to the number of active connections in the coverage area for the conventional OMA.

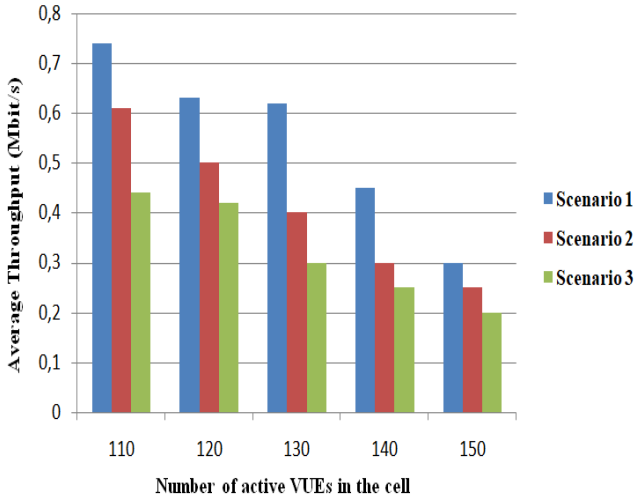


Figure 4: Average throughput analysis for OMA (Rural Macro-cell)

It's obvious that overall throughput decreases when VUEs increases. From that, it is clear that NOMA outperforms OMA even in all scenarios, which verifies that PSA based on NOMA can attain a higher cell throughput than OMA which has a gap to the greatest attainable cell throughput due to it uselessness on active V2X links selection. So, our proposed algorithm out performs OMA-based scheme due to a better resources allocation.

This is because PSA effectively preserves user diversity by its RB allocation metric. Unlike OMA scheduler, NOMA can simultaneously ensure optimal use a huge number of VUEs as well as performing better QoS requirements.

System model of our proposed algorithm does not consider only rural propagation scenario but also urban environment. Figure 5 illustrates the average throughput varies with the number of connected vehicular users in the cell for urban macro-cell propagation scenario evaluated for three different scenarios.

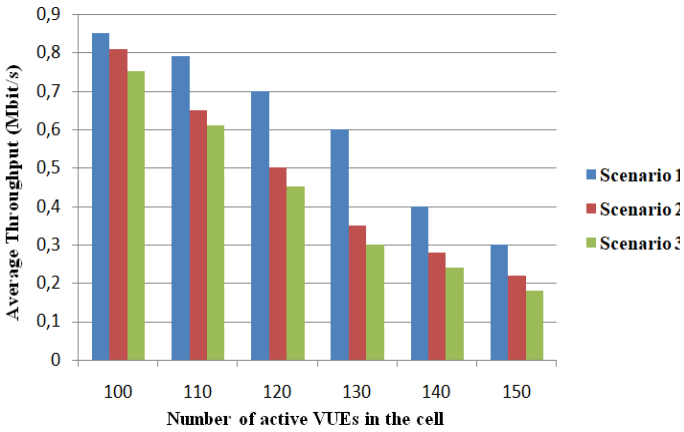


Figure 5: Average throughput analysis (Urban Mcro-cell)

Urban macro-cell propagation will face much path loss due to various properties, such as diffraction, reflection, refraction, absorption, free space loss, coupling loss and aperture-medium compared to the rural macro-cell propagation. This is confirming that the average sum rate for rural macro-cell outperforms the urban macro-cell average throughput as shown in figure 5. Moreover, all propagation scenarios based on NOMA algorithms are much better compared to OMA one, which confirming the convergence and stability of our proposed algorithm and verifying that NOMA can obtain a higher system throughput than OMA.

In addition, Figure 6 displays the system throughput of the developed resource allocation approach PSA compared with the attainable system throughput in [21] versus of the number of vehicular UEs in the system. Our algorithm is compared to [21], already detailed in related work section, because we find that this reference present the most high throughput compared to other studied related works.

Tree-based Search power Allocation algorithm (TSA) can both solve the link selection problems, user scheduling and power allocation with a low computational complexity for Device to Device underlaying cellular network based on NOMA techniques to increase the average throughput [21].

For FIxed (channel-Independent) Power Allocation algorithm (FIPA), the set of users is sorted in the decreasing normalized channel gain order, the system reaches to assign more power level to users experiencing high channel qualities based on the fixed weighted sum of the instantaneous throughput [33].

For FRactional Power Allocation (FRPA), the transmission power control depends on channel condition variations among users. Respective users must be informed by allocated power through explicit control signaling, which differentiates this allocation from the fixed power allocation [33].

For our proposed PSA algorithm and TSA defined in [21], the cell throughput increases with a varying speed that increasingly slows down due to increasing the number of connected VEUs in the system.

From Figure 6, the increment of the average throughput, PSA is better than TSA and therefore outperforms the other three algorithms FRPA and FIPA.

Figure 6 also clearly demonstrates the superiority of PSA that outperforms the compared scheme in terms of sum rate since the proposed algorithm takes into account the service classification and the channel gains of the respective non-orthogonally multiplexed VUEs.

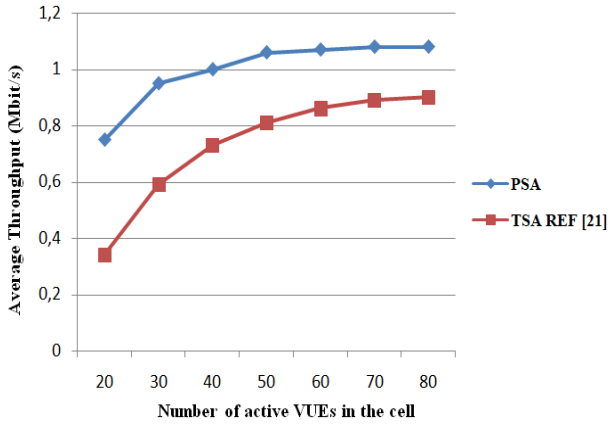


Figure 6: Average throughput analysis

6.3. Fair Index

The developed solution focuses on priority management to serve users fairly and increase the number of served VUEs. Fairness is a key factor for evaluating the simulation results of scheduling and resource allocation schemes, its calculates the equity index among VUEs. As PSA aims to maximize the minimum achievable instantaneous data rate, this can ensure fairness among users. To measure the equity of the system, we exploite the well-known techniques, it is Jain's fairness index [34] given by equation "(11)".

$$f(D_1, D_2, D_3, \dots D_K) = \frac{\left[\sum_{K=1}^{N_s} D_K\right]^2}{N \sum_{K=1}^{N_s} (D_K)^2} \quad (11)$$

where N_s is the number of served VST users and D_K is the throughput of the Kth VST.

We can observe on Figure 7 that the designed scheme has a good level of equity index for VST because it maximizes the use of the available RB between different V2X communications. Moreover, PSA tend to minimize the rate gap among scheduled users, thus maximizing fairness among them. Comparing with OMA algorithm, the proposed algorithm achieves not only higher system throughput but also better fairness between subscribers.

It was shown that the fairness index of the proposed resource allocation scheme with NOMA system is higher than others paper [18]. The equity of all algorithms decreases when the number of connected links raises. The authors simulate this factor of performance with limited active connections in the system. This demonstrated that the proposed algorithm can ensure the best fairness performance by the different number of connected vehicular users comparing with resultas shown on [18].

Simulation results have illustrate that the proposed algorithm PSA has various proprieties that maintain remarkable enhancement in terms of average throughput, Average Blocking Rate (ABR) and to guarantee fairness index between different class of VUEs comparing to the conventional OMA assisted relay scheme and some recent work for NOMA based allocation. According to [31] and [32], the simulation results guarantee the 3GPP recommendation when the minimum small packet size is 50 bytes. Hence, our proposed algorithm is capable to support variable packets sizes meeting the requirements imposed by 3GPP standard.

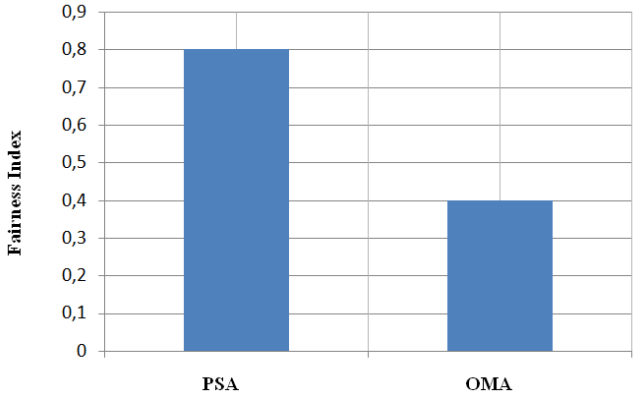


Figure 7: Fairness index analysis for ST connections

7. Conclusion

This paper, presents a novel scheduling and resource allocation approach named PSA for NOMA systems. PSA aims to maintain the highest average throughput by considering the heterogeneity of channels quality of each V2X link. The proposed work increases the number of served safety V2X links in the cell while a minimum requirement must be guaranteed on terms of throughput and maximize the average throughput of NST VUEs. The simulations results verify the optimality of our proposed algorithm in comparison with conventional scheme OMA that lack of caring fairness between vehicular users. The proposed scheme performs amelioration in term of throughput compared to other studied publication papers. A better compromise between cell throughput and fairness index has been achieved. As perspective for this research work, we plan to study the possibilities offered by Quanser platforms for reel environment implementation. In fact Quanser is an open-architecture system integrating plants with realistic dynamics sufficient for physically-relevant testing, real-time computation, and modeling for rapid testing of algorithms and concepts in MATLAB/Simulink and LabVIEW environments. We can also adopt such improvement to optimize the proposed scheme in terms of complexity.

References

[1] A. Al-Dulaimi, X. Wang, C. Lin , "Machine-Type Communication in the 5G Era: Massive and Ultrareliable Connectivity Forces of Evolution, Revolution, and Complementarity," Wiley-IEEE Press, 2018, doi:10.1002/9781119333142.ch14.

[2] Y. Han et al., "Periodic Radio Resource Allocation to Meet Latency and Reliability Requirements in 5G Networks," IEEE Vehicular Technology Conference (VTC Spring), 2018, doi:10.1109/vtcspring.2018.8417636.

[3] B. Di et al., "Non-Orthogonal Multiple Access for High-Reliable and Low-Latency V2X Communications in 5G Systems," IEEE Journal on Selected Areas in Communications, 2017, doi:10.1109/jsac.2017.2726018 .

[4] Z. MacHardy et al., "V2X Access Technologies: Regulation, Research, and Remaining Challenges," IEEE Communications Surveys & Tutorials, 2018, doi:10.1109/comst.2018.2808444.

[5] A. Masmoudi et al., "A Survey on Radio Resource Allocation for V2X Communication," Wireless Communications and Mobile Computing, 2019, doi:10.1155/2019/2430656.

[6] F.A. Giovanni, "An Overview of Vehicular Communications," Future Internet, 2019, dio:10.3390/fi11020027.
[7] B. Di et al., "V2X Meets NOMA: Non-Orthogonal Multiple Access for 5G-Enabled Vehicular Networks," IEEE Wireless Communications, 2017, doi:10.1109/mwc.2017.1600414.
[8] S. Rai et al., "Result Analysis of Efficient MIMO Framework for NOMA Downlink and Uplink Transmission based on Signal Alignment," International Journal of Computer Applications 2018, doi: 10.5120/ijca2018916907
[9] A. Agarwal et al., "Performance analysis for non-orthogonal multiple access (NOMA)-based two-way relay communication," the Institution of Engineering and Technology (IET Communication), 2019, doi:10.1049/iet-com.2018.5641.
[10] X. He et al., "Spectrum and Power Allocation for Vehicular Networks with Diverse Latency Requirements," IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 2019, doi:10.1109/smartgridcomm.2019.8909694.
[11] A. Bazzi et al., "Optimizing the Resource Allocation of Periodic Messages With Different Sizes in LTE-V2V," IEEE Access, 2019, doi:10.1109/access.2019.2908248.
[12] A. Nazari et al., "Resource allocation in power domain NOMA based cooperative multicell networks," the Institution of Engineering and Technology (IET Communication), 2020, doi: 10.1049/iet-com.2019.0624.
[13] N. Glei et al., "Power Allocation for Energy-Efficient Downlink NOMA Systems," international conference on Sciences and Techniques of Automatic control & computer engineering (STA), 2019, doi: 10.1109/STA.2019.8717240.
[14] K. Huang et al., "Energy efficient resource allocation algorithm in multi-carrier NOMA systems," International Conference on High Performance Switching and Routing (HPSR), 2019, doi:10.1109/HPSR.2019.8807992.
[15] Z. Yuan "Energy Efficient Power Allocation for Multi-carrier Non-orthogonal Multiple Access (NOMA) Systems with Proportional Rate Constraints," IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2019, doi:10.1109/ITNEC.2019.8729390.
[16] M. Zeng et al., "Spectral- and Energy-Efficient Resource Allocation for Multi-Carrier Uplink NOMA Systems," IEEE Transactions on Vehicular Technology, 2019, doi:10.1109/TVT.2019.2926701.
[17] M. Pischella et al., "NOMA-Relevant Clustering and Resource Allocation for Proportional Fair Uplink Communications," IEEE Wireless Communications Letters, 2019, doi:10.1109/lwc.2019.2897796.
[18] K. Long et al., "Spectrum Resource and Power Allocation With Adaptive Proportional Fair User Pairing for NOMA Systems," IEEE Access, 2019, doi:10.1109/access.2019.2908673.
[19] X. Wang et al., "Low-Complexity Power Allocation in NOMA Systems With Imperfect SIC for Maximizing Weighted Sum-Rate," IEEE Access, 2019, doi:10.1109/access.2019.2926757.
[20] R. Abozariba et al., "NOMA based Resource Allocation and Mobility Enhancement Framework for IoT in Next Generation Cellular Networks," IEEE Access, 2019, doi:10.1109/access.2019.2896008.
[21] J. Wang et al., "A Novel Resource Allocation Scheme in NOMA-Based Cellular Network with D2D Communications," Future Internet, 2020, doi:10.3390/fi12010008.
[22] L. Chen et al., "Proportional Fairness-Based User Pairing and Power Allocation Algorithm for Non-Orthogonal Multiple Access System," IEEE Access, 2019, doi:10.1109/access.2019.289618.
[23] H. Xing et al., "Sum-rate Maximization Guaranteeing User Fairness for NOMA in Fading Channels," IEEE Wireless Communications and Networking Conference (WCNC), 2018, doi:10.1109/wcnc.2018.8376949.
[24] J. He et al., "Design and Optimization of Scheduling and Non-orthogonal Multiple Access Algorithms with Imperfect Channel State Information," IEEE Transactions Vehicular Technology, 2018, doi:10.1109/tvt.2018.2869305.
[25] METIS D1.4, METIS Channel Models, ICT-317669-METIS/D1.4, METIS project, 2013.
[26] WINNER II D1.1.2, Channel models, V1.2, 2008.
[27] G. Dimitrakopoulos et al., "Vehicular Communications Standards, Current Technologies in Vehicular Communication. Springer, Cham, 2017.
[28] C.M. Silva et al., "A Survey on Infrastructure-Based Vehicular Networks," Hindawi Mobile Information Systems, 2017, doi:10.1155/2017/6123868.
[29] A.D. Trabelsi et al., "Dynamic Scheduling Algorithm based on Priority Assignment for LTE-V2X vehicular Networks," International Wireless Communications & Mobile Computing Conference (IWCMC), 2019, doi:10.1109/iwcmc.2019.8766632.
[30] G. Cecchini et al., "Localization-Based Resource Selection Schemes for Network-Controlled LTE-V2V," International Symposium on Wireless Communication Systems (ISWCS), 2017, doi: 10.1109/ISWCS.2017.8108147.
[31] 3rd Generation Partnership Project, Technical Specification Group Services and System Aspects: Study on LTE support for Vehicle to Everything (V2X) services(Release 14) , 3GPP TR 22.885 V14.0.0, 2015.
[32] 3rd Generation Partnership Project, Technical Specification Group Services and System Aspects; Study on enhancement of 3GPP Support for 5G V2X Services (Release 15) , 3GPP TR 22.886 V15.1.0, 2017.
[33] N. Otao et al., "Performance of Non-orthogonal Access with SIC in Cellular Downlink Using Proportional Fair-Based Resource Allocation," International Symposium on Wireless Communication Systems (ISWCS), 2012, doi:10.1109/iswcs.2012.6328413.
[34] R. Jain, The Art of Computer Systems Performance Analysis, Wiley, 1991.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1237-1243 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Creativity in Prototypes Design and Sustainability - The case of Social Organizations

Clara Silveira[*,1], Leonilde Reis[2], Vitor Santos[3], Henrique S. Mamede[4]

[1]*School of Technology and Management, Polytechnic Institute of Guarda, Guarda, 6300-559, Portugal*

[2]*School of Business Administration, Polytechnic Institute of Setúbal, Setúbal, 2914-503, Portugal*

[3]*NOVA Information Management School (NOVA IMS), Lisboa, 1070-312, Portugal*

[4]*INESC TEC, Universidade Aberta, Lisboa, 1269-001, Portugal*

ARTICLE INFO

Article history:
Received: 07 October, 2020
Accepted: 04 December, 2020
Online: 16 December, 2020

Keywords:
Creativity
Design
Sustainability
Social Organization

ABSTRACT

The role of creativity techniques in the design of prototypes is of particular interest given its potential for innovation. At same time, despite the efforts of decades in terms of policies and programs of action, humanity has not yet come close to global sustainability. Sustainability design must involve society and creatively employ all available knowledge sources for creating sustainable software. This paper proposes a prototype design approach rooted in employing creativity techniques, while being guided by the dimensions and principles of the Karlskrona Manifesto. This approach is applied to the development of a multidisciplinary aggregator for the optimization of social services. As a result. guidelines for the use of creativity in requirements engineering will be presented, as well as on how to include sustainability issues, namely the Sustainable Development Goals and the five dimensions of sustainability in the design of prototypes.

1. Introduction

This work is an extended version of the paper [1] originally presented in 15th Iberian Conference on Information Systems and Technologies (CISTI 2020).

The software industry has become extremely competitive. Often, a choice is made between multiple software systems that strive to meet the needs of users in the same application domain. Therefore, in order to be sustainable and grow in the market, a software product needs to differentiate itself from other similar products and captivate users with new and useful features. Given such context, the use of creativity techniques is a determining factor for those who develop software. In Grady Booch's opinion [2], the nature of the systems we build continues to change and as they cross collectively in our lives, we must take into account not only the technical elements of software development, but also human needs. The use of creativity techniques allows requirements engineers to deal with new problems and domains, unique contexts, new applications and a combination of existing methods and techniques [3], [4]. However, research on how the application of creativity occurs has been neglected. Various creativity techniques, such as Brainstorming, scenario generation and simulations, and ad hoc design can be used to support and promote creativity in the requirements engineering process [3].

The Karlskrona Manifesto [5] and other authors [6]-[11] gave visibility and increased awareness of the importance of sustainability in software and Information Systems (IS). Specifically, the Karlskrona Manifesto for Sustainability Design (KMSD) serves as a guide for designing and developing sustainable software systems. Thus, the study of sustainability in the area of software development and IS is of extreme importance in promoting good sustainability practices and the integration of the 17 Sustainable Development Goals (SDGs) defined by the United Nations [12].

This study aims to describe the Creative process and techniques to be used in the design of integrative prototypes for a Social Organization, taking into account sustainability issues and the SDGs.

The paper is organized into five sections. After the introduction, the fundamentals and motivation for using creative techniques in the design of software prototypes are presented, as well as the principles of the Karlskrona Manifesto. Section three

[*]Corresponding Author: Clara Silveira, Polytechnic of Guarda, mclara@ipg.pt

describes the proposed approach. Next, section four presents a case of applying the approach to a social organisation as well as the proposal for a prototype in relation to the specificity of the problem. Finally, in the last section, the conclusions are presented, as well as the perspectives of future works, showing the spaces still open in the field of work development.

2. Creativity and Sustainability Design

2.1. Creativity in Requirements Engineering

The field of software engineering, especially requirements engineering, is extremely creative. However, professionals and researchers seem to have different perceptions about creativity, making knowledge transfer difficult [13]. The types of knowledge that we find in the literature that impact creativity from the perspective of software engineers are classified as [14]: previous knowledge, analogies and dynamic knowledge.

Prior knowledge ensures useful information about previous projects for the early stages of ideas generation. This background information provides a starting point or motivation to start with ideas for solving the problem. Analogies help software engineers create a link between two objects in the same or another domain and create new ideas. Dynamic knowledge refers to the knowledge embedded in the tools and available at any time.

We also emphasize that knowledge itself is creative [15] and, therefore, should be encouraged to develop. Results from various experiments show that creativity during the design phase can be improved with the help of analogies [16]. Robert Sternberg [17] defines creativity as the ability to produce work that is simultaneously new (i.e., original and unexpected) and appropriate (i.e., useful and adaptable).

It is also considered that the design of the prototype must contemplate aspects of creativity, including also the concern with the level of comprehensiveness of the information in face of the organizational reality. One concern focuses on the excessive coverage underlying duplication of information [18]. Such a situation can occur when there are duplicates in the individual sources.

In this sense, prototypes must be created to stimulate the creativity of the actors. It may be useful to think about the possible causes of prototype coverage [18], when compared to the target population. Therefore, it is considered essential that a prototype is developed based on research, knowledge, analogies, creative techniques, sustainability factors and, ultimately, validated to answer if it works as specified [19].

Thus, given the specificity of the organization used in this study (of a social nature), it is considered that the purpose of a prototype [19], is to create a semi-realistic representation of something in order to be able to interact and be tested with end users. Its main function is to identify problems and opportunities for improvement long before implementation begins. As it is an agile and efficient method, prototyping will help to define and include more assertiveness in problem solving.

Within the scope of the works related to the thematic area and considering the objective of the paper, practical works developed in organizational context are cited, ([1], [3], [4]).

2.2. Sustainability Design for Software Development

The concept of sustainability is often extrapolated to software engineering with the KMSD [5]. The KMSD principles cover several aspects of sustainability, serving as a point of reference and guide during the software design process [20].

The KMSD manifesto appeared at the Third International Workshop on Requirements Engineering for Sustainable Systems (RE4SuSy), held in Karlskrona, Sweden. It is considered a guide to design and develop more sustainable software systems. The KMSD includes nine principles ([5], [21]):

- **P1.Sustainability is systemic**: Sustainability is never an isolated feature. Systematic thinking should be the main focus to include interdisciplinarity.
- **P2.Sustainability has multiple dimensions**: Advocates the inclusion of the five dimensions (Table 1) in our analysis to be able to understand the nature of sustainability in any situation.
- **P3.Sustainability transcends multiple disciplines**: Sustainability exceeds several disciplines, working on challenges from several perspectives.
- **P4.Sustainability is a concern independent of the purpose of the system**: Regardless of the system's development context, sustainability must always be included.
- **P5.Sustainability applies to both a system and its wider contexts**: Concerned with considering at least two spheres in the design of the system: the sustainability of the system itself and how it may affect the sustainability of the context of the entire system.
- **P6.System visibility is a necessary precondition and enabler for sustainability design**: Concerned with sharing the vision of the system and its context in different perspectives and with the various levels of abstraction, allowing an informed choice. In this principle, it is important to cite the application of the *SharedClearVision* pattern [22].
- **P7.Sustainability requires action on multiple levels**: Consists of researching measures for sustainability that are considered the most effective way to intervene in comparison with other alternative actions.
- **P8.It is possible to meet the needs of future generations without sacrificing the prosperity of the current generation**: Sustainability innovation can play a key role in defining present and future needs. Thus, it is possible to make options that benefit both the present and the future.
- **P9.Sustainability requires long-term thinking**: Assess, in the long term, the possible impacts, costs and benefits of decisions made.

The manifesto presents general and abstract principles as well as values of sustainability. The KMSD Manifesto also includes the dimensions (five) shown in Table 1.

Table 1: Karlskrona Manifesto - Sustainability Dimensions [5], [21]

Dimension	Focus
Human	It focuses on the well-being of the individual.
Social	Refers to social issues like groups of people and organizations in a community. Social

	responsibility, justice, job, and democracy are essential elements.
Economic	Economic sustainability aims at maintaining capital and added value.
Environ-mental	The main mission of environmental sustainability is to improve the well-being of the person, protecting natural resources and ecosystems.
Technical	Concerned with the longevity of information, systems, their evolution, maintenance and innovation.

The KMSD principles and commitments provide an overview of the various dimensions of sustainability and their relationships. In this way, we consider that they show the impact that concerns with sustainability can have in the design of an information system. We added a new vector, combining creativity with the various dimensions of sustainability, allowing to preserve the environment and achieve new technological innovations in social organizations.

3. Creative Approach in Prototype Design

In this section, we describe the creative approach used in prototyping within the framework of a more comprehensive methodology for requirements engineering. Among the various creativity techniques that can be used to support and promote creativity in the design of prototypes, the following were selected by the creativity expert [2], namely: Brainstorming, Brutethinking, Reversal and Ideabox. The selection of these techniques took into account the author's accumulated experience in conducting workshops.

Brainstorming is an idea generation technique, which consists of a meeting designed to encourage the full release of mental activity without restrictions. This technique works so well because, among many other reasons, "ideas call ideas". Other people's ideas are sometimes starting points for our best ideas [3].

The Brutethinking technique [3] is characterized by being a simple process, which develops in four stages: choosing a random word; choose things/elements associated with the randomly chosen word; force a link between the word and the problem and between associations and the problem; relate the ideas obtained and analyze them.

The Reversal technique originates from the transformations that generate ideas. In some cases it is better to think first about the negative and then invert the negative, indicate "the worst" and then revert to "the best" [3]. One of the ways is double inversion: initially the reversal identifies the ways to make the situation worse by realizing what can aggravate the problem, instead of improving it. Then, you make an inversion again, identifying the paths that can lead to the improvement of the situation.

IdeaBox is a tool that allows you to combine the parameters (characteristics, factors, variables or aspects) of a challenge (it can be a prototype) into new ideas.

In the creative approach, it is proposed to integrate multiple existing paradigms, in which the creative contribution results from the integration and combination of different approaches that can interrelate, in the search for the integration, in a single proposal, of different forms of thought about a phenomenon [23]. On the other hand, still in the context of the creative process, it is essential to mention the importance of the characteristics of the creative product, the different views on the models of creative processes, the role of the field of application, the personal characteristics found in highly creative people and, finally, the socio-organizational aspects involved [24].

And we have to note that [23] the multiple views of creativity are complementary and not mutually exclusive. This suggests that future models of creativity can integrate different sources and diverse individual perspectives. The creative approach proposed here is included in a more comprehensive methodology [3] and focuses on the step "Understanding user requirements and sustainability factors", which aims to apply Brainstorming, Brutethinking and reversal creative techniques to: identify the requirements of the user in a process of seeking a creative and sustainable solution; interaction with stakeholders to identify their needs and find out what needs to be built.

In this approach, existing applications and prototyping are used as a source for collecting and validating requirements. Sustainability principles and commitments are incorporated [5]. The preliminary list of requirements is drawn up, as well as the design of exploratory prototypes. In the analysis of prototypes it is proposed to use the Ideabox creativity technique.

4. Prototype Design in Social Organization

The Parish Social Center of the Parish of São Sebastião is a Social organization, founded in 1998 that develops strong intervention in the São Domingos Community (SDC). In his daily intervention integrates food distribution programs, monthly food aid (food products given by individuals and companies), distribution of clothing and hygiene products [25].

4.1. SDC Characterization

The organization under study, previously presented, has been the subject of work in the thematic area, contributing to the optimization of established practices [7]. In this sense, a brief characterization of the SDC (Figure 1) is presented to frame the problem.

Figure 1: SDC Characterization [7]

From the meetings between the Project Manager and the SDC Manager and also with the participation of the Software Architect, the organization's current situation was made. The development of the project also had the online collaboration of the Requirements Engineer element.

Figure 1 shows the various aspects that the prototype must include in Housing, Skills, Health, Training, Volunteering for the objective of Social Integration of the Citizen.

In view of the problem under study and the specificity of the social organization under analysis, the team used the Brutethinking technique, in three iterations, to identify the problem, the main objective, the primary causes, the actors of the system and their objectives, the functional and non-functional requirements, creative and non-creative requirements. The Reversal technique was used to prioritize requirements. The team also used the techniques of debate and Brainstorming, in order to evaluate and review artifacts and solutions found. In order to aggregate the information underlying the organization's activity, Figure 1 was conceived in order to highlight the themes involved in a transversal logic of support to the citizen inserted in a community, which, by its nature, has the collaboration of several official entities.

The development of the prototype for SDC was based on a multiple perspectives in order to describe each individual in the various components and taking into account the collaboration of several partners.The valence of volunteering has greater prominence (Figure 2) given the importance of collaboration / interaction of people in the community.

Figure 2: Volunteering characterization

Next, the other objectives and components of Figure 1 are characterized [26]:

- Social Center - support to activities –
  - Objective: aggregating skills and knowledge, so that the Social Center can offer high quality and integrated services. The objective is to provide a platform to the various actors, enabling them to access certain SDC data, depending on their access profile;
  - Characterization: the platform will allow: managing citizen profile, volunteer members, health professionals; nursing treatments; medication, physical activities; sharing goods and resources; share vehicle and register availability on the platform for volunteering.
- Official Entities – Partnerships
  - Objective: to list the entities with which they will establish partnerships, namely: parish council, health center, Public Security Police, Foreign Service and Borders. Collaboration with official entities will enable more effective responses to SDC people;
  - Characterization: the entity and the manager are identified, the mission, location, preferred contacts, partnership/protocol, association with other stakeholders are described.
- Experience - knowledge kills -
  - Objective: an SDC member's life experience (or that of their caregiver), can be valuable information to enhance cooperation with the SDC. Each member's level of literacy is also important so that their skills can be developed, culminating in the sharing of knowledge;
  - Characterization: the things a person did throughout their life; the things they like to do and would like to do; the activities they would like to participate in: these can be assessed from one's life experience.
- Resource/car sharing
  - Objective: to allow the sharing/donation of consumer goods or resources (equipment) and car sharing. Reducing carbon emissions and fuel consumption is promoted;
  - Characterization: application with resources and goods and availability; places and times to car sharing;
- Health and well-being
  - Objective: the aim is to thoroughly characterize the health status of community members, to help with specific and necessary care;
  - Characterization: in the various health valences, the importance of aggregating by areas is emphasized: health situation, rehabilitation/physiotherapy, motricity/biomechanics, medical and nursing care, taking medication.

The detailed characterization of the information previously exposed allows to present a multidisciplinary and aggregating vision in order to enhance the provision of services by the Social Center, corresponding to the expectations of its users.

4.2. Prototypes

The Social Center, as an entity that aggregates all information and considering the specificity of its mission and its social context, presents needs for collaboration with various entities.

The exploratory prototypes that were presented were conceived using the creative approach, namely the following techniques: Brainstorming, Brutethinking, Reversal and Ideabox. The following principles of the KMSD (described in section 2.2) have also been incorporated: P1; P2; P3; P4; P7 and P9.

Figure 3 shows the exploratory prototype for the main aspects of the Social Center. The prototype multidisciplinary enhances the aggregation of information. The project aims to define an access profile considering data protection concerns within the National Data Protection Commission, an entity that regulates citizens' access to and data protection in Portugal. Figure 3 presents as one of its components Partnerships representing the various entities

involved in the process, namely the Foreigners and Borders Service in order to support the organization of the process of inhabitants residing in the SDC district. The partners' name is also relevant in this context is social security since there is a regulatory and legal framework in Portugal that enhances the analysis of the economic and professional situation of citizens and families in order to enhance the creation of subsidies, namely the Social Income of Insertion. In this context, the prototype is of particular interest since it aggregates the information in various valences for each citizen and his family.

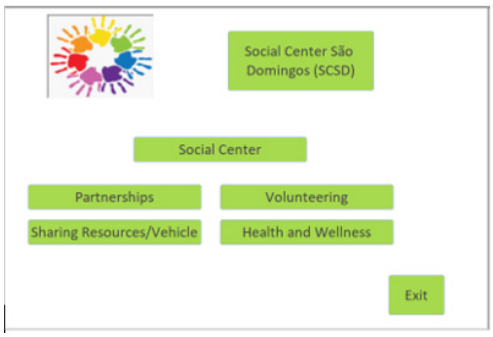


Figure 3: Main menu prototype

The Social Center São Domingos includes a low-level, low-economic community, mainly made up of elderly people with low economic levels, as they are retired people from the manufacturing environment. However, SDC also includes families with adolescents and children with economic needs. Thus, the creation of a prototype of this nature can enhance the centralization of community information in order to allow the delineation of support strategies. Figures 4 present prototypes for Citizen Profile.

Figure 4: Prototype Citizen Profile

The Citizen Profile characterization aims to collect and keep up to date a set of information referring to each citizen and his family in order to aggregate the various valences. Thus, it is intended a comprehensive characterization with regard to personal data in order to support the various sharing of information in order to create better living conditions. The medication and Health Problems includes the valences of improving the health care of each person, but also knowing the medication necessary to each person in the sense that in case of need the Social Center São Domingos provide its acquisition. Another of the concerns of the Social Center São Domingos is the mobility and conviviality of the elderly in which the Life Experience is an appeal to the sharing of knowledge promoting initiatives that allow to enhance the enormous knowledge of the elderly with other elderly and younger age groups through the organization of workshops and other events. Another aspect of the project focuses on recording in detail the Academic and Professional Training in which the practical experience of each one is emphasized by fostering this sharing through events. This need and also willingness to collaborate and explore the accumulated knowledge has allowed the create of availability for volunteering since it allows the taking of older people from home, enhances collaboration among community members increases the feeling of being useful and actively contributing in their community. Sometimes this collaboration is carried out in exchange for perishable and non-perishable goods increasing also of this were the resources of the family.

The regard to Volunteering - Actions to be promoted - it includes the concerns of involving SDC citizens in concrete voluntary actions. Particularly about the elderly, the objective is to enhance the coexistence and sharing of knowledge. The aspect willingness to receive volunteer support is carried out in actions such as: small reparations in housing, receiving non-perishable goods or other needs that are identified. In view of the sustainability concerns that are intended to be fostered, it is considered that in this context and in view of the specificity of the SDC Sharing – Resources/Vehicles has a view to promoting sustainability by appealing to the implementation of SDGs practices, in particular with regard to the re-use of resources and optimization of car sharing. Regarding Activities - Health and Well-being is intended to promote mobility and active life, as well as to have assertive monitoring particularly with regard to the senior population.

4.3. Discussion

The characterization of the SDC is of particular interest in the sense of having the opportunity, within the scope of a partnership, to study a community in order to test the applicability of the creative approach, contributing to the design of the prototype to meet the needs of the organization, although innovative and appropriate.

The prototype underlies the feedback from the team that includes the SDC manager. The creativity techniques Brutethinking, Reversal and Brainstorming were considered adequate for the design. The application of creativity techniques and the incorporation of concerns about sustainability allowed the inclusion of new services: car sharing; resource sharing; take medicine; volunteering; prototype integrating several valences and entities. This corroborates with the authors, in which the involvement and decision of the interested parties in the software project is an integral part of the constructions of software for effective use.

It is considered that the proposed prototype is an aggregator of information based on a multidisciplinary approach allowing the collaboration of several entities taking into account the specificity of intervention of each one so that the Social Center provides a more complete and integrating service.

In terms of sustainability concerns, it is noteworthy, for example, that volunteer actions allow contributing to SDGs as shown in Figure 5.

Figure 5: SDGs in the context of volunteering (adapted from [12])

The main contributions to the SDGs are highlighted [12]:

- SDG 4: The prototype under development aims to enhance the participation of young people from the Community of São Domingos to the extent that by getting involved in the various activities they acquire relational and technical skills in the context of their collaboration, even promoting their participation as entrepreneurs;
- SDG 5: The aim of this area is to promote women's empowerment by advocating a set of initiatives. A set of strategies have already been delineated, including short courses on the use of ICT and a financial education course in order to raise women's awareness of the importance of assertive management suggesting in this context the use of ICT to control expenditure;
- SDG 10: The Community of São Domingos has as its strategy inclusive policies given the multicultural provenance of its inhabitants. In this sense, it organizes social events serving the various cultures and religions existing in the neighbourhood.
- SDG 11: The Community collaborate in promoting efforts to protect the world's cultural and natural heritage organizes events;
- SDG 12: Aiming to promote sustainability policies, the Community of São Domingos promotes the distribution of non-perishable goods: clothing, specific health equipment such as wheelchairs and articulated beds and furniture. The Community not only aims to contribute to the well-being of its inhabitants, but also to promote the re-use and recycling of goods;
- SGD 14: Given its geographical location, the Community of São Domingos organizes awareness-raising actions about pollution mainly regarding the Sado River.

It should be noted that the SDGs cannot be analyzed in isolation, since the improvement of one SDGs will have a positive and/or negative impact(s) on another SDGs [27]. Integration adds valuable knowledge that can help society meet the SDGs [28].

5. Conclusions

This work presents a proposed approach to the design of prototypes using creativity techniques. It includes a study of a social nature, which contains a set of assumptions that constitute a challenge. Throughout the development of the work, conditions were created for the application and validation of techniques, namely in the field of creativity in IS. In this sense, Brainstorming, Brutethinking, Reversal and Ideabox techniques were applied. Another technique applied in the work was collaboration, where software engineers shared their ideas and thoughts with others. They build on each other's ideas and provide constructive feedback. In this way, it was possible to fill the knowledge gap of each one, creating a group. On the other hand, the principles of the Karslkrona Manifesto have already been applied in several areas [11], [27]-[31].

In this research was possible to apply the theoretical references mentioned to study a real problem, proposing an integrative and multidisciplinary solution in order to enhance the activity of the Social Center in the São Domingos Community in order to provide better services to its community.

As discussed in section 4.3, the results obtained show that creativity techniques can play an important role in transforming of existing Information Systems, bringing innovative and unexpected solutions, which without removing functionality or performance from Information Systems can make them more sustainable.

Therefore, in this context, the understanding that there is an important role to be played by creativity techniques in promoting SDGs and contributing to the development of more sustainable IS is reinforced.

The process was applied in another context [26] and a significant contribution was perceived in the field of sustainability, in the areas: social, individual, economic, environmental and technical [7].

As future work prospects, the case of the chosen Social Organization will be replicated in other communities/institutions in order to contribute to their sustainability. Also, the development of an application to make available to the Social Center is recommended, underlying concerns of a graphic nature, screen design, usability and accessibility. In this sense, solutions are being studied that address the specificities and constraints, particularly financial constraints implicit to social organizations. Thus, the choice of the platform for the development and hosting of information, licenses and maintenance are concerns about the importance of keeping the solution in full operation over time. For the validation of the application we´ll use inspection techniques, usability tests and accessibility.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] C. Silveira, L. Reis, V. Santos, H. Mamede, "Criatividade no Design de Protótipos - O caso das Organizações Sociais," in CISTI'2020 - 15th Iberian Conference on Information Systems and Technologies, Sevilha, Espanha, 2020, doi:10.23919/CISTI49556.2020.9140870.

[2] G. Booch, "The future of software engineering," in Proc. 37th International Conference on Software Engineering (ICSE 15), 2015.

[3] V. Santos, Criatividade em Sistemas de Informação, Lisboa: FCA, 2018.

[4] V. Santos, L. Amaral, H. Mamede, R. Gonçalves, "Creativity in the Information Systems Planning Process," in Handbook of Research on

Innovations in Information Retrieval, Analysis, and Management, IGI-Global, 2015.

[5] C. Becker, R. Chitchyan, L. Duboc, S. Easterbrook, B. Penzenstadler, N. Seyff, C. Venters, "Sustainability Design and Software: The Karlskrona Manifesto," in Proc. 37th International Conference on Software Engineering (ICSE 15), 2015.

[6] Software Sustainability Institute, "Manifesto," 2019. [Online]. Available: www.software.ac.uk/about/manifesto.

[7] L. Reis, C. Silveira, "Sustentabilidade Multidimensional em Sistemas de Informação," in XXX Jornadas Luso-Espanholas de Gestão Científica, Bragança, 2020.

[8] L. Castro, V. Santos, "A perceptional study on the role of development agents and global leaders in leveraging technology for sustainable development," in Multi Conference on Computer Science and Information Systems, 169-176, 2019.

[9] J. Meneses, "A sustentabilidade em 2019," 2019. [Online]. Available: www.bcsdportugal.org/noticias/a-sustentabilidade-em-2019. [Accessed 25 outubro 2019].

[10] S. Hettrick, Research Software Sustainability: Report on a Knowledge Exchange Workshop, The Software Sustainability Institute, 2016.

[11] C. Venters, R. Capilla, S. Betz, B. Penzenstadler, T. Crick, S. Crouch, E. Nakagawa, C. Becker, C. Carrillo, "Software sustainability: Research and practice from a software architecture viewpoint" Journal of Systems and Software, **138**, 174-188, 2018. https://doi.org/10.1016/j.jss.2017.12.026

[12] United Nations Development Programme, "Sustainable Development Goals," 2015. [Online]. Available: www.undp.org/content/undp/en/home/sustainable-development-goals.html. [Accessed 02 12 2019].

[13] R. Mohanani, P. Ram, A. Lasisi, P. Ralph, B. Turhan, "Perceptions of Creativity in Software Engineering Research and Practice," in 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 210-217, Vienna, 2017, doi:10.1109/SEAA.2017.21.

[14] R. Hegde, G. Walia, "How to Enhance the Creativity of Software Developers: A Systematic Literature Review," in International Conference on Software Engineering and Knowledge Engineering, SEKE 2014, 2014.

[15] T. Davenport, L. Prusak, Conhecimento Empresarial, Rio: Campus, 1998.

[16] N. Bonnardel, "Creativity in Design Activities: The Role of Analogies in a Constrained Cognitive Environment," Creativity and Cognition conference, pp. 158-165, 1999.

[17] R. J. Sternberg, Handbook of creativity, Cambridge University Press, 1999.

[18] A. Black, "The IDI prototype spine's creation and coverage," Statistics New Zealand Working Paper No 16–03, 2016.

[19] G. Silvestri, "Prototipação e a criação de produtos que resolvem problemas," 18 09 2017. [Online]. Available: https://uxdesign.blog.br/o-objetivo-de-um-protótipo-é-criar-uma-representação-semi-realística-de-algo-que-seja-possível-e0445d85ca93. [Accessed 18 02 2020].

[20] B. Penzenstadler, S. Oyedeji, "Experiences from Applying the Karlskrona Manifesto Principles for Sustainability in Software System Design," Proceedings of the 8th International Workshop on Requirements Engineering for Sustainable Systems, 2020.

[21] B. Ovelheiro, C. Silveira, L. Reis, "Sustainability Design Applied to the Digital Signature of Documents," in In Carvalho, L. C., Reis, L., Prata, A., & Pereira, R. (Ed.), Handbook of Research on Multidisciplinary Approaches to Entrepreneurship, Innovation, and ICTs, IGI Global, 2021, 349-374, doi:10.4018/978-1-7998-4099-2.ch016.

[22] S. Adolph, P. Bramble, Patterns for Effective Use Cases, Addison-WesleyPearson Education, 2003.

[23] R. Stenberg, "The development of creativity as a decision-making process," in Creativity and development V.J.-S. K.., New York, Oxford University Press, 2003, 91-138.

[24] L. Nguyen, G. Shanks, "A framework for understanding creativity in requirements engineering," Information and Software Technology, **51**, 655–662, 2009. https://doi.org/10.1016/j.infsof.2008.09.002

[25] CSPSS, "Centro Social Paroquial de São Sebastião," Documento interno, Setúbal, 2019.

[26] C. Silveira, L. Reis, "Sustainability in Information and Communication Technologies," in Handbook of Research on Multidisciplinary Approaches to Entrepreneurship, Innovation, and ICTs, In Carvalho, L. C., Reis, L., Prata, A., & Pereira, R. (Ed.), IGI Global, 2021, 375-396, doi:10.4018/978-1-7998-4099-2.ch017.

[27] L. Reis, C. Silveira, L. Carvalho, C. Mata, R. Madeira, "Design de Protótipo para Pessoa sem-abrigo - Contributos de Sustentabilidade," in IV International Forum on Management, Universidade Aberta. Lisboa, 2020.

[28] P. Shrivastava, M. Smith, K. O'Brien, L. Zsolnai, "Transforming Sustainability Science to Generate Positive Social and Environmental Change Globally," One Earth, **2**(4), 329-340, 2020. https://doi.org/10.1016/j.oneear.2020.04.010

[29] S. Oyedeji, A. Seffah, B. Penzenstadler, "A Catalogue Supporting Software Sustainability Design," Sustainability, **10**, 2296, 2018. https://doi.org/10.3390/su10072296

[30] B. Penzenstadler, S. Betz, C. Venters, R. Chitchyan, J. Porras, N. Seyff, L. Duboc, C. Becker, "Everything is Interrelated: Teaching Software Engineering for Sustainability," in IEEE/ACM 40th International Conference on Software Engineering: Software Engineering Education and Training, 2018.

[31] P. Lago, "Architecture Design Decision Maps for Software Sustainability," in IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society, 2019.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1244-1249 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Computer Vision for Industrial Robot in Planar Bin Picking Application

Le Duc Hanh*, Huynh Buu Tu

Mechanical Department, Ho Chi Minh City University of Technology, Vietnam National University, Ho Chi Minh city, 70000, Vietnam

ARTICLE INFO

Article history:
Received: 05 August, 2020
Accepted: 03 December, 2020
Online: 16 December, 2020

Keywords:
Bin picking
Computer vision
Industrial Robot
3D Point cloud
Watershed

ABSTRACT

This research presents an effectively autonomous method that can save time and increases productivity for an assembly line in industry by using a 6DOF manipulator and computer vision. The objects are flat, aluminum type and randomly stacked in a box. Firstly, 2D color image processing is performed to label the object, then using 3D pose estimating algorithm, the surface normal, angle and position of an object are calculated. To reduce the burden time of 3D pose estimation, a voxel grid filter is implemented to reduce the number of points for the 3D cloud of the objects. As known the 3D image object was obtained by camera in bin often involves both heavy noise and edge distortions, so to prepare for the assembly a manipulator will pick and place it at sub position then a 2D camera is used to estimate the pose of the object correctly. Through implement experiment, the system proved that it is stable and have good precision. Installing time and maintaining is fast and not complicated. It is applicable in production line where mass product is produced. It is also the good foundation for other deep researches.

1. Introduction

For many reasons, such as people, productivity, management and working environment; the robots gradually replace the people to carry out the tasks flexibly, accurately and safely. The tasks such as assembling and disassembling a machine part, arranging large quantities of goods requiring a lot of effort. In addition to exposure to hazardous environments, the risks of machinery such as billet cutting machines, corrugated iron stamping, turning, heavy chemical containers, grippers and leveling of embryos when welding can be dangerous to humans. Therefore, the participant of intelligent robot is indispensable because The robot can recognize the surrounding environment and objects correctly by the sensors as well as very high performance. Currently, there are many researches in the world about automatically supplying embryos in the factory or especially the task of picking up objects occluded each other inside a bin and placing it in the assembly area or welding. This task is called "Bin Picking". Casado and Weiss proposed method to apply algebraic relations between the invariants of a 3D model and those of its 2D image under general projective projection for estimating object 3D poses [1-2]. However, because of the occlusion, pose estimation using 2D image matching only works when their poses are limited to a few cases especially in automatic random bin-picking systems. Ban and Xu et proposed a laser vision system consists of using a projector and a fixed camera to project vertical and horizontal lines, thereby using the camera to determine the X, Y, Z coordinate and direction of objects to pick [3-4]. Chang recommended attaching a laser and projector on the hand of the mobile device to reconstruct the 3D image of the workspace and objects [5]. However, the system [3-5] has limited speed because it requires waiting times for projector operating and also of process complex images. Recently, the work has been done to use the stereo vision system to view the work area and to calculate the position and normal vectors of the gripping objects or using both stereo vision, projector and CAD models to determine the position and coordinates of the object to be picked up in a box [6-9]. However, CAD models of objects in actual production are often not provided. Recently, the structured-light vision system and Point cloud are also recommended as Schraft proposed an algorithm to calculate the position and coordinate of the object based on the structure light system and a sample point cloud [10]. After obtaining 3D working area point cloud the system will compare with the sample point cloud to determine the location of an object. The system has the disadvantage that the model points must be saved at first as a data system so the algorithm will take a lot of time to compare with this sample data point cloud. More over the database will be huged and complicated. Bousmalis [11] and Mahler [12] use the learning method to identify the object to be picked, but the algorithm is

*Corresponding Author: Le Duc Hanh, Email: ldhanh@hcmut.edu.vn

limited because the object is randomly stacked, so the objects can overlap each other [11-12]. When the algorithm performs learning, it will also learn the black overlap region, so the amount of data learned is very large and when performing the pick-up, the robot will have a lot of difficulties due to these black shadows data. Over past decade, deep machine learning has also been applied to object recognition such as Tuan which uses large data collection and neural network to identify and take rectangular objects in containers [13]. Chen re-modeled the 3D object to be picked up, simulating and training the position and gripping position of the object to be picked [14]. However, the technique of [13-14] need to be retrained the database if using other objects and training data is very large.

This research is different from discussed above. A Kinect sensor, in addition to returning the location data of the pixel, also produces the color of each location, so combining the location and color data will improve the algorithm accuracy. Firstly, all the point cloud information consists of pixel location and its RGB image are converted to 2D image for labeled data and then converted again to point cloud data. Secondly, the 3D pose of object is roughly estimated. The data is then transmitted to a Manipulator to perform a grasping task. The object will be placed on the V-groove and the system uses a 2D camera to process the image to refine the object's pose.

2. Segmentation

2.1. 2D image processing

Watershed is an algorithm used in image segmentation [15]. Algorithms are popular in fields such as biomedical and medical processing, computer vision. Firstly, due to the high density of 3D data acquired from Kinect, the location of point cloud inside the registered range are kept by using pass-through filter algorithm. By using filer algorithm, the number of points in cloud reduced from 307200 points to about 30000 points. After filtering, all the point cloud information consisting of pixel location and its RGB image are converted to 2D image. The canny and the distance image transform algorithm are applied to find the contour of the object. After that the watershed algorithm is applied to label object. The labeled objects contained pixel locations and labeled colors are then converted back to 3D point cloud data as shown in Figure 1.

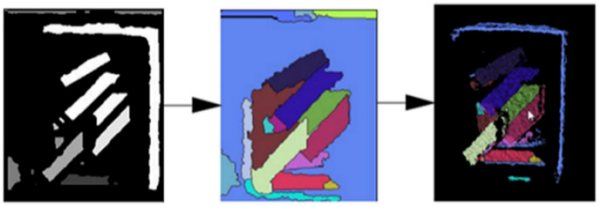


Figure 1: Watershed algorithm and colors assigning to Point cloud.

After objects are labeled, the data remains as unorganized objects data and will be segmented into each planar surface which represents for object cloud called cluster.

2.2. 3D image processing

When using 3D devices, the collected data often appears to be hovering and not in any plane. These points usually appear near the interface between surfaces, or near the edges of an object and. These points need eliminating, so the outlier filter is applied by considering the number of neighbor points K and the radius R of each point. If any point in point cloud whose number of neighbors in sphere of radius R is not greater than a threshold number, that point will be removed from point cloud. After that normal vector and the cluster segmentation are calculated. Normal vector is the most important point feature that distinguishes between different 3D cloud. the normal vector is the normative vector of a plane estimated from a given set of points. The problem of determining the normal to a point on the surface is approximated by the problem of estimating the normal of a plane tangent to the surface. To determine a plane equation in space, it is necessary to determine an x point in that plane and its normal vector $\vec{n}$. The point x is determined by focusing on nearby points.

$$x = \overline{p} = \frac{1}{k}\sum_{i=1}^{k} p_i \tag{1}$$

As discussed in [16], the solution for $\vec{n}$ is given by finding the eigenvalues and eigenvectors of covariance matrix $C \in R^3$ of P^k with:

$$C = \frac{1}{k}\sum_{i=1}^{k} \xi_i (p_i - \overline{p})(p_i - \overline{p})^T \tag{2}$$

The eigenvectors $\vec{v}_j$ form an orthogonal frame, corresponding to the principal components of P_i^k

$$C\vec{v}_j = \lambda_j \vec{v}_j, \quad j \in \{0,1,2\} \tag{3}$$

If $0 < \lambda_0 < \lambda_1 < \lambda_2$ the eigenvector $\vec{v}_0$ corresponding to the smallest eigenvalue λ_0 . Therefore the approximation of $\vec{n}$ is $\vec{n} = \{n_x, n_y, n_z\}$ or $-\vec{n}$. The calculation of the normal vector at the center will have errors, so to increase the reliability for the estimated vector, the maximum allowed error between the normal vector at the center and the normal vector at the adjacent point satisfy the following conditions:

$$\theta_{\overline{xi}} = ar\cos(\vec{n}_x \vec{n}_i) \tag{4}$$

Through conducting experiment, the $\theta_{\overline{xi}} = 5\ ^\circ$ is found to be appropriate.

Although the Kinect was placed perpendicular to the bin, the barrel edge point cloud still existed and connected to the real object in the bin formed an unexpected object in bin. When calculating the normal vector, the program will consider the barrel edge and the object are one object then the normal vector calculation and the position of the real object are not correct. Therefore, based on the normal vector, it is possible to separate the sides of the barrel with the idea that the angle between the two normal vectors on the side and the bottom edge will be larger than threshold value α. From experiment, the angle $\alpha = 50^0$ is appropriate. The results are shown in Figures 2 and 3.

The Euclidean clustering algorithm basically used to segment cluster and based on their neighbors [17]. Therefore, to estimate the normal vector to recognize that they belong to one plane, the Clustering procedure are through the following steps:

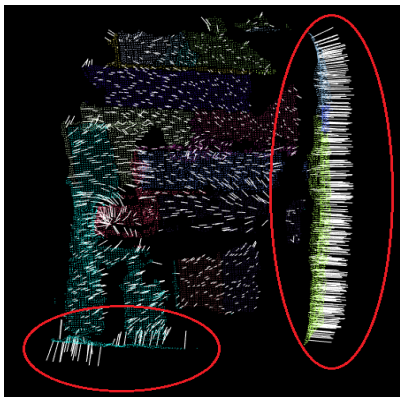


Figure 2: Normal vector of points

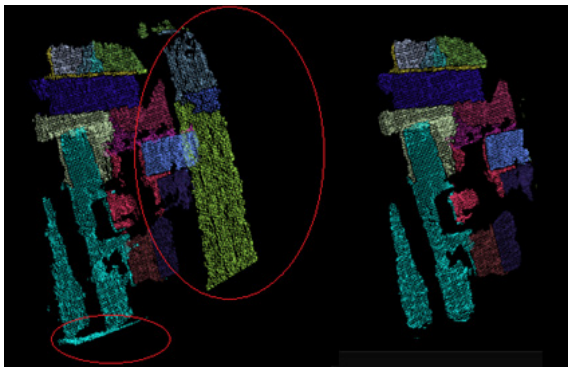


Figure 3: Image before and after removing barrel edge.

- Normal calculation of point cloud.
- Apply Kd-tree algorithm by define radius searching to find nearby points.
- Create list of cluster C and list of points to handle L.
- For every point p_i in P, do the following step:
- Add p_i in to L
- Search the list of neighbor n of p_i within a sphere of radius r using kd-tree algorithm.
- For every neighbor n_i, check if point exists in L. If it exists we skip to the next neighbor, otherwise we apply the following check.
- Calculate the angle between 2 point p_i and its neighbor n_i

$$\theta_{ij} = a\cos(\vec{p}_l, \vec{n}_l) \tag{5}$$

- Determine the color difference of 3 channels R, G, B of 2 points:

$$\left|R_p - R_n\right| + \left|G_p - G_n\right| + \left|B_p - B_n\right| < \text{Threshold} \tag{6}$$

- if the point p_i and its neighbor n_i satisfies (5) and (6) then add n_i to L, otherwise skip to the next neighbor.
- After checking all the points in L, put L into C and reset L.
- If the number of points in the cluster are in the range (Min_clust, Max_clust), it will save the cluster. Among the cluster that meet the conditions of color, normal, size, the cluster will be selected if it has the highest height as shown in Figure 4.

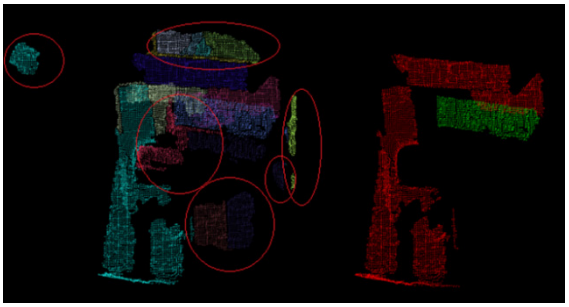


Figure 4: Image before and after extract cluster.

In Figure 4, the area circled to the left is the cluster having a smaller number of points than the Min_clust. Therefore, they were removed. Finally, the result is shown on the right with the green cluster which has the highest height. The robot will use this cluster to calculate the position and direction.

3. Manipulator and camera calibration

Because it is hard to determine the position of the camera coordinate that corresponds to the coordinate of the manipulator; therefore, using the method of measuring the distance from the manipulator to the camera will not give accurate results. Therefore, this research uses intermediate positions with known coordinates in both camera coordinates and manipulator to find transfer matrices. Suppose we have any point M in 3D real space, the coordinates of this point M in the camera coordinate system is denoted M_C and in the robot coordinate system is M_R, then we have the formula to transfer coordinates from the camera coordinate system to the robot coordinate system as follows:

$$M_R = RM_C + T \tag{7}$$

Where R is a rotation matrix and T is the translation matrix containing the translational vector from the robot origin to camera origin. If $M_R = \begin{bmatrix} X & Y & Z \end{bmatrix}^T$ and $M_C = \begin{bmatrix} x & y & z \end{bmatrix}^T$ then (7) becomes:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix} + T$$

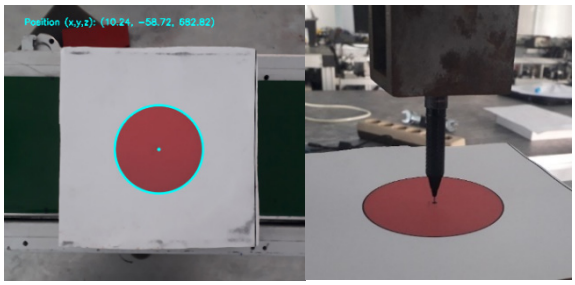
$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \tag{8}$$

From (8), there are 12 variables to find, so it is necessary to have 12 equations corresponding to 12 variables, each point coordinates can only provide 3 equations so we need to take 4 points to be able to find two matrices R and T. After getting the necessary points, the system is solved as linear equation

$$A = BX \Leftrightarrow X = B\text{-}1A \tag{9}$$

With Kinect cameras, both 2D and depth images have a resolution of 640x480 and each pixel in 2D images corresponds to a point obtained from the depth camera. Kinect's true coordinate

system is defined to coincide with the color camera's coordinate system, so when the pixel coordinates of a point in an RGB image are obtained, the x, y, and z values of that point can be retrieved in the camera coordinates. So, the method to get coordinates is as follows: Use the Hough Circle algorithm to determine the center of a circle. From the pixel coordinates of the center of the circle in the color image, the center of the circle in the camera coordinate system is retrieved as shown in Figure 5a. Then, use the Teach-pendant to control the manipulator's head to the center of the circle to get the coordinates of this point in the robotic coordinate system. The work head needs to be small enough to increase accuracy Figure 5b. The data table of this process and error checking are shown in Table 1 and Table 2



(a) (b)

Figure 5: (a) Coordinate from camera, (b) coordinate from robot

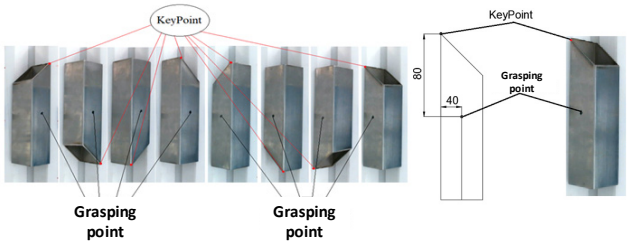
Table 1: Coordinates from camera and robot

Point	Camera Coordinate (mm)	Robot Coordinate (mm)
1	-11.238; -95.1482; 798	298.218; -339.686; 0.0484
2	64.4799; -101.508; 680	289.673;-419.271; 120.996
3	-35.7193; -88.948; 746	307.759 -314.414 50.322
4	22.9576; 28.523; 741	427.611; -374.943; 51.3479

Table 2: Calibration error checking table

Camera coordinate (mm)	Robot coordinate (mm)	Absolute error (mm)		
		X	Y	Z
48.2097; 15.5754; 790	414.531; -399.18; 1.72738	0.08498	1.2113	1.3723
-5.20496; 33.4604; 792	432.75; -344.522; 1.66852	0.7737	0.7121	2.4602
-29.8731; -62.924; 677	330.144; -319.86; 119.984	1.9182	2.0250	0.3202
14.6488; -16.0439; 743	378.403;-367.813; 51.472	2.4365	1.1975	0.2991
Mean error (mm)		1.3033	1.2865	1.1130

After identifying separated object, object is grasped and placed on the V-groove and there are eight case of placing as shown in Figure 6a. Since it is placed on a notched groove, the position of the object just changes along the axis of the groove. To determine the position O for grasping, it is only necessary to determine the coordinates of one point called the key point on the object as shown in Figure 6b and then using the dimension of the object to calculate the O position.



(a) (b)

Figure 6: (a) eight cases placing on V-groove, (b) Key point and grasping point.

To determine the key point, Firstly, the template matching algorithm are performed then the Hough Line detection applied to find the line equation. Then based on line equation, the intersection point coordinate of each two lines are calculated. By combining each case result of template matching and calculated points, the appropriate coordinate of the key point is determined.

4. Experiment

Working space for experiment includes one RGB-D camera fixed above the working space, a monocular camera 2D attached at the end effector of Manipulator, the 6DOF Nachi Manipulator and pneumatic suction cup with maximum 1 kg payload and the V-groove for fine tuning location of object as shown in Figure 7. The 3D Camera is the Microsoft Kinect v1 with a resolution 640x480 combined with 2D camera Logitech. Both of them are parallel fixed with table surface having rectangular box objects. The specification of computer is: Laptop, RAM 16GB, Core I7, CPU 3.1GHZ. Six objects, representing for six random cases, are aluminium type and randomly stacked inside the box. They will be picked up and put at registered position.

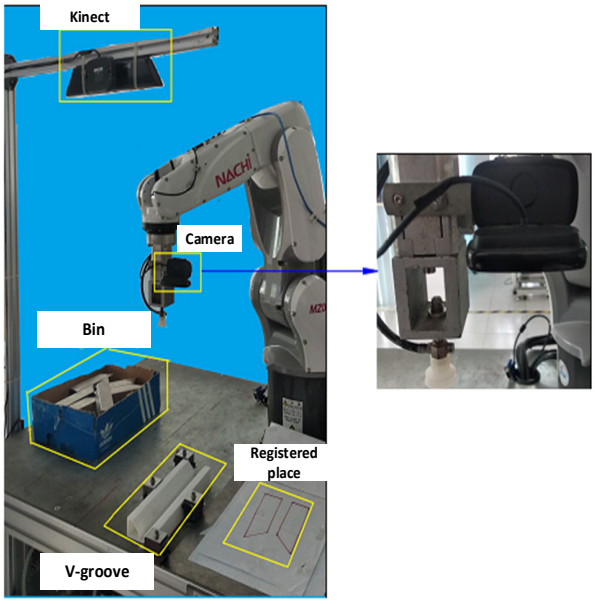


Figure 7: Experiment apparatus

By using 2D and 3D image processing discussed above, the positions of the objects are calculated as shown in Fig.8. The cluster having the highest height, green colour, is chosen to be grasped

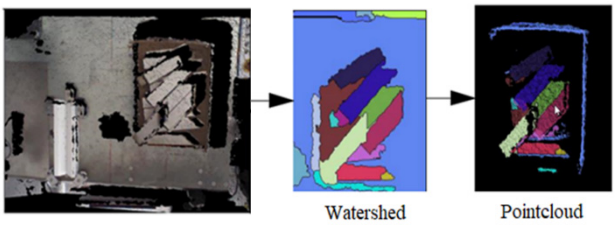


Figure 8: Image processing procedure

After having the coordinates and directions of object in the bin, the manipulator will perform the grasping task and then place it on the V-groove for the fine-tuning step Figure 9. After placing the object in the V-groove, the robot will use the second monocular camera to take the image to find the location and side of object to be picked Figure 10. Then the robot will put it to the registered position as shown in Figure 11.

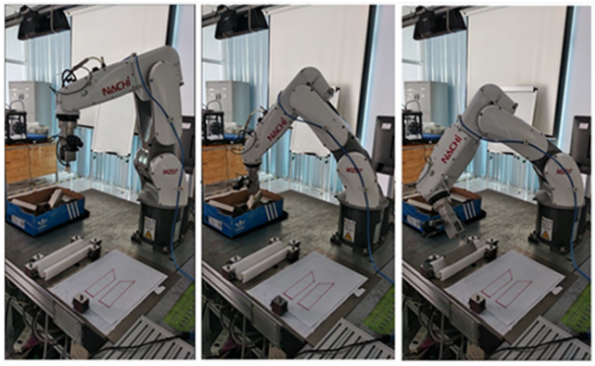


Figure 9: Process of grasping the object in the bin and putting it on V-groove

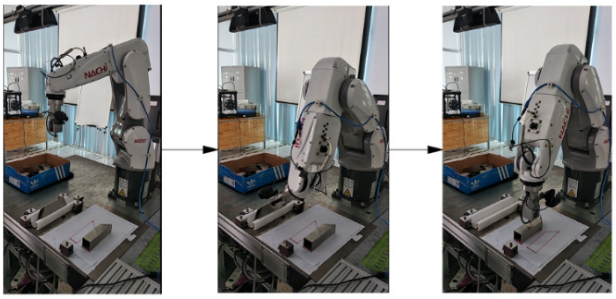


Figure 10: Process of grasping object on V-groove and putting it at registered

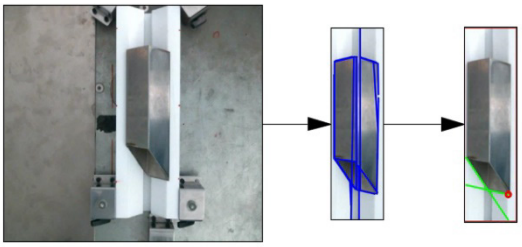


Figure. 11: Finding the key point of the object

After finish grasping task, the angular error, position error of the objects and the desired position are compared as shown in Figure 12. The real position of objects is obtained by reading the coordinate value on teach-pendant when moving robot manually. From that position the error of angle and position are calculated. The data are taken many times to calculate the mean error of the system as shown in following Table 3. The processing time is also calculated Figure 13

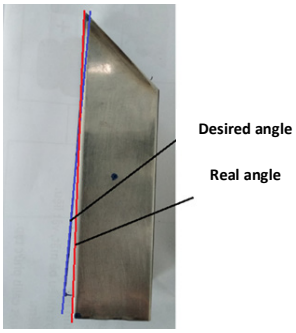


Figure 12: Desired position and real position

Table 3: Error position and angle calculation

No	Position Error (mm)			Angle error (degree)
	X	Y	Z	
1	1.00	1.73	0.08	1.1243
2	1.25	1.62	0.11	1.6973
3	1.01	1.52	0.19	1.0981
4	1.18	1.91	0.36	1.8352
5	0.50	1.15	0.34	1.9559
6	1.61	1.27	0.52	1.0600
Mean Error	1.0917	1.533	0.2667	1.4618

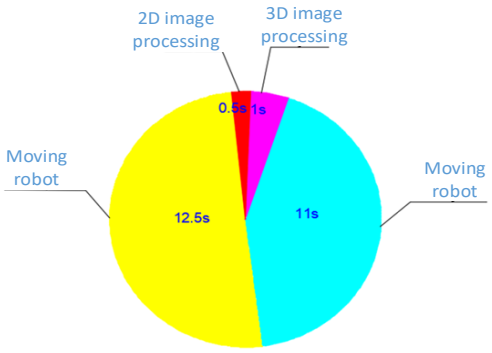


Figure 13: Processing time of system

From Table 3 it can be seen that the error of angle and position compared to the desired position is small and acceptable. The time to move the robot takes up most of the time by 23 seconds. Time to process images and data transmission is small about 1.5 seconds. The total time of implementation is 25 seconds, which is less than 30s. The image processing time is just about 6% of total time. To improve speed of system, the speed of robot could be increased to 100%. But for safety reasons it should only be done at 50% speed.

5. Conclusion

The paper proposed an effective method for the bin-picking problem. The image processing time is small and the accuracy is high. The method consists of two steps: the first step is to use the

3D camera and to combine 2D and 3D image processing for object segmentation and the second step is to use 2D camera for fining the grasping task. The effectiveness of the proposed approach was confirmed by a series of experimental results. The proposed method is also easy for installation and maintenance and applicable for assembly task in real industry. The comparing performance of proposed system and existing ones which are discussed in the introduction section are summarized in below Table 4.

Table 4: Comparing performance of Systems

	Proposed System	Existing system
Complexity	Not complicated	Very complicated
Time executing	Fast	Time consuming
Quality	Good and stable	Good and stable
Quantity	Mass product	Mass product
Setup	Easy	Complicated

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research was supported by Saigon Hi-Tech Park Training Center and Vietnam National University, HoChiMinh City

References

[1] F. Casado, Y.L. Lapido, D.P. Losada, A.S. Alonso, " Pose estimation and object tracking using 2D image," In 27th International Conference on Flexible Automation and Intelligent Manufacturing, Modena, Italia, 63-71, 27-30 June, 2017, doi: 10.1016/j.promfg.2017.07.134.

[2] I. Wesis, M. Ray, ""Model-based recognition of 3D objects from single images," IEEE Transactions on Pattern Analysis and Machine Intelligence, **23**(2), 116-128, 2001, doi: 10.1109/34.908963.

[3] K. Ban, F. Warashina, I. Kanno, H. Kumiya, " Industrial intelligent robot," FANUC Tech. Rev., **16**(2), 29-34, 2003.

[4] J. Xu, " Rapid 3D surface profile measurement of industrial parts using two-level structured light patterns," Optics and Lasers in Engineering, **49**(7), 907-914, 2011, doi: 10.1016/j.optlaseng.2011.02.010.

[5] W.C. Chang, C.H. Wu, " Eye-in-hand vision-based robotic bin-picking with active laser project," International Journal of Advanced Manufacturing Technology, 85, 2873–2885, 2016, doi: 10.1007/s00170-015-8120-0

[6] K. Rahardja, A. Kosaka, "Vision-based binpicking: recognition and localization of multiple complex objects using simple visual cues," In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 100-107, 1996, doi: 10.1109/IROS.1996.569005.

[7] M. Berger, G. Bachler, S. Scherer, " Vision guided bin picking and mounting in a flexible assembly cell," In Proeeding of the 13th International Conference on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems, 109-118, 2000, doi: doi.org/10.1007/3-540-45049-1_14

[8] M.Y. Liu, O. Tuzel, A. Veeraraghavan, Y. Taguchi, T.K. Marks, R. Chellappa, "Fast object localization and pose estimation in heavy clutter for robotic bin picking," The International Journal of Robotics Research, **31**(8), 951–973, 2012, doi: 10.1177/0278364911436018

[9] A. Adan, S.V. Andrés, P. Merchán, R. Heradio, "Direction Kernels: Using a simplified 3D model representation for grasping," Machine Vision and Applications, **24**(1), 351-370. 2013, doi: 10.1007/s00138-011-0351-y

[10] R. D. Schraft, T. Ledermann, "Intelligent picking of chaotically stored objects," Assembly Automation, **23**(1), 38-42, 2003, doi: 10.1108/01445150310460079

[11] K. Bousmalis, A. Irpan, P. Wohlhart, Y. Bai, M. Kelcey, M. Kalakrishnan, L. Downs, J. Ibarz, P. Pastor, K. Konolige, S. Levine, V. Vanhoucke, "Using simulation and domain adaptation to improve efficiency of deep robotic grasping," In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), 4243 – 4250, 2018, Australia, doi: 10.1109/ICRA.2018.8460875.

[12] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J.A. Ojea, K. Goldberg," Dex-net 2.0: deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," In Robotics: Science and Systems, 245-251, 2017, doi: 10.15607/RSS.2017.XIII.058

[13] L. Tuan-Tang, L. and Chyi-Yeu, "Bin-Picking for Planar Objects Based on a Deep Learning Network: A Case Study of USB Packs," Sensors, **19**(16), pp. 3602-3618, 2019, doi: 10.3390/s19163602

[14] J. Chen, T. Fujinami, E. Li, "Deep Bin Picking with Reinforcement Learning," In Proceedings of the 35th International Conference on Machine Learning, 1-8, 2018, Sweden.

[15] S. K. Anton, V.S. Ilia, "An Overview of Watershed Algorithm Implementations in Open Source Librarie," Journal of Imaging, **4**(10), 123-138, 2018 doi: 10.3390/jimaging4100123.

[16] L.D. Hanh, L.M. Duc, "Planar Object Recognition For Bin Picking Application," In Proceedings of 5th NAFOSTED Conference on Information and Computer Science (NICS), 213-218, 2018, doi: 10.1109/NICS.2018.8606884.

[17] Conditional Euclidean Clustering. Available online: http://pointclouds.org/documentation/tutorials/conditional_euclidean_clustering.php(accessed on 12 January 2020).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1250-1257 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Crystallinity and Hardness Enhancement of Polypropylene using Atmospheric Pressure Plasma Discharge Treatment

Oscar Xosocotla[1,2,*], Horacio Martinez[1], Bernardo Campillo[2]

[1]*Laboratorio de Plasmas Atmosféricos y Espectroscopia Raman, Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, 62220, México*

[2]*Facultad de Química, Universidad Nacional Autónoma de México, Ciudad de México, 05000, México*

ARTICLE INFO

Article history:
Received: 10 October, 2020
Accepted: 06 December, 2020
Online: 16 December, 2020

Keywords:
Polypropylene
Crystallinity
Hardness
Atmospheric pressure plasma

ABSTRACT

Atmospheric pressure plasma was used to treat polypropylene (PP) surfaces. Optical emission spectroscopy (OES) was used to determine the chemical species formed in the plasma as well as the electron temperature and density of the plasma. The flux of species (O and OH) produced during the plasma treatment interact with the polymer surface creating polar groups on the PP surface, which were evaluated by X-ray, Raman, and Attenuated total reflectance-Fourier transform infrared spectroscopy (ATR-FTIR) analyses. Moreover, the thermal properties of PP were investigated by thermogravimetric analysis (TGA) and differential scanning calorimetry (DSC) analysis. The plasma treatment increased the crystallite size and microhardness. This enhanced effect was produced from high concentrations of O and OH that induced functional polar groups containing C–O, C=O, and C=O–OH bonds appended to the PP surface. The polar groups produced by the air plasma can be attributed primarily to the oxygen radicals in the air plasma hitting the PP surface and to heat-induced oxidation rather than the incorporation of oxygen radicals or UV-induced oxidation from the plasma. The increase in hardness is attributed to the introduction of carbonyl and hydroxyl groups, cross-linking, annealing effects, and chemical etching.

1. Introduction

Atmospheric pressure plasma discharge treatment has become important because of its relatively simple process, low cost, easy operation, and applicability to vacuum-incompatible materials and processes compared with the low-pressure plasma process [1]. Typically, two categories of plasmas exist based on their gas temperatures: thermal and nonthermal (cold) plasmas [2, 3, 4]. In thermal plasmas, electron temperature is very close to those of heavy particles (both charged and neutral), and therefore, it cannot be used for the surface modification of heat-sensitive materials. However, nonthermal plasma electrons have a temperature that is 2–3 orders of magnitude higher than the temperature corresponding to the random motion of the background gas molecules. As a result of the very low heat capacity and density of electrons, plasma temperature is almost the same as room temperature. Subsequently, the ions remain rather cold, and a nonthermal plasma does not produce any thermal damage to the treated surfaces [2-4]. These characteristics are particularly advantageous for the treatment of heat-sensitive and fragile materials, such as polymers [5, 6].

Surface modification can be applied either directly or indirectly. Direct surface modification methods are used to modify an inert surface with free radicals for specific applications, whereas indirect surface modification methods are based on polymer grafting [5, 6]. The efficiency of plasma treatment methods depends on surface type, the gas phase of the plasma, and treatment conditions such as discharge type, pressure, frequency, discharge voltage and current, and treatment time. To modify surface properties without any change in the bulk material, it is essential to use successful surface modification methods such as atmospheric pressure plasma [5, 6].

Surface modification of polypropylene (PP) has been extensively studied to improve its properties, such as wettability (surface free energy), hydrophilicity, biocompatibility, and adhesion, using different atmospheric pressure plasma reactors [5,

*Corresponding Author: Oscar Xosocotla, Email: hm@icf.unam.mx

7-11]. However, only two studies have shown PP hardness improvement: the first study [12] involved beta irradiation at several doses, showing substantial changes in PP structure and mechanical and micromechanical properties, and the second study [13] utilized atmospheric pressure plasma treatment by diffuse coplanar surface barrier discharge using ambient air as the working gas. They observed no significant changes in the mechanical properties of the films after plasma treatment. Therefore, the purpose of this study was to characterize and test the impact of the commercial plasma device on the PP surface. Therefore, optical emission spectroscopy (OES) was used to investigate chemical species that cause plasma surface modification, electron temperature (T_e) and density (n_e) were used to characterize the plasma, differential scanning calorimetry (DSC) was used to determine the melting points of materials, thermogravimetric analysis (TGA) was used to evaluate the degradation of materials subjected to plasma treatment, X-ray diffraction (XRD) was employed to analyze crystallite orientation and size, Raman spectroscopy and Attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy were performed to investigate the conformations of polymer chains, and atomic force microscopy (AFM) with a micro-durometer was used to study the morphology of the surfaces and determine the microhardness of PP.

2. Experimental

2.1. Plasma Source

A schematic illustration of the APC 500 plasma torch source (Diener Electronics, Germany) used for surface modification is shown in Figure 1. The plasma source [14] was operated at a voltage of 20 kV, discharge current of 25 mA, frequency of 40 kHz, and gas pressure of 1500 torr. The plasma gliding discharge was generated by two electrodes. The discharge extended from the electrode region because of air flow, and the active plasma covered an area of ~10 cm^2. The samples were placed in the most homogenous region, which was at the center of the gliding plasma. OES was used for optical characterization of the plasma. For OES measurements, an Ocean Optics spectrometer (model HR2000CG-UV-NIR) was used in the spectral range 200–1100 nm, and the data were obtained with a step of 0.35 nm and an optical resolution of 1.0 nm (FWHM).

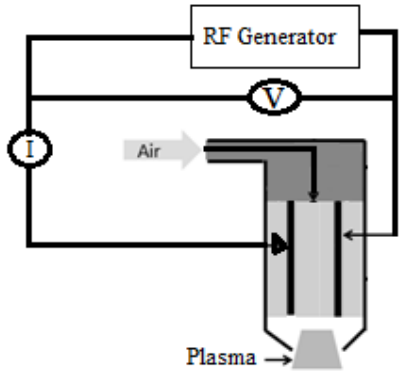


Figure 1: Diagram of the radio frequency atmospheric plasma generator

2.2. Materials

A commercially available isotactic (Lyondell Basell) PP cylinder with a density of 0.91 grams per cubic centimeter (g cc^{-1}) was cut into 1.5 × 1.5 × 0.5 cm^3 pieces. One side of each PP sample was polished to achieve uniform roughness using silicon carbide sandpaper with a grain size of 1,500. The samples were then washed in an ultrasonic bath for 20 min. To remove the residual liquid, the samples were placed in an oven at 30 °C for 30 min.

2.3. X Ray analysis

The XRD patterns of PP samples were recorded by a Rigaku D MAX-2000 diffractometer using Cu K_α (λ = 0.154 nm) radiation with 2θ in the range 10°–20° using a step size of 0.06° and step time of 10 s. The diffraction patterns of untreated and treated PP samples consisted of well-defined peaks. The high-intensity peaks were chosen to estimate the crystalline size and lattice strain using Scherrer's formula and Williamson–Hall (W–H) model [15, 16].

2.4. Raman analysis

Raman spectra were obtained using a SENTERRA II spectrometer (Bruker). The characteristic peaks of the crystalline and amorphous phases in the spectra of treated and untreated PP samples were then evaluated. The Raman spectrometer was equipped with a 785-nm laser and was operated at a power of 100 mW. The integration time was 1,000 ms.

2.5. DSC and TGA

DSC and TGA were performed to analyze the thermal properties of the samples, which were placed in aluminum specimen holders. To correct for their thermal histories, the treated PP samples were heated from 10 °C to 250 °C at a rate of 10 °C/min in a nitrogen atmosphere.

2.6. ATR FTIR analysis

The chemical compositions of the PP surfaces were analyzed before and after plasma treatment using a Spectrum 400 ATR-FTIR spectrometer (PerkinElmer). The vibrational modes appearing in the ATR-FTIR spectra were consistent with those reported in the literature [17-20].

2.7. Roughness

A NX10 atomic force microscope (Park) was used to quantitatively analyze the surface roughness of the PP samples due to plasma treatment [21-23].

2.8. Microhardness

A micro-durometer (Matsuzawa) was used to measure PP microhardness under a 300 g load. The penetration time was set to 20 s, and five indentations were made on each PP surface to obtain an average value [12, 24].

3. Results and analysis

3.1. Plasma characterization

The efficiency of plasma treatment depends on the surface type and treatment conditions. Therefore, the plasma discharge

was produced by applying a voltage (V) of 10 kV, and the electrodes were separated by a distance (d) of 2 cm. The resulting electric field strength (E) was equal to V/d. E was calculated to be 5.0×10^3 V cm^{-1}, and E/P was equal to 6.58 V cm^{-1} torr^{-1}.

The electron drift velocity (v_d) and E/N, where N is the number density, were obtained from the literature [25], considering the experimental conditions used. We obtained a v_d of 3.2×10^6 cm s^{-1} and E/N of 0.19 Td (1 $Td = 10^{-17}$V cm^2). Thus, N is 2.63×10^{19} cm^{-3}.

The electron density (n_e) was estimated from the measured current density (J) [26]:

$$n_e = \frac{J}{e v_d}. \quad (1)$$

The electron density was found to be $(1.95 \pm 0.24) \times 10^{12}$ cm^{-3}, and the order of magnitude was consistent with reported values [27, 28]. The electron density, n_e, was assumed to be equal to the ion density n_i in the plasma bulk [29].

3.2. Optical emission spectroscopy (OES)

OES was performed to characterize the plasma discharge. The OES spectrum obtained from the discharge is shown in Figure 2. It is well known that atomic and molecular transitions emit photons with specific wavelengths and energies that correspond to the difference between the upper and lower energy levels involved in the transitions. The OH($A^2\Sigma^+ - X^2\Pi_i$) molecular emission near 309 nm indicated that the plasma was not in equilibrium because of strong transitions and trace amounts of water vapor in the air. It is often difficult to prevent water from affecting the plasma. The peak at 337 nm was due to a $C^3\Pi_u \rightarrow B^3\Pi_g$ transition in N_2 and was the most intense peak in the region of 330–530 nm. Moreover, the first negative system, N_2^+ ($B^2\Sigma_u^+ \rightarrow X^2\Sigma_g^+$) generating a 0–0 band between 388.5 and 391.5 nm was observed together with two low-intensity peaks at 777.19 and 844.71 nm, which were attributed to O ($^5S^0 - {}^5P$) and O ($^3S^0 - {}^3P$) transitions, respectively.

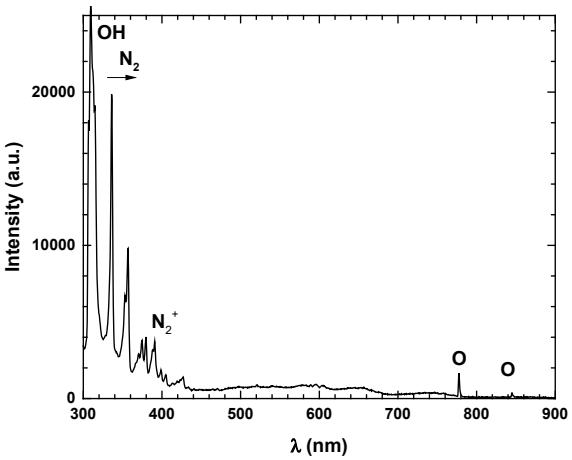


Figure 2: Optical emission spectroscopy (OES) spectrum.

OES is often used to determine the energy of electrons in the gas phase [30-33]. Thus, it can be applied to plasmas that are not in equilibrium at atmospheric pressure, which is often the case during the analysis of complex mixtures. The optical characterization of the discharge was based on the ratios of the O atomic line intensities in the emission spectrum. The electron energy was found to be (0.36 ± 0.04) eV, which was consistent with the results of previous atmospheric plasma studies [34-36]. The electron temperature (Te) can be several orders of magnitude higher than the temperature of the ions (Ti) or the neutral gas (Tg). Therefore, the heavy particles stay close to ambient temperature (27^0C) [29, 37, 38].

The species observed by (OES) in the plasma are formed due to electrons are accelerated by the electric field and can induce electron impact excitation, ionization and dissociation of the primary gas phase (humid-air containing 79% N_2/20% O_2/ 1%) reactions, these are:

$$e + N_2 \rightarrow N_2^* + e, \quad (2)$$
$$e + O_2 \rightarrow O_2^* + e, \quad (3)$$
$$e + N_2 \rightarrow N_2^+ + e + e, \quad (4)$$
$$e + O_2 \rightarrow O_2^+ + e + e, \quad (5)$$
$$e + N_2 \rightarrow N + N + e, \quad (6)$$

The formation in OH• radicals and H species during the plasma treatment may be attributed to the fragmentation/decomposition of H_2O molecules:

$$e + H2O \rightarrow H + OH + e \quad (7)$$

or

$$e + O_2 \rightarrow 2O + e \quad (8)$$
$$O + H_2O \rightarrow OH + OH \quad (9)$$

whereas the formation of atomic oxygen may be due to the dissociation of oxygen molecules by reaction (8).

3.3. X-Ray Diffraction analysis

The XRD patterns of the treated and untreated PP samples are shown in Figure 3. The peak characteristics of the α phase were observed at (2θ) 14.02° (110), 16.8° (040), and 18.44° (130), and between 20.80^0 and 22.40^0 for (111), (131), and (041). The same peaks appeared in the patterns of the treated PP samples, indicating that the plasma had no effect on the crystalline structure of PP. However, the peak intensity of the untreated sample at 16.8° differs from those of the treated samples.

The XRD patterns were used to calculate the crystallite size and orientation of planes in each sample.

The size of crystallites (D) in each PP sample based on the (110) and (040) planes was calculated using the Scherrer equation [16]:

$$D = \frac{0.9\lambda}{\beta \cos\theta}, \quad (10)$$

where λ is the wavelength, β is the full width at half maximum (FWHM), and θ is the maximum angle of dispersion.

The crystallite size (D) was compared with the values obtained from the Williamson-Hall (W–H) model, where the total peak broadening is represented by the sum of the contributions from crystallite size (D) and strain (ε) in the material. Assuming that the strain present in the material is uniform and considering the isotropic nature of the crystal, the W–H equation [15, 16] is given by

$$\beta \cos(\theta) = \frac{0.9\lambda}{D} + 4\varepsilon \sin(\theta). \quad (11)$$

The estimated values of the strain and crystallite size are given in Table 1.

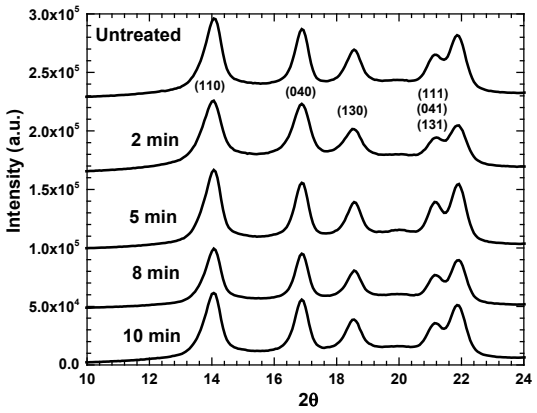


Figure 3: XRD patterns of PP samples subjected to plasma treatment for various periods.

Table 1: Geometric parameters for PP as a function of treatment duration.

Treatment duration	Scherrer´method	Williamson-Hall method	
min	D (nm)	D (nm)	ε (10^{-7})
0	11.7	10.7	4.4
2	10.5	8.0	9.9
5	12.4	10.5	5.6
8	12.9	11.9	3.4
10	12.5	11.8	2.8

We observed that increasing the treatment duration increases the crystal size, except for 2 min of treatment, which may be due to the rearrangement of the orientation parameters of the crystal planes due to the introduction of polar groups on the PP surface by the plasma treatment. In addition, the strain at 2 min of plasma treatment increases almost 100% and then decreases as the treatment duration increases. That behavior may be attributed to dislocations that increase the volume of grain boundaries (increase in the crystal size) due to the plasma treatment and incorporation of the polar groups on the surface of PP.

To evaluate the orientation parameters of the (040) crystal plane, A_{110} (040), and (110) crystal plane, C_{040} (110), we used Zipper's equation, which takes into account the intensities of the peaks observed [39-41]:

$$A_{110}(040) = \frac{I_{110}}{I_{110}+I_{111}+I_{131+041}}, \quad (12)$$

$$C_{040}(110) = \frac{I_{040}}{I_{110}+I_{040}+I_{130}}, \quad (13)$$

where I_{110}, I_{111}, I_{104}, I_{040}, I_{130}, and $I_{131+041}$ correspond to the peak intensity of the respective crystal plane.

Figure 4a shows the orientation indices of A_{110} (040) and C_{040} (110) crystal planes as a function of treatment duration. The orientation indices appear to increase as the duration of plasma treatment increases due to the surface functionalization by the plasma treatment. Changes in crystallite sizes in the (110) and (040) planes during plasma treatment are illustrated in Figure 4b. Crystallites of the samples became smaller during the first 2 min of treatment. The crystallite size increased as the duration of plasma treatment increased. It can be explained because the plasma treatment acts as a nucleating agent on the PP surface, which promotes the formation of perfect crystals [12, 39-41].

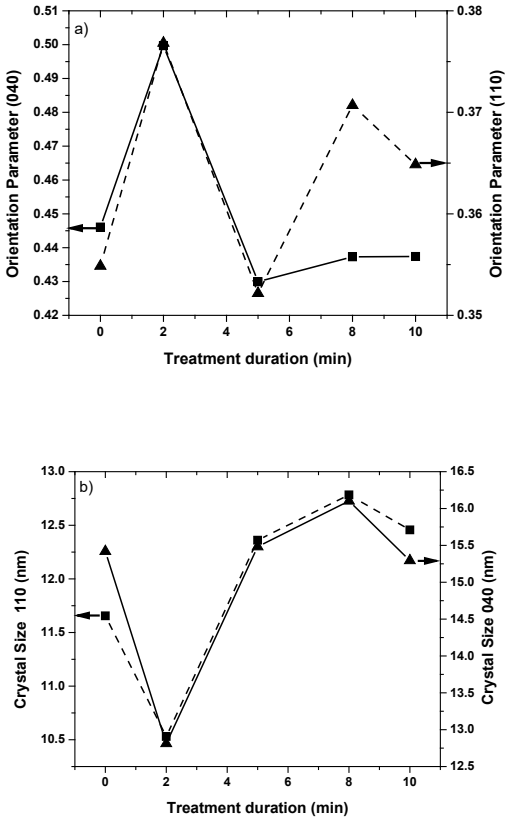


Figure 4: Crystallographic parameters of the treated PP samples. 4.a) Orientation indices of the crystal planes and 4.b) crystallite size.

3.4. Raman analysis

Raman spectra from 809 to 841 cm^{-1} are shown in Figure 5. The peak attributable to the crystalline phases in the PP samples was observed near 809 cm^{-1}, and the peak near 841 cm^{-1} was ascribed to the phases that contained helical chains with isomeric defects [17, 25, 42, 43].

The intensities of the observed peaks (I_{809} and I_{841}) were used to evaluate dispersion changes that were not related to crystallinity. The sum of the intensities ($\bar{I} = I_{809} + I_{841}$) was

independent of crystallinity within the range of measurement. The crystallinity of the treated PP samples was estimated using equations 14 and 15. The proportions of phases with defects and crystalline are denoted by x^b and x^c, respectively [40, 41]:

$$x^b = \frac{I_{841}}{\bar{I}}, \tag{14}$$

$$x^c = \frac{I_{808}}{\bar{I}}. \tag{15}$$

Variations in the proportions of phases with defects (x^b) and crystalline phases (x^c) in the samples are shown in Figure 6. Plasma treatment for 10 min increased the crystallinity of the PP samples, although the x^c values of samples treated for 2, 5, or 8 min slightly decreased. Variations in the proportions of phases with defects indicated that phase defects were associated with crystallinity. Regular chains in the samples subjected to plasma treatment for the greatest amounts of time reorganized to generate helices and local order. The amorphous phases contained irregular atactic chains that could not be reorganized into helices because of the random orientations of their side groups. Those changes are correlated with the increase in crystallite size and orientation parameters due to the plasma treatment and the functionalization of the PP surface [12, 21, 22, 25, 39-43].

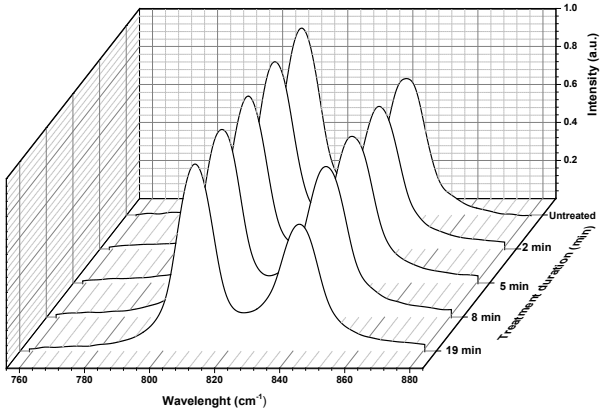


Figure 5: Raman spectra as a function of treatment duration.

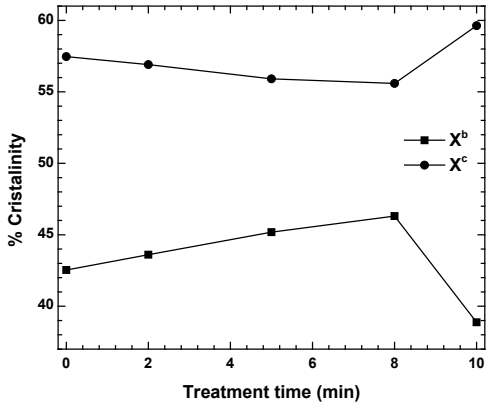


Figure 6: % crystallinity as a function of treated time.

3.5. DSC and TGA analysis

Changes in the thermal properties of the treated PP samples were Changes in the thermal properties of the treated PP samples were analyzed via DSC and TGA. Degradation is shown as a function of temperature in Figure 7. Plasma treatment did not affect mass loss under a nitrogen atmosphere because the percentages of mass lost by the untreated and treated samples did not differ. The first derivative of each mass loss curve indicated the largest change in mass. The highest rates of change in the weight loss curves of PP samples treated for 2 or 5 min occurred at 404 °C. In the weight loss curves of the untreated samples and samples treated for 8 or 10 min, the highest rates of change were observed at 416 °C (Figure 8).

Structural and compositional transformations in a polymer due to plasma treatment can be analyzed by evaluating the changes in its properties at different temperatures [39]. The DSC thermograms of the treated and untreated samples are shown in Figure 9. The untreated sample and samples treated for 8 or 10 min melted at 165 °C, whereas the samples treated for 2 or 5 min melted at 167 °C. Therefore, the melting points of the treated and untreated samples did not differ significantly [24].

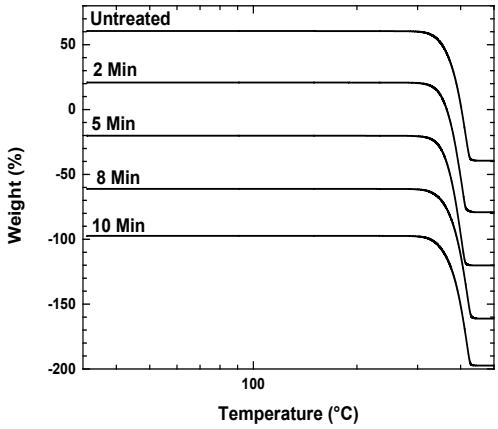


Figure 7: TGA thermograms of the treated and untreated PP samples, the curves have been shifted vertically for better comparison.

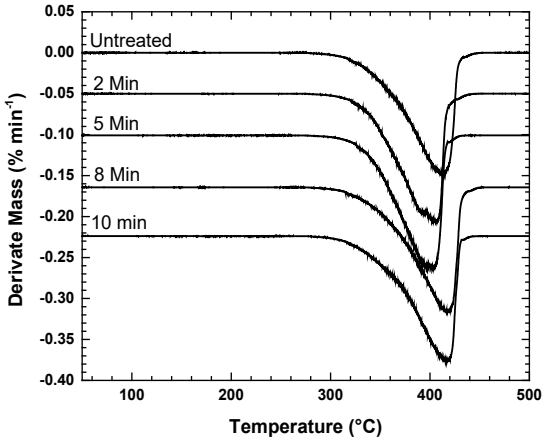


Figure 8: Weight loss curves of the treated and untreated PP samples, the curves have been shifted vertically for better comparison.

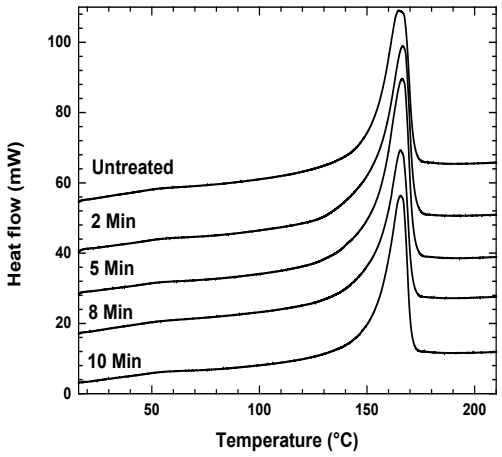


Figure 9: DSC thermograms of the treated and untreated PP samples, the curves have been shifted vertically for better comparison.

3.6. ATR-FTIR analysis

The FTIR spectra of the untreated and treated PP samples are shown in Figure 10. Chemical differences in the surface properties of the plasma-treated PP samples can be observed in Figure 10, indicating changes in their surface chemical compositions. The spectra confirmed the introduction of polar groups into all treated samples. Vibrational bands attributed to carbonyl (C=O) groups appeared from 1,500 to 1,700 cm^{-1} [18-21]. Moreover, the carbonyl (C=O) band was observed at 1,720 cm^{-1} in the spectra of the samples treated for 5 and 10 min but did not appear in the spectra of samples treated for 2 or 8 min. It means that the C content decreases, and the O content increases on the surface because of the plasma treatment. Additionally, the peak around 3400 cm^{-1} corresponding to the hydroxyl (–OH) group was produced by the plasma treatment. The relative intensities of all absorption peaks observed by FTIR were enhanced, indicating the increases in the carbonyl and hydroxyl groups on the surface [44]. In the present treatment, large amounts of O and OH polar groups were appended to the surface of PP because of the plasma treatment, improving the crystallinity and microhardness of PP.

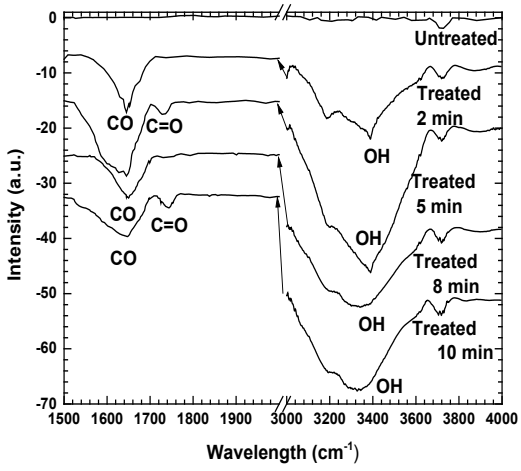


Figure 10: ATR-FTIR spectra as a function of treatment duration.

3.7. Roughness

The AFM images of the untreated and treated samples at 5 min are shown in Figures 11 and 12, respectively. The morphologies of the samples treated by plasmas were smoother than those of the untreated sample. Those figures were used to study the changes in the morphology of the PP samples, and the RMS roughness found for the untreated sample was 0.554 ± 0.064 pm, whereas for the samples treated at 2, 5, 8, and 10 min, the surface roughness values were 0.153 ± 0.031, 0.136 ± 0.013, 0.111 ± 0.024, and 0.090 ± 0.008 pm, respectively. The sample treated at 8 min shows the lowest value of RMS roughness. Those morphological variations may be due to the removal of the top few monolayers of the PP samples produced by the bombardment of plasma particles on the PP surface [13, 18], reducing the surface roughness and improving the microhardness of PP. Although their surfaces were smoothed, the ATR-FTIR spectra of the treated samples indicated that the introduction of polar groups was possible [13, 18-20, 22]

3.8. Microhardness

The microhardness values of the treated and untreated PP samples are plotted in Figure 13. The hardness values of the treated samples were higher than those of the untreated sample, and the hardness of the sample treated for 10 min increased by nearly 56%. An increase in hardness is attributed to annealing effects and chemical etching produced by the plasma treatment [45]. Also, the hardness increase may be related to either the hardening of interlamellar regions due to the polar group penetration or the hardening of the crystals themselves caused by the surface attachment of polar groups, as observed by Raman and ATR-FTIR. Plausible changes on the PP surface caused by the plasma treatment include cross-linking and oxidation. Cross-linking is due to the formation of tertiary carbon radicals. Tertiary hydrogen abstraction generates carbon radicals, which react with radicals in the nearest molecular chain. Oxidation primarily affects secondary carbon atoms. The introduction of oxygen-bearing groups, including –C–OH, –CO–OH, and –C=O moieties, occurs when the bonds between hydrogen and secondary carbon atoms on the PP surface are broken during the plasma treatment [13, 22-23].

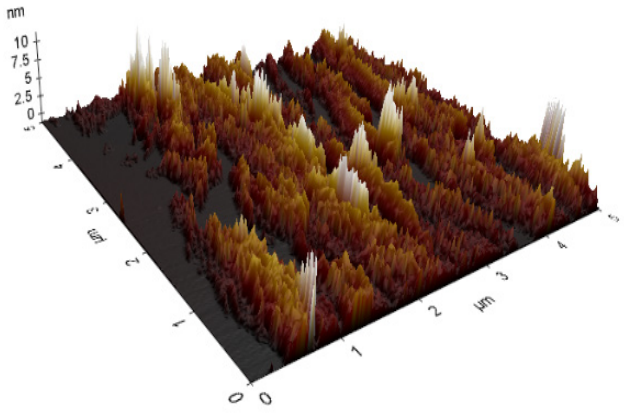


Figure 11 Roughness profiles of the untreated PP sample.

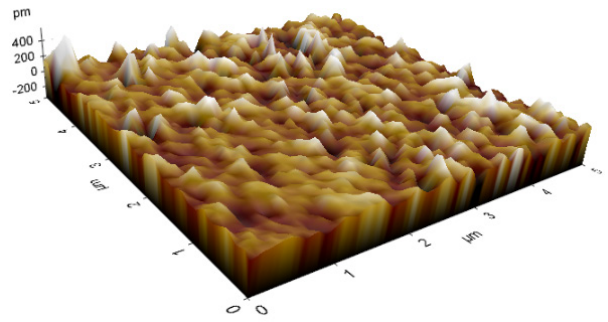


Figure 12: Roughness profiles of PP treated for 5 min.

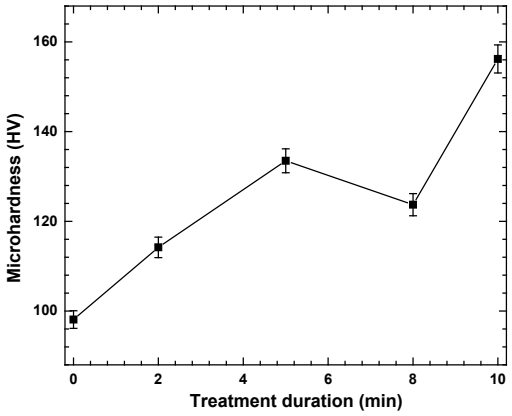


Figure 13: Microhardness values of the samples before and after plasma treatment

4. Conclusions

A commercial plasma device was used for the surface modification of PP. Using OES, we determined the chemical species formed in the plasma. High concentrations of OH and O radicals were observed in the plasma, and they led to the formation of functional polar groups, C–O, C=O, and C–O–OH bonds, on the PP surface. The plasma treatment increased the crystallinity size and microhardness. This enhanced effect is attributed to the formation of polar groups produced by the air plasma due to the oxygen radicals in the air plasma hitting the PP surface. The formation of polar groups in the PP surface was corroborated by Raman and ATR-FTIR spectroscopies. DSC and TGA indicated that plasma treatment had no effect on the thermal properties of PP. AFM analysis showed that plasma treatment for extended periods reduced the surface roughness of PP. Finally, the increase in PP hardness is attributed to the introduction of carbonyl and hydroxyl groups by the plasma treatment, cross-linking, annealing effects, and chemical etching produced by the plasma treatment.

Acknowledgment

The authors thank E. Vazquez, H. H. Hinojosa, F. Castillo, and O. Flores for their technical assistance. This work was partially supported by DGAPA project IN-105519 and CONACyT projects 225991 and 268644.

References

[1] E.E. Kunhardt, "Generation of large-volume, atmospheric-pressure, nonequilibrium plasmas", IEEE Trans. Plasma Sci. **28**, 189-200 (2000), https://doi.org/10.1109/27.842901

[2] P.J. Bruggeman, F. Iza and R. Brandenburg, "Foundations of atmospheric pressure nonequilibrium plasmas", Plasma Sources Sci. Technol. **26**, 123002, 2017, https://doi.org/10.1088/1361-6595/aa97af.

[3] A. Fridman, A. Chirokov and A. Gutsol, "Non-thermal atmospheric pressure Discharges", J. Phys. D: Appl. Phys. **38,** R1–R24, 2005, doi:10.1088/0022-3727/38/2/R01

[4] A. Schutze, J. Y. Jeong, S. E. Babayan, J. Park, G. S. Selwyn and R. F. Hicks, "The atmospheric-pressure plasma jet: a review and comparison to other plasma sources", IEEE Transactions on Plasma Science, **26**(6), 1685-1694, 1998, doi: 10.1109/27.747887.

[5] S.K. Nemani, R. K. Annavarapu, B. Mohammadian, A. Raiyan, J. Heil, Md. A. Haque, A. Abdelaal, and H. Sojoudi, "Surface Modification of Polymers: Methods and Applications", Adv. Mater. Interfaces **5**, 1801247, 2018, doi:10.1002/admi.201801247

[6] V. Palma, M. Cortese, S. Renda, C. Ruocco, M. Martino and E. Meloni, "A Review about the Recent Advances in Selected NonThermal Plasma Assisted Solid–Gas Phase Chemical Processes", Nanomaterials **10**, 1596, 2020, doi:10.3390/nano10081596

[7] M. Kehrera, A. Rottensteiner, W. Hartl, J. Duchoslav, S. Thomas, D. Stifter, "Cold atmospheric pressure plasma treatment for adhesion improvement on polypropylene surfaces", Surface & Coatings Technology in press (2020), https://doi.org/10.1016/j.surfcoat.2020.126389

[8] H.B. Baniya, R. P. Guragain, B. Baniya, and D. P. Subedi, "Cold Atmospheric Pressure Plasma Jet for the Improvement of Wettability of Polypropylene", International Journal of Polymer Science, **2020**, 3860259, 2020, https://doi.org/10.1155/2020/3860259;

[9] R. Janík, M. Kohutiar, M. Pajtášová, A. Dubec, J. Pagáčová and J. Šulcová, "The impact of DCSBD plasma discharge on polypropylene", IOP Conf. Series: Materials Science and Engineering **776,** 012090, 2020, 10.1088/175799X/776/1/012090

[10] F. Darvish, N. M. Sarkari, M. Khani, E. Eslami, B. Shokri, M. Mohseni, M. Ebrahimi, M. Alizadeh, C. Fu Dee, "Direct plasma treatment approach based on non-thermal gliding arc for surface modification of biaxially-oriented polypropylene with post-exposure hydrophilicity improvement and minus aging effects", Applied Surface Science **509**, 144815, 2020, https://doi.org/10.1016/j.apsusc.2019.144815

[11] Y. Akishev, G. Aponin, A. Petryakov and N. Trushkin, "On the composition of reactive species in air plasma jets and their influence on the adhesion of polyurethane foam to low-pressure polyethylene", J. Phys. D: Appl. Phys. **51** 274006, 2018, https://doi.org/10.1088/1361-6463/aac5fd

[12] D. Manas, M. Hribova, M. Manas, M. Ovsik, M. Stanek, D. Samek, "The effect of beta irradiation on morphology and micro hardness of polypropylene thin layers", Thin Solid Films **530**, 49–52, 2013, http://dx.doi.org/10.1016/j.tsf.2012.09.051

[13] A. Oravcova, I. Hudec, "The Influence of Atmospheric Pressure Plasma Treatment on Surface Properties of Polypropylene Films", Acta Chimica Slovaca, **3**(2), 57 - 62, 2010, http://www.acs.chtf.stuba.sk/?id=7&paper=68

[14] http://plasma.de/en/plasmasystems/plasma-apc500.html, accessed on 28.9.2015

[15] W. Burton, O. Kenneth, R. Thomas, I.Y. Chan, "On the estimation of average crystallite size of zeolites from the Scherrer equation: A critical evaluation of its application to zeolites with one-dimensional pore systems", Microporous Mesoporous Mater. **117**, 75-90, 2009, doi:10.1016/j.micromeso.2008.06.010

[16] G.K. Williamson and W. H. Hall, "X-Ray line broadening from filed aluminium and wolfram", Acta Metallurgica **1**, 22-31, 1953, https://doi.org/10.1016/0001-6160(53)90006-6

[17] E. Andreassen and J. Karger-Kocsis, "Infrared and Raman spectroscopy of polypropylene", in: Karger-Kocsis J. (Eds.), Polypropylene. Polymer Science and Technology Series, **2**, 320-328, 1999, https://doi.org/10.1007/978-94-011-4421-6_46

[18] M.R. Monzó Pérez. "Investigación de la influencia del tipo de gas utilizado en el plasma atmosférico, en la modificación superficial del polipropileno" (tesis de doctorado). Universidad Politécnica de Valencia. Valencia, España, 2015.

[19] M.J. Shenton and G. C. Stevens. "Surface modification of polymer surfaces: atmospheric plasma versus vacuum plasma treatments". Journal of Physics

D: Applied Physics, **34**, 2761-2768, 2001, doi:10.1088/0022-3727/34/18/308

[20] N.V. Bhat and D. J. Upadhyay, "Plasma-Induced Surface Modification and Adhesion Enhancement of Polypropylene Surface", Journal of Applied Polymer Science, **86**, 925-936, 2002. https://doi.org/10.1002/app.11024

[21] H.H. Šourková, G. Primc, and P. Špatenka, "Surface Functionalization of Polyethylene Granules by Treatment with Low-Pressure Air Plasma", Journal Materials, **11**, 885, 2018. 10.3390/ma11060885

[22] S. D. Lee, M. Sarmadi, F. Denes, and J. L. Shohet, "Surface Modification of Polypropylene Under Argon and Oxygen-RF-Plasma Conditions", Plasmas and Polymers, **2**, 177-198, 1997. https://link.springer.com/article/10.1007/BF0276613

[23] M. Kehrer, J. Duchoslav, A. Hinterreiter, A. Mehic, T. Stehrer, D. Stifter, "Surface functionalization of polypropylene using a cold atmospheric pressure plasma jet with gas water mixtures", Surface and Coatings Technology, **384**, 125170, 2020, https://doi.org/10.1016/j.surfcoat.2019.125170.

[24] A. García, M. Domínguez-Díaz, R. Paniagua, J. A. Hernández-Péreza, H. Martínez. "Microhardness modification of ultrahigh-molecular-weight polyethylene by oxygen plasma: Effect of the polymer crosslinking". Nuclear Inst. and Methods in Physics Research B, **445**, 8-12, 2019, https://doi.org/10.1016/j.nimb.2019.02.020

[25] Y. Hiejima, K. Takeda, and K-h Nitta, "Investigation of the Molecular Mechanisms of Melting and Crystallization of Isotactic Polypropylene by in Situ Raman Spectroscopy", Macromolecules, **50**, 5867−5876, 2017, https://doi.org/10.1021/acs.macromol.7b00229.

[26] J. Dutton, "A survey of electron swarm data", Journal of Physical and Chemical Reference Data **4,** 577, 1975, https://doi.org/10.1063/1.555525

[27] K.K. Trusov, "Dynamics of multichannel and quasi-homogeneous sliding discharge formation in rare gases", Journal of Physics D: Applied Physics, **40**(3), 786–794, 2007, doi:10.1088/0022-3727/40/3/015

[28] A. Yu Nikiforov, Ch Leys, M. A. Gonzalez and J. L. Walsh, "Electron density measurement in atmospheric pressure plasma jet: Stark broadening of hydrogenated and non-hydrogenated lines", Plasma Sources Sci. Technol, **24**(3), 034001, 2015, https://doi.org/10.1088/0963-0252/24/3/034001

[29] K. Yoshida, K. Nitta, H. Ohmi, K. Yasutake, and H. Kakiuchi, "Plasma parameters in very high frequency helium and argon plasmas at atmospheric pressure", J. Appl. Phys**.** **128**, 133303, 2020, https://doi.org/10.1063/5.0010195

[30] C. Laux, T. Spence, C. Kruger and R. Zare, "Optical diagnostics of atmospheric pressure air plasmas", Plasma Sources Sci. Technol*,* **12**, 125–138, 2003, https://iopscience.iop.org/article/10.1088/0963-0252/12/2/301

[31] S. Moon and W. Choe, "A comparative study of rotational temperatures using diatomic OH, O_2 and N_2^+ molecular spectra emitted from atmospheric plasmas", Spectrochim. Acta B, **58**, 249–257, 2003, https://doi.org/10.1016/S0584-8547(02)00259-8

[32] S. Astashkevich, M. Kaning, E. Kaning, N. Kokina, B. Lavrov, A. Ohl and J. Ropcke, "Radiative characteristics of 3p Σ, Π; 3d Π^-, Δ^- states of H_2 and determination of gas temperature of low pressure hydrogen containing plasmas", J. Quant. Spectrosc. Radiat. Transfer **56**, 725–751, 1996, https://doi.org/10.1016/S0022-4073(96)00103-3

[33] P.J. Bruggeman, N. Sadeghi, D. C. Schram and V. Linss, "Gas temperature determination from rotational lines in non-equilibrium plasmas: a review", Plasma Sources Science and Technology, **23**(2), 023001, 2014, https://doi.org/10.1088/0963-0252/23/2/023001

[34] P.J. Bruggeman, F. Iza and R. Brandenburg, "Foundations of atmospheric pressure non-equlibrium plasmas, Plasma Sources Sci. Technol. **26**(12), 123002, 2017, https://doi-org.biblioteca.ibt.unam.mx:8080/10.1088/1361-6595/aa97af

[35] H. B. Baniya, R. Shrestha, R. P. Guragain, M. B. Kshetri, B. P. Pandey, and D. P. Subedi, " Cold Atmospheric Pressure Plasma Jet for the Improvement of Wettability of Polypropylene", International Journal of Polymer Science, **2020**, 3860259, 2020, https://doi.org/10.1155/2020/3860259

[36] A. Czernichowski, H. Nassar H, A. Ranaivosoloarimanana A.A. Fridman, M. Simeek, K. Musiol, E. Pawelec and L. Dittrichova, "Spectral and Electrical Diagnostics of Gliding Arc" Acta Phys. Pol. A89, 595, 1996, doi:10.12693/APhysPolA.89.595

[37] K. Ollegott, P. Wirth, C. Oberste-Beulmann, P. Awakowicz, and M. Muhler, "Fundamental Properties and Applications of Dielectric Barrier Discharges in Plasma-Catalytic Processes at Atmospheric Pressure", Chem. Ing. Tech, **92**(10), 1542–155, 20208, doi:10.1002/cite.202000075;

[38] M. Keidar, A. Shashurin, O. Volotskova, M.A. Stepp, P. Srinivasan, A. Sandler, and B. Trink, "Cold atmospheric plasma in cancer therapy", Physics of Plasmas **20**, 057101, 2013, https://doi.org/10.1063/1.4801516;

[39] S.H. Jafari, A. Kalati-vahid, H. A. Khonakdar, A. Asadinezhad, U. Wagenknecht, D. Jehnichen. "Crystallization and melting behavior of nanoclay-containing polypropylene/poly (trimethylene terephthalate) blends", Express Polymer Letters, **6**(2), 148–158, 2012, 10.3144/expresspolymlett.2012.16

[40] M.R. Abadchi, A. Jalali-Arani, "Crystallization and melting behavior of polypropylene(PP) in (vulcanized nanoscale polybutadiene rubberpowder/PP) polymer-nanocomposites", Thermochimica Acta, **617**, 120-128, 2015, https://doi.org/10.1016/j.tca.2015.08.027

[41] M.M. Favaro, M. C. Branciforti, R. E. S. Bretas, "A X-ray Study of β-Phase and Molecular Orientation in Nucleated and Non-Nucleated Injection Molded Polypropylene Resins", Materials Research, **12**(4), 455-464, 2009, https://doi.org/10.1590/S1516-14392009000400014

[42] A.S. Mielsen, D. N.Batchelder and R. Pyrz, "Estimation of crystallinity os isotactic polypropylene using Raman spectroscopy", Polymer **43**(9), 2671-2676, 2002, https://doi.org/10.1016/S0032-3861(02)00053-8

[43] T. Kida, Y. Hiejima, K.-H. Nitta, "Molecular orientation behavior of isotactic polypropylene under uniaxial stretching by rheo-Raman spectroscopy", eXPRESS Polymer Letters, **10**(8), 701–709, 2016, 10.3144/expresspolymlett.2016.63

[44] N. L. Singh, A. Qureshi, N. Shah et al., "Surface modification of polyethylene terephthalate by plasma treatment," Radiation Measurements, vol. **40**(2-6), 746–749, 2005, https://doi.org/10.1016/j.radmeas.2005.01.014

[45] F.J. Baltá Calleja, Microhardness relating to crystalline polymers. In: Kaush H.H., Zachman H.G. (Eds), "Characterization of Polymers" in the Solid State I: Part A: NMR and Other Spectroscopic Methods Part B: Mechanical Methods. Advances in Polymer Science, **66**, 117-148, 1985, https://doi.org/10.1007/3-540-13779-3_19

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1258-1266 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

LEACH Based Protocols: A Survey

Nour Najeeb Abdalkareem Qubbaj*, Anas Abu Taleb, Walid Salameh

Computer Science Department, Princess Sumaya University of Technology, Amman, 11941, Jordan

ARTICLE INFO

Article history:
Received: 10 October, 2020
Accepted: 06 December, 2020
Online: 16 December, 2020

Keywords:
Fuzzy Logic
Energy efficiency
Hierarchical routing protocols
LEACH
CHs (clusters Head)
WSN

ABSTRACT

Advances in the world of communications and information technology, as well as the urgent necessity to monitor particular areas and regions, have led to a considerable and influential development in the world of wireless sensor nodes. As they are small, low-cost multi-purpose nodes with limited energy and capabilities. The most important points that deserve research and solution is the problem of energy conservation and increasing the lifetime of the network as a whole. Whereas, many routing protocols have been proposed in the wireless sensor network, some of which are based on homogenous networks while others are not, some are location-based, some are based on hierarchical orientation, some are based on the hybrid approach and many more. Despite the long time that has passed since the discovery and activation of the Low Energy Adaptive Clustering Hierarchy (LEACH) protocol, it and its descendants still have the lion's share of attention, research, and development. In this research, we will be presenting the most prominent routing protocols while pointing to some pros and cons of each protocol.

1. Introduction

"This paper is an extension of work originally presented in "11th International Conference on Information and Communication Systems (ICICS). IEEE, 2020" [1].

Wireless sensor networks (WSN) can be defined as collection of thousands or hundreds of thousands of small self-configuring sensor nodes that bond together and collaborate to aggregate information and send it to a central station. The purpose simply is to get rid of the human aspect [2]. Usually, (these nodes being distributed randomly by aircrafts to serve military interests, detecting natural disasters like fires and earthquakes, also it plays a huge role in some peaceful applications such as detection of a specific disease in a specific crop, health care, electronic sales, and others [2]. On the other hand, routing is defined as the process of exactly determining the path from the main source to the final destination, and it must be fast and reliable as well.

As a reminder, these nodes are thrown from aircraft randomly and to places that are difficult for humans to reach, so they must remain functional as long as possible. Unfortunately, the only energy source for them is the battery and it is known that the battery depletes quickly and the node dies at that time. Therefore, the most important challenge is to keep the nodes alive as much as possible and distribute the total energy wisely, and not lose their energy, especially in the early stages. Researchers have provided a massive number of routing protocols and classify them on several levels, considerations, and parameters [1].

The rest of our research is split into the following sections: in section II the problems of routing are presented. In section III. Most important WSN routing methods. In section IV. Furthermore, Low energy adaptive clustering hierarchy (LEACH), the before the last section V. Different routing protocols based on (LEACH) protocol, the last VI. Conclusion and future work subsequently reference.

2. The Problems of Routing

In WSN, routing protocols can be considered to be dependent on the application, the data to be collected, and the good use of limited energy. A good routing protocol is a protocol that has several advantages, including simplicity, ability to know its energy level, ability to expand and adapt under the limited capabilities of the nodes, to name a few: power, memory, computation capabilities, and bandwidth allocated to the network according to [3]. The main purpose of the network design is to

*Corresponding Author: Nour Najeeb Abdalkareem Qubbaj, Email: nou20188107@std.psut.edu.jo

deliver all the information from the deployed nodes to the base station (BS) with the lowest possible power. This design faces several obstacles, including [1,3]:

2.1. Deployment of nodes

Deployment of nodes relies on the type of application that massively affects the efficiency of the routing protocol. Where the deployment of the sensor nodes may be random (self-organizing), by plane, for example, or be in a specific and thoughtful place. The node may be fixed, not moving, and it may be mobile, or all of them are stationary and the main station is mobile.

Location and movement play an important bit part in energy efficiency.

2.2. Conservation of energy

Data and information are transmitted by radio through the network, the power of the radio is inversely proportional to the square of the distance traveled, and the power consumption will be affected in the event of obstacles. Therefore, it is better to use the multi-hop approach rather than one-hop. It is worth noting that the one-hop orientation is suitable for a network with a small number of nodes and close to each other, which confirms that energy is affected by the paths that you take.

2.3. Tolerance of Faults

Error tolerance can be described as the ability of the network to continue its operation in the presence of obstacles. Such as the nodes death. For example, the routing protocol must quickly find an alternate route to cancel an insufficiency of information to the main (Base) station.

2.4. WSN Scalability

The number of nodes in the target area to be monitored will exceed hundreds or thousands. Thus the routing protocol must be able to accommodate this large density of nodes.

2.5. Constraints on hardware

All components of the nodes including, the power unit (battery), the processing unit (memory and the microprocessor unit), the sensor unit (sensor), and the radio communication unit, must be smaller in size, less expensive, and energy-consuming.

2.6. The surrounding environment

The node must be prepared for the difficult conditions in which it will monitor.

2.7. Transmission Media

From the name (WSN), we find that the mean of communication is wireless communication, as it is vulnerable to the rate of error and greatly influences the correct operation of the network.

2.8. Models for data delivery

Delivering the data to the main station is related to the type of application, it may be continuous, where all the data is sent to the station continuously and periodically. While if the connection is related to an event, the node sends its data when a specific event occurs. The communication may also be inquiry-oriented, so that when the station needs to inquire about a specific piece of information from the nodes, and the connection may be a hybrid connection consisting of two or more of the mentioned methods. Finally, it's worth pointing that the routing protocol being influenced by the type of that connection significantly.

2.9. Information aggregation (fusion)

The cluster head (CH) collects data from more than one node, so there is a possibility of duplicating data, especially adjacent nodes, so CH does this data fusion process to minimize the volume of the transmitted data to the main station.

2.10. Capacity of node

The capacity of a node this means that it is possible to assign a specific sensor node to one of the functions(tasks) (relaying, sensing, aggregation) according to the application, instead of giving the node all the functions(tasks) together to avoid consuming excess energy.

3. Most Important WSN Routing Methods

Route pathways may be defined in one of three forms i.e. reactive, proactive, or mixed, the reactive paths calculate the paths when the actual necessity of them. While the proactive route precedes the occurrence of the real event and stores the possible paths in a table for each node, it is heavy-duty due to the huge number of network nodes. The last type is the combination of the characteristics of both proactive and reactive see Fig 1.

Based on the design of network flow, the routing protocols are classified into flat-routing, hierarchical and location-based routing. In flat-based routing, all nodes execute the same task such that data is transferred from each sensor node with a strong data repetition [3].

Hierarchical Protocols, the clusters are formed in this type to raise energy efficiency and accommodate a large number of nodes and cover a larger area of the network with better efficiency by applying two main steps. The first one is determining the head of the cluster. The second is the path from the head of the cluster to the main station. [3].

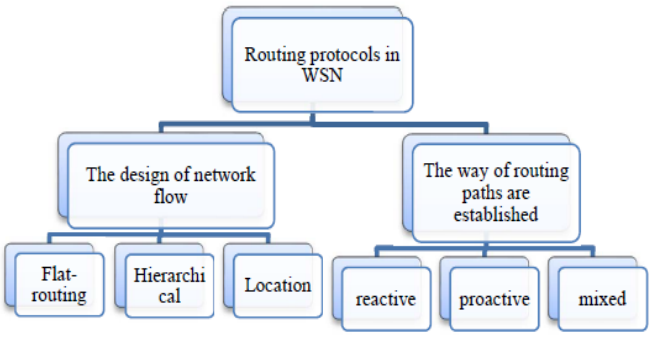


Figure 1: Routing protocols kinds [1].

Location-based protocols, some applications may require the

location of the node to calculate the distance between the nodes. There are several ways to calculate this, including (Global Positioning System) GPS or other ways [3]. See Fig 1.

The most rudimentary way to transfer data from the nodes to the base station is that each node does this by itself without an intermediary, which is known as Direct Transmission (DT), as it is considered one of the flat and simple protocols. And another example of it too is Minimum Transmission Energy (MTE). Whereas, the nodes that are farther from the base station and the nodes with a large volume of data require more energy than others. Therefore, the hierarchical protocols are distinguished from the flat and simple protocols, such as the ones mentioned.

The working principle of the clustering methods of WSN, is based on dividing the entire network into groups called the clusters, and in each cluster, there is a header that gathers data from the rest of the nodes, merges them, reduces their size, and sends them at once to the Base Station (BS) and the role of this head is exchanged with the rest of the nodes in the clustering according to the followed protocol [4]. Cluster formation has several characteristic and advantages, so it excels over flat protocols, including [5]:

- Easily handle the complexity of the network: The network is composed of nodes that act as the head of the cluster and that communicate directly with the main station on behalf of all the nodes of the cluster, and other nodes that collect data from the surrounding environment and send it to the head of the cluster that does not have another load, thus the master station receives data from a small number of nodes rather than all of the cluster members and thus facilitates control of the network as a whole see Figure 2.
- Aggregation / Fusion of information: Where the CH integrates and collects information from the rest of the members and reduces its size before sending it to the base station, the base station receives data from a small number of nodes rather than all of the cluster members.

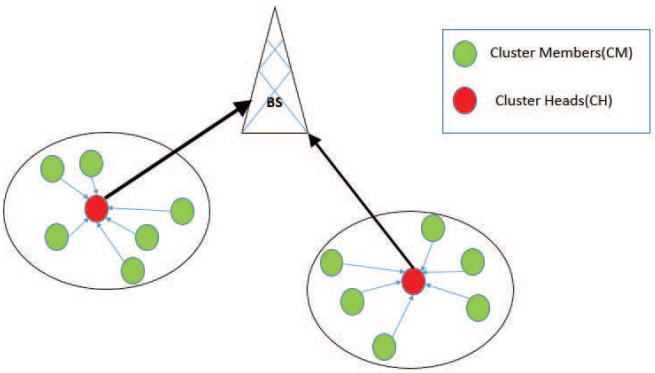


Figure 2: Creation of clusters [6].

- The life of the network is stronger and longer: By rotating the head of the cluster in every round to distribute an equal distribution of energy among the network members.
- Reducing collisions and traffic: In a WSN with a flat communication model, all the nodes must send their information to the main station at the same time. In the clustering organized, the communication with the main station is limited only to the head of the clusters.

4. Low Energy Adaptive Clustering Hierarchy (LEACH)

It is the parent protocol for hierarchical routing based on Clustering proposed by W.B. Heinzelman [7]. As has been reported, the clustering protocols have proven effective in distributing energy consumption among members and increasing the life of the network as a whole.

LEACH works on the principle of dividing the network into random groups (clusters) and then choosing a node from each group to be the head node to represent its group while connecting to BS. The role of the cluster head rotates to keep the energy consumption distributed among the network members. It's worth pointing that the head nodes have the heading role as an additional role to their original tasks (tasks of the non-cluster head nodes). It is also a kind of one-step approach (one-hop routing) to the BS. The LEACH principle is divided into several rounds, and each round contains two basic stages: One is the Setup stage, and the other is the Steady stage as shown in Figure 3.

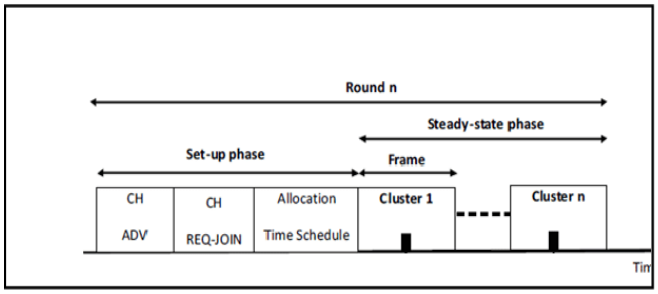


Figure 3: Operation of LEACH [8].

The node must initially pick a random number (n). If this number (n) is less than the threshold specified in the first equation T(n), then the node is selected as the cluster header (CH).

$$T(n) = \begin{cases} \dfrac{p}{1 - p.\left(r \bmod \frac{1}{p}\right)} & if\ n \epsilon\ G \\ 0 & otherwise \end{cases} \quad (1)$$

while p is the possibility of a cluster head (CH), G is a list of nodes that will never be chosen as cluster-head (CH) nodes until the 1 / p level.

After selecting the heads of the cluster according to the number (n), each cluster head will broadcast its status to all other nodes via CDMA (Code Division Multiple Access). After receiving broadcast messages from CHs, each normal node uses the Received Signal Strength (RSS) to determine which CHs are closest to it. Then the CH creates TDMA (Time Division Multiple Access) to divide a schedule for members of the same group (cluster) so that the normal node is activated and connected to CH at the time allotted to it according to the TDMA schedule. It will

be inactive at other times to preserve its energy, which means that CH deals with all nodes from the same cluster and is the only one who deals with BS from cluster members [9]. Note that the period of the Steady stage is longer than the Setup stage.

The significant drawbacks of LEACH are the following, according to [1,2,10]:

- Since LEACH (one-hop routing) to the BS is never suitable for a large network of sensors.
- Not suitable for time-compulsory applications due to delays.
- Unacceptable distribution for cluster heads: as have been stated earlier, a random selection of CH that is not based on clear criteria such as the distance between CH and BS, causes an imbalance in the power consumption of the network and this is considered a significant drawback.
- Unacceptable selection of cluster head: The randomness in choosing the head of the cluster so that the primordial energy of each node is not taken before it is considered as the head of the cluster.
- Note that the types of data delivery to the base station may be continuous, event-related, directed to an inquiry, or a mixture between them. The LEACH protocol is more appropriate with continuous delivery only.
- Different sizes for each cluster: The randomness of choosing the cluster causes the existence of large-numbered clusters and other small-number without specific study due to randomness only.

Several routing protocols have been proposed to solve some of these problems faced by the LEACH protocol, the most prominent of which will be presented in section Five.

5. Different Routing Protocols Based On (LEACH) Protocol

5.1. Low Energy Adaptive Clustering Hierarchy Centralized (Leach-C)

LEACH-C is one of the LEACH versions, which is a centralized protocol. It is divided into several rounds as the LEACH protocol and each round into two basic stages. This protocol assumes that each node can communicate directly with the BS and can calculate its energy and current position.

Unlike LEACH, LEACH-C take into consideration the distance from the node to the base station. The base station determines who is the CH in this round based on its energy and the distance between the node and the base station.

Where in the setup stage all nodes send their energy and location to the base station, the station calculates the average value of the energy values, and the node that contains energy higher than the average value is elected as CH for this round. In the Steady stage, the non-head nodes send their information to the head of their cluster, and the head of the cluster, in turn, collects the data and sends it back to the base station [5].

5.2. LEACH-Balanced (LEACH-B)

The LEACH-B protocol overcomes the original LEACH protocol by taking into account the remaining energy of the node and also strives to maintain the optimum number of cluster heads per round of the protocol. The optimum value for the number of cluster heads varies from 3 to 5% based on what was mentioned in [11]. Thus reduce energy consumption, it is a decentralized protocol, LEACH-B has improved energy performance compared to the parent LEACH protocol see Figure 4.

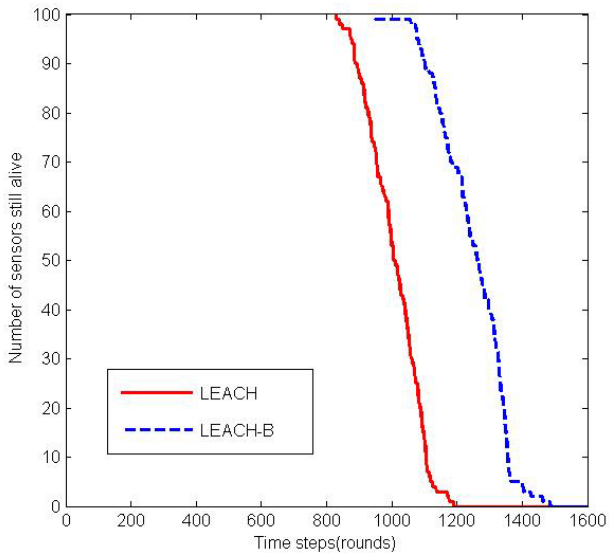


Figure 4: System life of LEACH and LEACH-B [11].

5.3. Fixed Number of Clusters LEACH (LEACH- F)

This protocol is a centralized approach, like the LEACH-C protocol. clusters are set, and rotation is just for the CH inside the same cluster. The steady-state is identical to the original LEACH since the number of clusters has been determined in advance, the energy wanted for re-clustering is saved. It is not elastic in the case of extension, elimination, or death of a sensor node [12].

5.4. Advanced Low Energy Adaptive Clustering Hierarchy (A-LEACH)

As a result of the overload on CH, it is likely to die early before the other nodes, A-LEACH prolongs the duration of stability (the time before the death of the first node). It worth mentioning mention that A-LEACH is a heterogeneous protocol.

Its working principle is as follows: Every sensor knows how to continue each round utilizing the synchronized clock. Let's assume that the total number of nodes is (n), and (m) is the number of nodes that contain more power than others (also called (CGA) nodes that chosen as gateways or CH), so the number of nodes other than nodes can be found using equation 2, see Fig 5.

$$head = (1 - m)\,.\,(n) \qquad (2)$$

A-LEACH is defined as a Distributed Algorithm because the selection of clusters is independent of the base station. CAG nodes can start to transmit data even when all the non-cluster head nodes collapse [12].

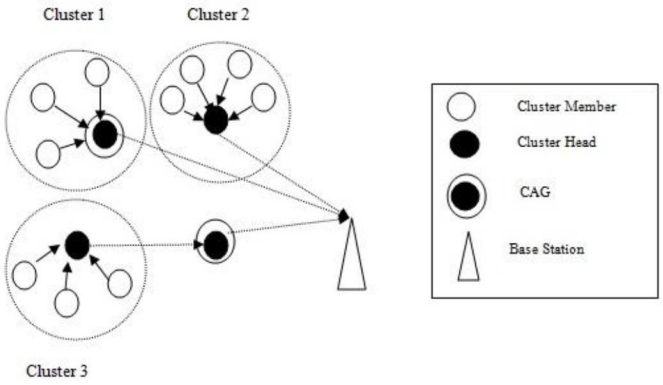


Figure 5: LEACH -A system [1].

5.5. Energy Low Energy Adaptive Clustering Hierarchy (LEACH-E)

E-LEACH is a variant of the LEACH protocol that takes into account the residual energy of a node selected as a cluster head. So the node with the maximum and most appropriate energy is chosen. In other words, E-LEACH is concerned with the number of CHs where the large number of cluster heads consumes higher energy while the small number of cluster heads will cover a greater distance. If the nodes are scattered away from the BS, it will consume more energy. The most appropriate number of cluster heads should be determined, their location and energy. The minimal spanning tree of cluster heads is used in the E-LEACH see Figure 6 [1,2].

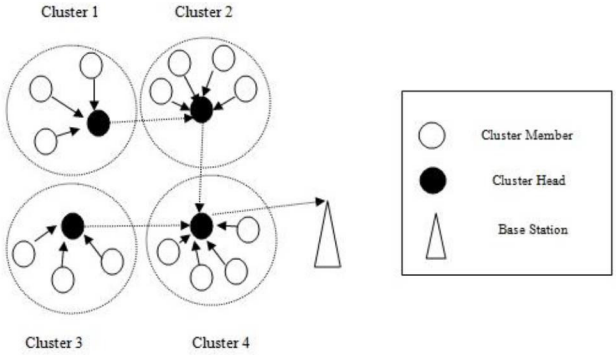


Figure 6: LEACH -E system [1].

5.6. Vice Cluster Head LEACH (V- LEACH)

V-LEACH is another version of the LEACH protocol. It is known that the load on the cluster head is greater because it aggregates data from the rest of the nodes and performs simple operations to reduce its size and transfer it to the main station.

If the CH energy is reduced, it will die thus, the whole cluster is useless because it does not communicate its information to the main station. V-LEACH suggested that the presence of a deputy head of the cluster in the event of his death. The vice-CH is responsible for transferring the information to the main station thus, the whole cluster remains connected with the main station, but in case of death of the vice-CH, there is no solution in this protocol see Figure 7 [10].

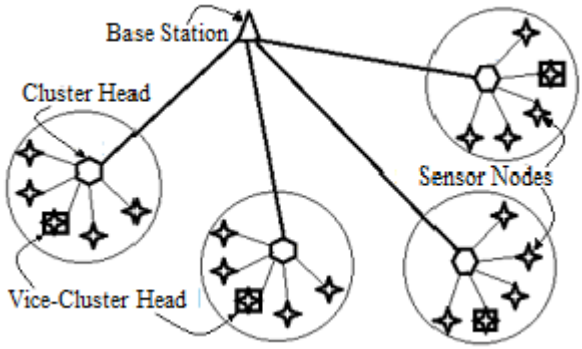


Figure 7: LEACH -V system [13].

5.7. I-LEACH

A new algorithm, I-LEACH, has been proposed, which takes into account the following: the geographic location of the CH, the number of neighbors, and the residual energy to improve network performance and lengthen its life. So the node with the most residual energy and closest to the BS is chosen. This protocol overcame a set of drawbacks of the original LEACH protocol some of them: Different size for each cluster, Unacceptable distribution for cluster heads, and Unacceptable selection of cluster head.

I-LEACH has been evaluated and proven to be effective in average energy consumption, the number of data packets received, and the flexibility to change the location of the BS in the network. Where it reduced energy consumption by approximately 62%, increased the percentage of data packets received by approximately 56%, and increased the life of the network each by 65% at least [14].

I-LEACH, has some of the following assumptions:

- All nodes have the same primary energy, are not rechargeable, and in the same random position when the nodes are thrown from the airplane. They remain in place and never move.
- All nodes have their identifier and can know their exact location and the remaining energy in its battery.
- It is known that each node contains a processing unit and a memory unit, and it can also define the sensed and the aggregated data.
- A node can sense different types of data.
- Each node can communicate directly with the BS and has the power required for this.

5.8. Cell Low Energy Adaptive Clustering Hierarchy (Cell-LEACH)

Cell-LEACH is another version of the LEACH, as in the original LEACH protocol, the entire network is split into groups (clusters). Here too, the network is divided into groups where each group is divided into 7sub-groups called cells. Each group has a cell head whose task is collecting the information from the members of the cell and communicates directly with the head of the cluster. The cell is randomly assigned and then chosen based on its remaining energy to be cell head see Figure 8 [15]. Cell-LEACH has proven its ability to lengthen the network life without

reducing efficiency as can be seen in Figure 9.

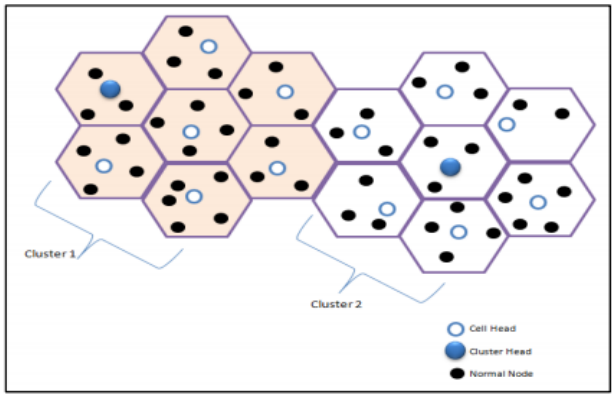


Figure 8: LEACH -Cell system [16]

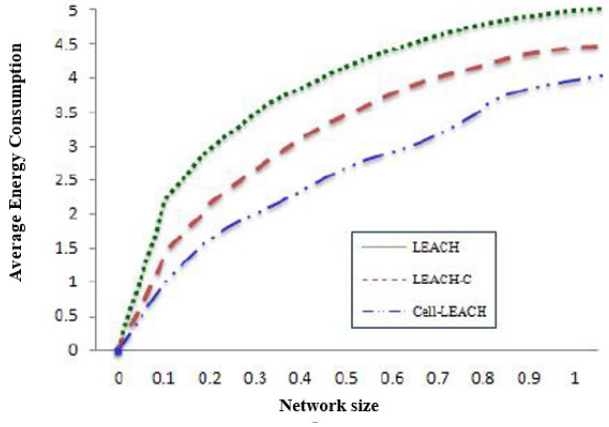


Figure 9: Average power consumed with network size [15].

5.9. Multi-Hop LEACH

5.9.1. TL-LEACH (Two levels LEACH)

Two levels LEACH) was suggested in [17]. Instead of aggregating and transmitting data in one hop, the data in CH being aggregated at a certain level and transferred to a closer CH level to BS thus, reducing the transmission power, which enhances energy efficiency, and extending the network life. To illustrate the idea further, we have Fig 10 that shows that there are 3 CHs at the top level is closest to BS and also 5 CHs on the second level closest to the normal nodes sensor, and the other 10 SN (Sensor Node) normal nodes at the final level. It has been shown that TL-LEACH increases network life by nearly 30 percent and also increases the percentage of packets received by 20 percent [17].

5.9.2. Multi-Hop LEACH:

Due to the disadvantage in " LEACH (one-hop routing) to the BS", and as mentioned earlier, the task of the cluster head is to transfer the collected data from the non-head nodes to the BS. This implies that if the distance is large, then the energy of the cluster head runs out quickly. So a midway cluster head has been considered to reduce distance, thus reducing energy consumption this is known as multi-hop communication [6].

SN	**Sensor Nodes**
CH_{ij}	**Cluster Head to second level**
CH_i	**Cluster Head to first level**
BS	**Base Station**

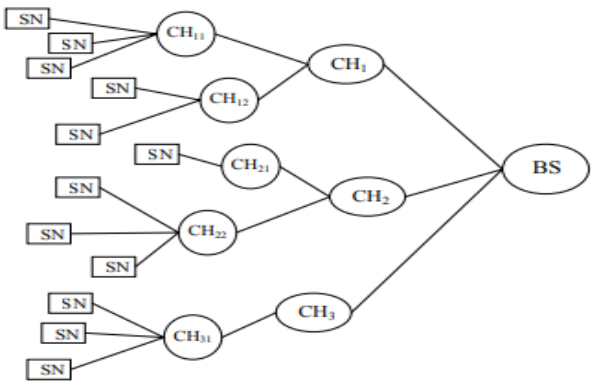


Figure 10: The TL-LEACH network format after the setup phase [17].

The original LEACH protocol does not take into account the primitive energy of the head of the cluster, CH may lose its energy more quickly than others and die, and this cluster loses its connection with the main station. Therefore, it has been proposed to amend the basis of CH selection to take into account energy.

A term was added to the original formula for the original LEACH protocol as in Equation 3 [6].

$$T(n) = \begin{cases} \dfrac{p}{1 - p.\left(r \bmod \frac{1}{p}\right)} \cdot \dfrac{E_{rem}}{E_{totnet}} & if\ n \epsilon\ G \\ 0 & otherwise \end{cases} \quad (3)$$

while p is the possibility of a cluster head (CH), G is a list of nodes that will never be chosen as cluster-head (CH) nodes until the 1 / p level. E_{rem} is the remaining energy level of the node, E_{totnet} is the total energy network. This happens in the setup phase.

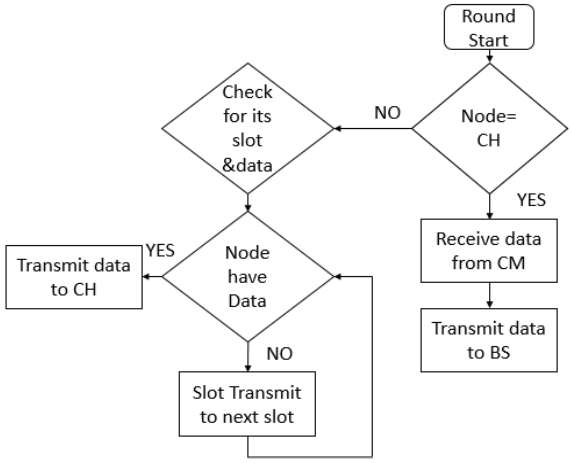


Fig 11: Freshly revised TDMA description [6].

In several protocols associated with an event, the most appropriate thing to do is to take advantage of the free time of the nodes. For example, when a node does not have information to send to the main station, an allocation technique activates to allocate the time assigned to the free node in the TDMA schedule to another node that has information to send. Thus, reducing energy and time consumption as well, as can be seen in Fig 11 [6].

5.10. Clustering Process Based on Fuzzy Logic

Fuzzy logic simulates human decision-making and experience to judge things and deals with decisions better than probabilistic approaches. Also, it is more suitable for real-time applications [18]. Due to the suspicion and doubt in the WSN environment, some protocols use Fuzzy Logic technology, as it simulates human logic less than others, which include computational complexities and the need for memory. Fuzzy Logic protocols used to choose CH more effectively and flexibly due to its flexibility and error tolerance choose CH more. The Fuzzy Logic model requires four steps to accomplish, namely:

- Fuzzification: In it, the inputs are inserted with clear values and are distorted into a fuzzy group, and this happens in the fuzzifier.
- Rule evaluation (Fuzzy Rule base): This is storing Rule IF-THEN.
- Fuzzy Inference Engine: This engine handles both input and law IF-THEN to produce a fuzzy group.
- Defuzzification: Converts the fuzzy group to clear values.

These steps are shown in Figure 12. Most of the Fuzzification methods use Mamdani's Method Fuzzy Inference technique to select the CH.

This block is called fuzzy logic controller (FLC). Not too long ago, the authors began leveraging Fuzzy Logic (FL) and incorporating it into WSN to lengthen the life of the network and improve the method for selecting the appropriate CH. Some of the popular Clustering Process Based on Fuzzy Logic will be featured in this section:

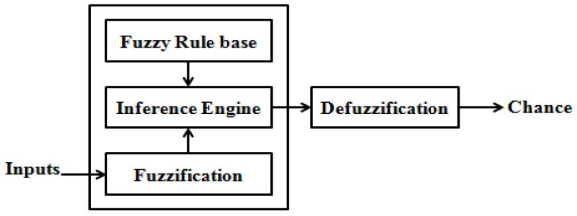


Figure 12: The Fuzzy Inference Process structure diagram.

5.10.1 Cluster Head Election using Fuzzy Logic (CHEF):

In this type, CH is chosen by entering two inputs: power and local distance. (The local distance is the total distance between the temporary CH in that round and the cluster members (normal nodes) within a predetermined fixed competition radius) [19]. So that the likelihood of the closest and more energy nodes to be selected as CH is higher than others, As the experiments of [18] show that CHEF is 22.7% more effective than LEACH.

However, some of the disadvantages of this type are that not all sensor nodes have a GPS device and sometimes are not able to know their geographical location [18]. Besides, this type is not suitable for network sized about 200m × 200m or more [19,20]. clusters closer to BS, run out of energy faster because traffic becomes congested the closer to BS, so an unequal clustering has been suggested to balance energy consumption [19].

5.10.2 Energy Efficient Unequal Clustering (EEUC):

This type divides all nodes into clusters that are not equal in number so that the clusters close to the BS contain fewer nodes. The distant clusters contain a greater number because the closer we get to the base station, the congested the traffic becomes, causing the closest nodes to consume more energy from the farthest [21]. As in the LEACH, the choice of CH is random here is random as well, a random number between 0 and 1 assigned for each node. The number indicates whether the node is going to be a CH or not [19]. Experiments proved that EEUC is more efficient than LEACH see Figure 13.

5.10.3. Energy-Aware Unequal Clustering with Fuzzy (EAUCF):

Battery level goes down to the CH [19]. In other words, when the residual energy of the temporary CH and the distance to the BS is large then the competition radius is large and vice versa [21]. Let us say that each temporary CH will forward its remaining energy and also if it is within the competition radius of this CH, another temporary CH will be compared between their energy, and the least energy will be out of competition and not CH [20]. In the experiments. It has been proven that EAUCF is more efficient than LEACH, CHEF, and EEUC see Figure 13. While this type does not pay attention to the degree of the node, it means the density of the internal cluster, as the density of communication inside the cluster costs energy and reduces the life of the wireless network as a whole [20].

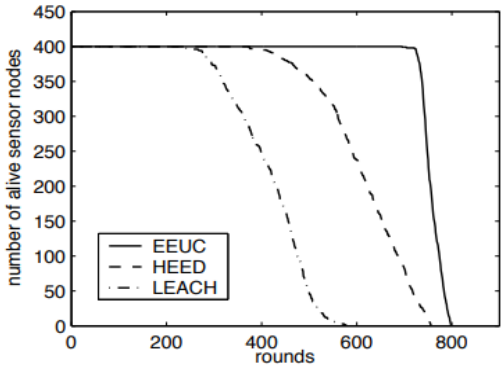


Figure 13: Over time, the number of live sensor nodes [22].

5.10.4. Fuzzy Cluster Head Selection Algorithm based on LEACH:

The authors of [16] noticed that when more than one entry is taken into account to determine one output (appropriate CH selection), Fuzzy Logic, is more accurate and appropriate than others. This type requires three fuzzy input variables. The first variable expresses the number of surviving neighbors, the second

variable expresses residual (remaining) energy, the last variable expresses the distance to BS. The one variable of fuzzy output is named a chance [20] as shown in Figure 15. The CH is selected with the most living neighbors, the highest residual energy, and the least distant to BS [20].

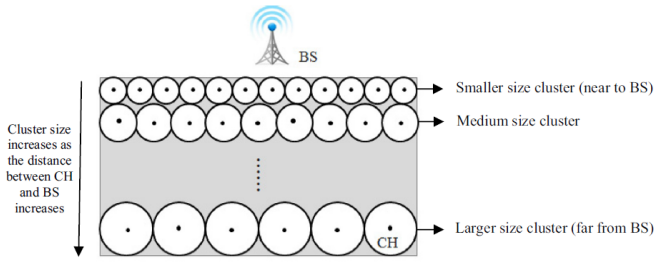


Figure 14: Unequal clustering architecture at WSN [22]].

Table 1: At round 250, the total remains energy for each algorithm [19]

Algorithm	Total Remaining Energy (J)
LEACH	27.14
CHEF	43.66
EEUC	44.78
EAUCF	46.41

Just like the LEACH, in each cycle, every node picks a random number between 0 and 1, and if the number was chosen is less than the threshold T, then it becomes within the Concurrent CH. Then each Concurrent CH node calculates its chance from the output of the Fuzzy and shares it with the rest of the Concurrent CH who are within its competition radius. Consequently, each Concurrent node CH has a list of all CH node and its chance, where the highest is elected as CH node, and the rest within the competition radius is a normal node and the elected CH node joins. Also, as usual, the elected CH node is required to create a TDMA schedule and to publish it to the rest of the cluster. The normal nodes send their information to the selected CH node in the time allotted to them and turn off their transmitter device at other times to conserve energy [20]. Compared to LEACH, this form improves network life by 58.38 percent see Table 1.

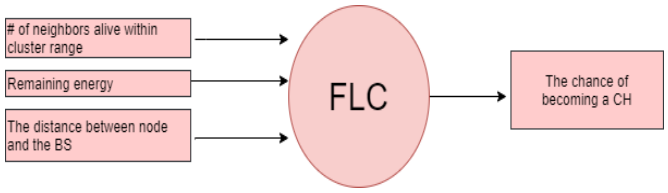


Figure 15: The Fuzzy-Logic system which controls the clustering

5.10.5. A Fuzzy-Logic based Energy-Efficient Clustering algorithm (FLEEC):

The authors of [21] noticed that when more than one entry is taken into account to determine one output (appropriate CH selection), Fuzzy Logic, is more accurate and appropriate than others. It should be noted that the process of calculating the distance between the node and the BS is done based on the Received Signal Strength (RSS).

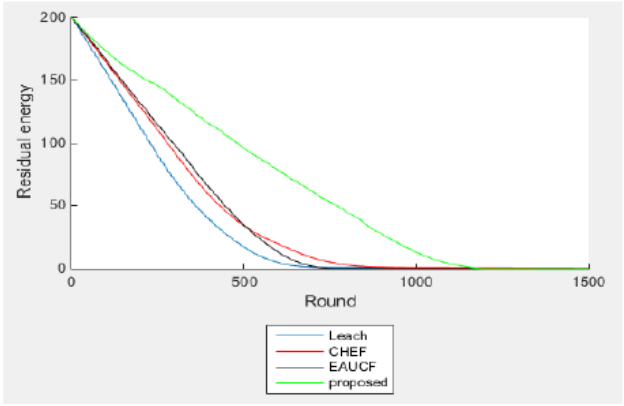


Figure 16: The full remaining energy of every cycle [20].

If clusters of the same size are created without taking into account the distance to the BS, then a "hot spot problem" begins to arise. As was evident from Figure 14, it requires more energy for the cluster closer to the BS. To create uneven clusters, the first fuzzy logic controller (FLC) is constructed, which takes two inputs (The distance to the base station and Node Density), and one output (cluster Radius) see Figure 15. Then the head of potential clusters measures the total distance of the neighbors that are in their radius [21]. The total distance to the CH and Residual energy are inputs to the second FLC, and the output is the probability of each node to be CH. The node that has the highest likelihood to become CH is when the node has the largest residual energy and more neighbors.

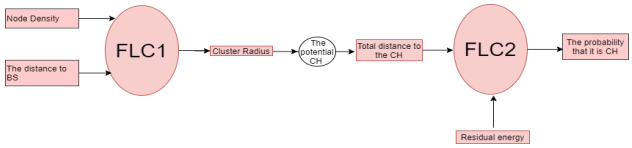


Figure 17: The Fuzzy-Logic system which controls the clustering in (FLEEC)

FLLEC is composed of three major stages [21]:

- Determination of cluster size process: First of all, the BS node broadcasts a welcome message to every node in the region, and then each node would like to calculate its distance from BS via Received Signals Strength (RSS). Each node sends to the BS node an ACK message, which contains its ID, residual energy, geographic distance from the BS node. Then the BS computes the density distribution from the information it obtained from every node. Here comes the role of the first FLC by calculating the corresponding cluster radius for each sensor node, and this information is saved in the memory of each node.
- Determination of likelihood to be a cluster head process: After the cluster radius determined successfully, all nodes would like to broadcast a message of exploring the neighbors around it, and each would like it to receive that message followed an ACK message. After that, the nodes calculate the total distance of their neighbors according to RSS. Here comes the role of the second FLC to calculate the likelihood to be CHs depending on the fuzzy rules reserved in the fuzzy rules base.

- Determination of CH process: Each Node broadcasts the HELLO message to all of its neighbors within the competition radius that it has both CH and its ID. During this information exchange process, the CH that has the greatest probability is the one who becomes CH and the rest is a normal node. CH broadcasts an advertisement message (ADV). If the normal node receives more than the ADV, then it determines the most suitable one that will belong to it, then each node sends a request message to the most appropriate CH.

Comparing FLEEC to LEACH (the main clustering algorithm) and CHEF (the main fuzzy-logic-based algorithm) algorithms. As can be seen from Figure 18, the results showed the superiority of FLEEC over both LEACH and CHEF, especially in the issue of energy consumption.

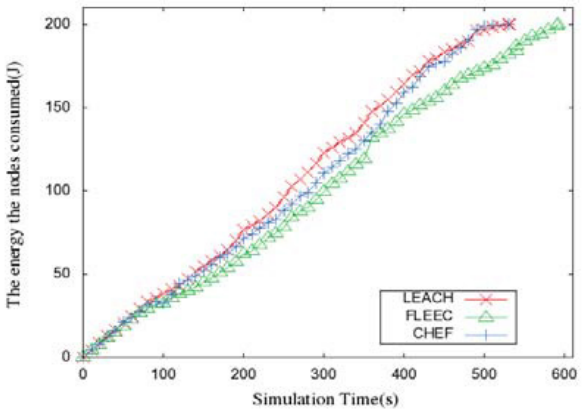


Figure 18: The energy consumed by the three main algorithms is comparable [21].

6. Conclusion and future work

Due to the emergence of nanotechnology and the persistence of multiple computing fields, the field of WSN and the optimal management of node energy remains a fertile field for research and experiment. This study covered two main topics, presenting and discussing some of the most paramount routing protocols and describing their key types and issues, and going through the most prominent hierarchical protocols and their development process.

We concluded from this research that the clustering design is one of the most energy-saving designs. Where we noticed that Multi-Hop LEACH saves energy because it reduces the distance over the CH and also excels and the fuzzy logic strategy based on clustering is superior to other clustering strategies because it is in line with human logic and its ability to convert more from one input to one output, and as it was presented in the research, it was saving more energy than any other algorithms.

In future work, machine learning with the wireless sensor network can be studied to further rationalize energy consumption and merge data before sending it to BS.

References

[1] N. Qubbaj, A.A. Taleb, W. Salameh, "Review on LEACH Protocol," 2020 11th International Conference on Information and Communication Systems, ICICS 2020, 414–419, 2020, doi:10.1109/ICICS49469.2020.239516.

[2] K. Sohraby, D. Minoli, T. Znati, Basic Wireless Sensor Technology, 2007, doi:10.1002/9780470112762.ch3.

[3] P. Kumar, M.P. Singh, U.. Triar, "A review of routing protocols in wireless sensor networks," 2008 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2008, **1**(4), 1–14, 2012, doi:10.1109/WiCom.2008.946.

[4] A. HNINI, A. EZZATI, M. FIHRI, A. HAJAMI, "Time Slots Investment of a dead node in LEACH protocol on Wireless Sensor Network," International Journal of Advanced Computer Science and Applications, **4**(2), 8–12, 2014, doi:10.14569/specialissue.2014.040202.

[5] A. Kumar, R.. Saini, E. Brijbhusahn, "Comparison and Review of LEACH ," (June), 2016.

[6] C.T. Sony, C.P. Sangeetha, C.D. Suriyakala, "Multi-hop LEACH protocol with modified cluster head selection and TDMA schedule for wireless sensor networks," in Global Conference on Communication Technologies, GCCT 2015, 2015, doi:10.1109/GCCT.2015.7342720.

[7] J.S. Leu, T.H. Chiang, M.C. Yu, K.W. Su, "Energy efficient clustering scheme for prolonging the lifetime of wireless sensor network with isolated nodes," IEEE Communications Letters, **19**(2), 259–262, 2015, doi:10.1109/LCOMM.2014.2379715.

[8] N. Nejah, "Propagation Approach for Clustering in Wireless Sensor Network," (March), 2018.

[9] S. Nasr, M. Quwaider, "LEACH Protocol Enhancement for Increasing WSN Lifetime," 2020 11th International Conference on Information and Communication Systems, ICICS 2020, 102–107, 2020, doi:10.1109/ICICS49469.2020.239542.

[10] S. Varshney, R. Kumar, "Variants of LEACH Routing Protocol in WSN: A Comparative Analysis," 199–204, 2018.

[11] T. Mu, M. Tang, "LEACH-B: An improved LEACH protocol for wireless sensor network," in 2010 6th International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2010, 2010, doi:10.1109/WICOM.2010.5601113.

[12] H. Dhawan, S. Waraich, "A Comparative Study on LEACH Routing Protocol and its Variants in Wireless Sensor Networks: A Survey," International Journal of Computer Applications, **117**(12), 27–33, 2015, doi:10.5120/20608-3231.

[13] P. Maurya, A. Kaur, "A Survey on Descendants of LEACH Protocol," International Journal of Information Engineering and Electronic Business, **8**(2), 46–58, 2016, doi:10.5815/ijieeb.2016.02.06.

[14] Z. Beiranvand, A. Patooghy, M. Fazeli, "I-LEACH: An efficient routing algorithm to improve performance & to reduce energy consumption in wireless sensor networks," in IKT 2013 - 2013 5th Conference on Information and Knowledge Technology, 2013, doi:10.1109/IKT.2013.6620030.

[15] A. Yektaparast, F.-H. Nabavi, A. Sarmast, "An improvement on LEACH protocol (EZ-LEACH)," Conference Proceedings of 2015 2nd International Conference on Knowledge-Based Engineering and Innovation, KBEI 2015, 956–960, 2016, doi:10.1109/KBEI.2015.7436173.

[16] M. Al-Shalabi, M. Anbar, T.C. Wan, A. Khasawneh, "Variants of the low-energy adaptive clustering hierarchy protocol: Survey, issues and challenges," Electronics (Switzerland), **7**(8), 1–28, 2018, doi:10.3390/electronics7080136.

[17] V. Loscrì, G. Morabito, S. Marano, "A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH)," IEEE Vehicular Technology Conference, **3**, 1809–1813, 2005, doi:10.1109/VETECF.2005.1558418.

[18] P. Nayak, A. Devulapalli, "A Fuzzy Logic-Based Clustering Algorithm for WSN to Extend the Network Lifetime," IEEE Sensors Journal, 2016, doi:10.1109/JSEN.2015.2472970.

[19] H. Bagci, A. Yazici, "An energy aware fuzzy unequal clustering algorithm for wireless sensor networks," in 2010 IEEE World Congress on Computational Intelligence, WCCI 2010, 2010, doi:10.1109/FUZZY.2010.5584580.

[20] W. Abidi, T. Ezzedine, "Fuzzy Cluster Head Election algorithm based on LEACH protocol for wireless sensor networks," 2017 13th International Wireless Communications and Mobile Computing Conference, IWCMC 2017, 993–997, 2017, doi:10.1109/IWCMC.2017.7986421.

[21] Q. Wang, D. Lin, P. Yang, Z. Zhang, "A Fuzzy-logic Based Energy-efficient Clustering Algorithm for the Wireless Sensor Networks," 2018 26th International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2018, 424–428, 2018, doi:10.23919/SOFTCOM.2018.8555848.

[22] C. Li, M. Ye, G. Chen, J. Wu, "An energy-efficient unequal clustering mechanism for wireless sensor networks," 2nd IEEE International Conference on Mobile Ad-Hoc and Sensor Systems, MASS 2005, **2005**, 597–604, 2005, doi:10.1109/MAHSS.2005.1542849.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1267-1277 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Towards a Documents Processing Tool using Traceability Information Retrieval and Content Recognition Through Machine Learning in a Big Data Context

Othmane Rahmaoui*, Kamal Souali, Mohammed Ouzzif

Computer Science Department, Hassan II University, Casablanca, 20000, Morocco

ARTICLE INFO

Article history:
Received: 18 November, 2020
Accepted: 07 December, 2020
Online: 16 December, 2020

Keywords:
Document Management
Data traceability
Information Retrieval
Machine Learning
Recommendation
Spark

ABSTRACT

In 1980, an application was developed to track, manage and store documents in electronic format. Scan technology has enabled organizations to digitize papers for easier document storage and tracking. Document management tools have since developed by introducing new functionalities, related to security, users services, workflow and audit. Our research is part of the context of improving the efficiency of document processing by proposing an approach using information traceability retrieval and content recognition techniques through machine learning. In this sense, we started by proposing the exploitation and extraction of relationships between documents based on the traceability links and the calculation of similarity using information retrieval. Then, in order to improve the processing of documents, we proposed a contribution of the use of recognition of content techniques by machine learning approaches. Thus, the visualization of the results, according to user profiles, motivated us to offer recommendations dedicated to the document management system. A Big Data environment is proposed because of the exponential growth of data and also to our needs concerning the analysis and the distributed calculation of the voluminous masses of data.

1. Introduction

The definition of a document according to ISO (International Organization for Standardization); "Is the set consisting of an information medium and the data recorded on it, in a generally permanent form and readable by humans or by a machine". A document in electronic format is a file created, modified and readable by a dedicated computer application; it can be an audio, an image or any data file. Electronic documents are the precious treasure and the important part for the reuse and preservation of valuable data in storage based on operators of compression, decompression, encryption and decryption [1], [2]. From these definitions, we notice that any document, no matter how electronic or classic, it keeps two important properties; the information which represents the content processed at the document level and the metadata part which offers the possibility of introducing the notion of object at the level of the design of a document [3]. In the Information Technology (IT) domain, traceability is becoming an obligation for companies offering applications, systems or services that meet market requirements. At the IT project level, changing needs is essential in order to develop a good product that meets all customer requirements, in which case traceability is an essential element for managing changes as well as analyzing its impacts. However, traceability must be proposed to show that a methodological development process has been respected and to justify that the developed tool is ready for use [4]. Using what we have called traceability information retrieval, for the purposes of this article, is a process of producing a list of relevant documents as a response to a user's request by comparing the request to an automatically generated index of the content of documents in the tool [5]. These documents can be consulted subsequently to be processed in the same tool. Today everyone uses tools and applications using information research techniques, for example the search engines Google, Yahoo and Bing.

Nowadays, organizations are constantly seeking to be more flexible and optimized to cope with the revolution in the digital world. Each organization aims for organizational excellence through different means to satisfy either the project teams or their customers or their suppliers [6]. To meet expectations, an effective policy and a strategic vision must be created in a clear way through the structuring of processes, documents and resources. According to management standards [7], the updating of documents plays a very important role in the continuous improvement of the

*Corresponding Author: Othmane Rahmaoui, Hassan II University, ESTC RITM Lab., ENSEM, Casablanca, Morocco, othmane.rahmaoui@gmail.com

company, because the documentation, updated, guarantees optimal efficiency unlike documentation that discards outdated information and so that will make it obsolete. Documentation management is one of the important levers for building quality management; it allows the organization of a set of information, documents and records used in projects or in the various departments of the 'business [8]. A document can be in paper or electronic format.

In order to protect and ensure the validity of easily accessible data as well as keep traceability on the operations and activities carried out, we have thought of improving the processing of documents. However, we started by researching and formalizing methods and procedures necessary and sufficient to control the information and documents useful for the entire organization (entity, service, project, etc.) as well as the technologies that can be used in this direction. The proposed approach targets important issues in order to limit the risks of loss of documents and to make work collaborative with the sharing and capitalization of resources as well as communication between employees and accessibility to information, without forgetting security and reliability as well as traceability management [9], [10]. Our approach saves time in the search for information or documents, increases efficiency, increases the satisfaction of project teams (users) as well as reduces paper volumes, while using modern and efficient techniques through which we called ; IR Traceability Recommendations: Recommendations based on data traceability and information retrieval. Our approach deals with the case of document management that concerns the business domain, as we can generalize it in order to apply it to any sector and in any organization.

For electronic document management, improving the flow of documents has been widely studied in several research works based on the integration of one or more techniques through the implementation and proposal of tools for monitoring and storage as well as document management. The document management administrator is always required to monitor and control the flow of documents processed [3], [11]. Few works offer tools for automatic and incremental improvement of electronic document management. These contain unstructured data and are difficult to operate in another system and even the creation of relationships between information from different documents is difficult to manage, which is why we sought to propose a transformation of documents into data in order to facilitate the processing and development as well as the analysis of data and therefore improve the management of documents in an automatic manner.

In recent years, Big Data with information retrieval has become a more popular area of research [12]. Big Data is a collection of heterogeneous unstructured and structured data. The heterogeneity, volume and speed at which data is generated make the processing and analysis of big data problematic [13]. The traditional database system, warehouses and analysis tools fail to process this type of data [14]. Big Data with information retrieval techniques is an emerging approach not only because of the sheer volume of data, but also because of the kind of unstructured nature. Data related to the user's request should be retrieved with information research. According to experts, employees spend around 11 hours a week dealing with document-related issues, resulting in low productivity and poor customer service. Hence we came up with the idea of harnessing the potential of machine learning, so that companies can revamp and improve the traditional document management system. Machine learning can also be a powerful data mining tool [15], [16]. Using this technology, companies can develop a way to help their employees extract relevant data, classify and process documents automatically, improve data quality and analyze data. Document management by machine learning will therefore save a lot of time and effort. Recommendation techniques also provide new opportunities for the search for personalized information on this work. The recommendation, on the other hand, also serves to reduce information overload, which is a very common problem in information retrieval tools and allows users to have access to services or products that are not readily available to users on the tool [17]-[19].

In this article, we present an improvement of a proposed document management system using traceability and information retrieval with content recognition techniques in a big data context.

Our contribution consists in the improvement of document management systems. The remainder of this article is organized as follows ; The Sect. II presents a state of the art about document management, traceability information retrieval and content recognition through machine learning research area. The literature made was chosen based on recent research and discussion of important topics related to document processing, in order to be useful in relation to the objective of our proposed approach. We present the use of traceability and information retrieval in a big data environment in Sect. III. Machine learning techniques and recommendation are described in Sect. IV. The Sect. V describes the suggested document management system and we conclude in Sect. VI.

2. State of the art

2.1. Document management system

Document management is defined as "the set of techniques for organizing, managing and distributing documentary information in electronic form" [20]. Document management is actively developing especially within companies, because it covers the administration of the document flow within all departments of the company. In the literature [11], [21], [22], we can define document management system as seamlessly automating the document lifecycle process and providing tools to securely manage a large range of documents. Document management system uses statistical algorithms to classify and search documents. Document Management is actively developing especially within companies, because it covers the administration of the document flow within all departments of the company. The objective of this work is to propose techniques to improve the efficiency of document management on the one hand and to minimize the total cost of selecting and recommending documents on the other hand. For document management, improving the flow of documents has been widely studied in several research works based on the integration of one or more techniques through the implementation and proposal of tools for monitoring and storage as well as document management [8], [1].

2.2. Information retrieval based on traceability

In the world of computer science research, traceability is a modern and very recent term, in dictionaries this term appeared

only around the year 2000 while traceability has been known for a long time, its definition differs according to the fields and sectors (Industry, Food, Production, Logistics ...) [23]. The ISO-9001 standard of the international organization for standardization, defines traceability as: 'the ability to trace the history, application, use and location of a product or its characteristics using data from 'recorded identification', Indeed, this definition generates two important phases: the first concerns the identification of the item by a notation and the second concerns the existence of a recording of the data relating to this item on a support which has itself a traceability. The literature shows that there are problems related to the definition and understanding of traceability. the definition of which changes depending on the way it has been proposed. According to the dictionary, traceability is the "possibility of identifying the origin and reconstructing the journey of a product, from its production to its distribution". Researchers say that traceability is not a new concept but a practice that we must implement in order to respect the norms and rules of law [24]. However, in recent years, it has become a necessity in areas where the safety and security of consumers have challenged, particularly in the medical and food sectors [25]-[27]. Traceability is not a new concept but it represents the necessity of results which imposes the replacement of papers and pens by means and techniques more reactive and especially more efficient, traceability comprises two essential elements; tracing and tracking; The purpose of traceability is to provide proof at all times of the conformity of an element and its components as well as of its origin. Tracking represents the way to be able to save and know the paths of elements at the level of the data flow and the tracing is the fact of identifying the origin of a specific data at the level of the flow.

Information retrieval is a field allowing the search for relevant information from any type of data (audio, video, document, article or image). Traditional methods of finding information involve breaking down data into subsets or clusters across dimensions depending on the given problem. The Information Retrieval (IR) approach reduces the effort and time prescribed to create traceability links between artifacts [28] [29]. A search engine is an example of using the IR approach; it can search billions of web documents to answer a very specific query. Information retrieval techniques make it possible to solve several problems related to the processing carried out in a particular domain as well as to its optimization [30]. We have, in our case study, a large remarkable mass of documents and objects, and therefore to find an element adaptable to a need for information expressed by a user using a search query, we have thought of using the algorithms of the traceability information retrieval with recovery of traceability links.

2.3. Machine learning and recommendation

Machine learning has become one of the key words in recent research, machine learning represents an influential source of applications to automate the tasks of information retrieval. the latter offers several features that would fit into machine learning, but really what drives applications of machine learning to IR is not so much developments in machine learning technology as changes in our work and environment which requires new modes of operation [31]. Machine learning is an important discipline due to the digital transformation which has led to the production of large volumes of data of different types and formats. Machine learning can be seen as a branch of artificial intelligence. Indeed, a system which finds it difficult to learn can hardly be considered intelligent. The ability to learn from experiences is essential in a system designed to adapt to an environment that changes often. Artificial intelligence, defined as 'the set of techniques used to build machines capable of demonstrating behavior that can be described as intelligent' [32], also calls on cognitive sciences, to logic, to engineering to electronics, and much more. Artificial intelligence measurements are used by different tools and applications to collect and process and also to communicate and share useful information from datasets. Arthur Samuel defined machine learning, in 1959, as a "field of study that gives computers the ability to learn without being explicitly programmed" [33]. Machine learning (ML) is a concept that allows a computer program to be learned as well as to adapt to new information and new data without any human intervention. With the gradual use of technologies like cloud computing and advanced storage capabilities, the huge amount of data becomes easily accessible and available. Machine learning is the discipline of the computer science that concerns the analysis and interpretation of models and data to enable learning and processing as well as reasoning and decision making without thinking of human intervention. In a simple way, machine learning allows the user to feed an immense amount of data into an algorithm and ask the machine to analyze and make decisions as well as recommendations based on the entered data [34]. If corrections are identified, the algorithm can incorporate this information to improve future decision making. Decisions about data are increasingly making the difference between staying up to the competition and falling behind. Machine learning can be the key to unlocking the value of business data and customer and make decisions that allow a company ahead of the competition.

2.4. Literature and related work

Several works talk about traceability or information retrieval and also about content recognition through machine learning in a big data context, in literature we are finding :

About documents classification in [35], author proposed a review about a wide variety of methods on document classification task for German text with an evaluation of different approaches; Perceptrons, Discrete Naïve-Bayes, MC4, 3 Nearest-Neighbors, Racchio Centroid and Support Vector Machines. In [36], the author have proposed a framework for modeling link distributions by using a structured logistic regression model capturing both content and links in order to improve text classification. In [37], the author have presented TRAIL (TRAceability lInk cLassifier) as an approach for maintaining traceability information (traceability links between pairs of artifacts) to train a machine learning classifier which is then able to classify the links between any new or existing pair of artifacts as valid or invalid (two artifacts are related or no). So, document classification is an example of machine learning in the form of natural language processing (NLP), in the real world numerous more complex algorithms exist for classification such as support vector machines (SVM), Naïve Bayes and Decision Trees. There are also works that use information retrieval with several traceability methods; in [38], the author have presented a study to analyze the equivalence of information retrieval methods for automated traceability link recovery; Jensen-Shannon (JS), Vector Space Model (VSM),

Latent Semantic Indexing (LSI) and Latent Dirichlet Allocation (LDA). The result show that while JS, VSM and LSI are almost equivalent. In [39], the author have proposed an approach for improving requirements tracing based on framing it as an information retrieval (IR) problem, they have started with a classical vector space model algorithm but they found that this algorithm does not outperform analysts or existing tools in terms of recall or precision. In [40], the researchers describe a new technique called DROPT (Document Ranking Optimization) developed for documents retrieved from a corpus with respect to document index keywords and the query vectors, the purpose of DROPT id to reflect how human users can judge the context changes in IR result rankings according to information relevance. In [41], the author represent an algorithm based on classical mathematical expressions for calculating similarity between groups, known as the cosine, jaccard and dice formulas. In [42], an application of information retrieval to software maintenance and in particular to the problem of recovering traceability links between the source code of a system and its free text documentation by using an IR method (Vector Space), and in [43] a source code parser has been described to create a syntax tree as a processing step for information retrieval to recover traceability links between source code and documents. Several works of recovering of traceability links between documentation and source code by using Latent Semantic Indexing (LSI) like an information retrieval technique.

About the use of Machine learning, in [44] the author have proposed a predictive model for classifying potential traceability links in a system as either valid or invalid, by using a Machine Learning approach and features such as text retrieval (IR), rankings and query quality (QQ) metrics. In [45], the author also uses a machine learning approach for generating requirements traceability relations, this approach is based on a learning algorithm that produces traceability rules which are able to capture link between requirements statements specified in natural language and object models.

3. Improving document management using traceability and information retrieval in a big data environment

3.1. Context

The main goal of this work is to build a scalable model for classifying documents and extracting information from them. In a corpus of documents, the model extracts the text from each document and applies vectorization to it. Vectorization consists of counting the number of occurrences of each word appearing in a document. The vectorized representation is then entered into the model for prediction. Vectorization includes several algorithms and techniques in order to extract as much information as possible. Once this information is extracted and after dividing the data into a test set and a validation set, we apply different classification models on the data (VMS, KNN and decision trees). With arbitrary tests, the measurements showed very attractive results: precision, recall and F-measure between models, but the best result was recorded using TF / IDF with an accuracy of 0.934, an accuracy of 0.942, a recall of 0.905, and an F measure of 0.921.

When a user initiates a query, the query has multiple terms, and some of those terms may be more important than others. TF-IDF is a popular technique for measuring and getting an idea of the importance of a term in relation to a document. The TF is used to measure how often a word appears in a document. The IDF measures the importance of a word (Table 1). The TF-IDF value increases if a term appears in the document a number of times while taking into account the number of existing documents in the corpus that contain the same term.

Table 1: Example of frequency of appearance of a term in a document

	Document 1	Document 2	Document 3
Term 1	5	2	1
Term 2	1	0	0
Term 3	3	1	0
Term 4	0	4	3

The following formula will allow us to calculate the weights of the terms in each document:

$$\mathrm{Poid}(t_i, d_j) = \mathrm{TF} * \mathrm{IDF} \quad (1)$$

with:

$$\mathrm{TF} = \frac{f(t_i,d_j)}{N} \quad (2)$$

$f(t_i,d_j)$ is the number of occurrences of the term t_i in the document d_j and N is the total number of terms in the document d_j.

and

$$\mathrm{IDF} = \frac{\log\left(f(t_i, d_j)\right)}{M} \quad (3)$$

$f(t_i,d_j)$ is the number of occurrences of the term t_i in the document d_j and M is the total number of documents in the corpus.

We say that two objects are similar if their vectors are confused, so the similarity used in our work is the cosine similarity, this measure uses the complete vector representation based on the frequency of objects or words. The formula is defined by the ratio between the scalar product of the X and Y vectors and the product of the norm of X and Y:

$$\mathrm{Sim}_{\cos}(X, Y) = \frac{\sum_{i=1}^{n} x.y}{\sqrt{(\sum_{i=1}^{n} x^2)}.\sqrt{(\sum_{i=1}^{n} y^2)}} \quad (4)$$

A document is made up of a joint agreement of terms that have various patterns of occurrence. Text classification is an important task especially for document processing, but with the expenential growth of data, there are algorithms that can improve classification efficiency while ensuring the high precision desired. The purpose of classification is to organize documents by sets of documents that are similar, and theoretically the documents are said to be close. This phase aims to classify documents for the sake of research and extraction of useful information in an efficient manner. In this phase we propose to use VMS with the KNN (K Nearest Neighbor) algorithm and the decision tree algorithm in parallel, which makes it possible to put a document in a dominant set among its k nearest neighbors. A similarity measure makes it possible to determine the documents closest to the request launched by the user or, in the

case of classification, to determine the sets closest to the document to be classified.

KNN is a popular algorithm in the field of classification. Its use as well as its results place it among the best classification techniques in [46] [47]. As for the decision tree, it is an algorithm that makes it possible to perform an iterative partitioning of the entire feature space; The tree predicts the same label for each lower category and each category is chosen by selecting the best division from a set of possible divisions, which maximizes the information gain at a tree node [48]. However, a decision tree in the sense of our work will allow to classify data or observations already labeled as well as the visualization of the processed data; The tree starts with a source, where there are all the observations, then several branches in series show the nodes and at the end we find the leaves which represent the classes to be predicted. The major goal of a decision tree is to produce user-understandable classifications.

When we have a request, we must find the relevant objects, however the way to evaluate a document, whether relevant or not, is to calculate the similarity between the request and the object. Before calculating the similarity, it is important to index, clean and format all the contents of all the documents by using the techniques of the recovery of the traceability of the data (links, who, when, how, where ... etc.) as well as the classification using a machine learning. The vector representation will also allow us to facilitate the use of the requests made where each element of the vector represents the weight of each term or concept in the document or in the query. The objective is to extract all the terms (t) or concepts from our corpus and for each document (d) we build a vector (v) which represents it, if a term exists in the document we calculate its weight, otherwise we put 0, at the end of the operation we will have a vector for each document to calculate the similarity between the documentation and the purpose of the research done.

The traceability processed in this work represents an entire history of the operations carried out on the documents. Once an operation is performed by a user, this action is tracked (consultation, modification, deletion…). The traced data makes it possible to know, at any given moment, the life cycle of the documents as well as the associated operations. Each document with unstructured data must be converted into a format that can be analyzed in order to structure it and extract useful information. The content of a document can be represented as a set of terms such as words, sentences or other entities, each term will have its weight, which gives its utility in relation to this document. The main objective of document exploration is then to allow administrators to extract information from textual resources and to manage operations such as retrieval of traceability links and classification.

3.2. Traceability Information retrieval

In terms of our approach, we take into consideration any operation, any action and any event such as a trace stored in the database. We present a traceability link by a triple (source document DS, target document DC and similarity S). In order to apply IR techniques for the proposed approach, we consider all artifacts as a text document first. The IR allowed extracting all terms from a DS source document by calculating the similarity to another DC target document, depending on the document type and after processing, in order to extract the information. The high similarity between two documents suggests a potential link between them. For each DS, preprocessing done according to the type of document (image, video, text, etc.) for which we have thought of using machine learning techniques, we will generate the necessary traceability information in the database.

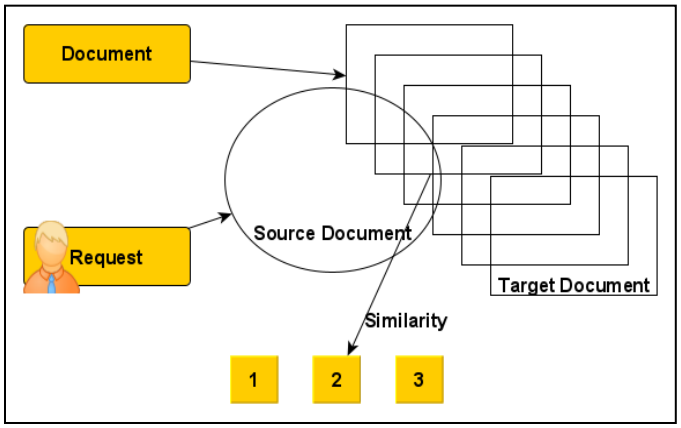


Figure 1: Similarity calculation and recovery of traceability links

We take as input, the construction of a matrix which concerns the terms per document (m × n), with m the number of all the terms appearing in the documents and n the number of documents. For each cell c_{ij} of the matrix, with i represents the term number and j the document number. We calculate the weight of each term and by applying the IR we calculate the similarity between two documents. The traceability links are created for each two documents with the value of their similarity. The latter is calculated by the positive cosine. For the recovery of traceability links, we aim to trace everything related to documents, even the traceability links in the document itself if it is divided into several parts.

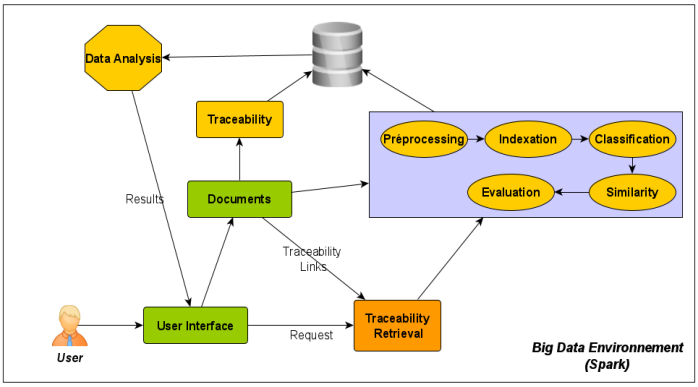


Figure 2: Process of the proposed approach

The interest in the need in Big Data is increasing day by day. In the age of big data, variety dominates volume and speed. Several difficulties and challenges may arise in acquiring, storing, processing and retrieving data. Companies like Google, Amazon, Facebook and YouTube have an advantage over others in personal development and competition due to their large amount of data. It should be remembered that processing data and finding information is as important as storing data. At the enterprise level, information can be structured or unstructured. In the first case, the information is presented in the form of a database and in the second case; the information is presented in the form of documents. Documents capable of keeping secrets that sometimes occupy special confidentiality in the company.

The architecture of the proposed approach is represented as shown in Figure 2.

The traceability of operations as well as the recovery of traceability links between documents is an important phase in order to reduce the number of documents processed when the user launches a search query. All of the document techniques algorithms help automate the process to have better results and identify relevant documents with very good precision. The Spark environment takes into account the sheer mass of data and its exponential growth, as well as the analysis done, in order to meet needs and help make decisions.

For the implementation of our approach we have carried out tests on tools to show what the approach can guarantee for the improvement of document management. We used KNIME (KoNstanz Information MinEr) [49], a software developed in 2004 by the German University of the city of Constance (Konstanz). It is open source software written in java. KNIME is a tool that allows the design and execution of workflows using nodes that are already predefined as components of the platform. What is also huge is that this platform also allows the processing of large volumes of data depending on the resources of the user's machine (storage capacity, RAM, etc.); That is, there are no restrictions with regard to the volume of data or with regard to the platform itself. KNIME is an easy-to-use application, what you just have to respect is the order of operations as well as the logic of the process so you have to choose the node corresponding to the task seeking to be executed in order to make the execution of the workflow successful ; In KNIME, a node represents an entity that performs a well-defined task.

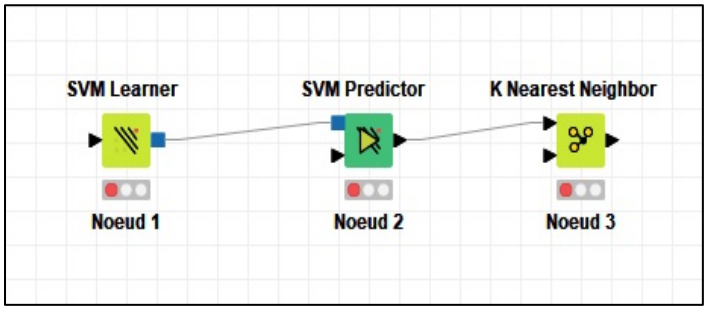


Figure 3: Example of node in KNIME

Figure 3 shows an example of nodes corresponding to the use of the VSM model with the KNN algorithm; The first node is used to train an SVM model against the input data. The second node uses the SVM model generated by the first SVM node to predict the output for given values. The third node classifies a dataset based on the KNN algorithm using the training data. A node consists of input, output and its state; if the state is in red it means that the node is not configured, if the state is in yellow it means that the node is well configured and if it is green, it means that the node is well executed, otherwise there is an error.

The visualization of the precision statistics shows in figure 4, the good precision taken with respect to the sets processed using the three types of VSM algorithm, Decision tree and KNN.

This approach is presented in the form of examples of tests in order to show the interest and the advantages brought to improve the management of documents by a system which allows to have a global vision on the history of the operations made on the documents. , as well as the relationship between them. On the other hand, it also makes it possible to facilitate the processing and application of information retrieval algorithms in order to extract adequate information according to specific needs. The use of the big data environment, especially Apache Spark, will allow to perform distributed calculations on a large volume of data and in a reasonable time. Our approach does not differ much from other approaches, but the proposed process has never been discussed in relation to electronic document management systems. On the one hand, we offer an important management of the traces made on the documents as well as the links between them, and on the other hand, by increasing the speed and the ease of treatment we offer the techniques of information retrieval as well as analyzes the distributed computing offered by Spark. While, the other works generally offer a classic system for managing documents manually. The works that offer tools for the electronic management of semi-automatic documents are rare, and those already existing are limited to private and commercial use.

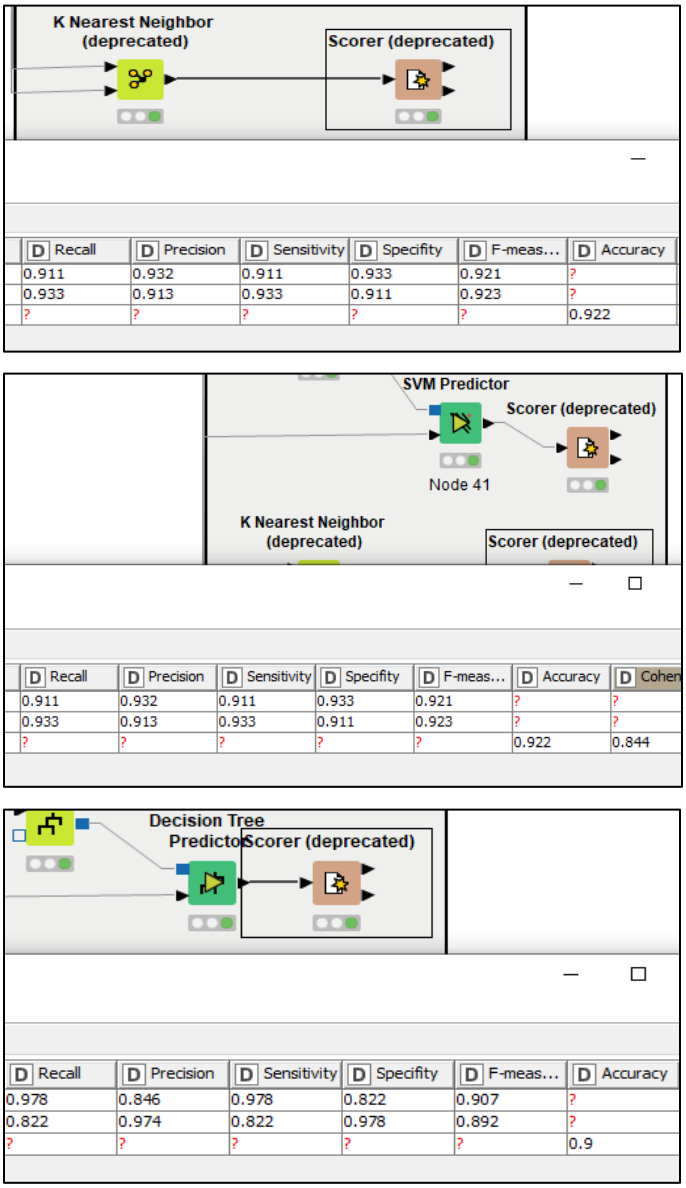


Figure 4: Precision statistics

4. Using Contents Recognition techniques in document management

4.1. Context

Several approaches to machine learning have been applied successfully to real world problems. Indeed, the choice of method in each case seems to be based largely on the experience and preference of the researchers involved. This success is also apparent in information retrieval applications, the different paradigms being used in similar proportions to those in other fields. Much of the information retrieval work can be automated. Processes such as document indexing and query refinement are usually done automatically, while document classification and term selection are more often done semi-automatically or manually. However, algorithms that classify documents by types to index information and model user interests to help them formulate queries, reduce workload, and ensure more consistent behavior. These algorithms are based on machine learning, a dynamic and growing field of computing.

The objective of this contribution is to propose an approach to improve document management based on information traceability retrieval but also on effective machine learning to extract data, create a model and provide a classification. Implementing such a solution could reduce administration costs, speed up the document delivery process and improve customer satisfaction in businesses. At the level of our proposal, we present a kind of multimedia management system; the use of machine learning also helps us in the automatic classification and extraction of data when a user adds any document (image, video, audio or text) to the corpus, using classification algorithms and content recognition algorithms. Until now, most companies simply scan these documents, index them with a date and number, and store them in a repository. There is currently a wide range of specialized algorithms for certain parts of document analysis. In large scale applications, these approaches must cope with the wide variety of documents. Machine learning algorithms are a possible treatment in this situation. From a set of documents, these algorithms are able to extract a lot of relevant data, and even they are able to weight this data in such a way that changing some information does not lead to a loss of performance. We aim in this work to propose an automatic classification of documents in predefined categories as well as by document type (image, video, audio or text). Supervised learning techniques are used for automatic classification, where predefined category labels are assigned to documents based on the probability suggested by a set of document training. Some of these techniques are described below. Several algorithms or combination of algorithms as hybrid approaches have been proposed for the automatic classification of documents. Among these algorithms there are VSM, decision trees and KNN (k-nearest neighbors) and their hybrid system with the combination of different other algorithms and techniques [50].

Computer systems that could learn to predict from a set of data and improve without having to be reprogrammed were a dream until recent years. But now, it has been made possible through machine learning. Today, machine learning is the most widely used branch of artificial intelligence which is adopted by major industries to benefit their businesses. Machine learning unveils a breakthrough science in how computers can learn and make predictions. It has applications in various industries and it is widely used. Machine learning has grown in popularity since its inception and it won't stop immediately. Data scanning is on the increase to help cut costs and save time, but simply converting documents using a scanner into an image or PDF file is not enough. A document in printed or handwritten form can be modified or edited at any time, which is not possible for a converted image. Thus, content recognition becomes a requirement to add search and edit functionality in a scanned document.

The most important documents and files today are stored in digital format. This is because digital storage is simple and does not involve physical storage space. In addition, the digital retrieval of documents is much easier than with filed paper documents. However, scanned documents are treated as image files and cannot be searched or edited because they are not machine readable. Optical Character Recognition (OCR) is a solution to convert these scanned documents into machine-readable text files. Files processed with OCR can be searched and edited as needed. Recognizing content in a document management system will therefore be a new approach that promises to transform the way businesses manage document processing. It will aim to recognize invoices, tax forms, survey forms and various other business and administrative documents which may be formal or loosely structured with proper storage and retrieval of these documents for business purposes.

4.2. Machine learning for document management

The essential goal of using machine learning in this work is to apply it to multimedia documents and also to have a classification, initially, by type of document (Image, video, audio or text), then and thanks to these content recognition algorithms we will try to extract the information that will be used on the information retrieval part. After this processing on the content of the documents, an automatic categorization is proposed. The automatic classification of documents is quite a complicated task. All the varied documents have different visual parameters: size, format, shade of paper, font size and type. The fact that some of the documents are handwritten complicates things and it was very difficult to distinguish the handwriting. Documents vary to a great extent and new documents are expected. This is why it is almost impossible to solve this task by software implementation of algorithmic solutions but by offering machine learning, this processing by learning algorithms seems reasonable.

Recognizing content in a document management system will therefore be a new approach that promises to transform the way businesses manage document processing. It will aim to recognize invoices, tax forms, survey forms and various other business and administrative documents which may be formal or loosely structured with proper storage and retrieval of these documents for business purposes.

4.3. Recommended documents

Spark MLlib is an Apache Spark module that provides machine learning primitives as APIs. Machine learning typically processes a large amount of data. Spark's basic IT framework is a huge plus. Additionally, MLlib provides most of the popular machine learning and statistics algorithms. This greatly simplifies the task of working on a large-scale machine learning project. MLlib offers recommendation techniques for unsupervised learning and

especially in the part of clustering that will be proposed in this work. Recommendation systems are software tools and techniques that provide suggestions of documents for a user to use. For our recommendation tool to be effective, it must have a comprehensive and properly cataloged resource repository. In addition, it must be precise to understand the user's needs and must essentially profile the user correctly. Ideally, you should take into account that a user's preferences are constantly changing. To address this need, we provide an interactive document recommendation tool that observes the different themes that arise in user work over time and continues to learn user preferences continuously based on feedback. of the user on the documents recommended to him. We then estimate the similarity between the query and each document using processed by the IR process presented in the first contribution. To take into account user preferences, we keep track of user documents to ensure traceability throughout the process.

Our recommendation phase depends on the user's profile. A profile can be made up of several types of information. In this proposition, we distinguish two types of information:

- A template of user settings, that is, a description of the types of documents that are of interest to the user. There are many other possible representations of this description, but a common representation is a function that, for any item, predicts the likelihood that the user will be interested in that document. For efficiency reasons, this function can be used to retrieve the n documents most likely to interest the user.
- A history of user operations with the document management system. This may include documents the user has viewed. The history also includes the recording of requests entered by the user.

5. The proposed system

5.1. Architecture

The architecture of the system proposed in this section covers most of the components involved in data processing

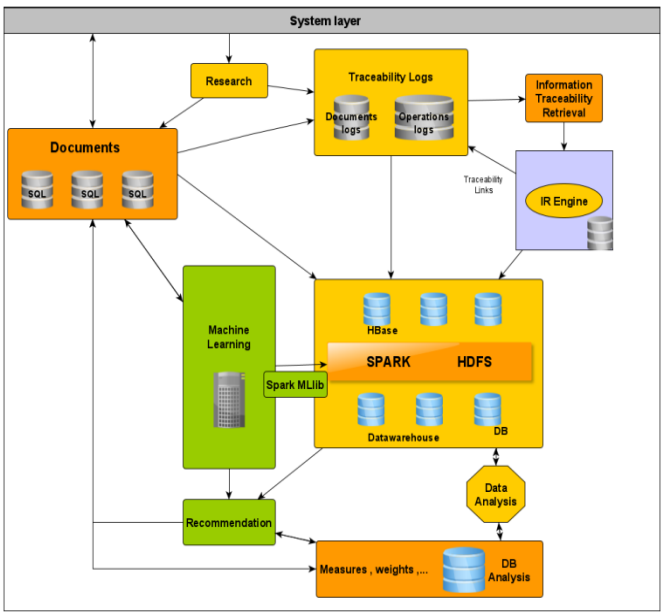


Figure 3: Architecture of the proposed system

Our system architecture consists of four main components:

- Corpus of documents: The documents coming from the corpus are heterogeneous. The user uses their interface to add new documents, even just with the action of "click and drag". Usually this is SQL data storage; (MySQL, Oracle, PostgreSQL, MongoDB, etc.). Log files will be used to manage traceability (user clicks (Action), user visits (Operation), activity, etc.).
- Data transformation: Data transformation involves the information retrieval part and converting documents into vectors, as well as data extraction so that the document can be indexed for easy searching. Indexing is a way to classify a document by adding terms to its metadata, such as tags, order numbers, or customer information. Mainly this phase is used to load the data to be processed. Logging and traceability management tools can also be considered part of data transformation, as they generate useful events from log files and present an important history to analyse.
- Data processing or integration: For data processing, this concerns the generation of usable data from a set of heterogeneous documents that can be consumed by machine learning and the big data environment; (Spark and its ecosystem (Spark / HDFS, Map-Reduce, HBase, MLlib, etc.). Spark offers in-memory scans, so this is an emergent task by reading directly from HDFS.
- Data visualization and recommendations: Visualization and recommendation represent forms of display of results and information to users or to other entities such as managers or decision-makers, in order to predict or estimate overall performance and also to generate recommendations, to using a set of algorithms, tasks defined by the user himself. (Reporting, personalized dashboards, interactive tables, statistics, research results, etc.).

5.2. Benefits and advantages

The improvement that is being made to document management systems today through this work is to offer new functionalities while keeping the classic functionalities of the system (Capture, Storage and distribution of documents):

- Traceability management: traceability management allows managers to control and track changes made to documents in the system. Traceability ensures that all actions and operations are tracked and recorded.
- Integration: Many document management systems integrate with programs such as messaging, a CRM (Customer Relationship Management) application or an ERP (Enterprise Resources Planning) database.
- Scalability: document management system should grow with the business so you don't have to change systems later.
- Security: The system must prioritize the security and protection of information and documents, as a data breach or storage problem could be disastrous for the entire organization.

- User-friendliness: The document management system should be easy to use. Users in the company should be able to easily access, manage and browse the necessary documents.
- Collaboration: The system must allow users to share and collaborate on documents.

The proposed improvement of the document management system supposed to make the work easier and more efficient, bringing many benefits: The system automates many aspects of document management. The system makes it possible to follow and see all the activities on a given document. It should apply built-in security and access controls to regularly control who can access which documents. The system ensures easy search of documents so as not to waste time and money. This makes it possible to find a document in seconds and also helps to retrieve some more relevant ones. It facilitates information sharing and collaboration, allowing documents to be accessed from multiple locations. The user can share documents, monitor workflows, grant or deny access to their documents and see what changes have been made. Also, the system offers users documents according to their profile and the history of their activities on the documents.

6. Conclusion

Using a document management system powered by machine learning, companies can modernize the way data is extracted. Such an advanced system can read information precisely to understand its intention. Based on specific models, the analysis of the document management system activated by machine learning and retrieves only the information useful to accomplish a particular job. Machine learning can also be a powerful tool for data mining and analysis in this kind of system, which will save a lot of time and effort. It is useful to examine the accuracy to 10 returned documents if one is interested in the system's ability to return relevant documents to the top of the list (which is a traditional concern of search engine users). The precision at 5, 10, 30,… returned documents nevertheless has limits: for example if a given request has only 8 relevant documents, and the processing does restore these 8 documents at the top of the list, the approach will have a precision of 10 returned documents equal to 0.8, which does not show that all the relevant documents available were found. In addition, in this example, a precision of 10 restored documents equal to 0.8 does not make it possible to determine where the two irrelevant documents are located among the ten returned. Document traceability management provides each user with proof that business processes have been respected, because if there is an error, it will easily be traced, as well as the associated time saving and improvement. Project management especially at the level of collaborative work. Several jobs are needed to improve the performance and accuracy of the document classification process. New methods and solutions are needed to obtain useful information from the growing volume of electronic documents:

- Use of semantics and ontology for document classification and information retrieval.
- For filtering and categorizing documents, the user may have folders and may want a classifier to classify each incoming document which automatically moves it to the correct folder. It is easier to find documents in sorted folders in a very large corpus.
- Reduce classifier processing and response time and improve classification accuracy, precision, and recall.
- An implementation of a meaning-based text classification procedure is required to retrieve meanings from words used in a specific context.

VSM stands out and occupies first place by its performance in all three cases. Thus, it confirms its good reputation in the literature as being one of the best classifiers. With the constant development and adaptation of machine learning, the proposed approach to document management systems will not only ensure better data analysis to meet customer needs, but also establish optimal research results. With the rapid pace at which technology and innovations invade industries with each passing day, it goes without saying that early adopters will also be the first to reap the myriad benefits that come with it.

In an electronic document management system, the majority of solutions on the market today are more concerned with the dynamic side of the system which concerns the workflow part of the document. In this work, the major objective that we tried to study is the static part by proposing for the improvement of this kind of system a multitude of techniques and approaches for this static part by attributing it a kind of fuel in order to automate it and make it as dynamic in relation to user expectations and its needs for the organization and readability of documents.

The first contribution in this direction, was the proposal of the use of traces throughout the system, thus the recovery of traces between documents in order to reduce the dimension of the treatment proposed also on documents using the techniques of information retrieval and the extraction of payload data. The voluminous mass of data has forced us to think about using a big data environment which will also help us in the analysis and distributed calculation of data.

The second contribution is also an important proposal, concerns the use of machine learning techniques in order to add to this system new types of documents (image, video, audio and text) which will be processed by proposing powerful algorithms. in this sense, and which are used to extract important data, and therefore offer users and businesses a new service in this type of system. An image or a document scanned in an electronic document management system, without it being processed manually by adding descriptions and fields to list it, it will be archived all the time. Hence the usefulness of our approach, in order to automate these methods and extract text content assembled from the metatdata of a document. This allows the IR engine and especially the indexing part which relies on the texts, the processing and the indexing of the images of the document. Because it is very easy to work on text than to work on an image. That is why we have proposed the use of content recognition algorithms which greatly assist in automatic document processing. An important point also, regarding the use of machine learning is that the preprocessing done on the images using OCR allows to extract high value textual data added during the training of an algorithm, and the The extraction of this data generates a serious increase in the volume of data, hence the aim of using a big data environment. This automated processing of multimedia documents prompted us to think of the user by offering an integrated module for recommending multimedia documents according to two criteria;

their profile settings and the history of their operations in interaction with the system.

The system proposed in this article is an improvement of the modern electronic management systems offered today. Thanks to this proposal, documents management tools also becomes easy to use, especially with an environment such as big data.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] A.W. Qurashi, V. Holmes, A.P. Johnson, "Document Processing: Methods for Semantic Text Similarity Analysis," in 2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), 1–6, 2020, doi:10.1109/INISTA49547.2020.9194665.

[2] T. Dash Roy, S. Khatun, R. Begum, A.M. Saadat Chowdhury, "Vector Space Model based Topic Retrieval from Bengali Documents," in 2018 International Conference on Innovations in Science, Engineering and Technology (ICISET), 60–63, 2018, doi:10.1109/ICISET.2018.8745587.

[3] Jian-Jun Zhou, "Study on several confidentiality protection technologies for electronic document," in Proceedings 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC), 2282–2285, 2013, doi:10.1109/MEC.2013.6885423.

[4] N. Mustafa, Y. Labiche, "The Need for Traceability in Heterogeneous Systems: A Systematic Literature Review," in 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), 305–310, 2017, doi:10.1109/COMPSAC.2017.237.

[5] C. Benkoussas, P. Bellot, A. Ollagnier, "The Impact of Linked Documents and Graph Analysis on Information Retrieval Methods for Book Recommendation," in 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 385–392, 2015, doi:10.1109/WI-IAT.2015.200.

[6] A. Boulmakoul, Z. Besri, Enterprise Organization Assessment through Structural Analysis Framework, 2013.

[7] R. Wohlrab, J.-P. Steghöfer, E. Knauss, S. Maro, A. Anjorin, "Collaborative Traceability Management: Challenges and Opportunities," in 2016 IEEE 24th International Requirements Engineering Conference (RE), 216–225, 2016, doi:10.1109/RE.2016.17.

[8] X. Jia, Z. Ma, "A Topic-based Document Retrieval Framework," in 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, 860–864, 2012, doi:10.1109/FSKD.2012.6234372.

[9] A. Demuth, R. Kretschmer, A. Egyed, D. Maes, "Introducing Traceability and Consistency Checking for Change Impact Analysis across Engineering Tools in an Automation Solution Company: An Experience Report," in 2016 IEEE International Conference on Software Maintenance and Evolution (ICSME), 529–538, 2016, doi:10.1109/ICSME.2016.50.

[10] S. Palihawadana, C.H. Wijeweera, M.G.T.N. Sanjitha, V.K. Liyanage, I. Perera, D.A. Meedeniya, "Tool support for traceability management of software artefacts with DevOps practices," in 2017 Moratuwa Engineering Research Conference (MERCon), 129–134, 2017, doi:10.1109/MERCon.2017.7980469.

[11] S.K. Shivakumar, Digital Asset Management and Document Management, IEEE: 253–271, 2017, doi:10.1002/9781119206842.ch8.

[12] S. Irfan, B.V. Babu, "Information retrieval in big data using evolutionary computation: A survey," in 2016 International Conference on Computing, Communication and Automation (ICCCA), 208–213, 2016, doi:10.1109/CCAA.2016.7813720.

[13] A. Gandomi, M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," International Journal of Information Management, **35**(2), 137–144, 2015, doi:10.1016/j.ijinfomgt.2014.10.007.

[14] Big Data, Fast Data and Data Lake Concepts, Procedia Computer Science, **88**, 300–305, 2016, doi:10.1016/j.procs.2016.07.439.

[15] Introduction au Machine Learning, Machine Learnia, 2019.

[16] S. Gopal, S. Raghav, "Automatic document retrieval using SVM machine learning," in 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), 896–901, 2017, doi:10.1109/SmartTechCon.2017.8358501.

[17] F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," Egyptian Informatics Journal, **16**(3), 261–273, 2015, doi:10.1016/j.eij.2015.06.005.

[18] T. Badriyah, S. Azvy, W. Yuwono, I. Syarif, "Recommendation system for property search using content based filtering method," in 2018 International Conference on Information and Communications Technology (ICOIACT), 25–29, 2018, doi:10.1109/ICOIACT.2018.8350801.

[19] Y. Afoudi, M. Lazaar, M. Al Achhab, "Impact of Feature selection on content-based recommendation system," in 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), 1–6, 2019, doi:10.1109/WITS.2019.8723706.

[20] J.S. Mary, S. Usha, "Web based document management systems in life science organization," in 2015 Online International Conference on Green Engineering and Technologies (IC-GET), 1–3, 2015, doi:10.1109/GET.2015.7453826.

[21] T.V. Khronusova, S.V. Kruchinin, E.V. Bagrova, "Implementation of Electronic Document Management in Russian Education. Quality Assessment," in 2019 International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT QM IS), 608–610, 2019, doi:10.1109/ITQMIS.2019.8928356.

[22] S.V. Kruchinin, E.V. Bagrova, "Systems of Electronic Document Management in Russian Education. Pros and Cons," in 2019 International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT QM IS), 628–630, 2019, doi:10.1109/ITQMIS.2019.8928315.

[23] K. Souali, O. Rahmaoui, M. Ouzzif, "An overview of traceability: Definitions and techniques," in 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt), 789–793, 2016, doi:10.1109/CIST.2016.7804995.

[24] S.F. Königs, G. Beier, A. Figge, R. Stark, "Traceability in Systems Engineering – Review of industrial practices, state-of-the-art technologies and new research solutions," Advanced Engineering Informatics, **26**(4), 924–940, 2012, doi:10.1016/j.aei.2012.08.002.

[25] D. Asioli, A. Boecker, M. Canavari, "Perceived Traceability Costs and Benefits in the Italian Fisheries Supply Chain," International Journal on Food System Dynamics, **2**(4), 357–375, 2011, doi:10.18461/ijfsd.v2i4.242.

[26] H. Cui, Z. Chen, Y. Xi, H. Chen, J. Hao, "IoT Data Management and Lineage Traceability: A Blockchain-based Solution," in 2019 IEEE/CIC International Conference on Communications Workshops in China (ICCC Workshops), 239–244, 2019, doi:10.1109/ICCChinaW.2019.8849969.

[27] X. Chen, J. Hosking, J. Grundy, "A combination approach for enhancing automated traceability: (NIER track)," in 2011 33rd International Conference on Software Engineering (ICSE), 912–915, 2011, doi:10.1145/1985793.1985943.

[28] S.M. Bakhshavesh, A. Mohebil, A. Ahmadi, A. Badarnchi, "A New Subject-Based Document Retrieval from Digital Libraries Using Vector Space Model," in 2018 Federated Conference on Computer Science and Information Systems (FedCSIS), 161–164, 2018.

[29] E. Wahyudi, S. Sfenrianto, M.J. Hakim, R. Subandi, O.R. Sulaeman, R. Setiyawan, "Information Retrieval System for Searching JSON Files with Vector Space Model Method," in 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), 260–265, 2019, doi:10.1109/ICAIIT.2019.8834457.

[30] D.V. Rodriguez, D.L. Carver, "Multi-Objective Information Retrieval-Based NSGA-II Optimization for Requirements Traceability Recovery," in 2020 IEEE International Conference on Electro Information Technology (EIT), 271–280, 2020, doi:10.1109/EIT48999.2020.9208233.

[31] S. Pandey, I. Mathur, N. Joshi, "Information Retrieval Ranking Using Machine Learning Techniques," in 2019 Amity International Conference on Artificial Intelligence (AICAI), 86–92, 2019, doi:10.1109/AICAI.2019.8701391.

[32] Q. Shuai, R. Wang, L. Jin, L. Pang, "Research on Gender Recognition of Names Based on Machine Learning Algorithm," in 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), 335–338, 2018, doi:10.1109/IHMSC.2018.10182.

[33] P. Bhavsar, I. Safro, N. Bouaynaya, R. Polikar, D. Dera, Chapter 12 - Machine Learning in Transportation Data Analytics, Elsevier: 283–307, 2017, doi:10.1016/B978-0-12-809715-1.00012-2.

[34] S. Solanki, S. Verma, K. Chahar, "A Comparative Study of Information Retrieval Using Machine Learning," in: Sharma, H., Govindan, K., Poonia, R. C., Kumar, S., and El-Medany, W. M., eds., in Advances in Computing and Intelligent Systems, Springer, Singapore: 35–42, 2020, doi:10.1007/978-981-15-0222-4_3.

[35] C. Goller, J. Löning, T. Will, W. Wolff, Automatic Document Classification - A thorough Evaluation of various Methods, Undefined, 2000.

[36] L. Getoor, Link-based Classification, Springer, London: 189–207, 2005, doi:10.1007/1-84628-284-5_7.

[37] C. Mills, J. Escobar-Avila, S. Haiduc, "Automatic Traceability Maintenance via Machine Learning Classification," in 2018 IEEE International Conference on Software Maintenance and Evolution (ICSME), 369–380, 2018, doi:10.1109/ICSME.2018.00045.
[38] R. Oliveto, M. Gethers, D. Poshyvanyk, A. De Lucia, "On the Equivalence of Information Retrieval Methods for Automated Traceability Link Recovery," in 2010 IEEE 18th International Conference on Program Comprehension, 68–71, 2010, doi:10.1109/ICPC.2010.20.
[39] J.H. Hayes, A. Dekhtyar, J. Osborne, "Improving requirements tracing via information retrieval," in Proceedings. 11th IEEE International Requirements Engineering Conference, 2003., 138–147, 2003, doi:10.1109/ICRE.2003.1232745.
[40] K.K. Agbele, E.F. Ayetiran, K.D. Aruleba, D.O. Ekong, "Algorithm for Information Retrieval optimization," in 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 1–8, 2016, doi:10.1109/IEMCON.2016.7746242.
[41] (PDF) Algorithm for calculating relevance of documents in information retrieval systems | IRJET Journal - Academia.edu, Nov. 2020.
[42] Information retrieval models for recovering traceability links between code and documentation - IEEE Conference Publication, Oct. 2020.
[43] S. Nagano, Y. Ichikawa, T. Kobayashi, "Recovering Traceability Links between Code and Documentation for Enterprise Project Artifacts," in 2012 IEEE 36th Annual Computer Software and Applications Conference, 11–18, 2012, doi:10.1109/COMPSAC.2012.10.
[44] C. Mills, S. Haiduc, "A Machine Learning Approach for Determining the Validity of Traceability Links," in 2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C), 121–123, 2017, doi:10.1109/ICSE-C.2017.86.
[45] G. Spanoudakis, A.S. d'Avila Garcez, A. Zisman, "Revising Rules to Capture Requirements Traceability Relations: A Machine Learning Approach," in Proceedings of the Fifteenth International Conference on Software Engineering & Knowledge Engineering (SEKE'2003), Hotel Sofitel, San Francisco Bay, CA, USA, July 1-3, 2003, 570–577, 2003.
[46] I.D. Schizas, "Online Data Dimensionality Reduction and Reconstruction Using Graph Filtering," IEEE Transactions on Signal Processing, **68**, 3871–3886, 2020, doi:10.1109/TSP.2020.3003423.
[47] J. Laaksonen, E. Oja, "Classification with learning k-nearest neighbors," in Proceedings of International Conference on Neural Networks (ICNN'96), 1480–1483 vol.3, 1996, doi:10.1109/ICNN.1996.549118.
[48] K. Taunk, S. De, S. Verma, A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," in 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 1255–1260, 2019, doi:10.1109/ICCS45141.2019.9065747.
[49] KNIME | Open for Innovation, Oct. 2020.
[50] O. Harrison, Machine Learning Basics with the K-Nearest Neighbors Algorithm, Medium, 2019.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1278-1285 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Time Granularity-based Privacy Protection for Cloud Metering Systems

Hesham Aly El Zouka[*,1], Mustafa Mohamed Hosni[2]

[1]*Computer Engineering Department, Arab Academy for Science, Technology, and Maritime Transport, Alexandria, 22206, Egypt*

[2]*Managing Director of OMIKRON Technologies, Electrical Engineering Department, Alexandria Faculty of Engineering, Alexandria, 22206, Egypt*

ARTICLE INFO

Article history:
Received: 22 August, 2020
Accepted: 25 September, 2020
Online: 16 December, 2020

Keywords:
Smart Meter
Zero Knowledge Proof
Authentication
Security
Privacy
Cloud

ABSTRACT

Due to the advent of cloud computing and Internet of Things, smart meters have become a crucial part of smart cities. Smart meters generate vast amounts of fine-grained data that can immediately provide useful information to electricity consumers, such as automatic billing, load monitoring, and dynamic time pricing. This will make the electricity consumers more energy-efficient and self-organized in IoT and cloud computing technology. Smart meters also enable two-direction flow of energy and balance the measurement of inflows and outflows of energy between the meter and electric utility's central system through remote reporting. Besides, they help designers integrate renewable energy generation, such as solar rooftops to cover remaining energy needs. However, smart meters entail some severe security challenges. From a security point of view, these meter readings have a potential threat on personal privacy and the collected meter data. They may also reveal some details about consumer's lives. Thus, the data acquisition from those meter readings is expected to incorporate privacy requirements and application demands. In addition, the collected raw data from various meter readings needs to be collected from IoT smart metering sensors and stored in the cloud to be accessed and analyzed by electricity consumers. Subsequently, this considerable information Is prone to privacy threats and security vulnerabilities as the transmission of data to the cloud is most likely to cause a data breaching. In this paper the privacy and confidentiality of meter readings is protected through trusted platform module, which will allow the electricity provider to get the energy consumption at a periodic time granularity.

1. Introduction

The future smart meters are expected to provide details on individual appliance consumption and enable customers to access electricity safely and economically [1]. These technologies lead the trend to shift from conventional power grid to smart grid that uses sensors, smart meters, advanced communication systems and information technologies to create monitoring and visualizing real-time electricity information. But these technologies pose threats to individual's privacy as they can simply reveal one's daily activities [2]. Fine-grained electricity consumption aggregation scheme for smart meters can be used to reveal some information about consumers' lives and their activities. It can be also used to detect the presence or absence of occupants [3], [4]. In addition, transmitting the smart meter data to private or public cloud storage will allow the household consumers to monitor their usage, change their energy habits, adjust their electric power consumption, and give them better management to their energy costs. Real-time monitoring, then, can be accessed via the cloud using mobile browser, or any PC/Mac browser in the world. Subsequently, these meter readings and information are prone to privacy and security threats as the transmission of data to the cloud is most likely to cause a data breaching. In this paper, a storage outsourcing in the cloud is considered secure where the external storage is often untrusted or only semi-trusted. In addition, privacy-preserving analytics for IoT and cloud-based smart metering system is ensured by using a trusted hardware module which is used in conjunction with generated random number generator. The proposed framework guarantees an

[*]Corresponding Author: Hesham El Zouka, AAST, P.O. Box: 1029 Alexandria, Egypt, helzouka@aast.edu

improved secure data storage of IoT smart meter reading and provides real-time measures of electricity consumption and the estimations of the consumption during operations.

In this system, the consumer grants the service company a meter reading request service at a time of charging and the granted utility company is only allowed to get the power readings during that granted time. This system also provides practical security measures and protects the privacy of consumers who have smart meters at their homes. Also, a privacy model is provided here to achieve the privacy requirement and experience that the proposed system may have a significant effect on privacy-preserving systems if the system is built under the assumption that the utility provider is semi-honest.

In addition, strong access configuration and authentication management protocols should be constructed in IoT - cloud outsourcing process, and any unauthorized access should be utilized and reported. Moreover, due to the lack of privacy and security issues, sensitive personal data can be revealed to third parties such as utility companies and cloud service providers. Today, energy might be collected and stored for future use as per demand. Additionally, data analysis programs can be used to improve the efficiency of power consumption of a particular meter.

It is true, of course, that adding memory or a processor to the smart meter will enhance its security services capabilities, but contradictorily, the cost of implementing these smart meters will be increased dramatically. The proposed security system in this paper is designed to support multiple time segment granularities and let the meter readings to be stored securely on the cloud. It also comprises the design, construction, and operation of proposed two-way communication scheme, which ensures secure and reliable connections between the smart meters and getaways.

The rest of this paper is structured as follows. Section II discusses the architecture and characteristics of smart metering system with the cloud computing. The related work is given in Section III. The system design and modeling are presented in Section IV. Results in terms of system implementation, authentication, security and privacy are presented and discussed in Section. V. The performance analysis of the proposed system is discussed in Section VI. In the final section, the conclusion and future work are presented.

2. Cloud Metering Infrastructure

Smart technologies like the IoT, artificial intelligence (AI), and other digital technologies are providing new opportunities to smart grid network, and can be used to control, monitor, maintain production, transmission, and consumption of electricity in smart metering systems. Many benefits could be obtained through the deployment of the smart grid services in cloud computing, in order to achieve energy savings. Cloud computing is also used to speed up communication with telemetry devices [5]. Moreover, power grid technology can be combined with smart meters to reduce the energy consumption in the peak demand and, hence, improve cost savings.

However, fine-grained technical approaches can cause security and privacy concerns for smart meter users. Current smart/interval meters can record electricity consumption in 30-minute intervals; however, the new generation of smart meters use more advanced technology and could convert all time units to seconds. Mechanisms for large-scale distributed sensor networks and other complex issues are likely to be addressed in order to prevent personal privacy from being breached and prevent revealing detailed information of one's daily activities. It should be made clear that smart meters record the energy usage exactly in the same way as the standard meters, and send this data to the consumers and the utility provider, on a required time interval, which allows consumers to monitor their power consumption at very fine time granularities. As opposed to the conventional method of electricity billing [6], the proposed approach does not require a person from the electric companies to read the number of units of energy consumed in the meter. The proposed system can play an important role in minimizing overall energy consumption and, hence, making significant contribution to energy efficiency and power saving.

One of the most important aspects of the proposed privacy-preserving energy scheme is its capability of providing a fine-grained smart meter access, while preserving the privacy of consumers based on an encryption scheme.

Therefore, the proposed fine-grained anonymous metering system will encourage consumers to use limited electricity during peak hours, especially when energy demand is high and inform them with their consumption patterns, and challenge them to fully utilize potential while saving costs. Finally, a Load Monitoring Center (LMC) will be introduced in attempt to access the sum of meter readings from various smart meter-time resolutions (e.g. half-minute, one minute, five minute, etc.).

Some previous studies have focused on privacy, security, and accuracy issues of plug load monitoring system in smart grid environment [6]. Other related studies have focused on payment predication and automatic bill generation. Various privacy-preserving models were proposed to protect the individual's privacy and minimize the energy consumption [7]. However, most of the corrupted meters will disclose the activities of users by measuring their usage frequency. Furthermore, cyber security can be utilized to characterize the surrounding magnetic field of smart meter, which in turn will expose some vulnerabilities with the data being transmitted from smart meters. To solve such problems, many researchers have developed and investigated a wide variety of approaches over the years in order to prevent disclosing the consumer living activities and invade their privacy. In this paper, cloud computing offers an alternative storage model to the traditional in-house smart meter model. This is a quite effective way, so that the data stored in the cloud will be secured by using multiple levels of encryption and, hence, only authenticated consumers would be getting access to it. All meter readings will be read at specified time intervals and readings will be stored in a secure cloud-based server which will be accessible from anywhere through a web browser or any mobile device. The meter reading privacy will, then, be protected within the cloud environment, and,

as a result, only authorized household consumers will get permission to access the sensitive content in the cloud.

The Cloud metering system (CMS) proposed here consists of three main parts: trusted identity module, cloud storage unit, and the utility service provider. The trusted identity module is embedded on a smart meter for the purpose of securing the metering services. The external cloud storage allows consumers to do their reading remotely, with no need to buy expensive equipment or hire large IT staffs. The energy service provider pools consumers' services, and demand different time granularities [9]. The proposed model has a mechanism to manage and grant meter readings to the utility company at specific time and date.

The utility company, then, can get energy readings at specific time intervals by analyzing meter data stored in the cloud. Meter readings are securely sent over the mobile network to be stored in the cloud for monitoring and analysis purposes. The meter construction and two-way communications capability will enable the system to process remote connections and disconnections of power service [10], [11]. The smart meters have the ability to measure, store, and report both energy inflows and outflows [12], [13]. The utility provider then, aggregates the consumption data from the storage system on the granted date in order to balance between the privacy requirements and the functionality of the system. Also, a privacy model is considered, as the meter readings should not be exposed to unauthorized persons or processes during transmission, storage and retrieving [14], [15]. Figure 1 illustrates the construction of the proposed CMS model.

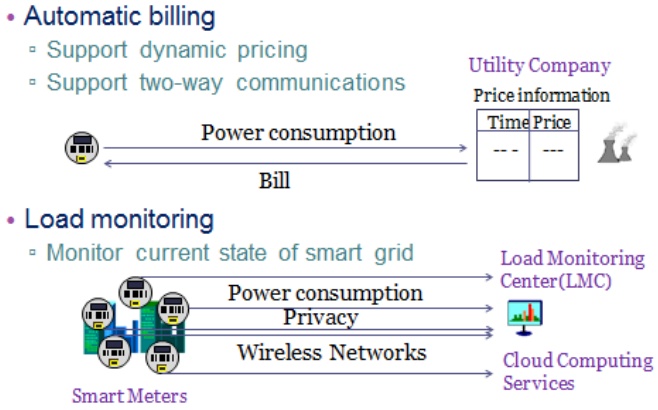


Figure 1: CMS Architecture

Clearly, the privacy is preserved by hiding smart meter's individual readings. Thus, sensitive data should be inaccessible to eavesdroppers. For example, data utilization, visualization, detection of emerging failures, centralized and decentralized operations and controls, all need to be considered when designing a secure network. There are complex security requirements to be deployed either on the cloud or on the smart meter with potentially limited resources like authentication and ID management. Thus, strong authentication protocols should be implemented in the cloud network to provide protection against privacy violations. Meanwhile, unauthorized users are rejected through the use of access control or privacy features. Without cloud, smart meters cannot preserve consumer privacy and confidentiality. On the other hand, sensitive data or meter readings should be encrypted before being outsourced to the cloud. [16], [17].

3. Related Work

Different secure architectures and data aggregations have been proposed to reduce the cost of digital signatures in smart meters. All these schemes aim to protect the transferred readings of energy data. There are different types of smart metering architectures, some are of a meter and others are of many meters. The former contribute to the billing applications and the later contribute to applications such as load monitoring.

Aggregation and the role of Trusted Third Parties (TTP} has been discussed in [18]. The role of a trusted third party is to validate and safeguard protected meter readings on the basis of privacy policies of third party services. Others have used cryptographic primitives. [19]-[21], for example, smart meters can greatly benefit from accessing cloud processing and storage services if security measures, like homomorphic encryption, are applied. This can be achieved by encrypting smart meter data before storage takes place in the cloud.

This model is quite unique when it comes to the aggregation of meters' readings using zero knowledge protocol and tamper proof signature.

As for [22] and [23], they replaced tamper proof signature by TPM. Mc Laughlinet al [24] propose an approach to privacy-preserving data aggregation being transmitted using a random noise. This random noise must be chosen in a manner that is transparent and does not affect the validity of aggregated data from smart meters. They also presented a privacy-preserving technique that is based on inserting the result of three aggregation functions for any given neighborhood in a short time period ψt. Based on these functions, the aggregation technique is implemented in such a way that the aggregator can produce an accurate aggregate number while maintaining the confidentiality of individual meter reading.

Moreover, [25] have proposed an approach that can utilize privacy settings on smart meters by using secret sharing scheme combined with symmetric encryption. The paper [26] has proposed a hybrid approach that uses random noise and pseudorandom sequences to preserve the privacy of readings collected from smart meters. Their proposed system supports load monitoring and billing applications as they claimed in their paper. However, most of the privacy preserving approaches in smart metering systems concentrate on the sensitivity of meter readings in attempt to protect time-series in general and not designed for very fine time granularities. Also, it is not clear which reading is considered sensitive and how it should be handled. While privacy-preserving solutions have been studied, users' privacy concerns have been not fully defined and the privacy of smart meter readings has not been formally addressed. Currently, many gaps exist between linking the meter readings to the status of specific applications. Appropriate regulatory protections vary from one application to other.

In this paper, the privacy of smart meter users is preserved by providing a flexible platform that supports multiple time granularities. In addition, the provided storage services of these smart metering systems are highly dependent on the cloud infrastructure. To ensure authenticity and integrity, a trusted platform module will be employed in this paper. The suggested module will be used to create signatures over the data to be signed.

4. System Design and Modeling

In this section, time scale notation, metering system and the proposed security model are described along with a brief introduction to the proposed privacy requirement.

4.1. Time Scale Notations

At the lowest level of time granularity S1, time is divided into several time units T1, T2, ….,Tn of 1 ms. Each time interval unit Ti is allocated to weight wi. At the startup routine, all the weights are equal, and the system entity structures extension to integrate weight hierarchies and time granularities into real time workflow. Automatically, a time granularity S is defined to support different applications. An example of three different time granularities, S1 , S2 and S3is shown in Figure2.

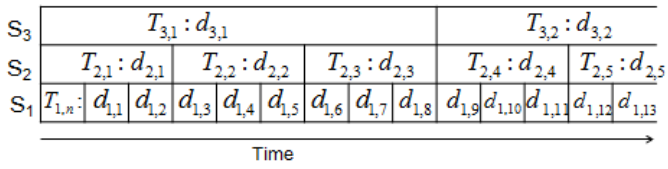


Figure 2: Time Granularities System

4.2. Smart Metering Process

The proposed metering system consists of a meter, an external cloud storage platform and the cloud service provider SPi which has a key pair: a public key (PK1) and a private key (SK1).

Figure3 shows an example of the proposed platform with a cloud computing architecture. Supposedly, a meter is fully trusted and the storage platform as well as the cloud service provider are honest, as they follow specific procedures in extracting individual meter readings from storage devices and communication links. If a consumer has a meter SM that records energy consumption d_1 at time unit T_1, H agrees time granularities Si with different service providers SP_i, and constructs SM by storing S_1.

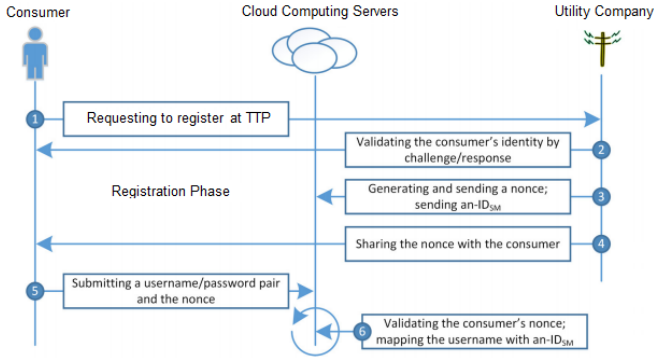
Meter M, then, encrypts the meter reading d1 as C_1 and stores C_1 into the cloud storage system. By getting the access right at a given time granularity S_i, an energy consumption di at time interval T_i is a sum of meter readings:

$$d_i = \sum_{S=(j-1)/i+1}^{j} d_1 \qquad (1)$$

The proposed model consists of 3 layers, hardware layer, application layer and Kernel layer. The hardware layer consists of all hardware modules including the precision measurement unit, the platform module and a communication and processing module. As for the kernel layer, it has a set of modules that interacts with hardware layer. Then comes the application layer that is constructed upon the kernel layer to offer facilities such as cloud network management service, transaction service and web edge for monitoring current meter readings which can be stored on the cloud. This allows easy access to such readings from anywhere through any web-based interface application.

5. Trust Assumptions and Security Requirements

The meter readings can be transferred directly to the cloud using a Trusted Third Party (TTP) encryption, which is responsible for the registration of consumers and servers. TTP can manage meter data using standard Internet connections and has the capability to provide data confidentiality, data integrity and fraud prevention. Due to these capabilities, the consumers can utilize the best authentication and registration features. Hence, home residents can easily monitor and control home electricity facilities through computer and smart phones over the TTP. The whole registration, authentication method is illustrated in Figure 3.



Consumer Authentication Process

The residential home has local renewable resources, energy storage and electrical appliances that are able to adjust power consumption behavior and predict demand-capacity. It then works with a set of service providers as each service provider primarily delivers access of meter readings at a time periodically. In addition, each network/service provider SP_i with a key pair (private and public) has a demanded time granularity Si, and the startup time at the initial time granularity $T_{i,j}$ can be approximated by the following equation :

$$T_{i,j} = \left(\sum_{v=(j-1)/si+1}^{j} T_i \right) modN \qquad (2)$$

Therefore, a network provider can compute the energy consumption di at time Ti by editing the encrypted meter readings as long as it knows a random value ri of sometime unit Ti . Then, the energy consumption di can be easily calculated by the following equation:

$$T_{i,j} = \left(\sum_{v=(j-1)/si+1}^{j} \mathrm{d}_i \right) modN \qquad (3)$$

The consumer appliances, local renewable energy generators, smart grid utility will all be connected by the cloud and controlled by smart home meter which has the skill of getting consumption information from cloud services.

This smart-home meter is also capable of static scheduling and runtime scheduling to reduce the total cost of electricity energy while sustaining power requirements [27].

From security point of view, the consumer appliances should give the corresponding random numbers rito SPi in order to permit the right access of meter reading at a specific time Si to a network provider SP_i. In order to reduce the communication cost of random number ri derives the next one $r_{i,j+1}$ by using a shared PRNG where a random number is generated so that the equation would be:

$$r_i = \left(\sum_{v=(j-1)/si+1}^{j} T_i \right) mod N \quad (4)$$

Where the time granularity Si is derived from r1 using the pseudorandom number generator g1. The consumer, then, will securely give ri, Si to the service provider SPi in order to grant the right access. SPi can consequently, derive other random numbers.

5.1. Trust Model

Let h be a crypto system and has h function where the encryption and decryption procedures used by the network provider SP_i to authenticate the communication parties and to create a secure connection. The consumer will, then, share random number generator g: with SP_i, correspondingly. Then, the consumer will configure a meter SM by setting initial state number of a specific time Si.

The TPM and SM will generate a master key K by the using a seed θ, the sequence number SN and the hash algorithm h. The session key k is securely saved in the TPM of SM which creates a seed key K_i by using the session key K and description D_i. M, then, computes r_i , and encrypts it by using public key PK_i as r_i and stores n_i to the external data storage system.

5.2. Usability Goal

Permanent data storage mechanisms are urgently needed for the development of smart-meter system which aims at reducing the electricity cost at a residential home by scheduling the power demands according to a rich set of power – related data that can describe the dynamic status of smart grid, local power generation, energy consumption of consumer appliances, sensor data, weather data and occupants' activity data.

All this data will be stored in the cloud by a data base that is implemented by MYSQL which will also support queries over data and metadata from the smart home systems.

$$c_i = \left(\sum_{v}^{j\pi i} c_1 \right) mod\ N \quad (5)$$

where $\eta_{i,[j/\pi i]} = enc^{i}_{pki}\ (d_{i,[j/\pi i]})$

Meaning that the service provider SPi gets the encrypted power assumption di of every time unit T_i , for $j \geq 1$ since $di < N$. In the next part, the privacy requirement that is proposed here is defined and it is proven that the proposed system meets it. Moreover, it is shown that managing meter data collections and billing are secure.

5.3. Privacy-Preserving Requirement

Privacy here requires that a network provider SP_i cannot get energy consumption di at specific time intervals S_v, where $j \geq 1$ and $v < I$. Paying into consideration that if SP_i cannot get energy consumption at specific time intervals S_{i-1}, it wouldn't get energy consumption at any time intervals S_v for $v < i - 1$. The privacy preserving requirement here is to keep the time granularity adapted to a prior time constraints when determining the maximum time boundary. For example, a demand phase continues first when A adoptively gets meter readings di at time intervals T_1, j as $j = 1$ and C returns back to A the encrypted meter readings C1. A, then, specifies to enter the experiment phase at $(j^* - 1) + 1$. A then chooses two challenge sets meter readings for a time $\in \{0, 1\}$. The intersection of the two sets is that the consequential di , j^* are the same as illustrated in the following equation:

$$\sum_{v=\pi_{i-1}(j^*-1)+1}^{\pi i-1 j^*} d_i^0 = \sum_{v=\pi_{i-1}(j^*-1)+1}^{\pi i-1 j^*} d_j^1 \quad (6)$$

After reading the encrypted meter data, A passes to the second phase. For a second time, A adaptively indicates a meter reading dj for subjective time interval Ti and C automatically returns the encrypted meter reading C1 back to A. Hence, a smart metering scheme satisfies the privacy preserving requirement against a network provider SPi for any probabilistic polynomial time function T bounded adversary A attacking the real process.

Table 1: Implementation analysis of the proposed system

Initialization phase	Computation cost (ms)	Storage cost
Registration phases	≤ 10 ms	1592 B
Generating a nonce number	1.2 ms	2 B
Encrypting a meter reading	< 12 ms	19
Session key exchange	4.3 ms	17 B
Data transmission phase	2.8 ms	68 B

5.4. Storage System Security

The meter readings are shown to be securely stored in the cloud storage devices, bearing in mind that each network provider has more information than the storage devices and that the proposed metering algorithm is securely integrated with privacy-preserving scheme.

The security requirement of cloud storage is the collective processes, technologies and controls that ensure that only authorized and legitimate users can store, access and use storage resources. The attacker A has the same role of the storage system, but without the knowledge of the stored pseudorandom number r. Hence, the metering systems satisfy the secure storage constraints by considering any probabilistic parameterized polynomial time algorithms which are employed by the users during the initialization and the security phases, assuming that all encryption algorithms used by network provider SPi are completely secure.

6. Performance Analysis and Results

The performance of the offered smart metering system is examined in terms of benefit-to-cost ratio and storage cost. Moreover, a specific scenario is considered where four time measures; years, months, hours, and seconds are used in the smart meter. The utilization of the system shows that the time of the generated key is the highest among the transmitted and authentication time.

In addition, the proposed system minimizes the overhead of the authentication process as compared to the other know authentication algorithms in smart metering networks because the key used in transfer time as well as the key used in the verification time produce a lower amount of overhead. As shown in Table 1, it is obvious that the encryption time is the highest among the authentication and transfer time.

It is also obvious, from the table, that the time complexity for key generation is in the order of O(log n) and it takes nearly less than 20 seconds on average to be accomplished. Despite the long generation time and relatively low encryption time, the total key transfer time range is 0.6 to 3 seconds, whereas the total key authentication time range is 0.8 to 5 seconds.

Noticing that the scheduling scheme produces a total of N sub-keys of the size of 32 bits and that the encryption system used by network providers is standard RSA encryption. The proposed system showed that the generated 2048-bit key of RSA algorithm takes significantly longer than the 1042-bit key. A commercial Information on Trusted Platform Module (TPM) chip is used to enhance the performance of cryptographic system in terms of space and speed, which is capable of supporting a 1024-bit signature in 50 ms and 2048-bit RSA signature in 250 ms.

The TPM can be initiated to cope with multiple security services, including the protection and secure storage of the keys. TMP can be used also to cope with RSA algorithm for key signing and encrypting data.

It is also assumed that calculating Si optimal additions is not slower than the estimated time complexity to factor 1024-bit RSA modulus, and the fastest known factorization algorithm requires around 265 times complexity to factor a 1024-bit signature. The task of generating a 1024-bit RSA signature, would take less than 100 ms to be generated.

Thus, the complexity of encrypting a meter reading takes less than 35 ms. Additionally, the storage cost of a meter is considered to be constant and the cost of storing data in the cloud system is analyzed as follows: for each reading, the cloud system stores an encrypted data reading of (log2 n) bits. In order to grant the access right to a network provider SPi, a cipher text is stored in the cloud storage system at the first time interval T1.

Consequently, for every Ti time unit at time granularity S1, an encryption of a nonce digits is stored. In the proposed model, the cloud storage efficiency of encrypted data reading of a month is around 325 MB, and as for the storage cost of encrypted data for a month is about 3.4 MB, which make the overall storage cost of the cloud storage system for a meter system per month is fewer than 126.3 MB.

Level of trust operations of cloud network providers includes encryption, modular arithmetic and the computation of nonce numbers are all within efficient range and work under all standard reference conditions required by secure service providers. The computation efficiency of meter system is analyzed as follows: large keys are converted into small keys by using MD5 hash algorithm and the data encryption takes place during the initialization process.

The calculation in the initialization process is subject to the computation of generated RSA keys. As the encrypting of a nonce takes 2 ms, the initialization itself takes 2 seconds. At T1, the meter reading is encrypted by using the nonce r1 that on average, takes between 10 – 15 ms to be completed.

On the other hand, the efficiency analysis and the security analysis show that the proposed protocol is secure and efficient ; the efficiency is verified by ns2 simulation results, which show a high authentication rate. The average time for the authentication phase is between 1 and 2 ms, and the average time of key exchange protocol is also utilized and was in the range of 1 to 7 ms.

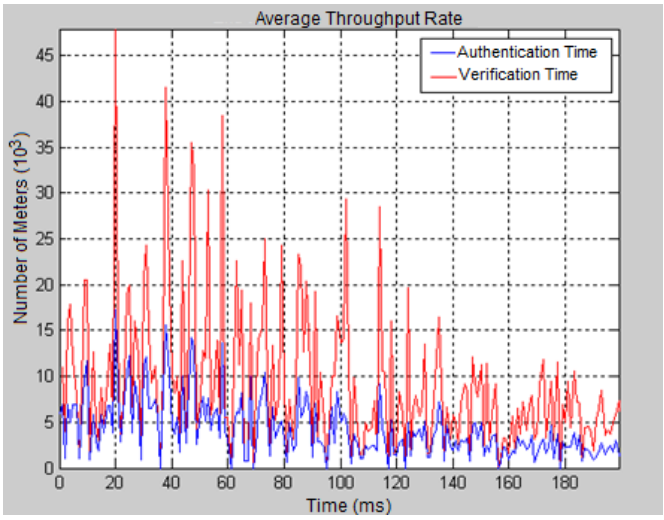


Figure 4: End-to-End Delay

The verification time has been calculated for various security measures in accordance with the key generation time and the key computation time to prevent intruders from decrypting the keys.

The simulation results showed that the elapsed time to complete the verification process including the key exchange protocol and the hash function is approximately equal to the total time required for exchanging of all smart meter readings within the communication channel.

The performance analysis of this system for varying number of meters is presented in Figure 4 and Figure 5, respectively. In the former figure, the measured end-to-end delay between the creation and aggregation of the meter readings at the cloud server and the readings response time at the utility company is illustrated by implementing the authentication protocols.

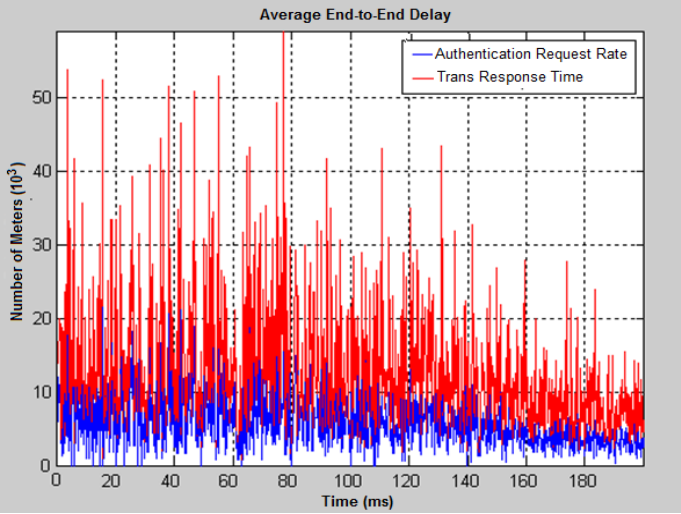


Figure 5: Throughput Rate

Cleary, the delay is affected by the time slot where the meter is assigned in the schedule of time granularity created by the users. Another reason for the increase in the message delay is the computational overheads taking place at the communication layers of the network. Furthermore, the suggested authentication scheme reduced the response time and the communication overhead in the proposed smart metering system because the verification time and the key transfer time produce minimum amount of data overhead. It is noted from Figure 5 that the generation time of the key pair is the lowest among the computation time and verification time. It is clear also from the figure that the generation time of the key-pair was performed in less than 9 ms. When considering the key exchange time over different network bandwidth, the key transfer time was quite high compared with key verification time. The average transfer time was less than one second and the verification time was the 6.2 ms on average. To avoid the risk of key-guessing attack and code replication attack, a strong crypto system together with a secure digital signature were used interchangeably to encode smart meter readings during transmission between the cloud server and the smart meters. As for the results, the proposed security and meter privacy preservation scheme in this paper have offered alternative ways for preserving the cloud storage and time granularity while still providing privacy in smart meters.

7. Conclusion and Future Work

In this paper a data processing and storage system that is applicable to smart meters with many modules that is connected as an internet of things via the public network and uses the full strength of cloud computing and secures users' privacy and the confidentiality of data exchanged on the grid by encrypting smart meter data before sending to the cloud for the sake of processing and storage with the help of a homomorphic asymmetric key cryptosystem. Each smart meter includes a set of public and private keys to which the grid operator can access and the household owner, on the other hand, can only access the decryption key corresponding to their own smart meter. With the help of the homomorphic feature of the cryptographic technique, several methods are created to enable most of the computing works to be done directly on the data encrypted by the cloud which is the main focus of this model. The smart reading system which is proposed in this paper considers an external storage cloud server and supports multiple time granularities in a privacy preserving manner against host-but-curious providers. The consumer gives different time granularities to different service providers. It is shown in this paper that a service provider gets power consumption of the consumer at only the granted time granularity. The security analysis in this paper confirms that the proposed security protocol preserves the overall privacy of the smart meter readings by allowing an efficient authentication and key management protocol for cloud-based smart metering systems. The performance of the proposed system is analyzed in terms of computation cost and storage cost of a meter. The results confirm also that the key generation time is the highest among the transfer and the verification time. Furthermore, the suggested scheme reduced the overhead of the authentication time as compared to the other existing authentication schemes in smart metering systems. The analysis also provides a solid step towards the practical construction of privacy in cloud-based smart grid technologies.

References

[1] G. Lobaccaro, S. Carlucci, and E. Lofstrom,"A review of systems and technologies for smart homes and smart grids," Energies, **9**(5), 1–33, 2016. https://doi.org/10.3390/en9050348

[2] Dark Reading. Smart meter hack shuts off the lights. Available at: http://www.darkreading.com/perimeter/smart-meter-hack-shuts-off-the-lights/d/d-id/1316242.

[3] B. J. Murrill, E. C. Liu, and R. M. Thompson,"Smart Meter Data:Privacy and Cybersecurity," CRS Report for Congress [Online]. Available: http://www.fas.org/sgp/crs/misc/R42338.pdf.(accessed on April 2018).

[4] M. Weiss, A. Helfenstein, F. Mattern, T. Staake, Leveraging smart meter data to recognize home appliances, in: 2012 IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, 190–197, 2012.

[5] S. Bera, S. Misra, and J. J. Rodrigues, "Cloud computing applications for smart grid: A survey," IEEE Transactions on Parallel and Distributed Systems, **26**(5), 1477-1494, 2015. https://doi.org/10.1109/ISGT-Asia.2012.6303140

[6] J. Zhou, R. Q. Hu, and Y. Qian, "Scalable distributed communication architectures to support advanced metering infrastructure in smart grid," IEEE Transactions on Parallel and Distributed Systems, **23**(7), 1632-1642, 2012.

[7] L. Ji, W. Lifang, and Y. Li, "Cloud Service based intelligent power monitoring and early-warning system," in Innovative Smart Grid Technologies-Asia (ISGT Asia 2012), IEEE, 1-4, 2012. https://doi.org/10.1109/ISGT-Asia.2012.6303140

[8] M. Akbar, D.Z.A. Khan, "Modified nonintrusive appliance load monitoring for nonlinear devices," in: Multitopic Conference, 2007. INMIC 2007. IEEE International, IEEE, , 1–5, 2007.

[9] X. Fang, D. Yang, and G. Xue, "Evolving smart grid information management cloudward: A cloud optimization perspective," IEEE Transactions on Smart Grid, **4**(3), 111-119, 2013.

https://doi.org/10.1109/TSG.2012.2230198
[10] N. Lu, P. Du, P. Paulson, F. Greitzer, X. Guo, and M. Hadley, "A multi-layer, hierarchical information management system for the grid," in Proc. of IEEE Conf. on Power and Energy Society General Meeting, 632 – 642, Sept. 2012.
[11] Fournet, C., Kohlweiss, M., Danezis, G., Luo, Z.: Zql: A compiler for privacy-preserving data processing. In: Proceedings of the 22Nd USENIX Conference on Security. 163–178. SEC'13, USENIX Association, Berkeley, CA, USA, 2013.
[12] P. Deng, L. Yang, A secure and privacy-preserving communication scheme for advanced metering infrastructure, in: Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES, IEEE, 1–5, 2012.
[13] A. Rial, G. Danezis, Privacy-preserving smart metering, in: Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society, ACM, 49–60, 2011.
[14] Jo, H.J.; Kim, I.S.; Lee, D.H. Efficient and privacy-preserving metering protocols for smart grid system. IEEE Trans. Smart Grid, **7**, 1732–1742, 2016.
[15] Kursawe, K.; Danezis, G.; Kohlweiss, M. Privacy-Friendly Aggregation for the Smart-Grid. In Proceedings of the 11th International Conference on Privacy Enhancing Technologies (PETS'11), Waterloo, ON, Canada, 27–29, July 2011.
[16] J.L. Hennessy, D.A. Patterson, Computer Architecture: A Quantitative Approach, Elsevier, 2012.
[17] G.E. Blelloch, B.M. Maggs, Parallel algorithms, in: Algorithms and Theory of Computation Handbook, Chapman & Hall/CRC, 25–27, 2010.
[18] Le Métayer, D.: Privacy by design: A formal framework for the analysis of architectural choices. In: Proc. of the Third ACM Conference on Data and Application Security and Privacy. 95–104. CODASPY '13, ACM, New York, NY, USA, 2013.
[19] Garcia, F., Jacobs, B.: Privacy-friendly energy-metering via homomorphic encryption. In: Cuellar, J., Lopez, J., Barthe, G., Pretschner, A. (eds.) Security and Trust Management, Lecture Notes in Computer Science, vol. 6710, 226–238. Springer Berlin / Heidelberg , 2011.
[20] N.P. Smart, F. Vercauteren, Fully homomorphic encryption with relatively small key and ciphertext sizes, in: Public Key Cryptography–PKC 2010, Springer, 420–443, 2010.
[21] F.D. Garcia, B. Jacobs, Privacy-friendly energy-metering via homomorphic encryption, in: Security and Trust Management, Springer, 2011, 226–238.
[22] J. Zhao et al., "Privacy Protection Scheme Based on Remote Anonymous Attestation for Trusted Smart Meters, IEEE Transaction, Smart Grid. 2016.
[23] Gong, Y.; Cai, Y.; Guo, Y.; Fang, Y. A privacy-preserving scheme for incentive-based demand response in the smart grid. IEEE Trans. Smart Grid, **9**(4), 3313-3320, 2016. https://doi.org/10.1109/TSG.2016.2626317
[24] McLaughlin, S.; McDaniel, P.; Aiello, W. Protecting consumer privacy from electric load monitoring. In Proceedings of the 18th ACM Conference on Computer and Communications Security, Chicago, IL, USA, 17–21 October 2011.
[25] S. Fahl, M. Harbach, T. Muders, and M. Smith. Confidentiality as a service–usable security for the cloud. In TrustCom, 153–162, 2012.
[26] Backes, M.; Meiser, S. Differentially private smart metering with battery recharging. In Data Privacy Management and Autonomous Spontaneous Security; Springer Science + Business Media: New York, NY, USA, 2014; 194–212. Available online: https://eprint.iacr.org/2012/183.pdf. (accessed on 26 August 2018).
[27] D.P. Rodgers, Improvements in multiprocessor system design, in: ACM SIGARCH Computer Architecture News, **13**, IEEE Computer Society Press, . 225–231, 1985.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 1286-1292 (2020)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

Analysis of the Operational Impact of ETC Lanes on Toll Station

Alimam Mohammed Karim*, Alimam Mohammed Abdellah, Seghiouer Hamid

MOSIL, University Abdelmalek Essaadi, 93000, Morocco

ARTICLE INFO

Article history:
Received: 15April, 2020
Accepted: 10 October, 2020
Online: 16 December, 2020:

Keywords:
ARENA
Electronic toll collection
Highway
Toll plaza
Trafic congestion

ABSTRACT

Traffic congestion is being experienced daily by many travellers both on the inner-city roads as well as on the highways. The toll plaza can become a real highway bottleneck, as vehicles must stop and manually pay the tolls at the toll booths. This stop-and-go process can seriously affect traffic fluidity, resulting in a lower quality of service at the plaza. One of the solutions used to improve the efficiency of toll plazas operation was the implementation of electronic toll collection (ETC), where vehicles can pass through the tolls without having to stop. In this research, a model that evaluates the benefits of the ETC system was developed using ARENA, a microscopic simulation tool. These benefits were visualized by conducting a comparative analysis of different scenarios. The Bouznika toll plaza, one of the major toll highways in Morocco, was selected for illustrating the improved operation realized from the use of ETC technology.

1 Introduction

One of the main causes of traffic jams on freeways is resulted from the conventional toll collection at toll gates, where each vehicle is required to stop to pay toll fees. The process of stopping to pay the toll reduces the traffic flow, increases the waiting time, queue length and thus reduces the efficiency of the toll plazas.

Therefore, in order to improve the level of service of toll plazas, a new technology has been introduced: The Electronic Toll Collection (ETC) Systems. These systems make it possible to pass through tolling location without being required to stop, and that it is thanks to a vehicle-mounted transponder that allows to uniquely identify each vehicle and to electronically collect the tolls.

With the development of ETC systems, various studies have been proposed to evaluate its performance and impacts.

Burris and Hildebrand had developed a discrete-event simulation model to examine the effect of ETC on traffic operations at the Halifax-Dartmouth Bridge in Nova Scotia, Canada. They simulated various scenarios of toll plaza configurations with and without ETC [1].

In [2], the authors used the INTEGRATION simulation model to study the impacts of ETC deployment on toll station at the Massachusetts Turnpike in Boston. Different scenarios were evaluated, and the simulation results showed an increase in the efficiency of traffic operations with the implementation of ETC. Throughput volumes, average queue length, and average waiting time spent in the system were used as measures of effectiveness.

Anthony A. Saka used a microscopic simulation model to evaluate the benefits resulting from the use of ETC technology by undertaking a comparative analysis of traffic operation condition before and after the ETC implementation [3] .

David Levinson and Elva Chang developed a model to maximize social welfare associated with toll stations. A payment choice model was used to estimate the share of traffic using ETC as a function of delay, price, and a fixed cost of acquiring the in-vehicle transponder [4].

Lu Jian made a qualitative analysis on the impacts of ETC and developed the performance index system of ETC benefits (2003) [5] . Jia Hong-fei estimated the monetary benefits of implementing ETC projects under different traffic volume conditions by using a cost-benefit analysis method [6].

Following this introductory section, a description of the proposed methodology is provided. Simulation results and performance evaluation are introduced in section 3. Finally, concluding remarks are stated in section 4.

2 Methodology

2.1 Study site selection

The selected study site is the Bouznika toll plaza. This plaza is one of the major toll highways in Morocco. It is located on the motorway axis that connects Casablanca, the nerve center of Mo-

*Corresponding Author: Alimam Mohammed Karim, Morocco, Tangier, karim.alimam@gmail.com

rocco's economic activity, to Rabat, the administrative capital of the country. This is the busiest section of the country's highways, with a throughput volume that could approach 50000 vehicles per day during peak periods.

The configuration of the toll facility consisted of nine manual lanes, two automated lanes and four ETC lanes.

2.2 Data collection and analysis

Data collection consisted of two major activities: (1) collection of service time (i.e., the time required to pay the toll and cross the barrier) data and (2) collection of traffic arrivals at the toll plaza [7]–[9].

The service time may be affected by a number of factors, such as the amount to be paid, the experience of the toll collector, the method of payment and the type of vehicle being serviced.

To determine service time distribution, we collect data over a given period of observation time. We record the time when the vehicle stops to pay the toll and the time after it crosses the toll barrier.

ARENA's Input Analyzer was used to find a probability distribution that best fits to the collected data. It is a tool designed to adjust theoretical distributions to observed data by estimating their parameters and measuring the quality of their adjustment.

With a sample size of 65 observations, we found that the service time distribution is 5.5 + Logn(16.7, 16.8) for the manned tollbooths and a triangular distribution with parameters 8, 10 and 12 for the automatic toll lanes (See figures 1 and 2). As for the ETC lanes, we assume an average toll transaction time of five seconds.

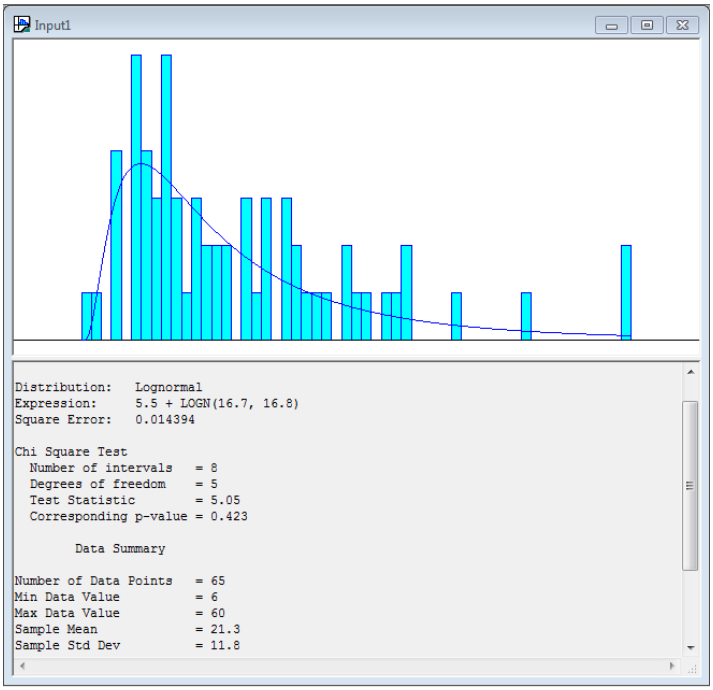


Figure 1: Results obtained for the service time distribution(Manual Tollbooth).

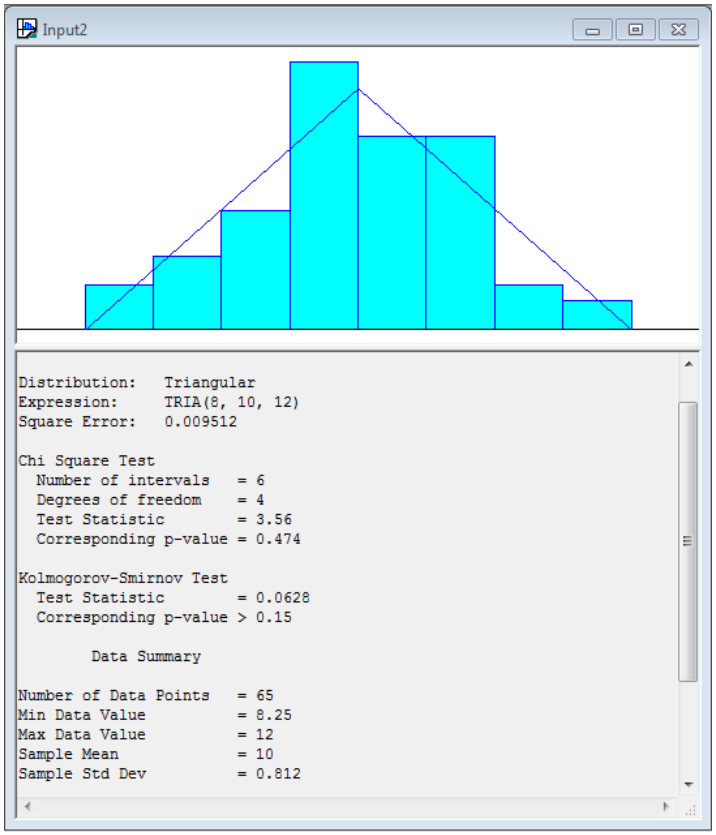


Figure 2: Results obtained for the service time distribution(Automatic Toll Lanes).

2.3 Model development

Simulation modelling provides a better understanding of the dynamic behaviour of systems, allows the user to compare different configurations and evaluate performance.

By using simulation, it is possible to analyse each of the system's component and to measure their relative impact on the overall performance of the system.

In our case, we use simulation in order to evaluate the impact of ETC on toll station. This evaluation was accomplished by utilizing collected field data and simulated scenarios using ARENA software. Arena is the most widely used discrete event simulation software in the world. It contains all the resources for modelling, process representation, statistical analysis and results analysis.

Furthermore, it combines the capabilities of simulation language in an integrated graphic environment, allowing for an easy user interface for model building.

Although the software was not specifically designed for toll plazas, its functionalities enable the user to model practically any type of process. It covers all discrete flows, all basic process functions: time, resources, information... at a very detailed level.

Arena provides a considerable number of statistics, automatically generated after each run, that estimate the performance measures of the system under study. These statistics are related to various components in the model, such as entities (Number in, number out, average total time in system...), resources (utilization, average number of busy resource units, number of times seized), queues (average queue delay, average queue size and minimal and

maximal observed values), and so on [10].

To construct this model, a great deal of empirical data were required, particularly traffic volumes, service time distribution, number of lanes and percentage of vehicles using each payment option.

Figure 3 shows the flow chart of the simulation model.

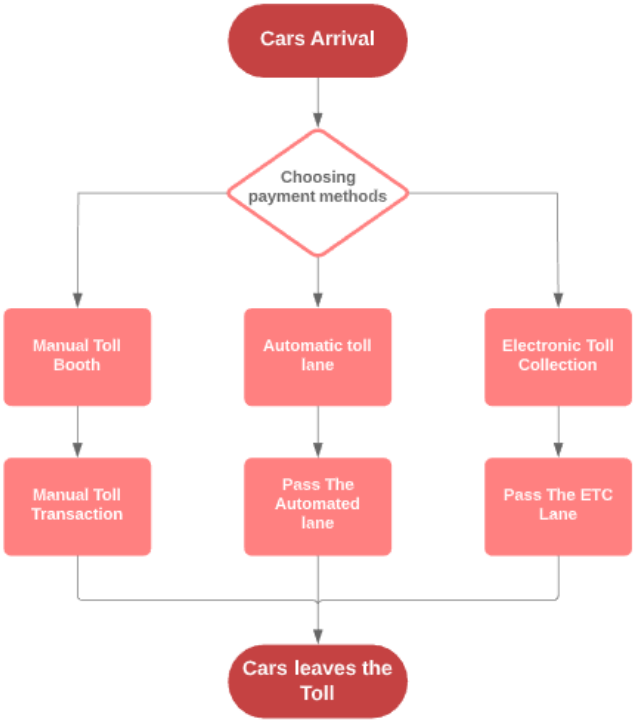


Figure 3: Simulation Flow Chart.

The simulation begins with the generation of entities (vehicles) according to a non-homogenous Poisson distribution.

The Poisson process is one of the most widely used counting processes. It is usually used in scenarios where we are counting the occurrences of certain events that appear to happen at a certain rate, but completely at random (without a certain structure).

A counting process is called a Poisson process if the following conditions hold:

- The number of entities that arrive in one interval is independent of the number of entities that arrive in any other non-overlapping interval.
- The number of occurrences in each interval can range from zero to infinity.
- The rate at which arrivals occur does not change over time.

If we look at our case study, only the two first conditions are true. In this case, we have a non-homogeneous Poisson process. Such a process has all the properties of a Poisson process, except for the fact that its rate is a function of time, which is the case for traffic arrival rate.

The traffic arrival rate varies over the course of the day, and is expressed in vehicles per hour. For the simulation, we used the hourly traffic volume obtained from the prediction results for the year 2030.

These traffic arrival rates are given in 1.

Table 1: Traffic arrival rates (vehicles per hour)

Hours	Traffic
1	1477
2	1160
3	930
4	730
5	802
6	797
7	1160
8	1576
9	1865
10	2300
11	2670
12	3184
13	3489
14	3419
15	3426
16	3419
17	4202
18	4487
19	5155
20	5238
21	6034
22	5116
23	4507
24	2860

Once generated, the vehicles were sent to either an ETC lane, automated lanes or manual lanes according to the rate of each payment option (The percentage of automated lanes users is fixed at 0.5%). Each vehicle would then join the queue that has the shortest length in its category and waited to be served. Once the toll lane become empty, the vehicle is serviced and sent to a model block where it exited simulation. Each vehicle was tracked from its creation in the model to its destruction, and statistics were recorded throughout the vehicle's evolution.

2.4 *Model calibration and validation*

The calibration and validation process consists of adjusting the input parameters for the simulation model in order to achieve reasonable correspondence between field data and simulation model output.

The input parameters for our simulation model are arrival rate (vehicles per hour) and service time. The arrival process follows a non-homogenous Poisson distribution and the service time follows a lognormal distribution.

The average waiting time at the toll plaza was selected as the measure of effectiveness of the model (index of comparison), which is part of the simulation output data.

The purpose of this study was to evaluate the performance of the toll plaza during heavy traffic conditions. Therefore, we started by configuring the model based on the data recorded on the busiest day for traffic of the year 2019.

Unfortunately, the waiting times at the toll plaza on this particular day are not available. We therefore decided to compare the model outputs with expert judgment, which indicate that the simulation results are relatively realistic.

3 Results and analysis

Scenarios with various ETC rates were simulated in order to evaluate the impact of electronic tolling system on toll plaza efficiency. Measures such as queue delay, queue length and capacity were observed when comparing the different scenarios.

Table 2 shows the experimental design of scenarios. Each scenario was simulated for 24 hours (a weekend : a sunday).

The figures 4, 5, 6 and 7 show the performance measures for the following four scenarios. They were summarized in terms of maximum, minimum, and average of queue length and waiting time.

The *Queues* section provide statistics for each of the fifteen toll-booth queues (nine cash receipt CR queues, two automated lanes AL queues and four electronic collection EC queues), such as average queue delay and average queue size.

As expected, in the case of manual service, the delay is longest compared to the ETC service. Increasing the ETC rate, without changing the configuration of the toll facility, has a direct effect in decreasing the queue length and the waiting time of the manned tollbooths, but on the other hand it increases the waiting time at the ETC lanes.

So we need to find a reasonable number of tollbooths for each payment method, considering different values of ETC rates, that minimizes the total time drivers spend waiting in the toll plaza [11]–[13].

The first scenario: We will consider a tolerable waiting time of 7 minutes and an ETC rate of 40%.

By adjusting the number of booths, the maximum value of total average waiting time obtained is 7 minutes, which happens when the number of active tollbooths is 22: 18 manual lanes, 3 ETC lanes

Category Overview

Toll Plaza

Replications: 1 Time Units: Minutes

Queue

Time

Waiting Time	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00	0,000000000	0.00	0.00
ALQ_2	0.00	(Insufficient)	0.00	0.00
CRQ_1	72.6931	(Correlated)	0.00	284.24
CRQ_2	72.0999	(Correlated)	0.00	293.27
CRQ_3	77.0382	(Correlated)	0.00	285.89
CRQ_4	81.1234	(Correlated)	0.00	285.85
CRQ_5	84.1970	(Correlated)	0.00	283.46
CRQ_6	86.3036	(Correlated)	0.00	286.51
CRQ_7	91.2500	(Correlated)	0.00	283.45
CRQ_8	95.8400	(Correlated)	0.00	286.73
CRQ_9	98.1664	(Correlated)	0.00	286.02
ETCQ_1	0.01584576	(Correlated)	0.00	0.2445
ETCQ_2	0.01204625	(Correlated)	0.00	0.2270
ETCQ_3	0.00884878	(Correlated)	0.00	0.1694
ETCQ_4	0.00693669	(Correlated)	0.00	0.1617

(a) Waiting time

Number Waiting	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00	(Insufficient)	0.00	0.00
ALQ_2	0.00	(Insufficient)	0.00	0.00
CRQ_1	344.72	(Correlated)	0.00	1483.00
CRQ_2	344.64	(Correlated)	0.00	1483.00
CRQ_3	344.57	(Correlated)	0.00	1483.00
CRQ_4	344.50	(Correlated)	0.00	1483.00
CRQ_5	344.41	(Correlated)	0.00	1483.00
CRQ_6	344.34	(Correlated)	0.00	1483.00
CRQ_7	344.27	(Correlated)	0.00	1483.00
CRQ_8	344.20	(Correlated)	0.00	1483.00
CRQ_9	344.13	(Correlated)	0.00	1482.00
ETCQ_1	0.1160	(Correlated)	0.00	3.0000
ETCQ_2	0.06580262	(Correlated)	0.00	3.0000
ETCQ_3	0.03451640	(Correlated)	0.00	3.0000
ETCQ_4	0.01832924	(Correlated)	0.00	2.0000

(b) Number waiting

Figure 4: Performance indicators: Scenario 1.

Category Overview

Toll Plaza

Replications: 1 Time Units: Minutes

Queue

Time

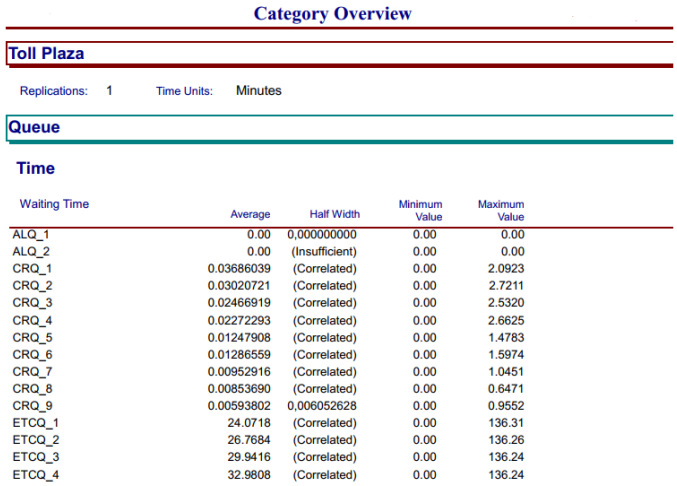
Waiting Time	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00010649	0,000155418	0.00	0.01905521
ALQ_2	0.00	(Insufficient)	0.00	0.00
CRQ_1	21.5830	(Correlated)	0.00	128.54
CRQ_2	23.1657	(Correlated)	0.00	134.41
CRQ_3	24.1926	(Correlated)	0.00	137.17
CRQ_4	26.1935	(Correlated)	0.00	130.33
CRQ_5	29.2717	(Correlated)	0.00	122.57
CRQ_6	29.6376	(Correlated)	0.00	136.93
CRQ_7	31.6353	(Correlated)	0.00	136.11
CRQ_8	33.4667	(Correlated)	0.00	139.00
CRQ_9	35.9886	(Correlated)	0.00	131.09
ETCQ_1	3.2501	(Correlated)	0.00	20.1479
ETCQ_2	3.8038	(Correlated)	0.00	20.0748
ETCQ_3	4.4917	(Correlated)	0.00	20.0791
ETCQ_4	5.2616	(Correlated)	0.00	20.0313

(a) Waiting time

Number Waiting	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00002381	(Insufficient)	0.00	1.0000
ALQ_2	0.00	(Insufficient)	0.00	0.00
CRQ_1	70.6170	(Correlated)	0.00	403.00
CRQ_2	70.5391	(Correlated)	0.00	403.00
CRQ_3	70.4706	(Correlated)	0.00	403.00
CRQ_4	70.4142	(Correlated)	0.00	402.00
CRQ_5	70.3592	(Correlated)	0.00	402.00
CRQ_6	70.3027	(Correlated)	0.00	402.00
CRQ_7	70.2541	(Correlated)	0.00	402.00
CRQ_8	70.1953	(Correlated)	0.00	402.00
CRQ_9	70.1430	(Correlated)	0.00	402.00
ETCQ_1	29.3340	(Correlated)	0.00	242.00
ETCQ_2	29.2129	(Correlated)	0.00	241.00
ETCQ_3	29.1274	(Correlated)	0.00	241.00
ETCQ_4	29.0338	(Correlated)	0.00	241.00

(b) Number waiting

Figure 5: Performance indicators: Scenario 2.



Category Overview

Toll Plaza

Replications: 1 Time Units: Minutes

Queue

Time

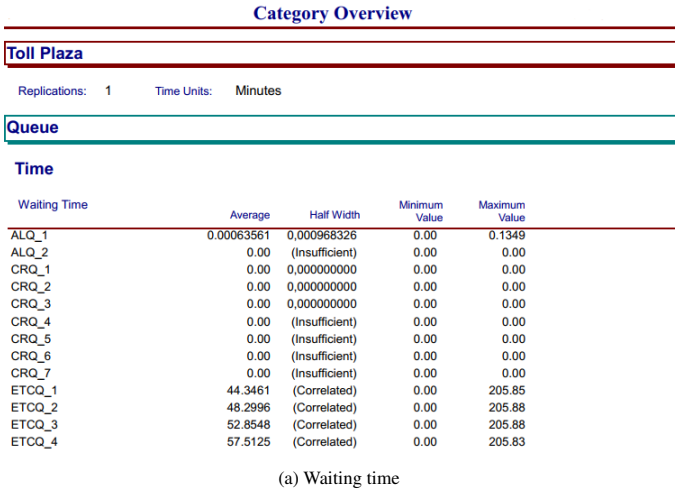
Waiting Time	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00	0,000000000	0.00	0.00
ALQ_2	0.00	(Insufficient)	0.00	0.00
CRQ_1	0.03686039	(Correlated)	0.00	2.0923
CRQ_2	0.03020721	(Correlated)	0.00	2.7211
CRQ_3	0.02466919	(Correlated)	0.00	2.5320
CRQ_4	0.02272293	(Correlated)	0.00	2.6625
CRQ_5	0.01247908	(Correlated)	0.00	1.4783
CRQ_6	0.01286559	(Correlated)	0.00	1.5974
CRQ_7	0.00952916	(Correlated)	0.00	1.0451
CRQ_8	0.00853690	(Correlated)	0.00	0.6471
CRQ_9	0.00593802	0,006052628	0.00	0.9552
ETCQ_1	24.0718	(Correlated)	0.00	136.31
ETCQ_2	26.7684	(Correlated)	0.00	136.26
ETCQ_3	29.9416	(Correlated)	0.00	136.24
ETCQ_4	32.9808	(Correlated)	0.00	136.24

(a) Waiting time

Number Waiting	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00	(Insufficient)	0.00	0.00
ALQ_2	0.00	(Insufficient)	0.00	0.00
CRQ_1	0.07453990	(Correlated)	0.00	2.0000
CRQ_2	0.05198157	(Correlated)	0.00	2.0000
CRQ_3	0.03541057	(Correlated)	0.00	2.0000
CRQ_4	0.02632073	(Insufficient)	0.00	2.0000
CRQ_5	0.01225376	(Insufficient)	0.00	1.0000
CRQ_6	0.00994403	(Insufficient)	0.00	1.0000
CRQ_7	0.00565794	(Insufficient)	0.00	1.0000
CRQ_8	0.00388310	(Insufficient)	0.00	1.0000
CRQ_9	0.00186800	(Insufficient)	0.00	1.0000
ETCQ_1	330.86	(Correlated)	0.00	1916.00
ETCQ_2	330.74	(Correlated)	0.00	1916.00
ETCQ_3	330.59	(Correlated)	0.00	1915.00
ETCQ_4	330.47	(Correlated)	0.00	1915.00

(b) Number waiting

Figure 6: Performance indicators: Scenario 3.



Category Overview

Toll Plaza

Replications: 1 Time Units: Minutes

Queue

Time

Waiting Time	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00063561	0,000968326	0.00	0.1349
ALQ_2	0.00	(Insufficient)	0.00	0.00
CRQ_1	0.00	0,000000000	0.00	0.00
CRQ_2	0.00	0,000000000	0.00	0.00
CRQ_3	0.00	0,000000000	0.00	0.00
CRQ_4	0.00	(Insufficient)	0.00	0.00
CRQ_5	0.00	(Insufficient)	0.00	0.00
CRQ_6	0.00	(Insufficient)	0.00	0.00
CRQ_7	0.00	(Insufficient)	0.00	0.00
ETCQ_1	44.3461	(Correlated)	0.00	205.85
ETCQ_2	48.2996	(Correlated)	0.00	205.88
ETCQ_3	52.8548	(Correlated)	0.00	205.88
ETCQ_4	57.5125	(Correlated)	0.00	205.83

(a) Waiting time

Number Waiting	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00014125	(Insufficient)	0.00	1.0000
ALQ_2	0.00	(Insufficient)	0.00	0.00
CRQ_1	0.00	(Insufficient)	0.00	0.00
CRQ_2	0.00	(Insufficient)	0.00	0.00
CRQ_3	0.00	(Insufficient)	0.00	0.00
CRQ_4	0.00	(Insufficient)	0.00	0.00
CRQ_5	0.00	(Insufficient)	0.00	0.00
CRQ_6	0.00	(Insufficient)	0.00	0.00
CRQ_7	0.00	(Insufficient)	0.00	0.00
CRQ_8	0.00	(Insufficient)	0.00	0.00
CRQ_9	0.00	(Insufficient)	0.00	0.00
ETCQ_1	740.40	(Correlated)	0.00	3590.00
ETCQ_2	740.27	(Correlated)	0.00	3590.00
ETCQ_3	740.09	(Correlated)	0.00	3589.00
ETCQ_4	739.96	(Correlated)	0.00	3589.00

(b) Number waiting

Figure 7: Performance indicators: Scenario 4.

and one automated lane (See figure 8).

Table 2: The experimental design of scenarios

Case study	No of toll booth	ETC penetration rate
Case-1	9 manual lanes, 4 ETC lanes, 2 automated lanes	40%
Case-2	9 manual lanes, 4 ETC lanes, 2 automated lanes	60%
Case-3	9 manual lanes, 4 ETC lanes, 2 automated lanes	80%
Case-4	9 manual lanes, 4 ETC lanes, 2 automated lanes	95%

The second scenario: We will consider a tolerable waiting time of 7 minutes and an ETC rate of 60%.

By adjusting the number of booths, the maximum value of total average waiting time obtained is 7 minutes, which happens when the number of active tollbooths is 17: 12 manual lanes, 4 ETC lanes and one automated lane (See figure 9).

The third scenario: We will consider a tolerable waiting time of 7 minutes and an ETC rate of 80%.

By adjusting the number of booths, the maximum value of total average waiting time obtained is 5 minutes, which happens when the number of active tollbooths is 13: 6 manual lanes, 6 ETC lanes and one automated lane (See figure 10).

The fourth scenario: We will consider a tolerable waiting time of 7 minutes and an ETC rate of 95%.

By adjusting the number of booths, the maximum value of total average waiting time obtained is 1.4 minutes, which happens when the number of active tollbooths is 10: 2 manual lanes, 7 ETC lanes and one automated lane (See figure 11).

From 3 summarizes the main results obtained from the simulation.

Category Overview

Toll Plaza

Replications: 1 Time Units: Minutes

Queue

Time

Waiting Time	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00550248	0,002848446	0.00	0.1748
CRQ_1	3.0862	(Correlated)	0.00	24.8171
CRQ_10	4.9395	(Correlated)	0.00	22.9470
CRQ_11	5.2368	(Correlated)	0.00	25.6058
CRQ_12	5.6181	(Correlated)	0.00	25.8467
CRQ_13	5.7738	(Correlated)	0.00	22.3388
CRQ_14	6.0671	(Correlated)	0.00	25.8133
CRQ_15	6.4121	(Correlated)	0.00	23.5318
CRQ_16	6.8475	(Correlated)	0.00	23.6201
CRQ_17	7.3041	(Correlated)	0.00	25.0478
CRQ_18	7.7169	(Correlated)	0.00	25.8032
CRQ_2	3.1836	(Correlated)	0.00	25.5149
CRQ_3	3.2326	(Correlated)	0.00	24.3351
CRQ_4	3.4704	(Correlated)	0.00	25.4010
CRQ_5	3.6993	(Correlated)	0.00	24.7151
CRQ_6	3.8440	(Correlated)	0.00	22.8341
CRQ_7	4.2608	(Correlated)	0.00	27.9945
CRQ_8	4.4391	(Correlated)	0.00	26.4704
CRQ_9	4.7953	(Correlated)	0.00	25.4164
ETCQ_1	0.4616	(Correlated)	0.00	5.4152
ETCQ_2	0.5675	(Correlated)	0.00	5.4111
ETCQ_3	0.7077	(Correlated)	0.00	5.4163

(a) Waiting time

Category Overview

Toll Plaza

Replications: 1 Time Units: Minutes

Queue

Other

Number Waiting	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00130684	(Insufficient)	0.00	1.0000
CRQ_1	7.4584	(Correlated)	0.00	60.0000
CRQ_10	7.2652	(Correlated)	0.00	59.0000
CRQ_11	7.2479	(Correlated)	0.00	59.0000
CRQ_12	7.2450	(Correlated)	0.00	59.0000
CRQ_13	7.2173	(Correlated)	0.00	59.0000
CRQ_14	7.2047	(Correlated)	0.00	59.0000
CRQ_15	7.1914	(Correlated)	0.00	59.0000
CRQ_16	7.1709	(Correlated)	0.00	59.0000
CRQ_17	7.1773	(Correlated)	0.00	59.0000
CRQ_18	7.1435	(Correlated)	0.00	59.0000
CRQ_2	7.4262	(Correlated)	0.00	60.0000
CRQ_3	7.3945	(Correlated)	0.00	60.0000
CRQ_4	7.3770	(Correlated)	0.00	60.0000
CRQ_5	7.3549	(Correlated)	0.00	60.0000
CRQ_6	7.3355	(Correlated)	0.00	59.0000
CRQ_7	7.3144	(Correlated)	0.00	59.0000
CRQ_8	7.2999	(Correlated)	0.00	59.0000
CRQ_9	7.2863	(Correlated)	0.00	59.0000
ETCQ_1	3.6993	4,68322	0.00	65.0000
ETCQ_2	3.5759	4,64659	0.00	65.0000
ETCQ_3	3.4807	4,60773	0.00	65.0000

(b) Number waiting

Figure 8: Performance indicators: The first scenario.

Category Overview

Toll Plaza

Replications: 1 Time Units: Minutes

Queue

Time

Waiting Time	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00345953	0,002067127	0.00	0.1470
CRQ_1	3.4967	(Correlated)	0.00	27.8493
CRQ_10	6.7617	(Correlated)	0.00	25.6551
CRQ_11	7.1176	(Correlated)	0.00	26.1772
CRQ_12	7.8593	(Correlated)	0.00	25.6907
CRQ_2	3.5548	(Correlated)	0.00	24.1704
CRQ_3	3.9318	(Correlated)	0.00	26.0083
CRQ_4	4.2310	(Correlated)	0.00	26.6180
CRQ_5	4.6566	(Correlated)	0.00	25.1830
CRQ_6	4.8155	(Correlated)	0.00	26.1134
CRQ_7	5.3093	(Correlated)	0.00	26.1084
CRQ_8	5.8271	(Correlated)	0.00	24.8218
CRQ_9	6.2634	(Correlated)	0.00	23.8368
ETCQ_1	3.9482	(Correlated)	0.00	23.1207
ETCQ_2	4.6241	(Correlated)	0.00	23.0741
ETCQ_3	5.4425	(Correlated)	0.00	23.0818
ETCQ_4	6.3124	(Correlated)	0.00	23.0698

(a) Waiting time

Number Waiting	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00078560	(Insufficient)	0.00	1.0000
CRQ_1	8.0132	(Correlated)	0.00	64.0000
CRQ_10	7.7430	(Correlated)	0.00	63.0000
CRQ_11	7.7206	(Correlated)	0.00	63.0000
CRQ_12	7.7010	(Correlated)	0.00	63.0000
CRQ_2	7.9761	(Correlated)	0.00	64.0000
CRQ_3	7.9401	(Correlated)	0.00	64.0000
CRQ_4	7.9038	(Correlated)	0.00	64.0000
CRQ_5	7.8742	(Correlated)	0.00	64.0000
CRQ_6	7.8553	(Correlated)	0.00	64.0000
CRQ_7	7.8164	(Correlated)	0.00	63.0000
CRQ_8	7.7978	(Correlated)	0.00	63.0000
CRQ_9	7.7641	(Correlated)	0.00	63.0000
ETCQ_1	35.6957	(Correlated)	0.00	278.00
ETCQ_2	35.5963	(Correlated)	0.00	277.00
ETCQ_3	35.4861	(Correlated)	0.00	277.00
ETCQ_4	35.3799	(Correlated)	0.00	277.00

(b) Number waiting

Figure 9: Performance indicators: The second scenario.

Category Overview

Toll Plaza

Replications: 1 Time Units: Minutes

Queue

Time

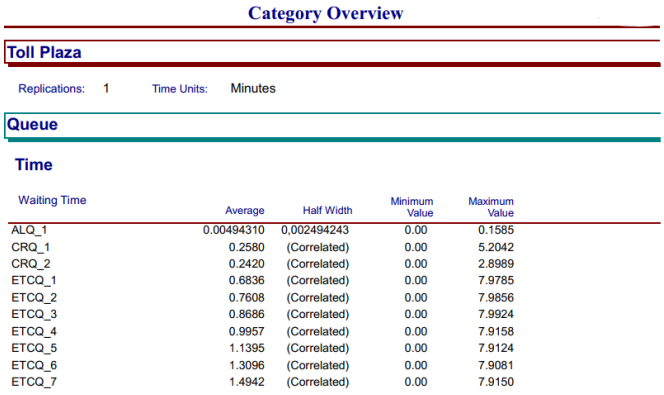
Waiting Time	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00333359	0,002532121	0.00	0.1707
CRQ_1	2.5841	(Correlated)	0.00	20.5455
CRQ_2	3.0176	(Correlated)	0.00	22.3880
CRQ_3	3.3551	(Correlated)	0.00	18.4739
CRQ_4	3.8355	(Correlated)	0.00	19.7483
CRQ_5	4.3805	(Correlated)	0.00	19.9995
CRQ_6	5.0772	(Correlated)	0.00	22.7152
ETCQ_1	0.6608	(Correlated)	0.00	7.2480
ETCQ_2	0.7563	(Correlated)	0.00	7.2418
ETCQ_3	0.8793	(Correlated)	0.00	7.2494
ETCQ_4	1.0303	(Correlated)	0.00	7.1692
ETCQ_5	1.2029	(Correlated)	0.00	7.1636
ETCQ_6	1.4019	(Correlated)	0.00	7.1556

(a) Waiting time

Number Waiting	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00078710	(Insufficient)	0.00	1.0000
CRQ_1	5.6420	(Correlated)	0.00	48.0000
CRQ_2	5.5700	(Correlated)	0.00	48.0000
CRQ_3	5.5057	(Correlated)	0.00	48.0000
CRQ_4	5.4496	(Correlated)	0.00	48.0000
CRQ_5	5.4026	(Correlated)	0.00	48.0000
CRQ_6	5.3522	(Correlated)	0.00	47.0000
ETCQ_1	6.0650	(Correlated)	0.00	87.0000
ETCQ_2	6.0136	(Correlated)	0.00	87.0000
ETCQ_3	5.9448	(Correlated)	0.00	87.0000
ETCQ_4	5.8984	(Correlated)	0.00	87.0000
ETCQ_5	5.8666	(Correlated)	0.00	86.0000
ETCQ_6	5.8307	(Correlated)	0.00	86.0000

(b) Number waiting

Figure 10: Performance indicators: The third scenario.



Category Overview

Toll Plaza

Replications: 1 Time Units: Minutes

Queue

Time

Waiting Time	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00494310	0,002494243	0.00	0.1585
CRQ_1	0.2580	(Correlated)	0.00	5.2042
CRQ_2	0.2420	(Correlated)	0.00	2.8989
ETCQ_1	0.6836	(Correlated)	0.00	7.9785
ETCQ_2	0.7608	(Correlated)	0.00	7.9856
ETCQ_3	0.8686	(Correlated)	0.00	7.9924
ETCQ_4	0.9957	(Correlated)	0.00	7.9158
ETCQ_5	1.1395	(Correlated)	0.00	7.9124
ETCQ_6	1.3096	(Correlated)	0.00	7.9081
ETCQ_7	1.4942	(Correlated)	0.00	7.9150

(a) Waiting time

Number Waiting	Average	Half Width	Minimum Value	Maximum Value
ALQ_1	0.00112936	(Insufficient)	0.00	1.0000
CRQ_1	0.3344	(Correlated)	0.00	9.0000
CRQ_2	0.2092	(Correlated)	0.00	9.0000
ETCQ_1	6.4671	(Correlated)	0.00	96.0000
ETCQ_2	6.4124	(Correlated)	0.00	96.0000
ETCQ_3	6.3624	(Correlated)	0.00	96.0000
ETCQ_4	6.3207	(Correlated)	0.00	95.0000
ETCQ_5	6.2869	(Correlated)	0.00	95.0000
ETCQ_6	6.2504	(Correlated)	0.00	95.0000
ETCQ_7	6.2281	(Correlated)	0.00	95.0000

(b) Number waiting

Figure 11: Performance indicators: The fourth scenario.

4 Conclusion

The benefits resulting from the use of ETC technology was evaluated in this paper. The analysis was accomplished by utilizing collected field data and simulated scenarios using ARENA simulation software. A comparison of different case studies was performed.

Results showed that an optimal toll plaza depends largely on the number of different types of tollbooths and the ETC share. An increase in the rate of ETC have a considerable impact in reducing the total delay caused at the manual lanes, but on the other hand it reduces the efficiency of the ETC lanes as they become gradually saturated with the increasing of its users.

Therefore, under a given time-dependent traffic flow and a set of different ETC rates, several scenarios were considered in order to find the optimal number of tollbooths that provide the best possible service.

As the proportion of the ETC user increases, number of tollbooths, queue length and waiting time are all decreased, resulting in high operating efficiency.

Table 3: Simulation results

Scenario	ETC rate	Optimal number of tollbooths	Maximum Average Waiting Time Obtained (min)
Scenario-1	40%	18 manual lanes, 3 ETC lanes and one automated lane	7
Scenario-2	60%	12 manual lanes, 4 ETC lanes and one automated lane	7
Scenario-3	80%	6 manual lanes, 6 ETC lanes and one automated lane	5
Scenario-4	95%	2 manual lanes, 7 ETC lanes and one automated lane	1.4

Conflict of Interest The authors declare no conflict of interest.

References

[1] Y. Liu, H. Liao, Z. Yu, M. Cai,"Analysis of the Operational Impact of ETC Lanes on Toll Station", in 11th International Conference of Chinese Transportation Professionals (ICCTP), 2011. DOI: 10.1061/41186(421)262

[2] H.M. Al-Deek, A.A. Mohamed, A.E. Radwan, "Operational Benefits of Electronic Toll Collection: Case Study", Journal of Transportation Engineering, **123**(6), 467–477, 1997. DOI: 10.1061/(ASCE)0733-947X(1997)123:6(467)

[3] H.M. Al-Deek, A.A. Mohamed, "Simulation and Evaluation of the OrlandoOrange County Expressway Authority (OOCEA) Electronic Toll Collection Plazas Using TPSIM", Florida Department of Transportation Research Center, April 2000.

[4] H.M. Al-Deek, A.A. Mohamed, E.A. Radwan, "New Model for Evaluation of Traffic Operations at Electronic Toll Collection Plazas", Transportation Research Record, **1710**(1), 1-10, 2000. DOI: 10.3141/1710-01

[5] Nezamuddin. "Developing microscopic toll plaza model using PARAMICS", MS Thesis, Department of Civil and Environmental Engineering in the College of Engineering and Computer Science at the University Of Central Florida Orlando, Florida, USA, 2006.

[6] L. Jian, D.I.N.G. Jiping, Y.E. Fan, "ETC System Benefit Evaluation and Performance Index System", Journal of Transportation Engineering and Information, **1**(1), 75-80, 2003.

[7] P. Tseng, D. Lin, S. Chien, "Investigating the impact of highway electronic toll collection to the external cost: A case study in Taiwan", Technological Forecasting and Social Change, **86**, 265-272, 2014. DOI: 10.1016/j.techfore.2013.10.019

[8] N.M.V. Dijk, M.D. Hermans, M.J.G. Teunisse, H. Schuurman, "Designing the Westercheldetunnel toll plaza using a combination of queueing and simulation", In Proceedings of the 31st conference on Winter simulation: Simulation–a bridge to the future, Association for Computing Machinery, New York, NY, USA, 1272–1279, 1999. DOI: 10.1145/324898.325061

[9] M. Burris, E. Hildebrand, "Using microsimulation to quantify the impact of electronic toll collection,"ITE Journal **66**(N7), 21–24, 2016.

[10] A.A. Saka, D.K. Agboh, S. Ndiritu, R.A. Glassco, "Estimation of Mobile Emissions Reduction from Using Electronic Tolls", Journal of Transportation Engineering, **127**(4), 327-333, 2001. DOI: 10.1061/(ASCE)0733-947X(2001)127:4(327)

[11] W.D. Kelton, R.P.Sadowski, D.A. Sadowski, "Simulation with arena", WCB/McGraw-Hill, 1998.

[12] G. Mackinnon, A. Kannan, M. Van Aerde, "Evaluating the impacts of implementing electronic toll collection using INTEGRATION", Towards an Intelligent Transport System. Proceedings of the First World Congress on Applications of Transport Telematics and Intelligent Vehicle-Highway Systems, November 30 - 3 December 1994, Paris, 1994.

[13] E.A. Gordin "Evaluation Of The Potential Benefits To Traffic Operations At A Toll Plaza With Express Etc Lanes." MS Thesis, Department of Civil and Environmental Engineering in the College of Engineering and Computer Science at the University of Central Florida Orlando, Florida, 2004.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 1293-1305 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Empirical Probability Distributions with Unknown Number of Components

Marcin Kuropatwiński*, Leonard Sikorski

Talking2Rabbit Ltd., R & D Department, 80-266, Poland

ARTICLE INFO

Article history:
Received: 28 August, 2020
Accepted: 09 December, 2020
Online: 16 December, 2020

Keywords:
probability mass function
probability density function
multinomial distribution
Laplace estimator
Gaussian mixture model

ABSTRACT

We consider the estimation of empirical probability distributions, both discrete and continuous. We focus on deriving formulas to estimate number of categories for the discrete distribution, when the number of categories is hidden, and the means and methods to estimate the number of components in the Gaussian mixture model representing a probability density function given implicitly in terms of its realizations. To reach the stated goals, we solve certain combinatorial problems for discrete distribution and develop methods to compute the expected Kullback-Leibler divergence for Gaussians. The last mentioned result is needed to develop the theory of continuous distributions. Sample applications and an extensive numerical study are given.

1 Introduction

This paper is an extension of work originally presented in 2019 Signal Processing Symposium [1]. The extension includes:

- measure theoretic account,
- derivation of the results first presented in the paper [1],
- new section devoted entirely to estimation of the number of components in continuous distributions,
- a numerical study of algorithms dealing with estimation of the continuous distributions,
- a specialized algorithm for counting number of monomials in a trace of covariance matrix raised to a power.

Empirical probability distributions are a crucial element of many applications such as machine learning [2], source coding [3], data compression [4], speech recognition [5], speaker recognition [6], image recognition [7], noise reduction [8], bandwidth extension [9], and many others. It is also a scientific discipline in itself, which is studied independently [10], [11]. In this paper, we deal with the means and methods to estimate probability distributions, both discrete and continuous. That being said, we also focus on a particular problem encountered during probability distributions estimation, which is the problem of an unknown number of components. In case of discrete distribution, the number of components is the number of categories, and in case of continuous probability distributions, it is the number of Gaussian components in the Gaussian mixture model (GMM) [2]. We provide estimators and methods to determine the number of components. Toward this, in case of GMMs, we formulate an information-theoretic criterion that allows the selection of an optimal, in some sense, number of components.

To strengthen the ties between discrete and continuous probability distribution, we provide a short account of measure theory and probability spaces. In particular, we show that the difference lies in the type of the underlying measurable space, which is a *standard space* with finite countable alphabets in case of discrete distributions and a *Polish Borel space* (where the *Polish space*, i.e., the complete separable metric space, is the sample space) in the case of continuous distributions. The name *Polish space* originates from the pioneering work by Polish mathematicians on such spaces.

1.1 Measure Theoretic Account

This account is based on the presentation provided by Robert M. Gray in an excellent book [12].

*Corresponding Author: Marcin Kuropatwiński, marcin@talking2rabbit.com

1.1.1 Measurable Space

A measurable space is a pair, $(\Omega, \mathcal{B})$ consisting of a sample space Ω with a σ-field $\mathcal{B}$ of subsets of Ω (also called the event space). A σ-field or σ-algebra $\mathcal{B}$ is a collection of subsets of Ω with the following properties:

$$\Omega \in \mathcal{B} \tag{1}$$

$$\text{If } F \in \mathcal{B}, \text{ then } F^c = \{\omega : \omega \notin F\} \in \mathcal{B} \tag{2}$$

$$\text{If } F_i \in \mathcal{B};\ i = 1, 2, \ldots, \text{ then } \cup F_i \in \mathcal{B} \tag{3}$$

The type of the measurable space is what differentiates discrete from continuous distributions. More precisely, it is the topology of the sample space. The discrete topology gives rise to discrete probability distributions, and the euclidean topology gives rise to continuous probability distributions.

Let $\mathcal{F} = \{F_i, i = 0, 1, 2, \ldots, n-1\}$ be a finite field of sample space Ω, that is, $\mathcal{F}$ is a finite collection of sets in Ω that are closed under finite set theoretic operations. To make it more concrete, let us give an example of such a field, which is the power-set of a finite, countable set of atoms (also called categories in the sequel). A set F in $\mathcal{F}$ will be called an *atom* only if its subset is also a field member of itself and the empty set, that is, it cannot be broken up into smaller pieces that are also present in the field. Let $\mathcal{A}$ denote a collection of atoms of $\mathcal{F}$. Then, one can show that $\mathcal{A}$ consists of exactly all nonempty sets of the form

$$\bigcap_{i=0}^{n-1} F_i^* \tag{4}$$

where F_i^* is either F_i or F_i^c. Let us call such sets *intersection sets*; we observe that any two intersection sets must be disjoint since for at least one i, one intersection set must lie inside F_i and the other within F_i^c. In summary, given any finite field $\mathcal{F}$ of the sample space Ω, we can find a unique collection of atoms $\mathcal{A}$ of the field such that the sets in $\mathcal{A}$ are disjoint, nonempty, and have the space Ω as their union. Thus, $\mathcal{A}$ is a *partition* of Ω. Such a sample space has a *discrete topology* with the basis given by the set of all atoms $\mathcal{A}$.

Now, we turn our attention to continuous distributions. As already mentioned, the underlying space is the *Polish Borel space*. An example of such a space, which will be used in this paper, is the *Euclidean space* endowed with the Euclidean topology. The basis for the Euclidean topology is the set of all open balls in that space. The *Polish space* is a complete, separable, metric space. We will explain what the listed properties of the *Polish spaces* mean.

A space is called *metric* if it is a set A, with elements called *points*, such that for every pair of points in A, there is an associated non-negative number $d(a, b)$ with the following properties:

$$d(a, b) = 0 \text{ if and only if } a = b \tag{5}$$

$$d(a, b) = d(b, a) \quad \textit{symmetry} \tag{6}$$

$$d(a, b) \leq d(a, c) + d(c, b) \text{ all } c \in A \quad \textit{triangle inequality} \tag{7}$$

A set F is said to be *dense* in A if every point in A is a point in F or a limit point of F.

A metric space A is called *separable* if it has a countable dense subset, that is, if there is a discrete set, say B, such that all points in A can be well approximated by points in B. This means that all the points in A are points in B or limits of the points in B. For example, n-tuples of rational numbers are dense in $\mathbb{R}^n$.

A sequence $\{a_n; n = 0, 1, 2, \ldots\}$ in A is called a *Cauchy sequence* if for every $\epsilon > 0$, there is an integer N such that $d(a_n, a_m) < \epsilon$ if $n \geq N$ and $m \geq N$. A metric space is *complete* if every Cauchy sequence converges, that is, if a_n is a Cauchy sequence, then there is $a \in A$ for which $a = \lim_{n\to\infty} a_n$.

1.1.2 Probability Spaces

A *probability space* $(\Omega, \mathcal{B}, P)$ is a triple consisting of a sample space Ω, a σ-field $\mathcal{B}$ of subsets of Ω, and a probability measure P defined on the σ-field; $P(F)$ assigns a real number to every member F of $\mathcal{B}$ so that the following conditions are satisfied:

Nonnegativity:

$$P(F) \geq 0, \text{ all } F \in \mathcal{B}, \tag{8}$$

Normalization:

$$P(\Omega) = 1. \tag{9}$$

Countable Additivity:

$$\text{If } F_i \in \mathcal{B},\ i = 0, 1, 2, \ldots \text{ are disjoint, then}$$
$$P(\bigcup_{i=0}^{\infty} F_i) = \sum_{i=0}^{\infty} P(F_i). \tag{10}$$

1.1.3 Densities

The probability density function (PDF) is a measure theoretic term defined through the *Radon-Nikodym theorem*.

A measure m is said to be *absolutely continuous* with respect to another measure P on the same measurable space, formally $m \ll P$, if $P(F) = 0$ implies $m(F) = 0$.

Theorem (*Radon-Nikodym theorem*)

Given the two measures m and P on a measurable space $(\Omega, \mathcal{F})$ such that $m \ll P$, there exists a measurable function $h : \Omega \to \mathbb{R}$ with the property that $h \geq 0$ such that

$$m(F) = \int_F h dP, \text{ all } F \in \mathcal{B}. \tag{11}$$

The function h is called the Radon-Nikodym *derivative* or *density* of m w.r.t. P and is denoted by $\frac{dm}{dP}$. If $\int f dP = 1$, then f is called a *probability density function*.

2 Discrete Distributions

For the statement of all results, as well as notations conventions, we refer the reader to the paper [1]. The results in the article [1] include discussion of the sun rise problem by Pierre-Simon Laplace from the essay [13] and some applications of the theory. We reference the paper [1] in its entirety. Here, only extensions of the material from the paper [1] are included, which have not found place there due to the form of a conference paper. However, for the readers convenience we provide a short summary of the notations used in the derivations:

- K - number of categories, possibly a hidden variable
- M - number of random experiments

- S - the partition, set of all categories
- $\{p_i\}_{i\in S}$ - multinomial proportions
- $X = \{x_i\}_{(i\in[1,M],x_i\in S)}$ - observations from the random experiments
- $|A|$ - cardinality of the set A
- UNIQUE(X) - unique elements in the multiset X
- $Z = |\text{UNIQUE}(X)|$ - the diversity index
- $P_i = |\{x_j : x_j = i\}|$ - counts of the observations

2.0.1 Derivations of the results for the uniform distribution

Assumption of uniform distribution is restrictive. However, it gives initial insight into the problem and, thus, is briefly presented here. As already introduced by X, we denote the sequence of observations. We can show that conditional probability of this sequence given the hypothetical K is equal to:

$$p(X|K) = \binom{K}{Z}\binom{M}{P_1 P_2 \cdots P_3}\frac{1}{K^M}. \tag{12}$$

It can be seen that the maximum likelihood estimate for the hypothetical number of categories, does not depend on the middle term, which includes the multinomial coefficient. Thus, the estimate can be obtained by,

$$K_{\text{ML}} = \underset{K}{\text{argmax}}\left[\binom{K}{Z}K^{-M}\right]. \tag{13}$$

It is convenient to introduce another quantity, which plays an important role in our theory. This quantity is the *generalization coefficient* N, which equals by definition:

$$N \equiv \frac{M}{Z}. \tag{14}$$

The ML estimate of K can then be obtained by solving the following equation:

$$\frac{1}{v}\ln\left(\frac{1}{1-v}\right) = N, \tag{15}$$

where:

$$v = \frac{Z}{K_{\text{ML}}}, \tag{16}$$

is the fraction of the number of observed categories to the number of all categories. The condition for the likelihood to be monotonically decreasing is as follows:

$$M > \log_{\left(\frac{Z+1}{Z}\right)}(Z+1). \tag{17}$$

If this condition holds, the ML estimate for K is equal to Z. The derivations of the above conditions are in place next. First, we note a property that establishes the link with the known in the statistical literature problem of coupon collector [14]:

$$\lim_{K\to\infty}\left(\frac{K\times H(K)}{\log_{\left(\frac{K+1}{K}\right)}(K)}\right) = 1, \tag{18}$$

where one can easily recognize that $K \times H(K)$ is the expected number of trials before the coupon collector collects the whole collection. In the above expression, $H(K)$ is the harmonic number, equal by definition:

$$H(K) \equiv \sum_{i=1}^{K}\frac{1}{i}. \tag{19}$$

The main vehicle of the derivation is the following expression that is valid for the harmonic numbers [15]:

$$\sum_{i=1}^{K}\frac{1}{i} = C + \ln K + \frac{1}{2K} - \sum_{i=2}^{\infty}\frac{A_i}{K(K+1)\cdots(K+i-1)}, \tag{20}$$

where C is the Euler-Mascheroni constant. It can be seen that asymptotically, as K approaches infinity, the terms after the logarithmic term vanish to zero. This leads to the following property:

$$\lim_{K\to\infty}\left(\sum_{i=1}^{K}\frac{1}{i} - \ln K\right) = C. \tag{21}$$

The derivation for the condition (15) begins with taking the logarithm of the considered expression (13):

$$\ln[p(X|K)] = \sum_{i=1}^{K}\ln i - \sum_{i=1}^{Z}\ln i - \sum_{i=1}^{K-Z}\ln i - M\ln K. \tag{22}$$

Suppose, i is a continuous variable, which setting follows from allowing that variable to take on non-integer values. It can be seen that the middle sum does not depend on K; therefore, derivative w.r.t that variable reads:

$$\frac{d}{dK}\ln[p(X|K)] = \sum_{i=1}^{K}\frac{1}{i} - \sum_{i=1}^{K-Z}\frac{1}{i} - \frac{M}{K} = \ln K - \ln(K-Z) - \frac{M}{K}. \tag{23}$$

The last expression allows us to immediately state the (15) and (16). Equations (15) and (16) allow us to conclude that the sample length needed to learn a given percent of the categories S is a multiple of K. Setting $Z = M$, we see that, indeed, this ($Z = M$) is the sufficient and necessary condition for optimal K approaching infinity. This is due to the following identity:

$$\lim_{K\to\infty}\frac{K}{M}\ln\left(\frac{1}{1-\frac{M}{K}}\right) = 1. \tag{24}$$

It remains to prove the (17). In this case, the maximum of the likelihood should be attained at $K = Z$. Thus, we have the following inequality:

$$\ln[p(X|Z)] > \ln[p(X|Z+1)], \tag{25}$$

which implies:

$$\ln(Z+1) - M\ln(Z+1) < -M\ln Z, \tag{26}$$

and, after some algebra, we attain at the (17). From the above reasoning all the results pertaining to the discrete, uniform distribution from the paper [1] can be deduced. In Figure 1, we illustrate the dependence of the data amount M needed to learn a given percent of all categories.

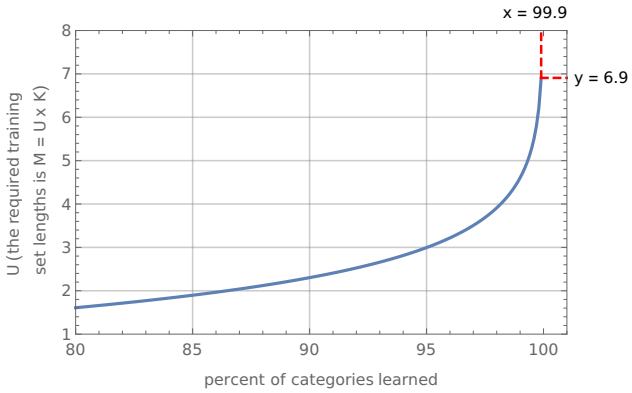


Figure 1: Data amounts requirement to learn a given percent of categories in case of uniform distribution

Next, we discuss the case of the non-uniform distribution, which is more relevant for real-world applications.

2.0.2 *Derivations of the results for the non-uniform distribution*

The next step will be the derivation of the conditions analogous to that introduced in the previous section, which are distribution-free (we relax the assumption of categories of equal probabilities). The desired effect will be achieved by using the earlier introduced multinomial proportions $\{p_i\}_{i\in S}$. Since we do not impose a constraint on the multinomial proportions, the considered probability function is now in the form,

$$p(X|K, p_1, \cdots, p_K) = \binom{M}{P_1 P_2 \cdots P_Z} \sum_{\{\mathbf{k}\,:\,\text{combinations of } Z \text{ objects out of } K \text{ objects}\}} \prod_{i=1}^{Z} p_{k_i}^{P_i}, \tag{27}$$

where $\mathbf{k} = [k_1, \cdots, k_Z]$. We integrate the above function over a unit simplex D:

$$D = \left\{ \mathbf{p} : \sum_{i=1}^{K} p_i = 1, p \in \mathbb{R}_+^K \right\}. \tag{28}$$

This corresponds to the assumption that all PMFs are equally likely, meaning we assume that we do not know the true PMF and attach to each possible $\mathbf{p} = [p_1, \cdots, p_K]$ an equal weight (we assume they are equally probable):

$$p(X|K) = \frac{1}{\text{vol}(D)} \int_D p(X|K, \mathbf{p}) d\mathbf{p} = \binom{K}{Z}\binom{M}{P_1 P_2 \cdots P_Z} \frac{1}{\text{vol}(D)} \int_D p_1^{P_1} \times \cdots \times p_Z^{P_Z} d\mathbf{p}, \tag{29}$$

where the expression (29) follows the fact, that the value of the integral does not depend on the choice and order of the probabilities in the monomial integrand. To proceed with the derivation, we compute the integral in (29) using Brion's formulae; please see [16]:

$$\frac{1}{\text{vol}(D)} \int_D p_1^{P_1} \times \cdots \times p_Z^{P_Z} d\mathbf{p} = (K-1)! \frac{\prod_{i=1}^{Z} P_i!}{(M+K-1)!}. \tag{30}$$

In light of the above expression, we get,

$$p(X|K, M) = \frac{K! M! \Gamma(K)}{Z!(K-Z)!\Gamma(K+M)}, \tag{31}$$

and next,

$$p(Z|K, M) = \frac{M\Gamma(K)\Gamma(K+1)\Gamma(M+Z)}{\Gamma(Z+1)^2\Gamma(K+M)\Gamma(K-Z+1)}. \tag{32}$$

To get the probability of the hypothetical number of categories, given M and Z, we apply the following derivation:

$$p(K|Z, M) = \frac{p(Z|K, M)p(K)}{p(Z|M)} = \frac{p(Z|K, M) \times C}{\sum_{K=Z}^{\infty} p(Z|K, M) \times C} = \frac{p(Z|K, M)}{\sum_{K=Z}^{\infty} p(Z|K, M)}, \tag{33}$$

where C is some constant (not to be confused with the Euler-Mascheroni constant used earlier in this document). Evaluation of (33) yields the following:

$$p(K|Z, M) = \frac{\Gamma(K)\Gamma(K+1)\Gamma(M-1)\Gamma(M)}{\Gamma(Z)\Gamma(Z+1)\Gamma(K+M)\Gamma(K-Z+1)\Gamma(M-Z-1)}. \tag{34}$$

From the reasoning presented above we can state the following:

$$s = \frac{u}{u+1}, \quad s = \frac{Z}{K}, \quad u = \frac{M}{K}. \tag{35}$$

Also, all results pertaining to non-uniform distribution, stated in the paper [1], are a consequence of the above reasoning. Based on (35) we can plot the dependence of the amount of data needed to learn given percent of all categories. This dependence is shown in Figure 2.

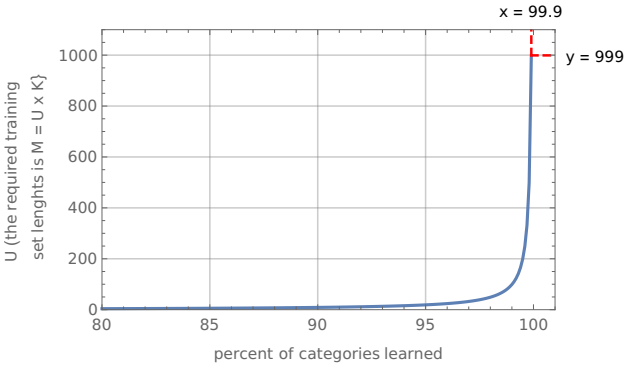


Figure 2: Data amounts requirement to learn a given percent of categories in case of non-uniform distribution

3 Continuous Distributions

The prominent method of density estimation for continuous distribution are Gaussian mixture models (GMM). The GMMs can approximate a PDF given by some training vectors drawn from that PDF. A wide-spread fitting procedure of the GMMs to the given training vectors is based on the Expectation-Maximization (EM) algorithm. Both, the mathematical definition of GMMs and the fitting procedure based on the EM is explained in detail in the book [2]. The neat property of the GMMs is that their expressive power allows modeling any PDF to any accuracy, given enough training

vectors are available. However, the traditional fitting procedure is susceptible to overtraining if too many components are chosen. Thus, in this paper, we study a fitting method that does not overfit, and, secondly, allows for the estimation of the number of components given the newly introduced information-theoretic criterion. To reach the goal of formulating the criterion, we first develop a mathematical approach to compute the expected Kullback-Leibler (KL) divergence for the single multivariate Gaussian, whereas the expectation is taken over given number M of samples used to estimate the multivariate Gaussian. In this section, we first develop the theory of expected KL divergence; next, we provide the modified training procedure and end the section with a numerical study.

3.1 Computation of expected Kullback-Leibler divergence

To the best of the authors' knowledge, the computation of the expected Kullback-Leibler divergence for multivariate Gaussian variables has been never previously explored in the literature. Thus, this paper deals with this problem, which can be formally expressed as,

$$\mathrm{E}_{\mathrm{KL}}(M) = \mathbb{E}_{x_i\sim p(x)}\left[D_{\mathrm{KL}}(p(x)\|\hat{q}(x,x_1,\cdots,x_M))\right], \tag{36}$$

where $p(x)$ and $q(x)$ are multivariate Gaussians.

The solution to the problem indicated above can be used in the context of the Minimum Description Length (MDL) cf. [17] or Minimum Discrimination Information (MDI), cf. [18]. MDI and MDL principles have been used often for solving computational learning problems (see [19]–[23]).

The well-known result from the statistics is the following formula for the KL divergence between two multivariate Gaussians:

$$D_{\mathrm{KL}}(p\|q) = \int p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x, \tag{37}$$

for $p(x)$ and $q(x)$ given as:

$$p(x) = \mathcal{N}(x|\mu_1,\Sigma_1) \quad \text{and} \quad q(x) = \mathcal{N}(x|\mu_2,\Sigma_2), \tag{38}$$

where:

$$\mathcal{N}(x|\mu,\Sigma) = \det(2\pi\Sigma)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right), \tag{39}$$

and finally (see [24]):

$$D_{\mathrm{KL}}(p\|q) = \frac{1}{2}\Big[\log\left(\frac{\det(\Sigma_2)}{\det(\Sigma_1)}\right) - d + \mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1)\Big], \tag{40}$$

where d is the dimension of PDFs p and q.

Next, let us assume that the positively defined and symmetric Σ_1 is given, μ_1 is a zero vector, and Σ_2 and μ_2 are estimated from samples x_i, where $x_i \sim \mathcal{N}(x|\mu_1,\Sigma_1)$:

$$\mu_2 = \frac{1}{M}\sum_{i=1}^{M} x_i \quad \Sigma_2 = \frac{1}{M}\sum_{i=1}^{M}(x_i-\mu_1)(x_i-\mu_1)^T = \frac{1}{M}\sum_{i=1}^{M} x_i x_i^T, \tag{41}$$

where M is the number of samples (length of the observation). We also assume that Σ_2 is positively definite and thus well-conditioned (otherwise KL divergence would go to infinity). This assumption leads to the requirement that M has to be greater than the dimension of Σ_1 when Σ_2 is non-diagonal.

As already indicated in (36), we are looking for the expectation of the Kullback-Leibler divergence over finite sample:

$$\mathbb{E}_{x_i\sim p(x)}\left[D_{\mathrm{KL}}(p(x)\|\hat{q}(x,x_1,\cdots,x_M))\right],$$

where $p(x)$ and $q(x)$ are given in (38).

We consider two cases: the first case is when the covariance matrix is diagonal, and the second case, which is a more complicated one, is when the covariance matrix is full.

3.1.1 Diagonal Case

At first, it is worth noticing that the expectation of the Kullback-Leibler divergence does not depend on a specific form of the covariance matrix Σ_1. Let us introduce the following substitution:

$$x_i = \Sigma_1^{\frac{1}{2}} y_i \quad \text{where} \quad y_i \sim \mathcal{N}(y|0,I), \tag{42}$$

which, in turn, under the assumption given by the (41), gives:

$$\mathbb{E}\left[\log\left(\frac{\det(\Sigma_2)}{\det(\Sigma_1)}\right)\right] = \mathbb{E}[\mathrm{tr}(\log\Sigma)], \tag{43}$$

where:

$$\Sigma = \frac{1}{M}\sum_{i=1}^{M}\mathrm{diag}(y_i\odot y_i), \tag{44}$$

where $\odot$ denotes the element-wise product, the Hadamard product. It is worth mentioning that we have used the following identity (see [25]):

$$\log(\det(\Sigma)) = \mathrm{tr}(\log\Sigma). \tag{45}$$

Since Σ is diagonal and y_i has a zero mean vector, we conclude,

$$\begin{aligned}\mathbb{E}[\mathrm{tr}(\log\Sigma)] = \\ \mathbb{E}\left[\sum_{k=1}^{d}\log\left(\frac{1}{M}\sum_{i=1}^{M} y_i(k)^2\right)\right] = \\ \mathbb{E}\left[\sum_{k=1}^{d}\log(\sigma_M(k)^2)\right] = \\ = d\cdot\mathbb{E}[\log(\sigma_M(1)^2)] = \\ d\left(\psi\left(\frac{M-1}{2}\right) + \log 2 - \log(M-1)\right),\end{aligned} \tag{46}$$

where $\psi(n)$ is the digamma function.

The next expression, which appears in (40), is:

$$\mathrm{tr}(\Sigma_2^{-1}\Sigma_1) = \mathrm{tr}(\Sigma^{-1}), \tag{47}$$

where Σ is defined as in (44). After a number of transformations, we get

$$\mathbb{E}[\mathrm{tr}\Sigma^{-1}] = \mathbb{E}\left[\sum_{k=1}^{d}\frac{1}{\sigma_M^2(k)}\right] = d\cdot\mathbb{E}\left[\frac{1}{\sigma_M^2(1)}\right] = d\frac{1}{M-2}. \tag{48}$$

The last expression, which needs to be calculated, is

$$\mathbb{E}[\mu_2^T \Sigma_2^{-1} \mu_2] = \mathbb{E}\left[\sum_{k=1}^{d} \frac{\bar{x}(k)^2}{\sigma_M(k)^2}\right] = d \cdot \mathbb{E}\left[\frac{\bar{x}(1)^2}{\sigma_M(1)^2}\right] = d\frac{M-1}{M\cdot(M-2)}. \quad (49)$$

And finally, our derived formula reads,

$$\mathrm{E_{KL}}(M) = \frac{d}{2}\left[\psi\left(\frac{M-1}{2}\right) + \log 2 - \log(M-1) - 1 + \frac{M}{M-2} + \frac{M-1}{M\cdot(M-2)}\right]. \quad (50)$$

3.1.2 Non-diagonal Case

In the non-diagonal covariance matrix case, we first introduce the same substitution as in the diagonal case, which is shown in (42):

$$x_i = \Sigma_1^{\frac{1}{2}} y_i \quad \text{where} \quad y_i \sim \mathcal{N}(y|0, I). \quad (51)$$

We start by calculating the logarithmic expression, which appears in (40). By using identity in (45), and transforming the expression using (41), as previously, we get,

$$\log\left(\frac{\det(\Sigma_2)}{\det(\Sigma_1)}\right) = \mathrm{tr}(\log \Sigma). \quad (52)$$

To calculate the expectation of this expression, we use the expansion of the matrix logarithm into the power series, see [26]:

$$\log A = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(A-I)^k}{k}, \quad (53)$$

and, after a number of transformations, we get,

$$\mathbb{E}[\mathrm{tr}(\log \Sigma)] = \mathbb{E}[\mathrm{tr}(\log m\Sigma - \log mI)] = \mathbb{E}\left[\mathrm{tr}\left(\sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k}\sum_{j=0}^{k}\binom{k}{j}(-1)^j m^{k-j}\Sigma^{k-j} - \log mI\right)\right], \quad (54)$$

where m is chosen so as to $\|m\Sigma - I\| < 1$ is satisfied.

We further develop our approximation by introducing the following auxiliary function:

$$z(k, k_{max}) = \begin{cases} \sum_{i=1}^{k_{max}} (-1)^{k+1} \cdot \binom{i}{k} \cdot \frac{1}{i} & \text{when } k = 0 \\ \sum_{i=k}^{k_{max}} (-1)^{k+1} \cdot \binom{i}{k} \cdot \frac{1}{i} & \text{when } k > 0 \end{cases}, \quad (55)$$

and, consequently:

$$\mathbb{E}[\mathrm{tr}(\log \Sigma)] = \sum_{k=0}^{k_{max}} z(k, k_{max}) \cdot m^k \cdot \mathbb{E}\left[\mathrm{tr}\left(\Sigma^k\right)\right] - d \log m, \quad (56)$$

where d is a dimension of the covariance matrix. To proceed further, we calculate the expectation of $\mathrm{tr}(\Sigma^k)$. It is worth noticing that the integral

$$\int_{-\infty}^{+\infty} t^n \mathcal{N}(t|0,1)\mathrm{d}t = \frac{1}{\sqrt{2\pi}} 2^{\frac{n-1}{2}}((-1)^n + 1)\Gamma\left(\frac{n+1}{2}\right), \quad (57)$$

vanishes for odd n. At this point, we need to calculate the number of monomials with only even exponents that appear in $\mathrm{tr}(\Sigma^k)$.

Now we can describe the algorithm for calculating the number of such monomials. First, we introduce the following auxiliary matrix:

$$A = \frac{1}{\sqrt{M}}\left[y_1 \middle| \ldots \middle| y_M\right]. \quad (58)$$

One can easily notice that:

$$\Sigma = \frac{1}{M}\sum_{i=1}^{M} y_i y_i^T = A \cdot A^T, \qquad \Sigma^n = \underbrace{(AA^T)\ldots(AA^T)}_{n}, \quad (59)$$

which leads to formula for the specific element, ${\Sigma^n}_{ij}$, of the covariance matrix raised to the n-th power:

$${\Sigma^n}_{ij} = A_{i*} \cdot A^T \ldots A \cdot A^T_{*j} = \sum_{s_1=1}^{d}\sum_{k_1=1}^{M} a_{ik_1} \cdot a^{(T)}_{k_1 s_1} \ldots \sum_{s_{n-1}=1}^{d}\sum_{k_n=1}^{M} a_{s_{n-1}k_n} \cdot a^{(T)}_{k_n j} = \sum_{s_1,\ldots s_{n-1}} \sum_{k_1,\ldots,k_n} a_{ik_1} a_{s_1 k_1} a_{s_1 k_2} \ldots a_{s_{n-1}k_n} a_{jk_n}, \quad (60)$$

The number of the above configurations with only even exponents is equal to the number of even partitions of $2n$. For such a partition, we have to calculate the number of monomials separately. The detailed description of the algorithm for counting monomials is presented in section A. Implementation of the algorithm for counting monomials and computing expected KL divergence is available (see [27]). Below, we have shown examples of the formulas derived by the above-mentioned algorithm, for the power of $n = 3$, where d is a dimension of the covariance matrix and M is the number of samples used to estimate the matrix:

$$\begin{gathered} \text{Partition: 2 2 2} \\ dM(M-1)(M-2) + dM(d-1)(d-2) + 3dM(M-1)(d-1) \\ \\ \text{Partition: 4 2} \\ 3dM(M-1) + 3dM(d-1) \\ \\ \text{Partition: 6} \\ dM \end{gathered} \quad (61)$$

Once we have calculated the number of monomials for each power n, it is relatively easy to calculate the expectation in (56).

The second expression, which has to be calculated, is

$$\mathrm{tr}\left(\Sigma_2^{-1}\Sigma_1\right). \quad (62)$$

We start in the same way as in the diagonal case:

$$\mathrm{tr}\left(\Sigma_2^{-1}\Sigma_1\right) = \mathrm{tr}\Sigma^{-1}. \quad (63)$$

The next step is to expand inversion of the covariance matrix into the Neumann series (see [28]):

$$\Sigma^{-1} = m \sum_{k=0}^{\infty} (I - m\Sigma)^k, \tag{64}$$

where m is selected so that $\|I - m\Sigma\| < 1$ is satisfied. After several transformations, we get

$$\Sigma^{-1} \approx \sum_{k=0}^{k_{max}} (-1)^k m^{k+1} \binom{k_{max}+1}{k+1} \Sigma^k, \tag{65}$$

which finally gives,

$$\mathbb{E}[\mathrm{tr}\Sigma^{-1}] \approx \sum_{k=0}^{k_{max}} (-1)^k m^{k+1} \binom{k_{max}+1}{k+1} \mathbb{E}[\mathrm{tr}\Sigma^k]. \tag{66}$$

We note that the resulting expression has a similar form to (56), with the difference of the k-dependent expression. The last expression for which the expectation needs be calculated is

$$(\mu_2 - \mu_1)^T \Sigma_2^{-1} (\mu_2 - \mu_1). \tag{67}$$

By using substitution (41), and subsequently (42), we get

$$\mu_2^T \Sigma_2^{-1} \mu_2 = \left(\frac{1}{M} \sum_{i=1}^{M} \Sigma_1^{\frac{1}{2}} y_i \right)^T \Sigma_2^{-1} \left(\frac{1}{M} \sum_{i=1}^{M} \Sigma_1^{\frac{1}{2}} y_i \right) = \left(\frac{1}{M} \sum_{i=1}^{M} y_i \right)^T \Sigma^{-1} \left(\frac{1}{M} \sum_{i=1}^{M} y_i \right), \tag{68}$$

then, given (58), and by expanding the inverse of the covariance matrix as in (64) into the Neumann series, we get:

$$\mu_2^T \Sigma_2^{-1} \mu_2 = \sum_{i,j} \left[\frac{1}{M^2} A^T \Sigma^{-1} A \right]_{ij} \approx \frac{1}{M^2} \sum_{i,j} \sum_{k=0}^{k_{max}} (-1)^k m^{k+1} \binom{k_{max}+1}{k+1} A_{i*}^T \Sigma^k A_{*j}. \tag{69}$$

Now, we need to calculate the following expression:

$$A_{i*}^T \Sigma^n A_{*j}. \tag{70}$$

By multiplying out we get:

$$A_{i*}^T \Sigma^n A_{*j} = \sum_{s_0=1}^{d} \sum_{s_n=1}^{d} a_{is_0}^{(T)} \left[\Sigma^n\right]_{s_0 s_n} a_{s_n j} = \sum_{s_0} \sum_{s_n} a_{s_0 i} \sum_{s_1,\ldots,s_{n-1}} \sum_{k_1,\ldots,k_n} a_{s_0 k_1} a_{s_1 k_1} a_{s_1 k_2} \ldots a_{s_{n-1} k_n} a_{s_n k_n} a_{s_n j} = = \sum_{s_1,\ldots,s_{n+1}} \sum_{k_1,\ldots,k_n} a_{s_1 i} a_{s_1 k_1} \ldots a_{s_{n+1} k_n} a_{s_{n+1} j}. \tag{71}$$

An important observation is in place, namely, the monomials cannot have only even exponents if $i \neq j$. This allows us to bring down the calculation of the expectation of the above expression to the expectation of the trace of Σ^{n+1}. Thus, finally, we get

$$\mathbb{E}[\mu_2^T \Sigma_2^{-1} \mu_2] \approx \frac{1}{M} \sum_{k=0}^{k_{max}} (-1)^k m^{k+1} \binom{k_{max}+1}{k+1} \mathbb{E}[\mathrm{tr}\Sigma^{k+1}]. \tag{72}$$

Obviously, the algorithm for calculating the above expression is the same as in the case of the two previously examined terms.

3.2 Procedure

Here, we show how to estimate the GMM using the modified training procedure. The introduced procedure solves the problem of overtraining. First, we focus on training the GMM with fixed K - the number of components in the mixture. Based on this development, we formulate a method to select the optimal number of components for a given training set without resorting to the development sets. The classical Expectation-Maximization algorithm, cf. [2], produces in each iteration more and more accurate estimates of the mean vectors, covariance matrices, and weights. We denote the trajectory of these parameters in subsequent iterations as,

$$\theta = \left[\theta^{(1)}, \cdots, \theta^{(J)}\right], \tag{73}$$

where J is the maximum acceptable number of iterations (or the number of iterations until convergence), and

$$\theta^{(j)} = \left[\Sigma_1^{(j)}, \cdots, \Sigma_K^{(j)}, \mu_1^{(j)}, \cdots, \mu_K^{(j)}\right], \tag{74}$$

with j the iteration number, Σ_i denoting covariance matrix for the $i-$th component, μ_i denoting the mean vectors for the $i-$th component, and K the number of components in the Gaussian mixture. Moreover, we denote $\theta_k^{(j)} = [\Sigma_k^{(j)}, \mu_k^{(j)}]$. To proceed further, we introduce and define some mathematical entities:

- the *entropy* in nats of the categorical distribution $\{r_i\}_i$ is defined as $H(\{r_i\}) = -\sum_i r_i \log r_i$
- the *kernel width* is the effective number of samples used to estimate a given component. Given the samples $\{x_i\}_{i=1}^M$, we compute the *kernel width*, $w(\theta, \{x_i\})$, also abbreviated by $w(\theta)$, using the following set of formulas:

$$r_i = \frac{\mathcal{N}(x_i|\theta)}{\sum_l \mathcal{N}(x_l|\theta)}, \tag{75}$$

$$w(\theta) = \exp\left(H(\{r_i\})\right), \tag{76}$$

- Gaussian *component differential entropy* in nats, $H(\theta)$, will be defined as:

$$H(\theta) = \frac{1}{2} \log\left(\det(2\pi e \Sigma)\right), \tag{77}$$

- the *score* or *cross entropy* for the component is expressed by the following formula:

$$s(\theta) = H(\theta) + \mathrm{E}_{\mathrm{KL}}\left(w(\theta)\right). \tag{78}$$

We see that the expression for *cross entropy* depends only on the observed quantities.

Equipped with the above definitions, we can formulate the algorithm,

Algorithm 1: train GMM

Require: starting parameters values $\theta^{(0)}$, maximal number of iterations J, initial scores $s\left(\theta_k^{(0)}\right) = \infty$, set all weights to $\rho_k = \frac{1}{K}$.
Ensure: $\theta^{(J)}$ and weights $\{\rho_k\}$.
1: **for** $j \in [1, \ldots, J]$ **do**
2: compute $\theta^{(j)}$ update according to EM (keeping weights unchanged)
3: **for** $k \in [1, \ldots, K]$ **do**
4: **if** $s\left(\theta_k^{(j)}\right) > s\left(\theta_k^{(j-1)}\right)$ **then**
5: $\theta_k^{(j)} = \theta_k^{(j-1)}$ (backtracking)
6: **end if**
7: **end for**
8: **end for**
9: run few EM updates of weights alone
10: **return** $\theta^{(J)}$ and weights $\{\rho_k\}$

Algorithm 2: train GMM

Require: starting parameters values $\theta^{(0)}$, maximal number of iterations J, initial scores $s\left(\theta_k^{(0)}\right) = \infty$, set all weights to $\rho_k = \frac{1}{K}$.
Ensure: $\theta^{(J)}$ and weights $\{\rho_k\}$.
1: **for** $j \in [1, \ldots, J]$ **do**
2: compute $\theta^{(j)}$ update according to EM (keeping weights unchanged)
3: **for** $k \in [1, \ldots, K]$ **do**
4: **if** $s\left(\theta_k^{(j)}\right) > s\left(\theta_k^{(j-1)}\right)$ **then**
5: $\theta_k^{(j)} = \theta_k^{(j-1)}$ (backtracking)
6: **end if**
7: **end for**
8: **end for**
9: ~~run few EM updates of weights alone~~
10: **return** $\theta^{(J)}$ and weights $\{\rho_k\}$

The mechanism behind the algorithm 1 is as follows:

- if the number of components is large in comparison to the number of training vectors, the precision of each component goes up with each iteration until it eventually reaches infinity; this means at the same time, $H(\theta)$ goes toward $-\infty$ (in practice, it will stop at a low value due to the constraint we put on the minimal eigenvalue of the covariance matrix)
- the second term in the expression for *score*, see 78, to the contrary, grows toward $+\infty$ as *kernel width* approaches dimension of the training vectors d. Thus, even if *componenent differential entropy* goes toward $-\infty$, the *cross entropy* has a minimal extreme point and at the very last grows toward $+\infty$.
- thus, the statements from 3-7 in the algorithm 1 prevent the collapse of the components to something resembling a dirac delta
- in effect, the algorithm does not overtrain

We also examine the algorithm 2, which does not run a few EM updates of weights (statement 9 in algorithm 1).

3.2.1 Number of components selection criterion

The *cross entropy* in 78 is the expression on number of nats needed to code from the component. Based on the *cross entropy*, we can get the number of bits needed to code with given fidelity using the Shannon-Lower-Bound [29].

Given the component *weights* $\{\rho_k\}_{k=1}^K$, the average number of nats needed to code from the GMM modeled source PDF is equal to

$$C_1\left(K, \theta^{(J)}, \{\rho_k\}\right) = C_1(K) = \underbrace{H\left(\{\rho_k\}_{k=1}^K\right)}_{\text{components indices entropy}} + \underbrace{\sum_{k=1}^{K} \rho_k s\left(\theta_k^{(J)}\right)}_{\text{mean per component cross entropy}} . \quad (79)$$

We also consider a second criterion, which seems to give meaningful results:

$$C_2\left(K, \theta^{(J)}, \{\rho_k\}\right) = C_2(K) = \underbrace{\sum_{k=1}^{K} \rho_k s\left(\theta_k^{(J)}\right)}_{\text{mean per component cross entropy}} . \quad (80)$$

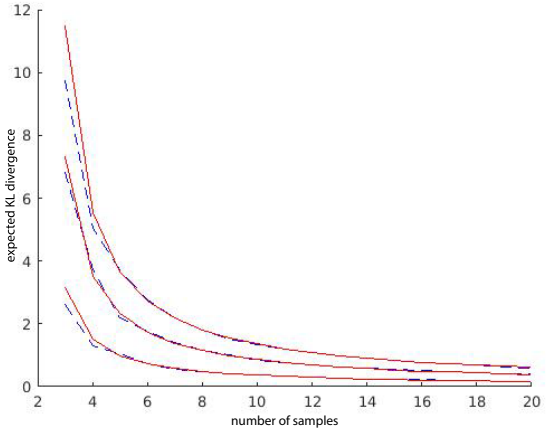


Figure 3: Comparison of the empirically and analytically obtained expectation of the KL divergence for Gaussians with the diagonal covariance matrices. Dashed lines represent the Monte Carlo result. Dimensions of covariance matrices from the beginning of the coordinate axes: 3, 7, and 11, respectively.

To find the optimal number of components K (which minimizes the average number of nats needed to code from the source), we proceed as follows:

- we swap the K from 1 to M
- for each K we run the algorithm 1 or 2.

- we smooth with smoothing splines, see [30], the resulting curve $C_*(K)$, obtaining a smoothed curve $\tilde{C}_*(K)$
- we select $\hat{K}$ for which $\tilde{C}_*(\hat{K})$ attains minimum and returns the optimal number of components.

Above, the asterisk means either 1 or 2.

3.3 Numerical experiments

3.3.1 Expected Kullback-Leibler Divergence

In this section, the comparison our analytical results with the Monte-Carlo obtained curves of expected Kullback-Leibler divergence will be presented. First, the results for the diagonal covariance matrix will be shown in Figure 3.

As can be seen, the analytical expectation shows high accuracy for the diagonal covariance matrices.

The results are a bit worse for the non-diagonal matrix. The following issues affected the accuracy of the result:

- the calculation of the higher degree series expansions was too hard as the computation of the number of even exponent monomials grew exponentially with n. We were able to compute the number of monomials for n as high as eight (the formulas for $n = 8$ when written on A4 page taking 44000 rows).
- The Monte-Carlo curves on the left are sensitive to the threshold for the detection of the semi-definiteness of the matrices. It sometimes happens that one of the eigenvalues of the covariance matrix is very small. Then value of term $\mathrm{tr}\Sigma^{-1}$ grows enormously. It is clear that the result depends on the choice of the threshold.

Figure 4 depicts the results for non-diagonal covariance matrices. To obtain the plot, we rejected matrices with the smallest eigenvalue less than 0.01 (which is around $\frac{1}{100}$ of the largest eigenvalue).

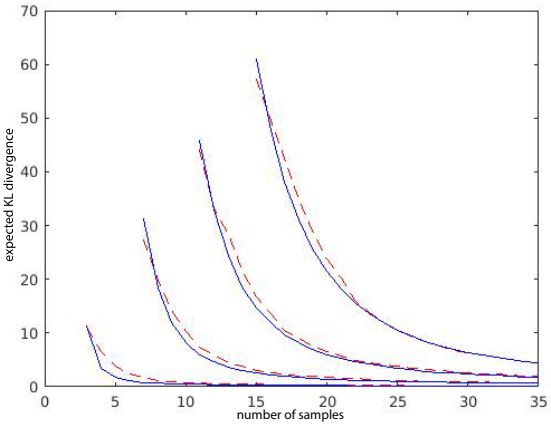


Figure 4: Comparison of empirically and analytically obtained expectation of the KL divergence for Gaussians with the non-diagonal covariance matrices. Dashed lines represent the Monte Carlo estimation. Dimensions of covariance matrices from the beginning of the coordinate axes: 3, 7, 11, and 15, respectively.

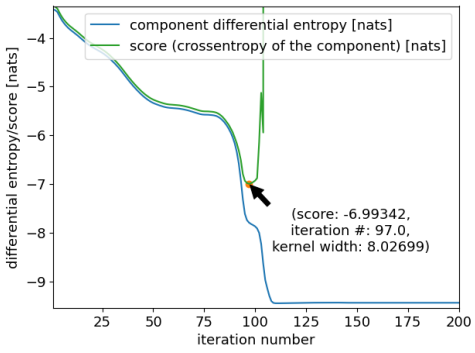


Figure 5: Illustration of the overtraining prevention mechanism. Plot shows that score prevents the overtraining, differential entropy of the Gausssian component goes down and the score starts increasing after kernel width exceeds 8 training points

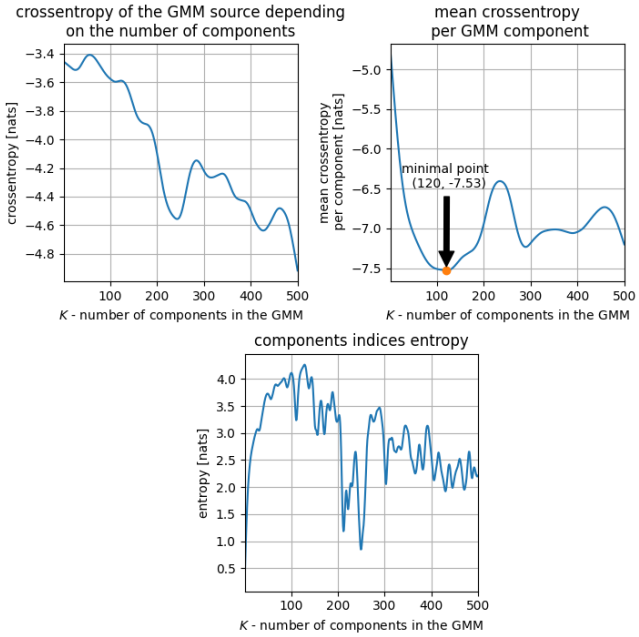


Figure 6: Results of the algorithm 1 for $d = 2$

3.3.2 Overtraining Prevention Mechanism

Here, we present a drawing (Figure 5) showing the trajectory of the *score* and *component differential entropy*. The plot has been obtained by by recording the parameters for subsequent iterations for one chosen component of the GMM. We see that the *score* attains minimum but *component differential entropy* goes to $-\infty$. This allows to use the backtracking procedure present in algorithms 1 and 2.

3.3.3 Numerical Analysis of the Number of Components Selection Criteria

All numerical experiments of this section has been carried out using the Line Spectral Frequencies (LSF) as the data to which the GMMs are fitted. The LSFs were computed from the LibriSpeech database. In all subsequent experiments the training set length M equals 1000.

First, the workings of the algorithm 1 will be presented. The Figure 6 shows the criteria C_1 and C_2 and the *components indices entropy* on pictures from left to right and from top to bottom, respectively. We see erratic behaviour of the *components indices entropy* that falls down, meaning it silences some components by driving some weights to zero. The criterion C_1 suggests that the optimal number of components is > 500, what is unlikely and probably caused by silencing of some components by zero weights. The C_2 criterion, instead, gives a sensible result. We see that for algorithm 2, selected by the C_2 criterion, number of components is similar.

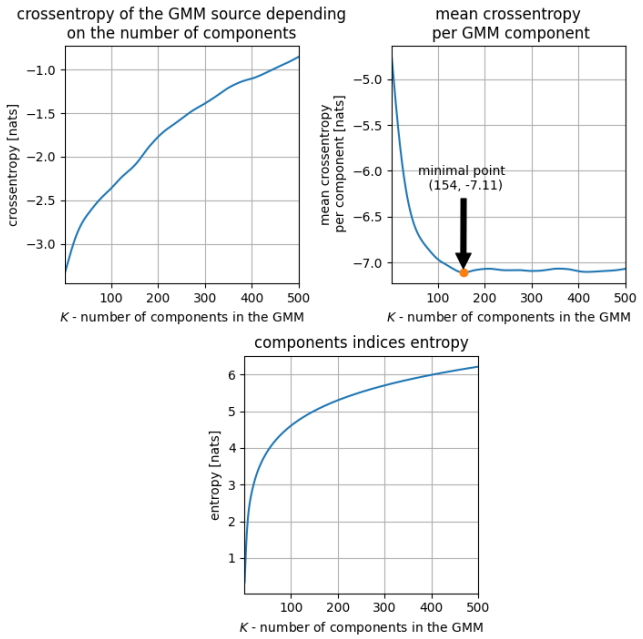


Figure 7: Results of the algorithm 2 for $d = 2$

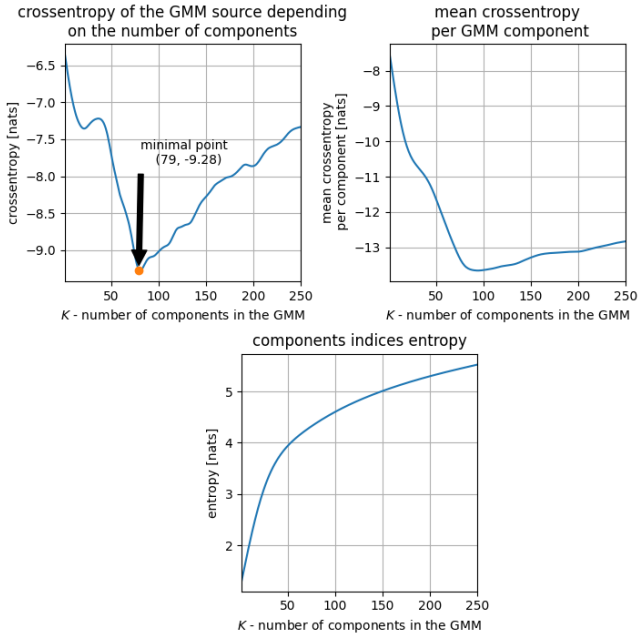


Figure 8: Results of the algorithm 2 for $d = 4$

As the second example, we show the workings of the algorithm 2. The plots in Figure 7 show the same curves as in the case of the Figure 6 (see previous paragraph). The C_1 criterion points as the optimal number of components exactly one component, that is, a single Gaussian. Above $K = 1$, the *components indices entropy* penalizes the *score* to the extent that the score curve is monotonically increasing. Another optimal number of components is suggested by C_2. We computed the GMMs for the number of components selected using the C_2 criterion, for which the contour plots are shown on Figure 10. We see that the GMM is tightly fitted to the data for the selection criterion C_2. However, the composite criterion C_1 suggests a single component, which seems to by trivial, but still optimal, as indicated by the methodology proposed.

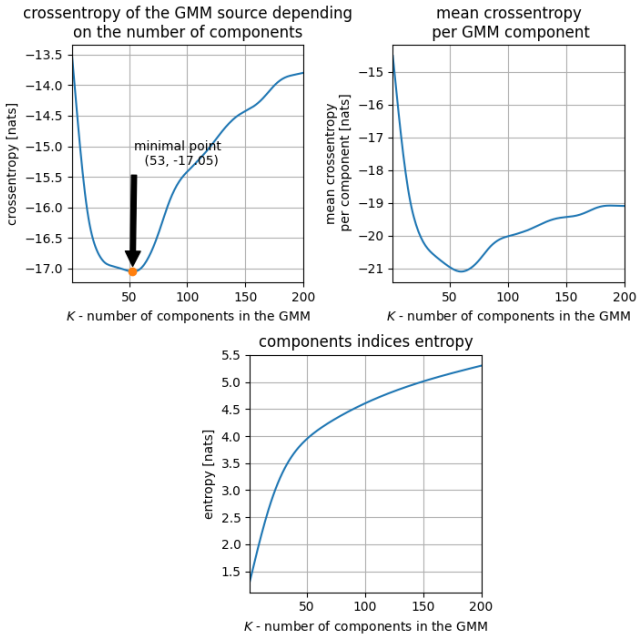


Figure 9: Results of the algorithm 2 for $d = 8$

As the third example, we show the working of the algorithm 2 for the problem dimension $d = 4$ (first four line spectral frequencies). In this case, the criterion C_1 returns trustworthy results, the minimum on the C_1 curve is clearly noticeable. The returned result is sensible and indicates that for $d = 4$, the training vectors distribution needs more components to be modeled accurately; the distribution is probably far from Gaussian. The curves are presented in Figure 8. This experiment is an indication that the proposed methodology is sound. To make the evidence even stronger we present more experiments for dimensions $d = 8$, Figure 9, and $d = 16$, Figure 12. As expected the optimal number of components decreases with the dimension of the problem. This is to maintain proper generalization.

The last result shows the behavior of the classical number of components selection criterion, that is the maximization of the likelihood on the development set. For this experiment, from 1000 samples of the training set, we excluded 100 samples as the development set. In Figure 13, we show the plot of the log-likelihood on the development set as a function of the number of components. It is evident that the plot is quite erratic of what decreases trust in this method of selection of the number of components. The results can

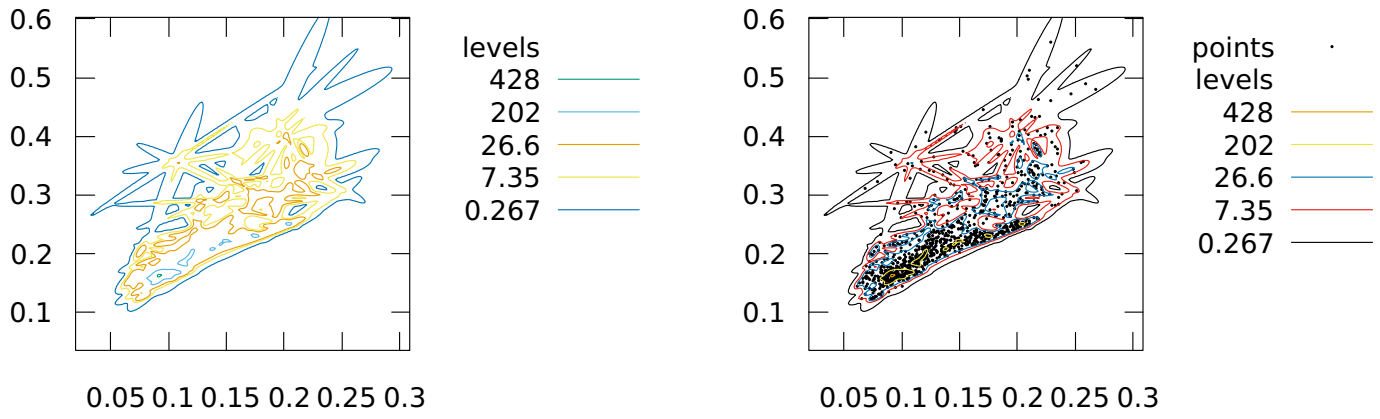


Figure 10: GMM fitted using algorithm 2 with number of components selected using the criterion C_2

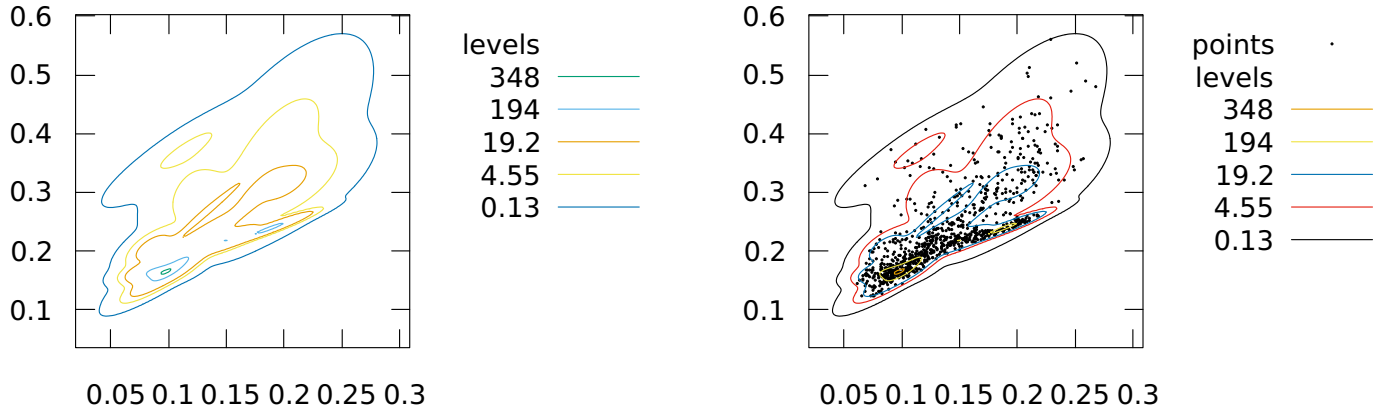


Figure 11: GMM fitted using traditional EM algorithm with number of components selected using tuning on the development set

be made more stable if we allow the development set to be larger. Such a measure will come at the cost of reducing the number of training points. Here, our proposed method shows the advantage of not using development sets. The contour plots for the GMM fitted using traditional method with the number of components tuned on the development set are shown in Figure 11.

4 Conclusions

In this paper, the theory of both discrete and continuous distribution with unknown number of components has been developed. The measure theoretic reason for the inherent similarity between PMFs and PDFs, with the difference in the structure of the underlying sample space, has been given. The difference causes the two distribution categories to be treated with much different mathematical tools. The main result of the paper are the means and methods of computing the number of components, that is, the number of categories in the PMF case and the number of Gaussian components in the GMM modeling in a PDF case, which is given implicitly in terms of its realizations.

Development of the theory required considering the expected Kullback-Leibler divergence - a difficult problem on its own rights. Especially, the algorithm for counting monomials with only even exponents in the expression for a trace of a matrix raised to a power has been developed. This algorithm allows to compute certain integrals analytically without resorting to the Monte-Carlo experiments. This algorithm can also be of interest on its own rights.

We believe that the theory presented in this paper will find many practical applications in diverse fields of science, technology, and engineering as a convenient tool of data analysis.

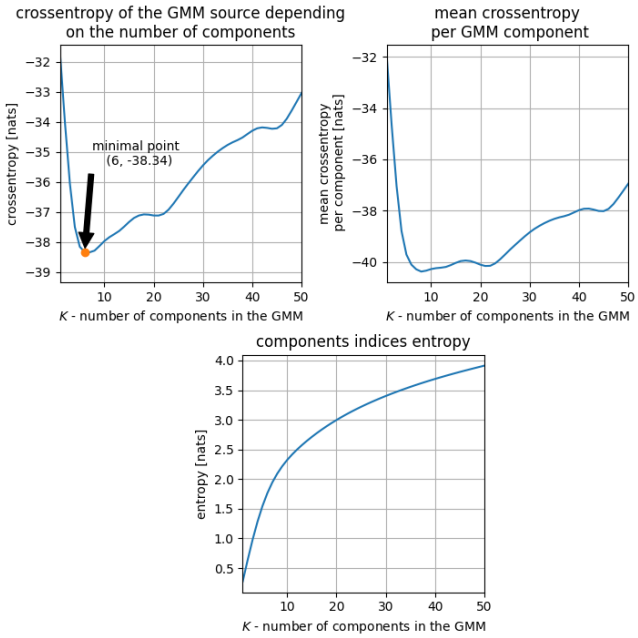


Figure 12: Results of the algorithm 2 for $d = 16$

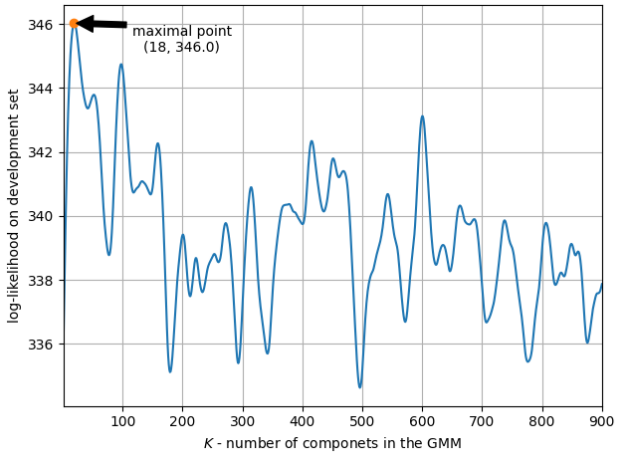


Figure 13: Results of the traditional EM algorithm $d = 2$

References

[1] M. Kuropatwiński, "Estimation of Quantities Related to the Multinomial Distribution with Unknown Number of Categories," in 2019 Signal Processing Symposium (SPSympo), 277–281, 2019, doi:10.1109/sps.2019.8881992.

[2] C. M. Bishop, Pattern recognition and machine learning, Springer, 2006.

[3] R. M. Gray, Source coding theory, Springer, 2012.

[4] R. M. Gray, "Gauss Mixture Vector Quantization," in IEEE International Conference on Acoustics, Speech, and Signal Processing, 1769–1772, 2001, doi:10.1109/icassp.2001.941283.

[5] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey et al., The HTK book, Cambridge University Engineering Department, 2002.

[6] D. A. Reynolds, T. F. Quatieri, R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," Digital Signal Processing, **10**(1), 19–41, 2000, doi:10.1006/dspr.1999.0361.

[7] H. Permuter, J. Francos, I. Jermyn, "A study of Gaussian mixture models of color and texture features for image classification and segmentation," Pattern Recognition, **39**(4), 695–706, 2006, doi:10.1016/j.patcog.2005.10.028.

[8] A. Kundu, S. Chatterjee, A. S. Murthy, T. Sreenivas, "GMM Based Bayesian Approach To Speech Enhancement in Signal/Transform Domain," in 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, 4893–4896, 2008, doi:10.1109/icassp.2008.4518754.

[9] M. Nilsson, H. Gustaftson, S. V. Andersen, W. B. Kleijn, "Gaussian Mixture Model Based Mutual Information Estimation Between Frequency Bands in Speech," in 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, 525–528, 2002, doi:10.1109/icassp.2002.1005792.

[10] B. W. Silverman, Density estimation for statistics and data analysis, CRC Press, 1986.

[11] L. Devroye, G. Lugosi, Combinatorial methods in density estimation, Springer, 2012.

[12] R. M. Gray, Probability, random processes, and ergodic properties, Springer, 1988.

[13] P. S. Laplace, Philosophical essay on probabilities, Springer, 1998.

[14] P. Flajolet, D. Gardy, L. Thimonier, "Birthday paradox, coupon collectors, caching algorithms and self-organizing search," Discrete Applied Mathematics, **39**, 207–229, 1992, doi:10.1016/0166-218x(92)90177-c.

[15] S. Gradshteyn, I. Ryzhik, Table of integrals, series, and products, Academic Press, 2014.

[16] V. Baldoni, N. Berline, J. De Loera, M. Köppe, M. Vergne, "How to integrate a polynomial over a simplex," Mathematics of Computation, **80**(273), 297–325, 2011, doi:10.1090/s0025-5718-2010-02378-6.

[17] J. Rissanen, "Modeling by shortest data description," Automatica, **14**(5), 465–471, 1978, doi:10.1016/0005-1098(78)90005-5.

[18] S. Kullback, Information theory and statistics, Courier Corporation, 1997.

[19] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, C. J. Taylor, "A minimum description length approach to statistical shape modeling," IEEE Transactions on Medical Imaging, **21**(5), 525–537, 2002, doi:10.1109/tmi.2002.1009388.

[20] R. M. Gray, J. C. Young, A. K. Aiyer, "Minimum Discrimination Information Clustering: Modeling and Quantization with Gauss Mixtures," in 2001 International Conference on Image Processing, 14–17, 2001, doi:10.1109/icip.2001.958039.

[21] M. H. Hansen, B. Yu, "Model selection and the principle of minimum description length," Journal of the American Statistical Association, **96**(454), 746–774, 2001, doi:10.1198/016214501753168398.

[22] Y. Ephraim, A. Dembo, L. R. Rabiner, "A minimum discrimination information approach for hidden Markov modeling," IEEE Transactions on Information Theory, **35**(5), 1001–1013, 1989, doi:10.1109/18.42209.

[23] R. M. Gray, A. Gray, G. Rebolledo, J. Shore, "Rate-distortion speech coding with a minimum discrimination information distortion measure," IEEE Transactions on Information Theory, **27**(6), 708–721, 1981, doi:10.1109/tit.1981.1056410.

[24] J. Duchi, "Derivations for Linear Algebra and Optimization," Technical report, University of California at Berkeley, 2007.

[25] C. S. Withers, S. Nadarajah, "log det A = tr log A," International Journal of Mathematical Education in Science and Technology, **41**(8), 1121–1124, 2010, doi:10.1080/0020739x.2010.500700.

[26] B. Hall, Lie groups, Lie algebras, and representations: An elementary introduction, Springer, 2015.

[27] L. Sikorski, M. Kuropatwiński, "Counting Monomials Algorithm," `https://www.codeocean.com/`, 2019, doi:10.24433/CO.3748956.v1.

[28] D. Zhu, B. Li, P. Liang, "On the Matrix Inversion Approximation Based on Neumann Series in Massive MIMO Systems," in IEEE International Conference on Communications (ICC), 1763–1769, 2015, doi:10.1109/icc.2015.7248580.

[29] T. Berger, Rate distortion theory: A mathematical basis for data compression, Prentice-Hall, 1971.

[30] C. De Boor, A practical guide to splines, Springer, 1978.

A The algorithm for counting monomials

The solution of the problem from section 3.1.2 is divided into six short algorithms

Algorithm 3: Counting monomials with only even exponents for given even partition of number n

Require: Even partition $\mathcal{P} = \{p_1, ..., p_k\}$ of the number n, problem dimension d,number of random samples m
Ensure: Number of monomials with only even exponents
1: $S := 0$
2: **for** every multipermutation of $\mathcal{P}$ **do**
3: Create $\mathcal{T} = \{\underbrace{1, ..., 1}_{p_1}, ..., \underbrace{k, ..., k}_{p_k}\}$
4: **for** every multipermutation of $\mathcal{T}$ **do**
5: Check if $\mathcal{T}$ is correct with Algorithm (7).
If Algorithm (7) returned false, go to the next multipermutation of $\mathcal{P}$.
6: Create matrices D and M for the table $\mathcal{T}$ with Algorithm (4).
7: If there are unfilled places in matrices D and M, create pairs of matrices D' and M' by filling unfilled places in D and M in every possible way.
8: From the set of every generated in previous step pairs of matrices, discard these which are incorrect. Correctness of the pair of matrices is checked with Algorithm (5).
9: For every pair of matrices calculate it's numerical value with Algorithm (6) and add it to S.
10: **end for**
11: **end for**
12: **return** S

Algorithm 4: Creating matrices D and M

Require: $\mathcal{T} = \{t_1, ..., t_n\}$, where $t_i \in \{1, ..., k\}$
Ensure: Filled matrices D and M
1: Create matrices D and M of a dimension k
2: **for** $i = 1$ to n **do**
3: **if** i is odd **then**
4: To $M_{t_i t_{i+1}}$ and $M_{t_{i+1} t_i}$ write "=".
5: To $D_{t_i t_{i+1}}$ and $D_{t_{i+1} t_i}$ write "$\neq$".
6: **end if**
7: **if** i is even **then**
8: To $D_{t_i t_{i+1}}$ and $D_{t_{i+1} t_i}$ write "=".
9: To $M_{t_i t_{i+1}}$ and $M_{t_{i+1} t_i}$ write "$\neq$".
10: **end if**
11: **end for**
12: **return** matrices D and M
Note: If there is an occurrence of writing "$\neq$" or "=" in previously filled place, stop the algorithm and return nothing.

Algorithm 5: Checking if given pair of matrices D and M is correct

Require: Matrices D and M
Ensure: *True*, if given pair is correct; *False* in other cases
1: **for** every pair of indexes i, j of the matrix **do**
2: **if** $D_{ij} = M_{ij} =$ "=" **then**
3: **return** *False*
4: **end if**
5: **if** the negation of any from implications listed below is true:
$[(X_{ij} =$ "=") $\wedge$ $(X_{ik} =$ "=")$] \Rightarrow X_{kj} =$ "=" or $[(X_{ij} =$ "=") $\wedge$ $(X_{ik} =$ "$\neq$")$] \Rightarrow X_{kj} =$ "$\neq$"
where X is a matrix D or M **then**
6: **return** *False*
7: **end if**
8: **end for**
9: **return** *True*

Algorithm 6: Calculating a component of sum from formula for number of monomials

Require: Pair of matrices D and M, problem dimension d, number of random samples m
Ensure: A component of sum from formula for number of monomials
1: $q = 1$
2: **for** $i = 1$ to dim D **do**
3: **if** $i = 1$ **then**
4: $q = q \cdot d \cdot m$
5: **else**
6: Calculate c_d oraz c_m with Algorithm (8) with input data: (D, i) and (M, i)
7: **if** there exists $i < j$, such that $D_{ij} =$ "$\neq$" **then**
8: $q = q \cdot (d - c_d)$
9: **end if**
10: **if** there exists $i < j$, such that $M_{ij} =$ "$\neq$" **then**
11: $q = q \cdot (m - c_m)$
12: **end if**
13: **end if**
14: **end for**
15: **return** q

Algorithm 7: Checking if table $\mathcal{T}$ is correct

Require: $\mathcal{T} = \{t_1, ..., t_n\}$, where $t_i \in \{1, ..., k\}$
Ensure: *True*, if $\mathcal{T}$ is correct; *False* in other cases
1: $A = \emptyset$
2: **for** $i = 1$ to n **do**
3: **if** there exists $s < i$, such that $t_s > t_i$ and $t_i \notin A$ **then**
4: **return** *False*
5: **end if**
6: insert t_i into A
7: **end for**
8: **return** *True*

Algorithm 8: Calculating auxiliary minuend

Require: Matrix X, number i
Ensure: Auxiliary minuend
1: **if** $i = 1$ **then**
2: **return** 0
3: **end if**
4: **for** each $j < i$ **do**
5: **if** If $X_{ij} =$ "$\neq$" **then**
6: Calculate c_j with Algorithm (6) with input data (X, j)
7: **end if**
8: **end for**
9: $c = \min_j c_j$
10: **return** $c + 1$

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1306-1311 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Basic Study of 3-D Non-Invasive Measurement of Temperature Distribution using Ultrasound Images during HIFU Heating

Ryosuke Sakakibara[*, 1], Yasuhiro Shindo[2], Kazuo Kato[3], Pak Kon Choi[4], Akira Takeuchi[5]

[1]*Graduate School of Science and Technology, Meiji University, 214-8571, Japan*

[2]*Department of Mechanical Engineering, Toyo University, 350-0815, Japan*

[3]*Department of Mechanical Engineering Informatics, Meiji University, 214-8571, Japan*

[4]*Department of Physics, Meiji University, 214-8571, Japan*

[5]*Department of Thermotherapy, Luke Hospital, 104-0045, Japan*

ARTICLE INFO

Article history:
Received: 16 September, 2020
Accepted: 01 December, 2020
Online: 16 December, 2020

Keywords:
HIFU
Ultrasound
Non-Invasive Temperature Measurement

ABSTRACT

High Intensity Focused Ultrasound (HIFU) was widely used for treating tumors non-invasively. In this treatment, ultrasound is focused on the target volume inside the human body to ablate cancerous tissues and Magnetic Resonance Imaging (MRI) is mainly used to grasp the target position and to measure the temperature distributions around the target. However, MRI is very expensive, and a large space is required.

In this paper, we presented a method for measuring the temperature distribution using an ultrasound diagnostic device, which is inexpensive and commonly used in many clinics, and actually showed the results of heating experiments on a human shaped agar phantom. The proposed method for measuring the temperature distribution around the heated target was conducted by performing image processing on two ultrasound images before and after heating. Furthermore, it was confirmed that it was possible to grasp the three-dimensional temperature distribution from the images in multiple layers. The effectiveness of the temperature distribution measurement results by the proposed method was shown by comparing the temperature measurement results with the infrared thermal camera. The error between the results was approximately 1 ℃.

It was found that the non-invasive measurement method of the three-dimensional temperature distribution around the target volume using the ultrasound images was useful for effective HIFU treatments.

1. Introduction

This paper is an extension of a previous work originally presented at the 21st edition of the International Conference on Electromagnetics in Advanced Applications (ICEAA 2019) [1].

High Intensity Focused Ultrasound (HIFU) devices are used for treating uterine fibroids, liver tumors, brain tumors, etc. as a minimal invasive treatment method [2, 3]. As shown in Figure 1, this treatment method focuses on ultrasound waves emitted from the outside of the body to a target inside the body and causes ablation necrosis of the tumor by the thermal energy generated at the focal point. This method is considered to be a treatment method that has a much smaller burden on the patient than conventional open surgery. Focused ultrasound treatment is performed at a high temperature of approximately 70 °C or higher, and therefore there is a risk of cauterizing normal tissue around the tumor. When performing this treatment, a device with MRI's guidance (MR-guided focused ultrasound surgery: MRgFUS) is commonly used. However, a large amount of introduction funds and a large space are required. In addition, this method is not considered to be suitable for cases in which there is metal inside the human body or MRI contrast media cannot be used because the inside of the body is monitored using MR images [4-6].

[*]Corresponding Author: Ryosuke Sakakibara, 1-1-1, Higashi-mita, Tama-ku, Kawasaki, Kanagawa, 214-8571, +81-44-934-7907, rsakakibara9@gmail.com

Therefore, in order to perform Focused Ultrasound Surgery (FUS) treatment, non-invasive temperature measurement using an ultrasound diagnostic imaging device that is cheaper and more convenient to move than an MRI device has been previously attempted [7-12]. In this study, the ultimate goal is to establish a method for non-invasively measuring three-dimensional temperature distribution in a living body with high resolution in subpixel units [9]. In the temperature measurement method using ultrasound images, it is a prerequisite that the used ultrasound diagnostic imaging system was designed to draw images with constant ultrasound velocity. The velocity of ultrasound waves changes depending on the temperature of the propagating tissue. Therefore, if there is a temperature change inside the tissue, the ultrasound velocity in that part locally changes, and it appears as an extremely small image displacement on the ultrasound image. The temperature could be measured by detecting this slight displacement with image analysis technology [8, 9].

In the present paper, first, the algorithm of two-dimensional temperature distribution measurement using ultrasound images is described. Next, the agar phantom was heated in a temperature-controlled water tank, the relationship between the temperature increase of the agar and the displacement of the ultrasound image was described. Next, the temperature distribution was measured from the ultrasound image inside the agar phantom when it was heated using the focused ultrasound heating device. The temperature distribution inside the agar phantom during heating was measured using two types of transducers with different focal lengths. In the previous research, we investigated the basic heating characteristics by measuring the temperature distribution using a cylindrical agar phantom [10]. In this study, we measured a wider region of temperature distributions using a human shaped agar phantom in order to confirm that the proposed temperature measurement method is effective in situations closer to clinical treatment.

From the examination of the above experimental results, the usefulness of temperature distribution measurement using an ultrasound diagnostic imaging system was clarified quantitatively.

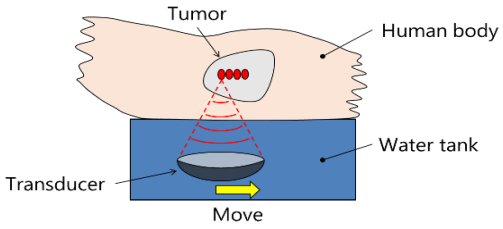


Figure 1: Illustration of Focused Ultrasound Surgery.

2. Materials and Methods

2.1. Principle of Temperature Measurement Using Ultrasound Images

As reported by Iseki et al., the temperature increases in living tissue can be obtained by solving (1) and (2) [8, 9].

$$\Delta T(x) = K_{tissue} \frac{\partial(\Delta d)}{\partial x} \quad (1)$$

$$K_{tissue} = \frac{1}{\alpha - \beta} \quad (2)$$

In these equations, ΔT is the temperature increase inside the tissue and K_{tissue} is the thermal constant that varies with the tissue. The subscript "tissue" is added to emphasize that it differs depending on the tissue. Δd is the displacement of the ultrasound image due to the temperature increase (the difference between the images before and after heating). The thermal constant K_{tissue} is expressed by (2). Here, α is the coefficient of thermal expansion of the tissue, and β is the coefficient of ultrasound velocity change due to the temperature increase inside the tissue. It is considered that the thermal expansion coefficient α and the ultrasound velocity change coefficient β change linearly in a relatively narrow temperature change region such as during focused ultrasound treatment [6].

2.2. Measurement System of Thermal Constant

Figure 2 shows the general-purpose ultrasound imaging system Nemio SSA-550A (Toshiba Medical Systems, Japan) with the attached ultrasound probe PLM-805AT (Toshiba Medical Systems, Japan) used in the experiment. The ultrasound probe used in the experiment is a linear probe as shown in Figure 2 (b). The number of elements is 128, the focus length is 20 mm, and the size of the contact part with the body is (12 mm × 67 mm). The size of the captured ultrasound image is (560 pixels × 450 pixels), which is a 256-gradation bitmap image. Here, 300 pixels were assigned to a distance of 70 mm on the captured ultrasound image.

Figure 3 shows an illustration of experimental setup for measuring thermal constant "*K*" to know the relationship between the temperature increase of the agar used as the heated object and the displacement of the ultrasound images before and after heating [8,9]. As shown in Figure 3, a cube agar phantom was fixed inside a water tank where the water temperature was constantly controlled using a thermostat and a stirring pump, and the ultrasound images inside the agar were taken from 20 ℃ to 30 ℃, because the experiment of temperature distribution measurement is in that temperature range.

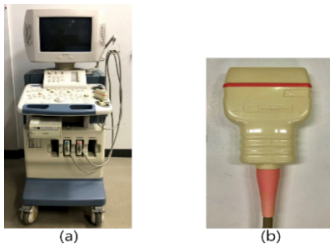


Figure 2: Ultrasound Diagnosis Device (a) Ultrasound Diagnosis Device, (b) Linear Probe.

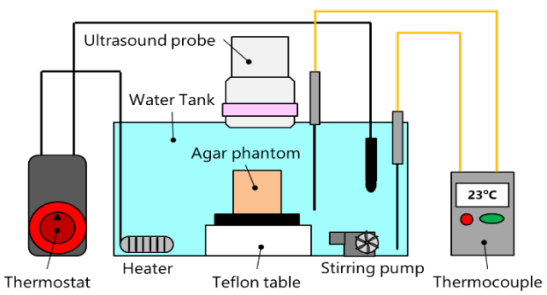


Figure 3: Illustration of Experimental Setup for Measuring Thermal Constant "*K*".

2.3. Image Analysis Algorithm

Figure 4 shows the image analysis algorithm proposed by authors for temperature distribution measurement [8, 9]. This algorithm is divided into 7 steps. First, the ultrasound images of the heated object before and after heating are input. Second, template matching is performed to measure the amount of displacement Δd between the ultrasound images before and after heating. Third, a median filter is applied to reduce the erroneous vector of displacement Δd on the ultrasound image. Fourth, since the displacement Δd distribution detected by template matching is expressed in pixel units, interpolation processing was performed to interpolate Δd distribution. Fifth, a moving average filter is applied to smooth the displacement amount of the distribution map subjected to this interpolation process. Sixth, the temperature increase ΔT during heating is calculated from the displacement Δd detected by applying the Sobel filter. Finally, a moving average filter is applied to the temperature increase distribution map to smooth the distribution. For the development of the image analysis program, general-purpose program Microsoft Visual Studio (Microsoft Japan) was used.

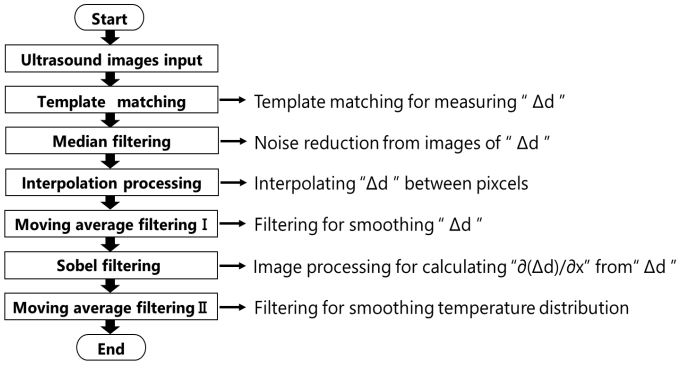


Figure 4: Flow Chart for Measuring Temperature Distributions.

2.4. Measurement System of Temperature Distributions

Figure 5 shows the device system for capturing ultrasound images before and after heating. This system mainly consists of two work sliders, an arm and a jig for fixing the ultrasound probe. As shown in Figure 5, the ultrasound probe was supported by the arm and fixed to the slider. Here, the ultrasound probe can be moved three-dimensionally by combining the two sliders. The range of motion of this slider is (0.0 to 66.0 mm), and the minimum movable interval is 0.05 mm.

Figure 6 shows an illustration of the focused ultrasound heating experiment. As shown in Figure 6, in order to measure the temperature distribution in the X-Z cross section using an infrared thermal camera, the human shaped agar phantom was previously cut at the center position. The transducer was placed underwater to focus on the position 65 mm from the bottom. Then, the ultrasound probe was moved up and down ± 15.0 mm from the center cross section every 5.0 mm along the Z direction, and ultrasound images were taken before and after heating in a total of 7 layers. These operations were performed for heating experiments with short focus and long focus transducers respectively. The heating time was 5 minutes and the heating power was 200 W. In addition, we compared the temperature distribution results by the present method using the ultrasound image and the thermal image taken by the infrared thermal camera in the central section of the human shaped agar phantom.

The focused ultrasound heating device used in the experiments was CZ901 (Mianyang Sonic Electronic, China). Two kinds of transducers were prepared for this device. Figure 7 shows the transducer used in the experiments. The short focus transducer as shown in Figure 7 (a) had an outer diameter of 170 mm, an inner diameter of 98 mm, an outer peripheral surface height of 75 mm, an inner peripheral surface height of 53 mm, and a focal length of 120 mm. The outer and inner diameters of the long focus transducer as shown in Figure 7 (b) were almost the same as those of the short focus transducer, however the inner peripheral surface height was 61 mm and focal length was 200 mm. The frequency of the focused ultrasound was 1 MHz. In the heating experiments, a frequency oscillator AWG1025F (Frequency Band: 1μHz – 25MHz, AS ONE, Japan) and a power amplifier L-400BM-HC (Frequency Band: 700kHz – 3.5MHz, JSE, Japan) were connected to the transducers to heat the human shaped agar phantom.

Figure 8 shows the human shaped agar phantom used in the experiments. The width of heating area was 160 mm and the height was 120 mm.

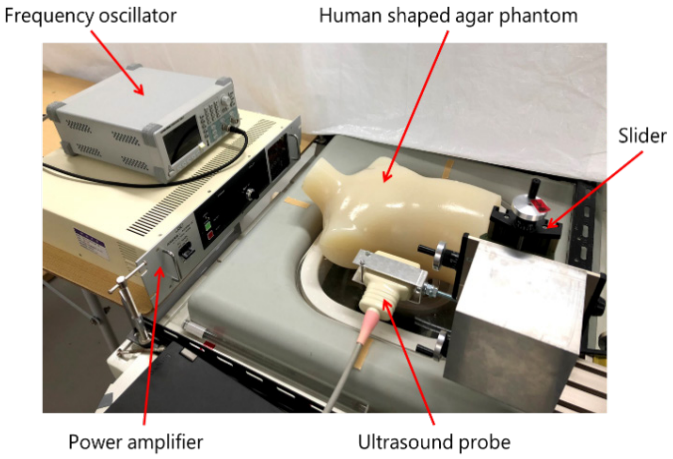


Figure 5: Experimental Setup for Capturing Ultrasound Images.

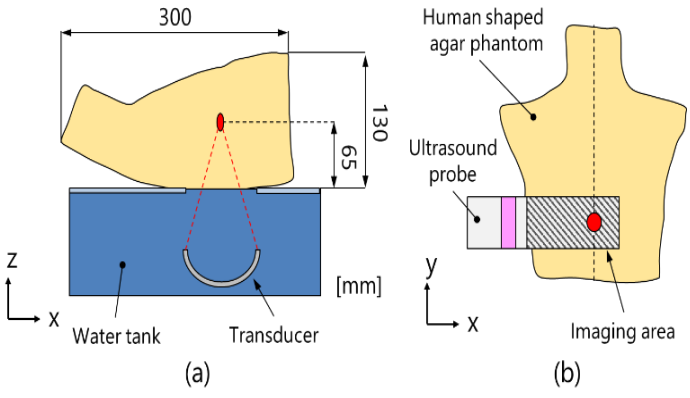


Figure 6: Illustration of Experimental Setup for Scanning Ultrasound Images. (a) Side View, (b) Top View.

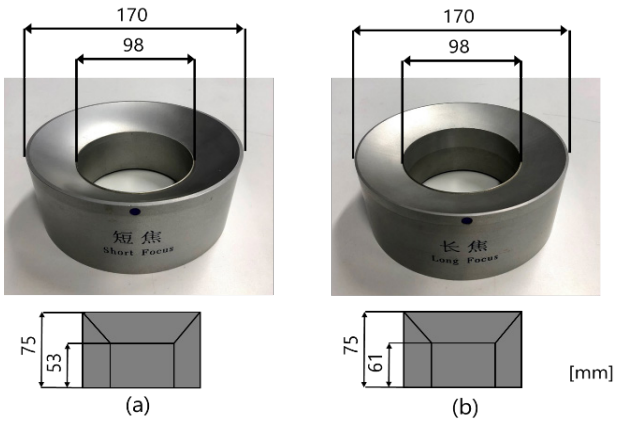


Figure 7: Transducers Used in the Experiments.
(a) Short Focal Length, (b) Long Focal Length.

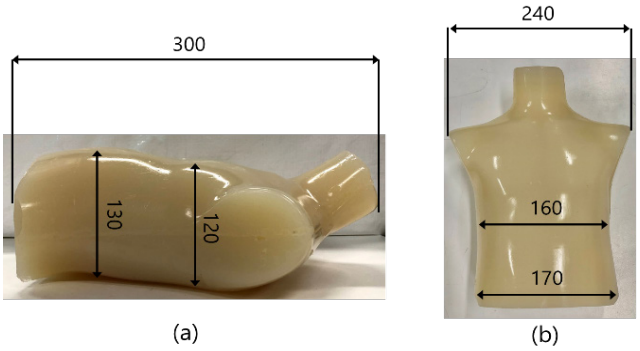


Figure 8: Human Shaped Agar Phantom.
(a) Side View, (b) Top View.

3. Results

3.1. Thermal Constant of Agar Phantom

Figure 9 shows the relationship between the temperature change (ΔT) of the agar phantom and the degree of ($\partial(\Delta d)/\partial x$) of the ultrasound image when heated as shown in Figure 3. The template size used in the image analysis as shown in Figure 4 was (13 pixels × 13 pixels). The total number of plots in Figure 9 is 45. As a result of finding the approximate straight line from these data by the method of least squares, the slope of the graph was 102.92 and the intercept was -0.5262. The thermal constant of the agar phantom used in this heating experiment was set to K_{agar} = 102.92. In addition, the coefficient of determination R^2, which is the degree of fitting of the approximate straight line, is 0.9941, which confirms that there is a high correlation.

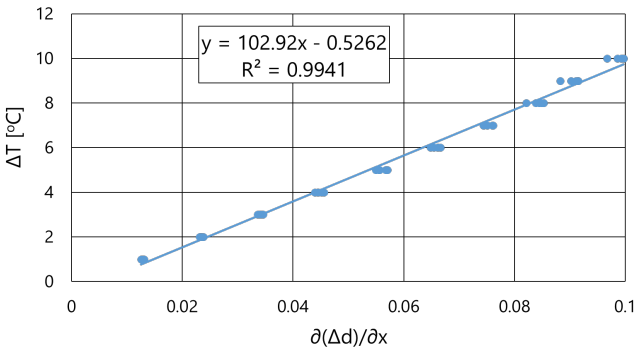


Figure 9: Relationship between ΔT and $\partial(\Delta d)/\partial x$.

3.2. 2-D Temperature Distributions

Figure 10 shows the comparison results of the temperature distribution in the central cross section of the human shaped agar phantom when heated with a short focus transducer. In the temperature distribution measurement result from the ultrasound image as shown in Figure 10 (a), the maximum temperature increase value ΔT_{max} was 10.1 °C. On the other hand, it was 10.8 °C from the infrared thermal camera as shown in Figure 10 (b). The measured temperature error by both methods was 0.7 °C, and the error rate was approximately 7 %. When comparing the absolute temperatures of both during and after heating, an error naturally occurs. However, it is useful to know the temperature distribution patterns.

Figure 11 shows the temperature in the center section of the agar phantom when the same heating was performed using the long focus transducer. In the temperature distribution measurement result from the ultrasound image as shown in Figure 11 (a), the maximum temperature increase ΔT_{max} was 7.2 °C. On the other hand, it was 5.9 °C from the image of the infrared thermal camera as shown in Figure. 11 (b). The measured temperature error was 1.3 °C, and the error rate was approximately 18 %.

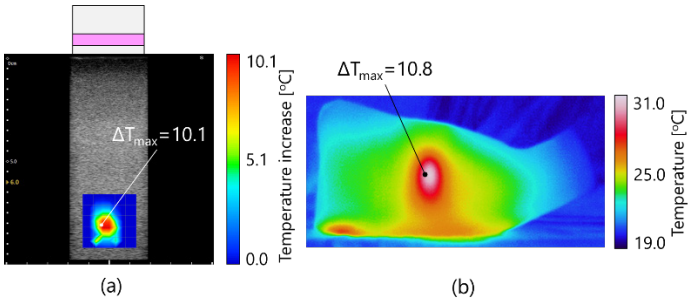


Figure 10: Comparison between Temperature Distributions inside Human Shaped Agar Phantom (Short Focal Length). (a) with Ultrasound Image, (b) with Thermal Camera

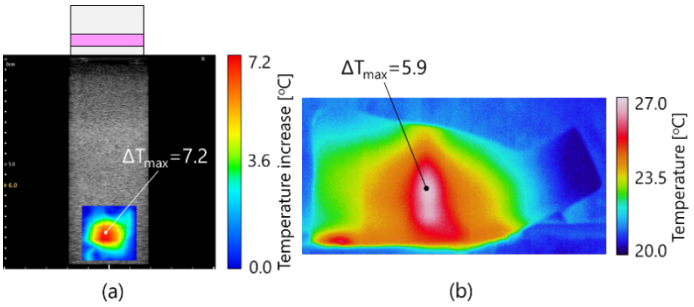


Figure 11: Comparison between Temperature Distributions inside Human Shaped Agar Phantom (Long Focal Length). (a) with Ultrasound Image, (b) with Thermal Camera.

3.3. 3-D Temperature Distributions

Figure 12 shows the two-dimensional temperature distribution results measured from the ultrasound images of the top and bottom (± 15.0 mm) taken at 5.0 mm intervals in the center section of the agar phantom. As shown in Figure 12, the two-dimensional temperature distribution results are three-dimensionally displayed by stacking them. Figure 12 (a) shows the case when using the short focus transducer. It can be confirmed that the temperature increase is the highest value in the central cross section and

decreases with increasing distance from the central cross section, and there is almost no temperature increase above ± 10.0 mm. On the other hand, Figure 12 (b) shows the case when using a long focus transducer. Also in this case, it can be seen that the temperature increase is the highest value in the central cross section. However, in the other image layers, unlike the case when the short focus transducer was used, it can be confirmed that the heating region is wide and the temperature increase value at ± 10.0 mm is similar to the value at the center cross section.

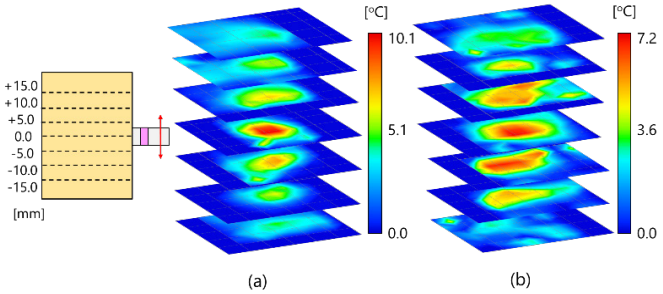


Figure 12: Temperature Distributions from Ultrasound Images. (a) Short Focal Length, (b) Long Focal Length.

4. Discussion

In order to improve the accuracy in the temperature distribution measurement method of applying ultrasound images, we have been working on, (i) taking the images before and after heating at the same position, and (ii) grasping the thermal constant K_{tissue} of the targeted object. Regarding item (i), as shown in Figure 5, the ultrasound probe could be moved three-dimensionally by using two sliders with a movement accuracy of ± 0.05 mm. Regarding item (ii), the value of the agar phantom was actually measured using the self-made device as shown in Figure 3. However, in the results shown in Figure 9, the variation of the measured 45 data was large. It is necessary to control the temperature of the object more accurately by using a higher quality thermostat.

In clinical HIFU treatments, the cauterization temperature of the tumor is mainly 70 °C - 80 °C. The human shaped agar phantom used in the HIFU experiments was prepared with the component ratio recommended by the Japan Hyperthermia Society. When the maximum temperature near the focal point, where ultrasound waves were focused, was 50 °C or higher, the agar phantom melted. Therefore, heating experiments of the agar phantom were conducted at 50 °C or lower. As shown in Figure 12 (a) and (b), heating results using transducers with different focal lengths had different maximum temperatures. Therefore, the normalized temperature (ΔT_N) shown in the following (3) was calculated, and the heating characteristics of both transducers were compared.

$$\Delta T_N = \frac{\Delta T - \Delta T_{min}}{\Delta T_{max} - \Delta T_{min}} \quad (3)$$

Here, ΔT_{max} is the maximum temperature increase value and ΔT_{min} is the minimum temperature increase value in each layer as shown in Figure 12.

Figure 13 shows the calculation results of the normalized temperature (ΔT_N). As shown in Figure 13, when a short focus transducer is used, the temperature increase at a position about ± 10 mm away from the focal point is 60 % or less of that focal point. That is, when the maximum temperature at the focal point is 80 °C, the temperature at the distance (± 10 mm) is approximately 50 °C. However, when a long focus transducer is used, the temperature at the same distance is approximately 60 °C, and the difference in temperature between them is 10 °C. From this result, careful attention is required at the boundary between the healthy tissue and the tumor tissue during HIFU treatments.

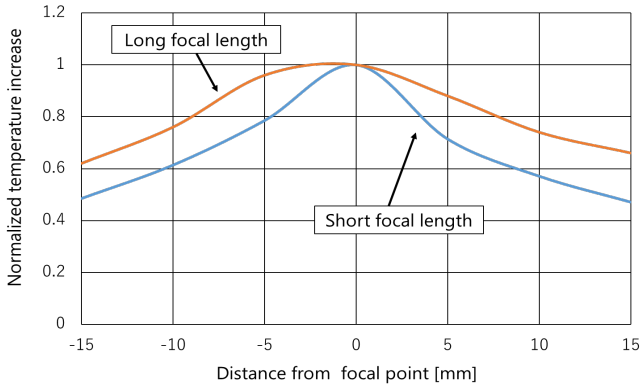


Figure 13: Normalized Temperature Profiles.

5. Conclusion

In this paper, we investigated the possibility of non-invasive measurement of the three-dimensional temperature distribution of the human shaped agar phantom during focused ultrasound heating from ultrasound images before and after heating. First, an algorithm of temperature distribution measurement using ultrasound images was presented. Second, the relationship between the temperature increase inside the heated object and the displacement of the ultrasound images before and after heating was described. Third, we measured the temperature distribution inside the agar phantom during HIFU heating using two types of transducers with different focal lengths. In the temperature measurement results, at the central cross section of the agar phantom, the error between the results of the proposed method using ultrasound images and the measurement results of the infrared thermal camera was approximately 1 °C. In addition, in order to discuss the heating characteristics of two types of transducers with different focal lengths, their three-dimensional temperature distributions were compared. Finally, we highlighted the problems in HIFU treatments using transducers with different focal lengths.

In the future, we have a plan to measure the three-dimensional temperature distribution inside heated animals and to discuss the problems of HIFU treatments.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Part of this research is supported by JSPS KAKENHI Grant Number JP18H03553.

References

[1] R. Sakakibara, K. Kato, P.K. Choi, A. Takeuchi, "Basic study of 3-D Non-invasive measurement of temperature distributions from ultrasound images

during hifu treatments," in Proceedings of the 2019 21st International Conference on Electromagnetics in Advanced Applications, ICEAA 2019, 2019, doi:10.1109/ICEAA.2019.8879194.

[2] S. Vaezy, R.W. Martin, U. Schmiedl, M. Caps, S. Taylor, K. Beach, S. Carter, P. Kaczkowski, G. Keilman, S. Helton, W. Chandler, P. Mourad, M. Rice, R. Roy, L. Crum, "Liver hemostasis using high-intensity focused ultrasound," Ultrasound in Medicine and Biology, 1997, doi:10.1016/S0301-5629(97)00143-9.

[3] G.O.N. Oosterhof, E.B. Cornel, G.A.H.J. Smits, F.M.J. Debruyne, J.A. Schalken, "Influence of high-intensity focused ultrasound on the development of metastases," European Urology, 1997, doi:10.1159/000480887.

[4] N. McDannold, Quantitative MRI-based temperature mapping based on the proton resonant frequency shift: Review of validation studies, International Journal of Hyperthermia, 2005, doi:10.1080/02656730500096073.

[5] K. Hynynen, "Feasibility of using ultrasound phased arrays for MRI monitored noninvasive surgery," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 1996, doi:10.1109/58.542049.

[6] J. Song, A. Pulkkinen, Y. Huang, K. Hynynen, "Investigation of standing-wave formation in a human skull for a clinical prototype of a large-aperture, transcranial mr-guided focused ultrasound (MRgFUS) phased array: An experimental and simulation study," IEEE Transactions on Biomedical Engineering, 2012, doi:10.1109/TBME.2011.2174057.

[7] C. Simon, P. Vanbaren, E.S. Ebbini, "Two-dimensional temperature estimation using diagnostic ultrasound," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 1998, doi:10.1109/58.710592.

[8] Y. ISEKI, K. NAKAMURA, D. ANAN, K. KATO, "Development of Ultrasound-Guided Resonant Cavity Applicator System: - Non-invasive Measurement of Temperature Distributions -," Japanese Journal of Hyperthermic Oncology, 2015, doi:10.3191/thermalmed.31.13.

[9] Y. ISEKI, D. ANAN, T. SAITO, Y. SHINDO, F. IKUTA, K. TAKAHASHI, K. KATO, "Non-Invasive Measurement of Temperature Distributions During Hyperthermia Treatments using Ultrasound B-mode Images," Thermal Medicine, 2016, doi:10.3191/thermalmed.32.17.

[10] E. MIURA, K. KATO, A. TAKEUCHI, "Basic Study of 3-D Non-Invasive Measurement of Temperature Distribution Using Ultrasound Images during FUS Heating (article in Japanese with an abstract in English)," Thermal Medicine, 2018, doi:10.3191/thermalmed.34.1.

[11] A.M. Pouch, T.W. Cary, S.M. Schultz, C.M. Sehgal, "In vivo noninvasive temperature measurement by B-mode ultrasound imaging," Journal of Ultrasound in Medicine, 2010, doi:10.7863/jum.2010.29.11.1595.

[12] M.D. Abolhassani, A. Norouzy, A. Takavar, H. Ghanaati, "Noninvasive temperature estimation using sonographic digital images," Journal of Ultrasound in Medicine, 2007, doi:10.7863/jum.2007.26.2.215.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1312-1321 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Wideband Active Switch for Electronic Warfare System Applications

Mahadev Sarkar*, Gaurav Anand, Sivakumar Ramadoss

Microwave-millimeter wave component Design Group, Product Development and Innovation Centre, Bharat Electronics Ltd. Bangalore,560013, Karnataka, India

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 23 November, 2020
Online: 16 December, 2020

Keywords:
Active SP2T Switch
Wideband SPDT
Broadband discrete switch
High speed SPDT switch
CPWG matching
Broadband equalizer
NAND Gate driver
Power Supply sequencing

ABSTRACT

Discrete component based ultra-wideband SPDT switch modelling, design and implementation challenges are reported in this article. The basic SPDT switch is planned to design using PIN diodes (bare die). Nonlinear parameters extraction steps of PIN diode from datasheets have been proposed in this article instead of using readymade nonlinear PIN diode models. Equalizers are designed as per requirement using discrete components and quarter wave transmission lines for different source and load impedances. Then all the designed components are integrated with a low noise amplifier (LNA). Extraction of PIN diodes' nonlinear parameters are carried out using diode equations along with datasheet parameters. Layout has been implemented using microstrip and CPWG techniques. Quickly responding driver is designed using transistors and logic gates. CPWG based layout achieves better isolation compared to microstrip line-based layout at highest frequencies. Measured and simulated responses are tallied and they are matching together respectively except few exceptions. Measured responses also validates the well agreement between the proposed nonlinear diode modelling and equalizer design with different input and output impedances.

1. Introduction

Modern miniaturized systems require compact, light-weight and low cost components with high yield rate [1]. Excellent ultra wideband electronic warfare (EW) receiver like RWR (Radar Warning Receiver) system requires compact, lightweight and faster SPDT (single pole double through) switch which will be used to switch between BITE and RF ports. This active SPDT switch is basically used to control the RF port isolations while examining the whole Radar in BITE (built in test) mode. As the operating bandwidth of this module is very wide, microwave devices always struggle with uncertain gain gradients and ripples. Thus, excellent class wideband receiver of an RWR system are placed far away from its distributed antenna arrangement systems and connected through a lengthy cable. Thereby, it experiences enormous gain flatness problems particularly at higher frequencies due to excessive transmission line losses. To maintain the best flatness response, it is important to compensate the gain by flattening the gain response over the operating frequency band. It is commonly carried out by using passive equalizers. This article describes various design and development concerns/ problems etc. of wideband discrete components like SPDT switch, passive equalizers and the active SPDT switch. This article also shows a detailed method of nonlinear PIN diode model parameter extraction. Then a suitable PIN diode model is made using those extracted nonlinear parameters and has been used to design the PIN diode based SPDT switch. A suitable wideband LNA (Low Noise Amplifier) has also been introduced at the output path to compensate the loss of the SPDT switch. Prior art survey is carried out over design and development procedures of SPDT switches [2-9]. Allanic et. al. proposed a new passive SPDT design using silicon substrate in [2]. It is band limited to 5GHz with unequal losses in Tx and Rx ports. Additional amplifier is required to make it active SPDT. A high-power passive SPDT design is reported in [3]. Here also an additional LNA is required to make it active switch. Yang et. al. proposed an active SPST switch in [4]. This proposed SPST is band limited to 4GHz using InGaP/GaAs HBT process. This proposed switch consists of seven transistors. This SPST (single pole single through) switch is active in nature but does not cover wide frequency band. A high isolation active absorptive SPST is reported in [5]. It is band limited to 10GHz. Values of various parameters like, minimum gain, gain flatness, isolations and VSWR are -0.4dB, ±1.5dB, 35dB and 3.6:1 max respectively. Single chip SiGe BiCMOS based active SPDT and SPST designs are reported in [6] respectively. This design is also

*Corresponding Author: Mahadev Sarkar, 8553551075, mahadevsarkar@bel.co.in

suffering from P1dB, gain, VSWR etc. Active wideband SPDT switch designs are proposed for V, D and Ku bands in [7–9] respectively. These proposed active SPDT switches are also suffering from gain, gain flatness, input output matching and isolations respectively. These reported designs are also lacking about P1dB and switching speed information except [9] (it provides only P1dB data). Present article uses a discrete LNA and integrates with PIN diode based SPDT switch. This article proposes nonlinear diode parameter extraction steps with an example using limited source of data (datasheet) and diode equations. Readymade nonlinear diode models have not been taken into consideration as vendor often do not extends support. Suitable diode models have been made using extracted nonlinear parameters and tuned up to 18GHz. Then those models have been used to design broadband switch. Equalizers are also designed separately with different input/output impedances to cope up with flatness issues. Only logic gates and discrete active components are used to design the driver where, NAND gate translates input logic to output logic control signals and transistor builds current over it. Active SPDT switch integrates separately designed SPDT, gain equalizers, LNA, attenuation pads, driver etc. Active SPDT switch was designed using microstrip lines initially, but isolation was not achieved 50dB. Then, co-planner waveguide (CPWG) approach is tried with one more extra shunt element for both the paths. This configuration achieves >50dB isolation at 18GHz. Switch isolation of microstrip line and CPWG based method is compared in the article. Switching speed of any switch is directly related to the signal throughput of the driver. Thus, how fast the driver acts switching speed will also be so faster. That is why it will be a combination of driver and switch's performance respectively [10, 11].

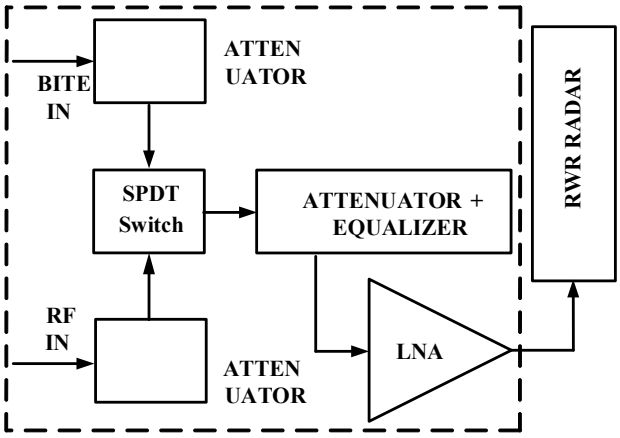


Figure 1: Block diagram of ultra-wideband high speed active SPDT Switch consists of basic SPDT, equalizers, attenuator pads and LNA

This present work advances P1dB performances, power dissipation, better input/output matching and superior packaging with reduced weight respectively from the previous work reported in [1]. Figure 1 shows the block diagram of active SPDT switch where, BITE and RF-in are the two ports of the SPDT switch. Attenuator pad/equalizer are kept at the input of the switch for better matching [1]. It increases noise figure very little as compared to the previous situation where, active switch was absent. An appropriate LNA is integrated just at the output port of the discrete SPDT switch. Equalizers are used before and after LNA to keep flatness within the specifications. Measured and simulated responses are compared in this article and analyzed the probable cause of discrepancy. An appropriate reason is explained for this discrepancy and conclude about limitations of the nonlinear diode modelling. This article has been organized as follows. In Section 2, PIN diode parameters extraction, diode modelling, SPDT switch, equalizer and driver design are discussed. In Section 3, design implementation and experimentation are shown along with simulated results. In Section 4, fabricated switch is shown and the conclusions are given in Section 5.

2. Individual Components Design

This section explains about design steps of three important components design likely, nonlinear parameter extraction of PIN diodes, modelling and comparison with measured response (plotted in the datasheets), equalizer design and driver design respectively. The major work is involved in nonlinear parameter extraction of PIN diodes from datasheet and then prepare a suitable model out of it. Here this extraction steps will be explained in details. PIN diode modelling is carried out both for the diodes used in series and shunt branches of SPDT switch. Discrete component based equalizer design explained in the next and at last NAND gate based driver design also shown in details.

2.1. *Nonlinear parameter extraction and PIN Diode Modelling*

SPDT block of the active switch shown in Figure 1, has been designed using PIN diode bare dies. Simple nonlinear model of PIN diode is costly and most often component vendors' do not support with exact nonlinear model parameter values in the datasheets. Thus it is necessary to make nonlinear model of PIN diode. It is possible to prepare nonlinear model of PIN diodes using data given in the datasheet [12]. Thus it has become designer's capability to build an acceptable model for diodes using existing data in the datasheet. Extraction of all nonlinear model parameters are extremely difficult using limited source. Approximate (mere accurate) model is possible to make using these extracted parameters by fine tuning other model parameters and that can tackle the issue [12].

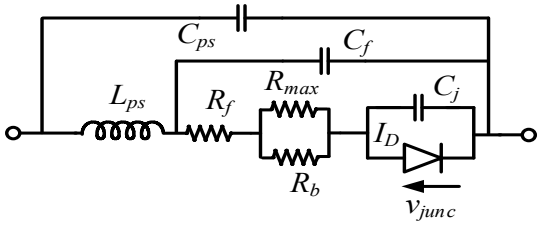


Figure 2: Classic diode model used to design circuit model of PIN diode

Adopting the above technique separate models have been prepared for shunt and series diodes respectively and compared with its measured responses. Figure 2 shows the basic diode model to be used to make equivalent PIN diode circuit model where, C_{ps}, C_f, C_j, L_{ps}, R_f, R_b, R_{max} I_D and V_{junc}, are package parasitic capacitance, beam lead parasitic capacitance, junction capacitance, package parasitic inductance, minimum resistance when diode is in forward bias, variable *i* region resistance on PIN diode, maximum resistance, diode current and voltage across the junction respectively. Equation mentioned below (in 1) describes

the current flowing through the diode where, I_s, q, n, k, T_j and I_{RBD} are, saturation current, electron charge, ideality factor, Boltzmann constant, temperature in kelvin and reverse breakdown current respectively [13].

$$I_D = I_s\left(e^{\frac{qV_{junc}}{nkT_j}} - 1\right) - I_{RBD} \quad (1)$$

and

$$I_{RBD} = I_{RBDS}e^{\frac{-q(V_{junc}+V_{RBD})}{kT_j}} \quad (2)$$

where, V_{RBD} and I_{RBDS} are reverse breakdown voltage and saturation current respectively. Datasheet does not provide and package parasitic like, C_{ps}, C_f and L_{ps} and to describe the bonding pad paracitics. Thus, it is very difficult to extract all the model parameters from manufacturers' data sheets, which only provide limited information. Fortunately, most of the critical parameters like, I_s, R_f, C_j, τ, α_c, δ etc. can be derived from the manufacturers' data sheets where, τ, α and ζ are transition time, arbitrary constant and built in potential respectively [14]. Datasheet provides the I-R and C-V characteristics of diodes. Thus, all the above parameters can be extracted using these characteristic curves.

$$\alpha_c = \frac{qV_{junc}}{nkT_j} \quad (3)$$

$$C_j = \begin{cases} C_{J0}\left(1 - \frac{V_{junc}}{\delta}\right)^{-\gamma(V_{junc})} + C_D & V_{junc} \le \mathrm{F}_c\mathrm{X}\delta \\ C_{J0}\mathrm{Xf}(V_{junc}) + C_D & V_{junc} > \mathrm{F}_c\mathrm{X}\delta \end{cases} \quad (4)$$

where C_D, C_{J0}, F_c and γ are capacitance due to the finite transit time (τ) through the *i*-region, depletion capacitance at $Vjunc = 0$, function of $Vjunc$, coefficient to divide the forward-bias, depletion capacitance into two regions and expressed by a 3rd order polynomial.

$$f(V_{junc}) = (1 - F_c)^{-(1+\gamma(V_{junc}))}\ \mathrm{X}\ \left\{1 - F_c\left(1 + \gamma(V_{junc})\right) + \gamma(V_{junc})\frac{V_{junc}}{\delta}\right\} \quad (5)$$

$$\gamma(V_{junc}) = \gamma_0 + \rho_1 V_{junc} + \rho_2 V_{junc}^2 + \rho_3 V_{junc}^3 \quad (6)$$

$$C_D = \tau\frac{\partial I_D}{\partial V_{junc}} \approx \frac{\tau I_D}{0.0256n} \text{ at } T_j = 300\mathrm{K} \quad (7)$$

The resistance in the i-region is a function of the forward current. At high forward current level, *Rb* becomes close to zero. Thereby the minimum resistance becomes *Rf*. At reverse bias condition, *Rb* is almost infinite, therefore the maximum resistance is *Rf* +*Rmax*. The resistance at low and moderate forward current level is written in (6) [14].

$$R_{Total} = \mathrm{R}_f + \frac{R_b X R_{max}}{R_b + R_{max}} \quad (8)$$

$$R_b = \begin{cases} \frac{\beta_1}{{I_D}^{\beta_2}} & I_D > 0 \\ \infty & I_D \le 0 \end{cases} \quad (9)$$

Variable resistance can be found by putting in (9) in (8) considering β_2=1, where R_m is the arbitrary RF resistance at current level I_m.

$$\beta_1 = I_m(R_m - R_f) \quad (10)$$

Table 1 shows the extracted parameters considering HPND4018 PIN diode datasheet using (1) to (10). Here few parameters are directly taken from datasheet, like forward resistance, breakdown voltage, junction capacitance, transition time etc. Few other parameters have been finalized using curve fitting technique to match the response with datasheet while simulation of model. PIN diode model simulation is performed in

Table 1: The extracted model parameters for HPND4018 Diode

Sl. No.	Parameters	Extracted Values
1	I_s	1.21X10^{-9} A
2	I_{RBDS}	1.0 X10^{-7} A
3	R_f	3.7 Ω @ 20mA
4	R_{max}	780 Ω
5	α_c	22. 464239
6	V_b	60 V
7	n	1.94
8	C_{J0}	0.0481 pF
9	τ	26 X10^{-9} S
10	τ_{rr} *(5V, 10mA)*	3.934 X10^{-9} S
11	δ	0.197 V
12	γ	0.221
13	β_1	1.257X10^{-2}
14	β_2	99.87X10^{-2}
15	ρ_1	1.61X10^{-4}
16	ρ_2	-1.613X10^{-5}
17	ρ_3	6.761X10^{-7}
18	C_p	0.095 pF
19	L_p	1.0 nH

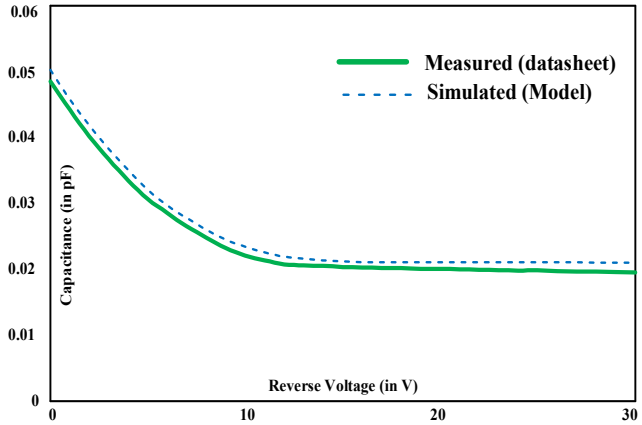


Figure 3: Comparison between modeled and measured DC C-V characteristic of HPND1084 PIN diode at 1MHz

2½D software taking parameters values from Table 1. Influence of bonding wires cannot be ignored while simulation. Thereby model is extracted including them. Tuning of the required model parameter values are carried out to match the model characteristics virtually similar as compared to the datasheet.

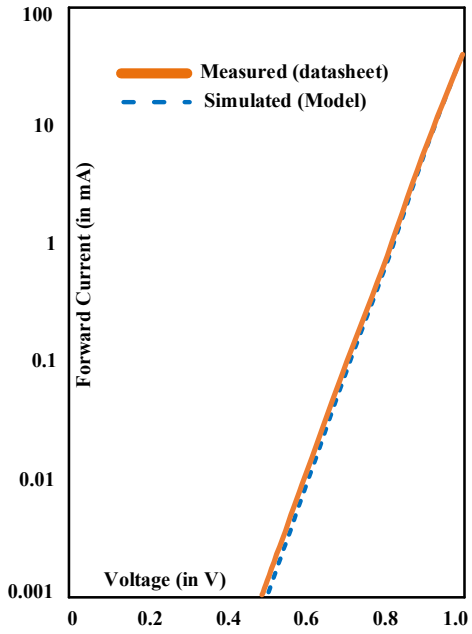


Figure 4: Comparison between modeled and measured DC I-V characteristic of HPND1084 PIN diode

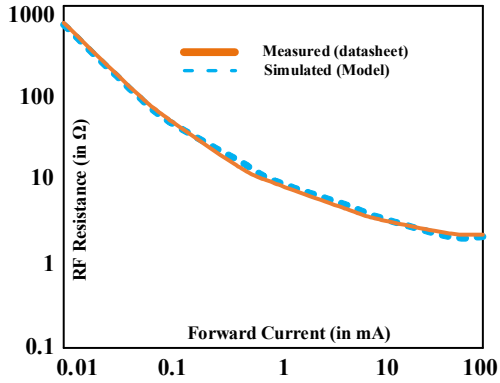


Figure 5: Comparison between modeled and measured RF resistance – Forward current at 100 MHz characteristic of HPND1084 PIN diode at 100 MHz

These models are purely made based on datasheet parameters and used for specific applications. Figure 3 shows the comparison between modeled and measured DC C-V characteristic of a HPND1084 PIN diode. Here modeled capacitance almost optimized as measured capacitance values. Figure 4 shows the comparison between modeled and measured DC I-V characteristic of the same PIN diode. Modeled current is showing a good agreement with measured current. Figure 5 shows the comparison between modeled and measured RF resistance characteristic. Here simulated resistance indicates that model parameters are optimized properly. Nonlinear model of MP6002 has also been carried out by following the above procedure. These two diode models are used for SPDT switch design, where, HPND1084 is used as series diode and MP6002 is used as shunt diode. Figure 6 shows the response of a modeled PIN diode (HPND1084) for 1– 20 GHz bandwidth. Maximum insertion loss is below 1dB at 18 GHz and it is well matched up to 18 GHz. Isolation is around 20dB at 18 GHz. The same loss characteristic is also calculated for shunt diode and kept ready for next level simulation. Here diode models are tuned several times to make it suitable and well matched with datasheet characteristics.

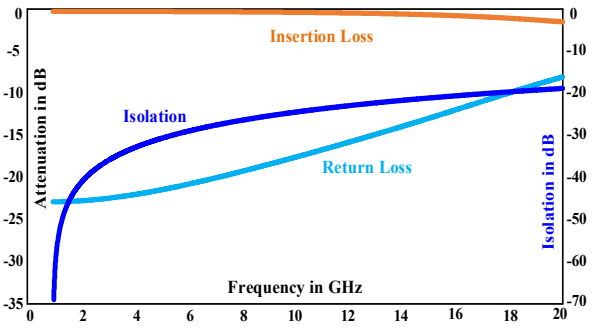


Figure 6: Simulated response of insertion loss, return loss and isolation characteristic of HPND1084 PIN diode

2.2. *SPDT Switch Design using PIN Diodes*

Simulated and verified nonlinear PIN diode models are used to design the SPDT switch where, HPND1084 is used in the series branch and MP6002 is used in the shunt branch. Schematic design

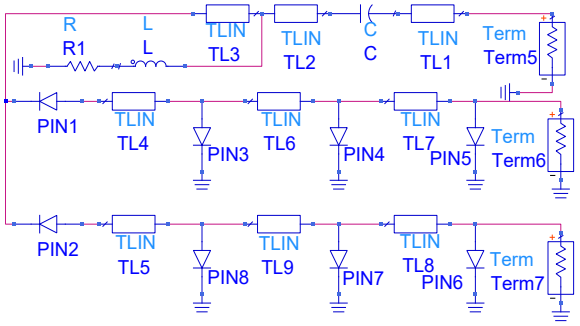


Figure 7: Schematics of basic SPDT switch using PIN diodes

of the switch is performed in 2½D software and shown in the Figure 7. Each branches of the switch having one series and three shunt diodes respectively. Here DC current return path is getting completed through inductor L and resistor R1 in the common path. Resistor R1 is used to limit the maximum current through the series diode. Bonding wires (2mil gold ribbon) have not been added in the schematics as it is already considered as parasitics in the diode modelling. The insertion loss (I_L) and isolation (I_{iso}) parameters of the SPDT switch are estimated using (11) and (12) respectively. Here, R_{fse}, R_{fsh}, Z_0, and X_C are forward resistance of series and shunt diode, characteristic impedance, and reactive impedance respectively. Tendency of the simulated isolation response is fluctuating around middle of the frequency band which will be recompensed in final layout simulation [14].

$$I_L = 10log\left[\left(1+\frac{R_{fse}}{2Z_0}\right)+\sum_{j=1}^{n}\left(\frac{Z_0+R_{fsh_j}}{2X_{C_j}}\right)\right] \quad (11)$$

where, $X_C = 1/2\pi f_L C_D$, $f_L = 1/2\pi\rho\varepsilon$, here ρ and Ɛ are the resistivity of the I region and dielectric constant of the silicon.

$$I_{iso} = 10log\left[\sum_{j=1}^{n}\left(1+\frac{Z_0}{2R_{fsh_j}}\right)^2 X\left(\frac{X_{C_j}}{2R_{fsh_j}}\right)^2 + \left(1+\frac{Z_0}{2R_{fse}}\right)^2\right] \quad (12)$$

Reverse breakdown voltage, power handling, switching speed, capacitance, forward resistance, carrier lifetime, thermal resistance etc. have been kept in mind while selecting diodes.

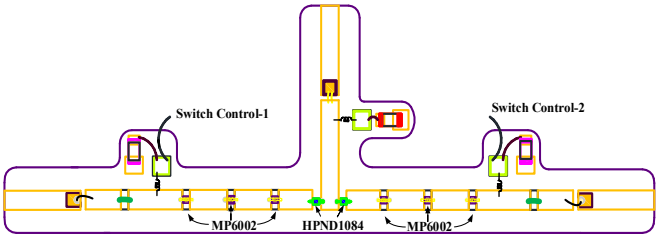


Figure 8: Proposed layout diagram of basic SPDT switch using PIN diodes

Final assembled layout of the proposed basic SPDT switch is shown in Figure 8. Here extra shunt diodes assembly option is kept in the layout so that fabrication iterations become less but, it has a disadvantage of higher insertion losses. This option will not affect at the overall performance as LNA is in the common path to compensate the extra losses. Initially, SPDT layout was carried

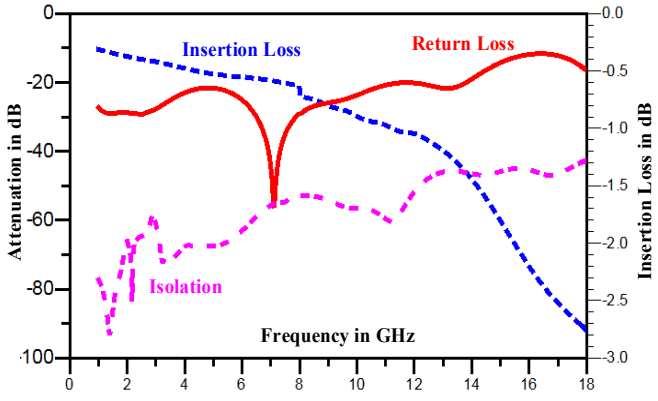


Figure 9: Simulated responses of proposed SPDT switch in 2½D software

out using microstrip lines and fabricated but, it was not meeting isolation specifications at higher frequencies (measured response is shown in the experimentation section). Thereby final layout of SPDT switch is made using CPWG lines in the first iteration. Figure 9 shows the simulated responses of the SPDT switch. Here insertion loss is about 2.8dB around 18GHz. Return loss is more than 20dB up to 14 GHz and it reduces to around 12 dB at 16–18GHz. Isolation is 50dB at 18 GHz. Now it is ready for next level integration with equalizer and LNA.

2.3. *Equalizer Design*

Equalizers are crucial parts for equalizing flatness of the module which, deteriorates with frequencies moving higher and higher. It is evident from Figure 9 that almost 2.5dB slope adjustment is must to flatten the switch response. Now there is an LNA in the common path that has its own gain slope (almost 2dB). Thus requirement specific equalizers are necessary in the various locations (common path as well as in the SPDT branches) to

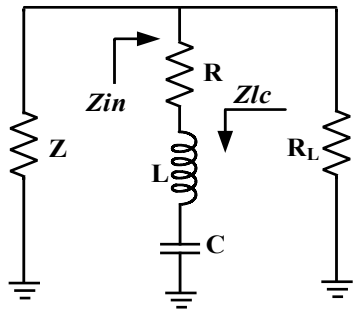


Figure 10: Schematic diagram of single section equalizer

compensate all the loss/gain slopes. Figure 10 shows elementary schematic diagram of a single resonator based equalizer [1]. This equalizer is comprising of discrete thin film resistor and a resonator of $\lambda_g/4$ length at the highest frequency. Slope of the equalizer has been finalized after analyzing the integrated switch response in the system level simulations. Thereby case to case final component twigging will be performed as per need basis. The ultimate discrete component values can be finalized at the final level of simulations to maintain tiniest gain flatness (within ±0.5dB). Driving point impedance of the R-L-C network can be calculated using (13) where, R, L and C are the shunt resistance, inductance and capacitance respectively [15-17].

$$|Z_{in}| = \sqrt{R^2 + \left|\frac{1-\omega^2 LC}{\omega C}\right|^2} \quad (13)$$

Final transfer characteristics of the equalizer containing n sections have been calculated by (14) where, Z_L, ω is the load impedance and angular frequency respectively where, $\omega = 2\pi f$ and f is the operating frequency.

$$H_{total}(\omega) = \sum_{i=1}^{i=n} \frac{2(j\omega C_i R_i - \omega^2 L_i C_i + 1)}{j\omega C_i Z_L + 2(j\omega C_i R_i - \omega^2 L_i C_i + 1)} \quad (14)$$

Although (14) has been calculated for n number of sections but the actual number of equalizer sections may be changed/decided according to flatness characteristics at the final

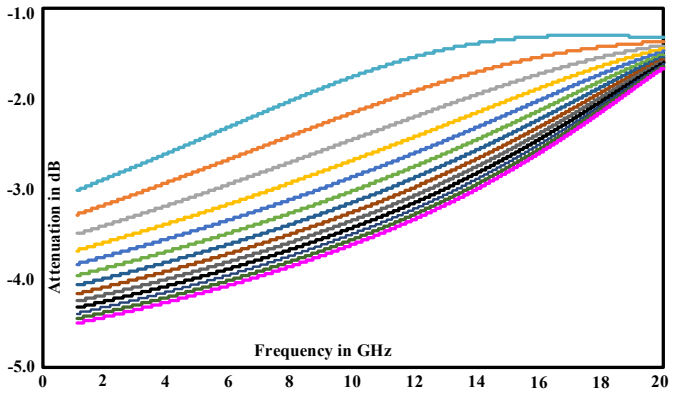


Figure 11: Insertion loss response of equalizer for different resistor values.

level of realization. Figure 11 shows insertion loss characteristics of passive equalizer varying numerous resistor (R) values where, upper and lower response curves are for lower and higher resistor values respectively. Now when it interacts with output impedance of SPDT switch as input impedance of it and input impedance of

LNA as arbitrary load then its input and output port matching also changes respectively. Figure 12 shows the input and output return loss characteristics of the equalizer for numerous R values. Here source impedance is almost 50Ω but load impedance is the input impedance of the next stage thus, input and output return loss response is not symmetric. In the system simulations various resistor values have been tuned to achieve different slopes.

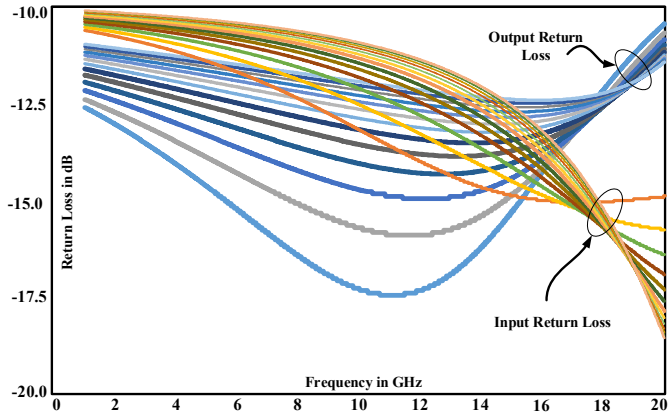


Figure 12: Input and output matching of equalizer for different resistor values

2.4. *Driver Design*

A suitable driver is designed to generate the required control signals to operate the switch and LNA (supply sequencing) from external control signals. As it is discussed that the switching speed is a combination of driver and switch's performance respectively [1]. Thereby design of appropriate driver is very crucial to achieve the best switching speed. Sequencing circuit is also required to feed supplies to the LNA. Here driver and sequencing circuit design is described in details. Logic gates, transistors, diodes etc. are used to design sequencing and driver circuit. External logic is passed through high throughput NAND gate and it passes through combination of transistors, zener diodes and resistor combinations to produce the output. Schematics level simulation of the driver is carried out using spice simulator tool. The design has also been verified by simulating the same circuit in 2½D simulator. The outcome of both the simulator tools are compared and they are very much satisfying**.** Figure 13 shows schematic diagram of a driver circuit comprises of logic gate (74LS00; dual-NAND gate) and zener diode (BGV90-C4V3) followed by transistors (current booster) and resistors in various combinations [1]. Simulated driver response is illustrated in the same Figure 13. In the Figure 13 the red color bit streams indicate the input control bits. The blue and violet color bit streams represent the output control–1 and control–2 bit streams respectively. These two output control bit streams are being used to alter the active switch RF and BITE paths respectively. It is also evident from the figure that, both the controls (control–1 and –2) are out of phase.

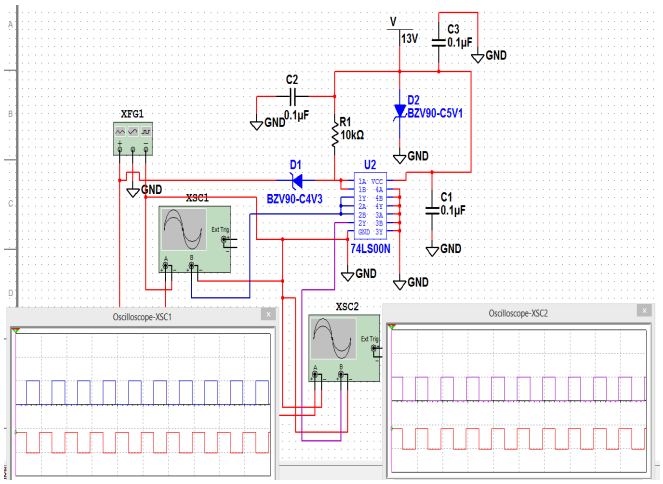


Figure 13: Schematic diagram and responses of switch driver circuit

3. Active SPDT Switch Design

Active SPDT switch design has finally been performed by employing scattering parameter (.sNp) files of all the individually designed modules. RF system analysis depends upon the precise design of all the discrete modules. Hence it will meet all the necessary specifications across all the environmental conditions over an extensive frequency spectrum. Layout is made using CPWG lines as basic SPDT switch. This technique has been adopted to rectify isolation and insertion loss characteristics of the channel respectively. The final schematic level simulation is performed using 2½ D simulator tool. Schematic diagram of the active SPDT switch for forward and reverse (isolation) channel respectively are shown in Figure 14 [1]. Tapers used to interconnect two different impedance lines are not CPWG lines also shown in this Figure 14. Figure 15 shows the simulated blue

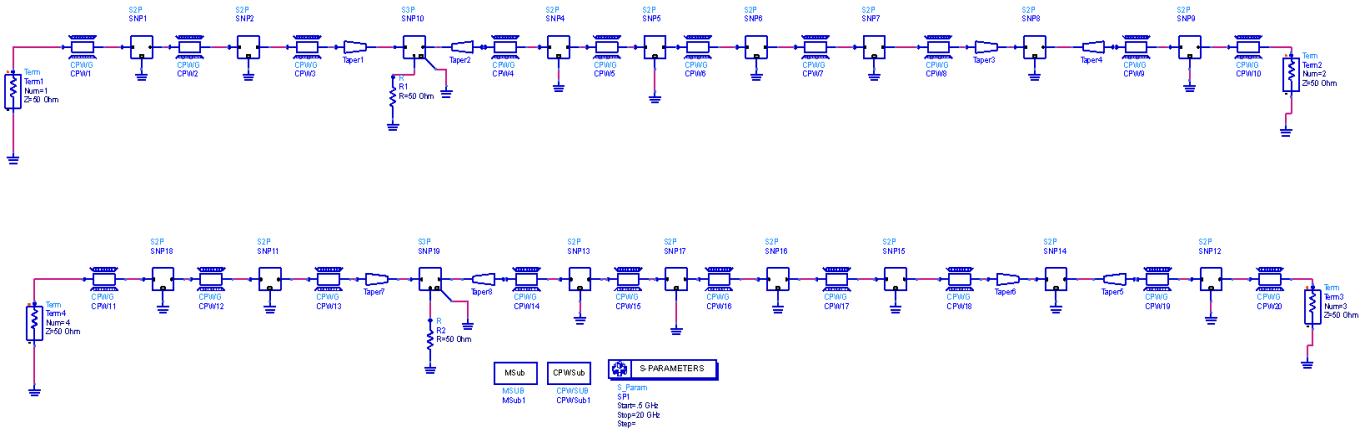


Figure 14: CPWG based schematic circuits staring from input port to output port of proposed SPDT switch for forward signal and Isolation path respectively

color insertion gain curve and saffron color isolation curves of the proposed active SPDT switch. Here the overall gain about 2 dB is achieved with flatness within ±0.5dB. More than 50 dB isolation has successfully been met at 18 GHz which is very crucial parameter. Almost 95dB isolations are also met at the lower frequencies. Appropriate LNA (die), discrete component attenuators and readily available (designed) equalizers are used in the output track of the active SPDT to match the gain flatness. Simulated VSWR response is shown in Figure 16. Budget analysis has also been performed at -7 dBm input power level and at mid-band frequency (10 GHz) using 2½D software. Parameters like isolation, dynamic range, SFDR (spurious free dynamic range) and cascaded P1 dB had been the prime concern and they have attained 54.4 dB, 64 dB, 53.24 dB and 20 .7 dBm respectively. 3D EM simulation is carried out for the layout and mechanical channels respectively. It ensures any cavity resonance is present up to 18GHz or not. Ultrathin soft substrates have been used for fabrications. Driver and control circuit is made over FR4 material. The SP2T switch and driver/control circuits are mounted on opposite sides of the same housing respectively. Switch control, sequencing and other supply connections are taken from driver circuit board to switch side through small holes, keeping in mind the EMI/EMC effects. Appropriate mechanical channeling for RF circuits are important for achieving expected responses. Microwave channel dimensions are kept narrow enough to avoid higher mode resonance (excitations), as it disturbs characteristics of the module at higher frequencies. Carrier plate are made to ensure robust ground effects at higher frequencies and very good mechanical support to the PCBs. 0.4 mm kover and copper-molly sheets having nicely nickel plated with minimum impurities have been used to make carrier plates [19].

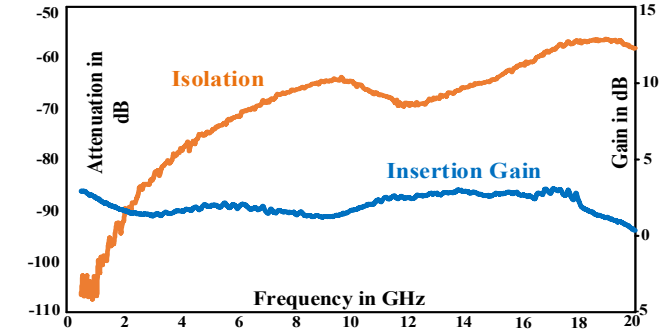


Figure 15: Gain and isolation simulation response of proposed SPDT switch in the transmission and isolation paths respectively

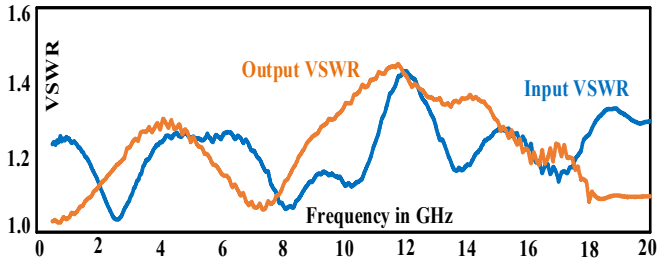


Figure 16: Input and output VSWR simulation response at input and one of the output ports of proposed SPDT switch

4. Experimentation

Active broadband SPDT switches have been developed using microstrip as well as CPWG lines respectively. Initially microstrip line based SP2T is developed and then CPWG line based switch due to noncompliance of isolation specification. The realized active SPDT switches are evaluated separately. Figure 17 shows the isolation responses of SPDT switch of microstrip line based approach. In this approach microwave channeling and air cavity sizes were not taken care precisely thereby output responses like isolation characteristics affected drastically. It is evident from the figure that isolation at 18 GHz is not meeting the requirement of 50dB. It is actually around 42dBc compared to gain plot. Gain plot of the microstrip line based SPDT is not shown as gain plot is identical to CPWG approach based SPDT. Comparison between measured and simulated responses of gain and isolation of CPWG approached based SPDT is shown in Figure 18 respectively. Here all measured and simulated responses are shown in firm and dotted lines respectively. Overall insertion gain met in this approached is 2.5dB except few frequencies. It is in good agreement with simulation responses. Each arm of this switch contains of sufficient gain equalizers to fit the gain flatness within ±0.4dB (maximum) over the entire spectrum at 27°C (Room Temperature; RT) as well as ±0.6 dB (maximum) across the wide temperature range -40°C to 71°C. Figure 18 also depicts the isolation response at RT which reached minimum 53dB (55dBc compared to gain response) in CPWG based approach.

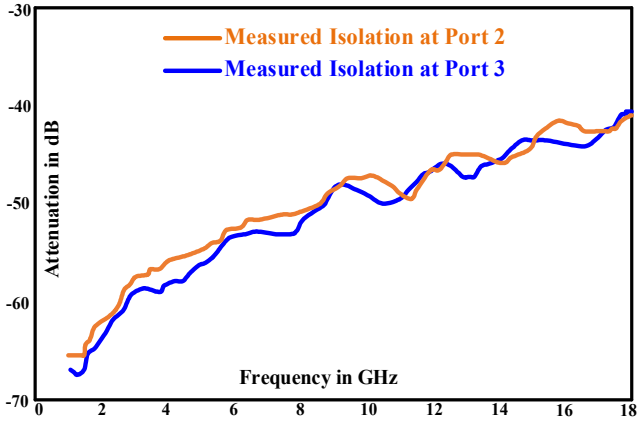


Figure 17: Measured isolation responses between common port and port-1 and port-2 of the developed SPDT switch following microstrip line approach

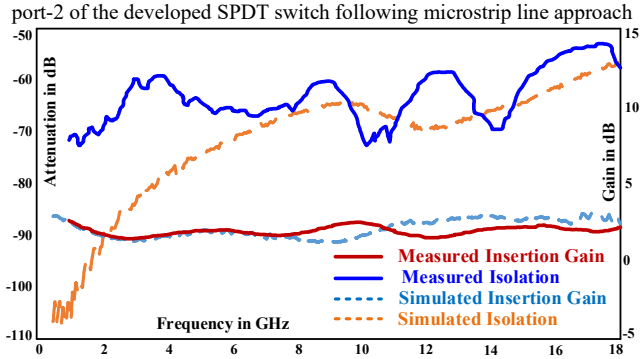


Figure 18: Comparison between measured and simulated responses of gain and isolation of common port – port-1 and common port – port-2 of the developed SPDT switch following CPWG line approach

Any input port out of two input ports of the active SPDT switch can be accessed at any instant of time according to the external logic signal. This process achieves ON/OFF input output port isolation of 55dBc at the highest frequency and more than 60dBc across 1–15.5 GHz frequency. Tested isolation response is

comparable with simulated response beyond 6GHz (6–18GHz). From 1–6GHz measured isolation is not matching to the simulation response due to more idealistic diode characteristics of the PIN diodes. Comparison between measured and simulated responses of VSWR of CPWG approached based SPDT is shown in Figure 19 respectively. The measured VSWR is 1.45:1 and 1.35:1 at the input and output ports respectively. Measured VSWR responses are in very good agreement of simulation responses. Better than 1.7:1 VSWR is measured across frequency band under ESS. About 50nS switching speed is measured with a very good rise and fall time characteristics.

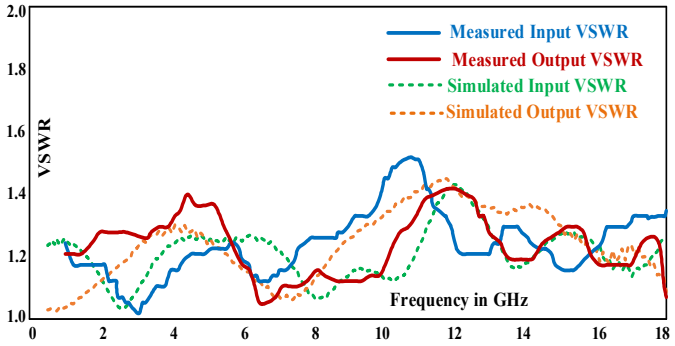


Figure 19: Comparison between measured and simulated responses of gain and isolation of common port – port-1 and common port – port-2 of the developed SPDT switch following CPWG line approach

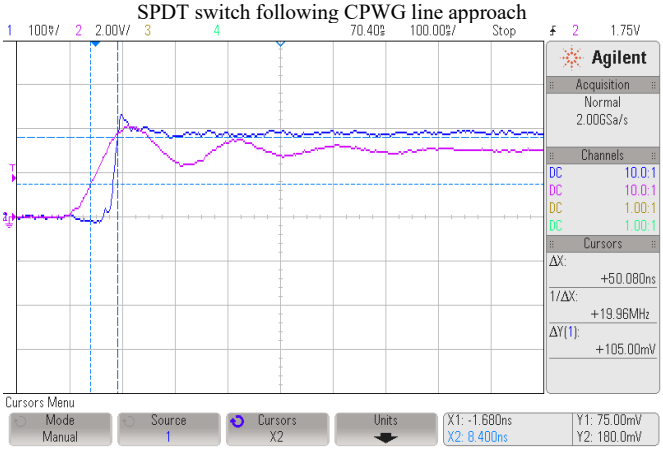


Figure 20: Response of measured switching speed at the rising edge of the developed CPWG approach based active SPDT switch

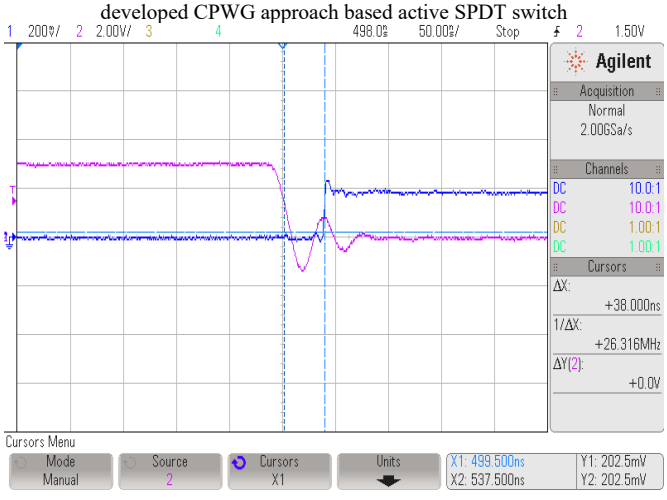


Figure 21: Response of measured switching speed at the trailing edge of the developed CPWG approach based active SPDT switch

Figure 20 and 21 describes the tested switching speed responses at rising and trailing edges against the input logic signal respectively. Here input switch control pulse is in 'volt' scale and detected output pulse in 'millivolt' scale although voltage scaling factor of input switch control and detected output pulses respectively is same (10:1). The detected output pulse will always be less than the input pulses as the input power level is fixed at -3dBm for switching speed measurement. This happens due to the absolute power level of the output pulses which is almost -1.5 – -1.0dBm (188–200mV) at 10GHz frequency and measured against a high value resistor (120Ω) is also shown in Figure 22 [18]. Thus final detected voltage after detector is very less compared to the external TTL signal. Achieved switch speed is 50nS including all RF and digital circuits. This module exhibits input P1 dB is >20 dB compared to prior work reported in [1]. This is achieved by replacing LNA with higher P1dB LNA, rectifying input/output port matching and twigging equalizer losses respectively. Developed active SPDT module also displays tremendous harmonics characteristics. Measured harmonics levels for 0dBm and 8dBm input signal for the active SPDT are 34dBc and 27dBc respectively. temperature cycling (TC) has been performed over -40° to +71° C and 3-axes random pre and post TC vibrations respectively.

Sl. no	Parameters	Specifications	Min	Max
1	Gain	2 ±2 dB	1.8 dB	2.3 dB
2	Isolation	50 dB min	54 dB	73.1 dB
3	I/P VSWR	2:1 (Max)	1.2	1.3
4	O/P VSWR	2:1 (Max)	1.2	1.5
5	P1 dB (I/P)	>10 dBm	>20 dBm	
6	2nd Harmonic	-25 dBc @ 0 dBm input -15 dBc @ 8 dBm input	-34 dBc -27 dBc	
7	Power handling	>1 watt	>1 W	
8	DC Power supply	-13V (±3%, -1%)/ 10 mA +13 (±1%, -3%)/ 50 mA	1 mA 21 mA	
9	Switching Speed	50 nS (Max)	50.08 nS	38 nS

The outcomes of environmental stress screening (ESS) of the active switch are steady. The developed module is light weight and measured weight is less than 25 grams. Figure 22 shows the

Table 3: Comparison between various parameter of recent works against present work

Ref.	Technology	f_0	BW	Gain	Isolation	Flatness	Speed	VSWR	P1dB
[2]	Sc. DDAs	5GHz	1.4 GHz	-2.38dB	43/49dB	±1.5dB	-	1.5:1	-
[3]	GaAs	30GHz	15GHz	-1.5dB	43dB	±0.35dB	4nS	1.8:1	35dBm
[4]	InGaP/GaAs HBT	22.8GHz	1.2GHz	-4.4dB	52.2dB	±3.2dB	100pS	2:1/ 3:1	-
[5]	SiGe HBT	20GHz	10GHz	3dB	<35dB	>±1.5dB	60pS	3.6:1	-
[6]	SiGe BiCMOS	22.5GHz	2.0GHz	-0.3dB	<45dB	>±1.5dB	100pS	2.3:1	0.5dBm
[7]	100nm GaAs HEMT	75.0GHz	30GHz	10dB	<15dB	>±1.0dB	-	2:1	-
[8]	100nm GaAs HEMT	145GHz	38.7GHz	3.3dB	<22dB	>±1.5dB	-	2.3:1	-
[9]	0.18µm CMOS	5.8GHz	5GHz	-1.1dB	27/40dB	>±0.3dB	-	2:1	20dBm
This work	PIN diode die	9.5GHz	17GHz	3dB	54/73dB	±0.25dB	50nS*	1.3:1 /1.5:1	>20 dBm

*switching speed has been measured including driver and others are without driver.

developed SPDT module and a classic arrangement for switching speed assessment [1]. It is clear form Figure 20 and 21, that, the switching speed is being measured between 50% of the external input control pulse and 90% or 10% of the final output pulse of the detector for rising and trailing edge respectively. Final achieved switching speed is nothing but the combination of active switch as well as doctor diode delay respectively. Thus the switching speed mentioned in the table 2 is more than the actual switching speed of the active SPDT switch. Signal delay of the detector diode has been measured and it showed almost 5nS at highest frequency. Table 3 shows a comparative study of various parameters of recent works and the present work. Switching speed shown in the above table is different from other work as additional driver delay is added to the active SPDT switch of the present work. VSWR, gain flatness, isolation etc. are excellent compared to other work listed in the table. Another significant change has happened in the DC supply circuits. Earline simple line regulators were used to produce required supply from ±13V thus power consumptions were very high (almost 1.2W) in [1]. In the present work DC-DC converters are being used to generate required supplies for active components thereby power consumptions reduce drastically (almost 270mW). Thus, it can be established that the design goal has efficiently been reached as projected in the initial calculations.

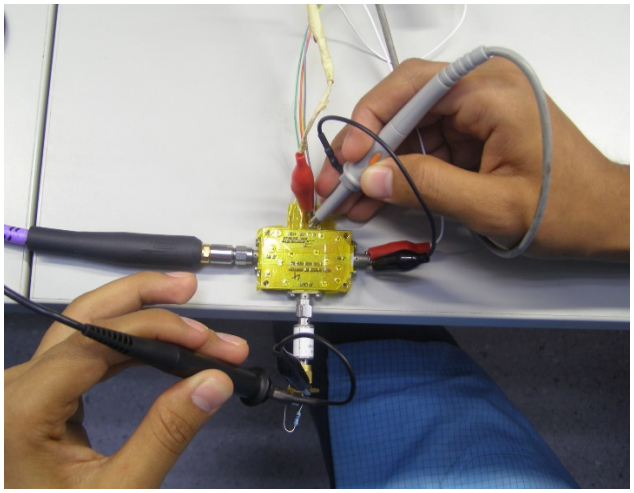


Figure 22: Measurement of switching speed by accessing input (TTL) signal to active SPDT switch and output detected pulse of the developed active SPDT switch module.

5. Conclusion

Ultra wideband active SPDT switch is designed using discrete modules like basic SPDT switch, equalizers, LNA, driver etc. Nonlinear PIN diode parameters extraction steps are shown in details using datasheet and diode equations. Nonlinear modelling is PIN diode is carried out and compared with measured responses (datasheet) for capacitance vs reverse voltage, forward current vs voltage and RF resistance vs forward current respectively. Extracted model parameters showed very good agreement with measured diode parameters. Separate modules are investigated, designed, assembled and evaluated separately. Passive equalizer with various component values are designed and finally most proper one has been selected for this active switch design. Final active switch parameters are plotted and tabulated against specification. Measured responses have been compared with simulated responses. Isolation characteristics of the developed module are not matching between simulation and measured responses from 1–5.7GHz. It happens due to limitations of PIN diode modelling. From the comparison of data, it is evident that the measured responses are in well in agreement with the simulated responses except few discrepancies. There are lots of future scopes regarding this work. LNA used in the present work has high gain which is not required as per present requirement. Thereby customized LNA can be designed to reduce power consumption further without affecting P1dB. This active SPDT switch can be compact by combining equalizers, PIN diodes, attenuators, driver and sequencing circuits as a multilayer (RF, DC and control circuit) single board solution as future work and further LNA can also be included within it using MMIC design solution as super compact light weight tiny chip.

Acknowledgment

Authors gratefully thank Mr. Manoj Jain, Head, Product Development and Innovation Centre and Mr. M.V. Rajasekhar, Director Research and Development, Bharat Electronics for their continuous inspiration to publish this work. Authors are also grateful to Microwave Super Component group, BEL for facilitating switch assembly, die bonding related processes and product design group for giving their support. They also thank all others who were involved directly or indirectly in this work.

References

[1] M. Sarkar, G. Anand, R. Sivakumar, "Light weight Active 1–18 GHz Ultra High Speed SPDT Switch for Radar Warning Receiver Systems," IEEE /MTT-S International Microwave and RF Conference (IMaRC), 1–6, 2020, doi:10.1109/IMaRC2019.9116744

[2] R. D. Allanic, L. Berre, C. Quendo, D. Chouteau, V. Grimal, D. Valente, J. Billoue, "A novel 5-GHz SPDT Switch Using Semiconductor Distributed Doped Areas," IEEE Microwave and Wireless Components Letters, **30**(4), 421-424, 2020, doi:10.1109/LMWC.2020.2978644.

[3] "24–35 GHz GaAs MMIC High Isolation SPDT Reflective PIN Switch," datasheet AP640R7–00, Alpha Industry Catalog.

[4] J. R. Yang, D. W. Kim, S. Hong "A 24 GHz Active SPST MMIC Switch with INGAP/GAAS HBTS," Microwave and Optical Technology Letters, **50**(8), 2155-2158, 2008, doi:10.1002/mop.23606

[5] I. Gresham, A. Jenkins, "A fast switching, high isolation absorptive SPST SiGe switch for 24 automotive applications", 33rd European Microwave Conference, **3**, 903–906, 2003, doi:10.1109/EUMA.2003.341108.

[6] T. M. Hancock, I. Gresham, G.M. Rebeiz, "A differential subnanosecond high-isolation absorptive active SiGe 24 GHz switch for UWB applications" Proc. IEEE Radio Frequency Integrated Circuits Symposium, 497–500, 2004, doi:10.1109/RFIC.2004.132065.

[7] A. Dyskin, S. Wagner, D. Ritter, I. Kallfass, "An Active 60–90 GHz Single Pole Double Throw Switch MMIC," Journal of Millimeter Terahz Waves, **35**, 412-417, 2014, doi:10.1007/s10762-014-0059-y.

[8] D. Muller, A. Tessmann, A. Leuther, T. Zwick, I. Kallfaaa., "Active Single Pole Double Throw Switches for D-Band Applications, IEEE/MTT-S International Microwave Symposium, 1–4, 2016, doi:10.1109/MWSYM.2016.7539969.

[9] M. Y. Chao, Z. T. Min, R. L. Chieh, K. Lin, Y. Chang, H. Wang, "Design and Analysis for a Miniature CMOS SPDT Switch Using Body-Floating Technique to Improve Power Performance," IEEE Transactions on Microwave Theory and Techniques, **54**(1), 31-39, 2006, doi:10.1109/TMTT.2005.860894.

[10] M. Sarkar, S. Jhariya, R. Sivakumar, "Correction of Nonunifirm Attenuation of Voltage Controlled Attenuator under Harsh Environment by Process Improvement" IEEE Transactions on Components, Packaging and Manufacturing Technology, **9**(8), 1632-1642, 2019, doi:10.1109/TCPMT.2019.2929093.

[11] M. Sarkar, K. L. Santha, " Designing of switch-limiter for Xband application in discrete components' topology" 9th International Radar Symposium (IRSI-2013), 1–6, 2013, https\\:radarindia.com/ProceedingdArchive/IRSI-13/13-FP-139.pdf

[12] G. Antognetti, Massobrio, Semiconductor device modelling with SPICE, McGraw-Hill, 1993.

[13] S. M. Sze, Physics of Semiconductor Devices, John Wiley & Sons, 1981.

[14] Microsemi Corporation, The PIN Diode Circuit Designers' Handbook, 1998.

[15] H. Wang, B. Yan, Z. Wang, R. Xu, "A Broadband Microwave Gain equalizer" PIER letters, **33**, 63-72, 2012, doi:10.2528/PIERL12052309.

[16] M. Sarkar, V. Kumar, R. Sivakumar, R. Manjunath, N. S. Charulu, "Performance stabilization of steep gain gradient class ultra-wideband direct conversion IFM receiver of homodyning based systems" Int. Journal of Appl. Sci. **3**(4), 27–31, 2017, issn: 2208-2182.

[17] T. F. Zhou, Y. Zhang, R. M. Xu, "Research on the millimeter wave gain equalizer," IEEE International Conference on Microwave echnology & Computational Electromagnetics (ICMTCE), 180-182, 2011, Doi:10.1109/ICMTCE.2011.5915195.

[18] M. Sarkar, V. Kumar, R. Sivakumar, S. Jhariya, B. R. Raghu, K. S. Sudheesh, "Some investigations on Some Investigations on Direct Substrate-Attachment Process for Wideband Instantaneous Frequency Measurement Receiver Development" IEEE Transactions on Components, Packaging and Manufacturing Technology, **10**(4), 704-716, 2020 doi:10.1109/TCPMT.2020.2966709.

[19] M. Sarkar, B. R. Raghu, K. S. Sudheesh, "Evaluation of Doubled-sided Substrate Attachment Technique for Round Shaped Large Metallic Housing under Airborne System Development," IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 191-194, 2020, doi:10.1109/ICCCA49541.2020.9250792.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1322-1327 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Evaluation of Simple Space Interpolation Methods for the Depth of Precipitation: Application for Boyacá, Colombia

Pedro Mauricio Acosta Castellanos[*,1], Alejandra Castro Ortegon[2], Hugo Fernando Guerrero Sierra[3]

[1]*Faculty of Civil Engineering, Santo Tomas University, Tunja, 150003, Colombia*

[2]*Faculty of Environmental Engineering, Santo Tomas University, Tunja, 150003, Colombia*

[3]*Faculty of Economics and Social Sciences, La Salle University, Bogotá, 110231, Colombia*

ARTICLE INFO

Article history:
Received: 04 September, 2020
Accepted: 01 December, 2020
Online: 16 December, 2020

Keywords:
Regionalization
Interpolation
Precipitation
Hydrology

ABSTRACT

Interpolation tools are used daily in hydrology and climatology. With the purpose to regionalize spot´s registration parameters, such as depth of precipitation, temperature, humidity, among others. The accuracy of these methods is not fully validated. This research presents a comparative study between the most used interpolation methods for the regionalization of hydrological and climatological parameters. Comparative analysis of spatial interpolation; it was carried out using the IDW, Kriging and Spline methods, for this the ARGIS software was used, because its widespread and widespread use. The depth of precipitation was considered as a parameter for the comparison. Accuracy was determined by cross validation. Also, for the coefficient of determination and the comparison for visual errors such as "bull's eyes". The research was spatially restricted to a politically delimited region; Boyacá, Colombia, South America. The best method for spatial interpolation in this case was shown to be Spline.

1. Introduction

Rainfall records are of great importance, because from this information can be obtained among other analyses of flows, design storms and changes in rainfall behavior. The records and their analysis usually come from a network of meteorological stations, each corresponding to a specific geographical point [1]. With this information it is possible to estimate and design structures for water control, economic evaluation of flood protection projects, planning and management of land use, water quality control, among others [2].

Because rainfall information represents the record of a single geographical point, it is common for information to be researched and developed for interpolation in the form of regionalization maps [3]. This information is more common than rainfall, due to its simplicity in the form of record, in both cases can be obtained from calculation treatments and analysis rainfall runoff models and obtain important information such as the curves intensity, duration and frequency [4]. In this sense, regionalization can be defined as the transfer of data or spatial information from one station to another [5]. The transfer can range from data characteristics to hydrological models, [6]. A factor which determines a better condition in regionalisation is homogeneity which essentially refers to the geographical proximity or contiguity and hydrological similarity of meteorological stations [7].

Geographical dispersion in measurement networks means that there are gaps in the spatiality of information, specifically in areas without station coverage [8]. It is evident that the same applies to rainy seasons, for this reason, it is usual that different investigations are presented that seek to find the best method of regionalization and interpolation of hydrological data in order to present a viable alternative for obtaining data in areas where there is no availability. This may influence better and more accurate hydrological models, especially in small and medium catchment areas [9].

This paper presents the evaluation of three interpolation methods; Interpolation with Weighted Inverse Distance (IDW), Interpolation by Gaussian Regression Processes (Kriging) and Segmental Interpolation (Spline), applied to the regionalization of rainfall data, looking for the best method for current and future research, determining the degree of reliability of each method. The interpolation and evaluation were carried out for the rainfall data of monthly and annual records, the application of the study was

[*]Corresponding Author: P. M. Acosta Castellanos, pedro.acosta@usantoto.edu.co

limited to the geographical area comprised by the entire department of Boyacá, Colombia. These three methods or interpolation techniques were evaluated because they are the most commonly used by consultancies and research [10].

2. Methodology

2.1. Interpolation

This research seeks to validate or give support to what is used daily, for this reason the ordinary kriging and the circular model were used because within the wide spectrum of the kriging method, these are the two most used, it is important to note that they can make and create the empirical models for semivariograms but the vast majority of references and even companies like ARGIS indicate that the ordinary kriging and the circular model are the most used [11–19].

Due to the fact that rainfall information is of a point type due to the need for a spatial range and not a point range, interpolation tools are often used, which can define regions with similar behavior [20]. The commonly used methods are Inverse Distance Weighting or IDW, Kriging and Spline, in this sense there are few comparative studies that seek to define the most suitable spatial interpolation method for climatic or environmental variables such as precipitation [21]. Spatial interpolation methods can be defined in four categories: regression method or model and trend surface, local methods, geostatistical methods and mixed methods [9], Only local methods (IDW, Spline) and a known geostadistic method Kriging were evaluated for this research. Likewise, only monthly and annual rainfall depth record values were interpolated for the region of Boyacá, Colombia, because it has a spatially acceptable coverage of rainfall stations, which has a direct bearing on the use of interpolation methods since it is necessary to have a regional coverage to carry out interpolation, since the techniques preside over the variable of interest in a specific place taking values from the surrounding region, resulting in homogeneity [22]. The project will be carried out in the department of Boyacá, Colombia, which has 123 municipalities and 116 rainfall stations of which 60 were useful, since they fulfilled a range of 20 years of historical records counting from 1998 to 2018.

2.2. IDW method

The IDW interpolation method is widely used because of its simplicity [23]. Which is based on the fact that the unknown or sought value of a point has the greatest influence of the nearby control points and the degree of influence or weight of them [24], this implies that it is directly proportional to the inverse of the distance between the points posed [25]. When defining a high power, greater emphasis will be on the nearest points and the result of the surface will be more detailed, but less smoothed. By specifying a lower power, greater importance will be given to the more distant points and therefore the resulting surface will be smoother. equation (1) represents the method

$$Z = \frac{\sum_{i=1}^{N} W_i Z_i}{\sum_{i=1}^{N} W_i} = \frac{\sum_{i=1}^{N}\left[\frac{Z_i}{D_i^P}\right]}{\sum_{i=1}^{N}\left[\frac{1}{D_i^P}\right]} \quad (1)$$

where Z is the interpolated value for a point with the unknown observed value, Wi is the weighting function that determines the importance and relationship of each control point Zi, while Zi is the observed value of the control point i (i=1,...,n), which is the closest zone to the interpolated point and n is the total number of points used in interpolation [26]. The surfaces obtained by this method tend to be symmetrical over the points and reflect the mean value of the dispersion points in the regions between the points. In this method it is possible to control the importance of known points over interpolated values based on distance.

2.3. Kriging method

The method of interpolation Kriging, has a large number of variants that have been developed in recent times, the common is the Kriging Ordinary (KO). We have in this method that the observed or not observed search points are calculated by the linear weighted average of the observed or existing points [23], as shown in equation (2).

$$\widehat{Z_0} = \sum_{i=1}^{n} \lambda_i Z_i \quad (2)$$

where $\widehat{Z_0}$ represents the point searched or not observed, Zi means the attribute of the value of an observed point i, λi is the weight or significance of Zi.

The KO method presents the best linear prediction for points not observed, this means that in theory the value sought or point not observed is equal or very similar to the true value and the variance in the error is minimal [27].

2.4. Spline method

This interpolation method determines the unobserved points using a mathematical function that minimizes the general curvature of the surface, which has the particularity of being a smoothed surface. Many tests performed with Spline show their accuracy which means that the interpolated regions pass through the observed points [28]. There are two types of Splines mostly diffused, the Regularized and Stress, the first refers to softened regions or surfaces and the second creates less smoothed surfaces that are closer to the observed values, both cases result from the general equation of the equation method (3).

$$Z_{(x,y)} = T_{(x,y)} + \sum_{i=1}^{n} \lambda_i Z_{ri} \quad (3)$$

where Z is the value sought or point not observed, n, the number of points, λi coefficient of the system of linear equations, Zi is the distance of the point (x,y) to the point I [29].. T and R, are defined according to whether it is regularized or tension.

2.5. Hydrological Information

The information used for the assessment of interpolation methods is of the monthly and annual rainfall type. We reviewed the records of 116 stations located at the site of the study (Boyacá, Colombia), of which 60 had continuous records between 15 and 20 years. It should be recalled that rainy seasons record only the precipitation depth measured in (mm). The place where the research was applied is located in Colombia, South America, in the department of Boyacá, with an area of 23189 km², which represents 2.03% of the Colombian territory [30]. It is a region with three main hydrographic basins on the rivers Magdalena, Arauca and Meta, in the central area of Colombia, the figure 1, shows the area of study.

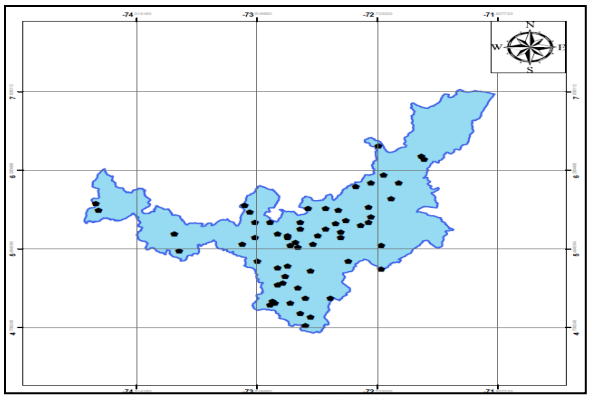


Figure 1: Location of rainfall seasons

3. Results and discussion

3.1. Validation of the methods evaluated.

The validation of the interpolated points using the three methods was performed by cross-validation, technique used to evaluate the results of a statistical analysis and to ensure that they are independent of the partitioning of the data obtained in this case from rainfall stations. The estimated value must be similar to the actual value, that is the measured value, so that the error committed by the model, that is, the difference between the two, is as low as possible [31].

The difference indicates how out-of-date the precipitation depth data calculated by the software are, when compared with the actual or measured information presented in the study area, that can be analyzed by the coefficient of determination or R2.

3.2. Cross-validation

Cross-validation is a way of predicting the fit of a model to a hypothetical test data set when the explicit test data set is not available. The cross-validation process gives a precise approximation of performance and error prediction [32]. Cross-validation in this study was performed by interpolating the totality of the rainfall stations and subtracting one weather station at a time, interpolating again, to verify by crossing the interpolated information against the actual or measured of each meteorological station. In other words, the interpolated data at the points where the measured data is held are verified and the actual or measured values and the interpolated values are compared.

3.3. Comparison of interpolation methods

Raster maps generated by a GIS were used in order to identify which interpolation method presents better results from different points of analysis; cross validation, coefficient of determination (R2) and the visual aspect, in the latter was sought to identify the error called "bull's eye". Based on the measured and calculated differences, statistical analyses were performed for the maximum annual and maximum monthly precipitation depth. The mean, median, minimum and standard deviation were compared, although it should be noted that greater importance was given to the statistical parameter of the mean, since it was considered to be the most relevant parameter in this case.

3.4. Interpolation of Annual precipitation depth.

It was observed that the Kriging technique has the least variation in the statistical parameter of the mean (54.70), of data standardization, but has the greatest dispersion of data with respect to the measured data (standard deviation), in addition it does not present an anisotropic behavior with spherical distribution behavior normalizing thus adjusting the plateau, data tolerance and bandwidth as shown in Figure 2.

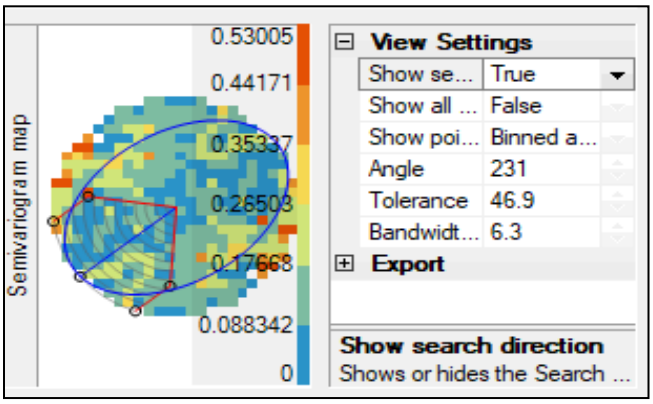


Figure 2: Data adjustment

In the standardization of data, a circular or spherical model was used in the semivariogram, adjusting the data to an exponential trend Figure 3, thus obtaining the highest reliability to perform the interpolation of IDW, ordinary kriging and spline as shown in Figure 4, 5, and 6 for 2008 year.

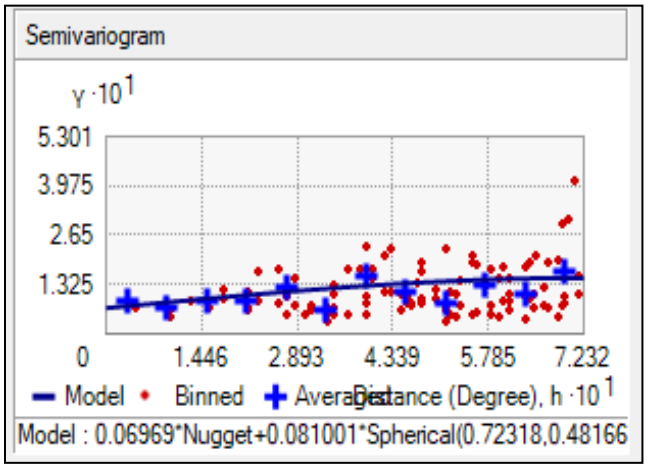


Figure 3: Semivariogram

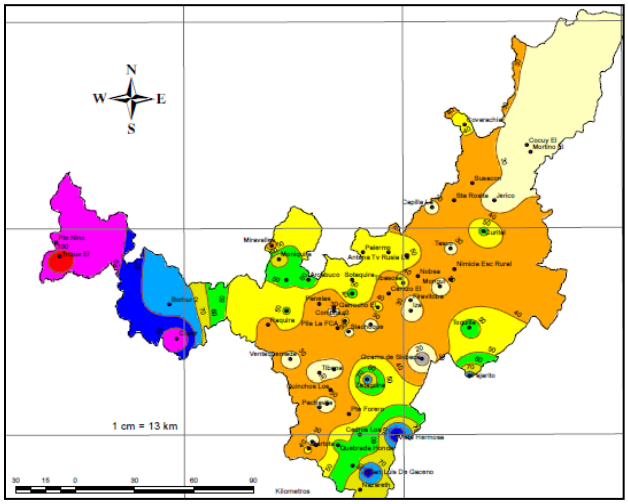


Figure 4: Interpolation by IDW, january 2008

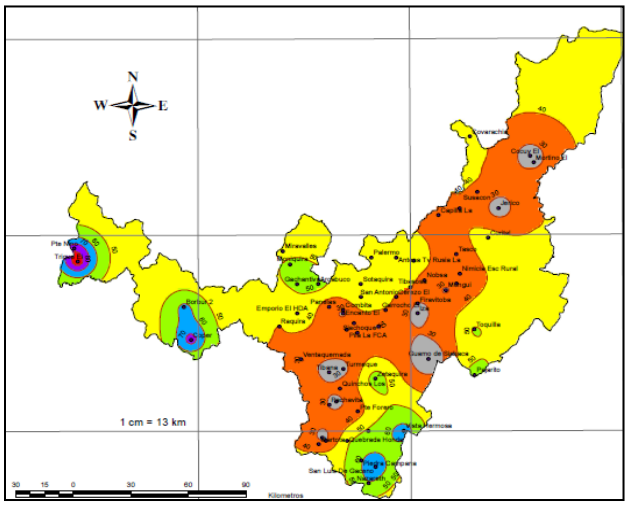


Figure 5: Interpolation by Kriging, january 2008

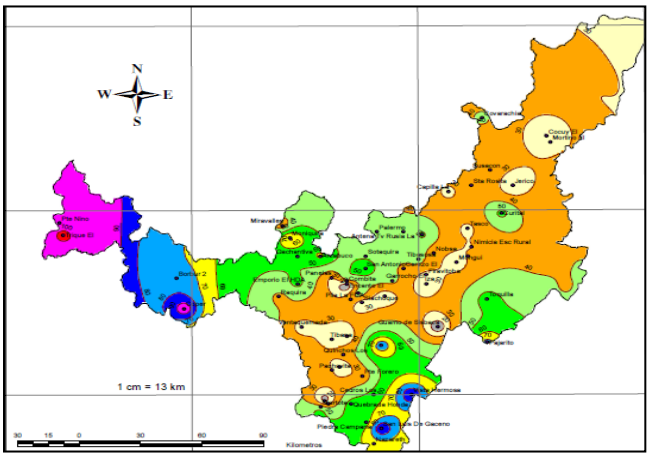


Figure 6: Interpolation by Spline, january 2008

On the other hand, the Spline technique also shows good performance in terms of average (54.66), but less than the Kriging value. The Spline method presented less dispersion than Kriging. Similarly, it was found that the IDW technique shows the greatest difference between the measured data and the interpolated data, in Table 1; the results of the analysis are evident.

Table 1: Statistical analysis of annual precipitation depth for 2008

Statistics	*Data measured*	*IDW*	*Spline*	*Kriging*
Mean	56.81	53.02	54.66	54.70
Median	47.1	49.04	48.40	51.04
Estándar deviation	30.51	20.12	20.99	16.26
Minimun	10.50	25.10	27.40	34.12

For the analysis of the interpolation of annual precipitation, it was determined that the kriging technique is the best fit in the statistical analysis of data, since the mean with the Kriging technique has a more appropriate behavior than the other two evaluated techniques.

3.5. Interpolation of monthly precipitation depth.

In the case of the interpolation of the monthly precipitation depth data, the Spline technique presented better results in terms of the statistical parameters of the mean and median, compared with the other techniques, although with great closeness between each method. In the second place, the Kriging technique is found in Table 2; the results of the analysis are evident.

Table 2: Statistical results of methods evaluated against mesuared monthy values, 2008

Statistics	Data measured	IDW	Spline	Kriging
Mean	66.31	62.22	63.95	63.86
Median	55.15	54.60	55.14	53.17
Estándar deviation	32.04	22.58	24.75	25.20
Minimun	28.80	35.13	31.70	34.44

3.6. Análisis de contornos.

Raster maps because they represent the space in a regular array of pixels, where the representation of the elements is made by pointing out the existence or not of data within each pixel [33], in the case of the regionalization of precipitation depth facilitates the visual behavior of the three interpolation techniques used in this investigation. Where we visually compared which of the three techniques has better contours and a more detailed surface in areas where there are no weather stations. One of the characteristics that is intended to show both in presence and absence is the typical error of "ox eyes" where interpolation techniques usually create concentration points or form islands around interpolated or measuring points [34], in this case the rainfall seasons. This anomaly is evident that with the IDW method where concentric circles are presented, as can be seen in Figure 1, this anomaly was presented as a visual trend a large percentage of the generated maps.

Ordinary kriging was the technique that presented the best results, since visually it presents different contours of precipitation depth, which is more accommodated to the recorded precipitation, consistent with the cross validation performed. In addition to performing the ordinary kriging, an additional raster is also interpolated to validate the interpolation of standard prediction errors, where the less data is those that may reliability and the larger data are those that should be resampled, as shown in figure 7.

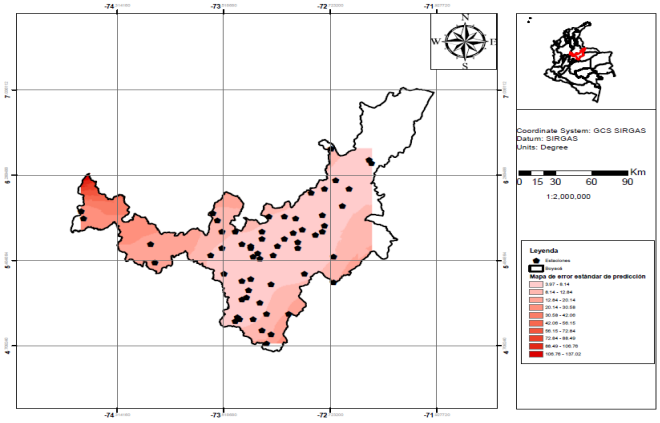


Figure 7: Standard Prediction Error Map.

3.7. Coefficient of determination R2

Raster maps because they represent the space in a regular array of pixels, where the representation of the elements is made

by pointing out the existence or not of data within each pixel [33], in the case of the regionalization of precipitation depth facilitates the visual behaviour of the three interpolation techniques used in this investigation.

3.8. Measurement of R2 in precipitation interpolation Annual

In this sense the method of interpolation of Kriging presents better correlation in the interpolation with depth of precipitation, the coefficient of determination is close to 1, with respect to the other models (Idw and Spline), thus yielding a high degree of reliability for choosing this model and carrying out the interpolation with which the regionalization of precipitation depth can be generated, due to the nature of this method it can be observed that it has a better behaviour with a higher data density, since analyzing interpolation with depth of monthly precipitation shows that the so-called spline model presents a greater correlation between its data, this allows choosing it to generate monthly regionalization.

Table 3: R2, para los métodos evaluados.

Interpolation Depth of Precipitation Max. Annual		Interpolation Precipitation Depth Max. Monthly	
Method	R2	Method	R2
Kriging	0,5762	Kriging	0,6717
Spline	0,5484	Spline	0,6861
IDW	0,5308	IDW	0,655

4. Conclusions

It was identified that visually the methods do not present an obvious difference, but in the case of the IDW technique generates the so-called portholes. For the regionalization of the maximum monthly precipitation depth it was possible to identify that the Spline method has a better behavior compared to the real data, since the R2, had a higher value (0.6861) compared to the other two methods evaluated.

In the case of the maximum monthly depth, the method that had a greater difference compared to the real data was the IDW, added to the visual errors (portholes) it is not advisable to use this method to regionalize hydrological data. Based on the analysis of both visual and cross-validation results, it can be stated that the kriging technique has a high degree of reliability when performing regionalizations with annual precipitation information.

In addition to the comparative results of the methods evaluated, the resulting information can be used as input for the determination of hydrological models in areas or locations that do not have their own information or measurement. In other words, they do not have a hydrological station. It could be shown that it is possible to use regionalized information with a high degree of confidence provided that it comes from data interpolated by the Spline method or the Kriging method

Twenty-four raster maps were generated, with maximum monthly and maximum annual rainfall depths, where twelve of these maps belong to the Spline technique which generates greater reliability with the maximum monthly values, and the remaining twelve were generated by the Kriging technique for maximum annual precipitation depth values. Maps generated by the IDW technique were discarded. It is evident that having a greater number of stations in optimum condition could make a more precise regionalization since interpolation would be more accurate, as well as having support stations would minimize the effect of "bull's eye".

References

[1] T.S. Kokkonen, A.J. Jakeman, P.C. Young, H.J. Koivusalo, «Predicting daily flows in ungauged catchments: Model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina», Hydrological Processes, **17**(11), 2219-2238, 2003, doi:10.1002/hyp.1329.

[2] M. Jessen, T. Einfalt, A. Stoffer, B. Mehlig, «Analysis of heavy rainfall events in North Rhine-Westphalia with radar and raingauge data», Atmospheric Research, **77**(1-4 SPEC. ISS.), 337-346, 2005, doi:10.1016/j.atmosres.2004.11.031.

[3] J.M. Mirás-Avalos, R.M. Mestas-Valero, P. Sande-Fouz, A. Paz-González, «Consistency analysis of pluviometric information in Galicia (NW Spain)», Atmospheric Research, 2009, doi:10.1016/j.atmosres.2009.05.011.

[4] G. Blöschl, M. Sivapalan, «Scale issues in hydrological modelling: A review», Hydrological Processes, **9**(3-4), 251-290, 1995, doi:10.1002/hyp.3360090305.

[5] Problems of rainfall-runoff modelling in arid and semiarid regions, oct. 2020, doi:10.1080/02626668809491261.

[6] R.J. Nathan, T.A. McMahon, «Evaluation of automated techniques for base flow and recession analyses», Water Resources Research, **26**(7), 1465-1473, 1990, doi:10.1029/WR026i007p01465.

[7] D.A. Post, A.J. Jakeman, «Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model», Ecological Modelling, **123**(2-3), 91-104, 1999, doi:10.1016/S0304-3800(99)00125-8.

[8] A. Agarwal, R. Maheswaran, V. Sehgal, R. Khosa, B. Sivakumar, C. Bernhofer, «Hydrologic regionalization using wavelet-based multiscale entropy method», Journal of Hydrology, **538**(538), 22-32, 2016, doi:10.1016/j.jhydrol.2016.03.023.

[9] C.C.F. Plouffe, C. Robertson, L. Chandrapala, «Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: A case study of Sri Lanka», Environmental Modelling and Software, **67**, 57-71, 2015, doi:10.1016/j.envsoft.2015.01.011.

[10] Y. Gyasi-Agyei, G. Pegram, «Interpolation of daily rainfall networks using simulated radar fields for realistic hydrological modelling of spatial rain field ensembles», Journal of Hydrology, **519**(PA), 777-791, 2014, doi:10.1016/j.jhydrol.2014.08.006.

[11] M.A. Oliver, R. Webster, «Kriging: A method of interpolation for geographical information systems», International Journal of Geographical Information Systems, **4**(3), 313-332, 1990, doi:10.1080/02693799008941549.

[12] Cómo funciona Kriging—Nuevas características y mejoras que se han agregado en esta versión. | Documentación, oct. 2020.

[13] A. Mazzella, A. Mazzella, «The importance of the model choice for experimental Semivariogram modeling and its consequence in evaluation process», Journal of Engineering (United Kingdom), **2013**, 2013, doi:10.1155/2013/960105.

[14] I.K. Tsanis, M.A. Gad, «A GIS precipitation method for analysis of storm kinematics», Environmental Modelling and Software, **16**(3), 273-281, 2001, doi:10.1016/S1364-8152(00)00068-2.

[15] D. Koutsoyiannis, «Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation», Hydrological Sciences Journal, **49**(4), 575-590, 2004, doi:10.1623/hysj.49.4.575.54430.

[16] E. Rojas, B. Arce, A. Peña, F. Boshell, M. Ayarza, «Cuantificación e interpolación de tendencias locales de temperatura y precipitación en zonas alto andinas de Cundinamarca y Boyacá (Colombia)», Corpoica Ciencia y Tecnología Agropecuaria, **11**(2), 173, 2010, doi:10.21930/rcta.vol11_num2_art:209.

[17] M. Sivapalan, G. Blöschl, «Transformation of point rainfall to areal rainfall: Intensity-duration-frequency curves», Journal of Hydrology, **204**(1-4), 150-167, 1998, doi:10.1016/S0022-1694(97)00117-0.

[18] R.J. Moreano Viteri, SISTEMA DE INFORMACION PARA LA INTERPOLACION ESPACIAL Y TEMPORAL DE DATOS SOBRE EL TIEMPO ATMOSFERICO Y EL CLIMA DEL ECUADOR , QUITO/ EPN/ 2008, 2008.

[19] Ana Carolina SantosAndrés Vargas-LunaAndrés Vargas-LunaNelson

Obregón-NeiraNelson Obregón-NeiraEDER CÁRDENAS, «Análisis de la distribución e interpolación espacial de las lluvias en Bogotá, Colombia», Dyna, 151-159, 2010.

[20] G. Gong, S. Mattevada, S.E. O'Bryant, «Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas», Environmental Research, **130**, 59-69, 2014, doi:10.1016/j.envres.2013.12.005.

[21] D. Zimmerman, C. Pavlik, A. Ruggles, M.P. Armstrong, «An experimental comparison of ordinary and universal kriging and inverse distance weighting», Mathematical Geology, **31**(4), 375-390, 1999, doi:10.1023/A:1007586507433.

[22] S.M. Vicente-Serrano, M.A. Saz-Sánchez, J.M. Cuadrat, «Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): Application to annual precipitation and temperature», Climate Research, **24**(2), 161-180, 2003, doi:10.3354/cr024161.

[23] F. Huang, D. Liu, X. Tan, J. Wang, Y. Chen, B. He, «Explorations of the implementation of a parallel IDW interpolation algorithm in a Linux cluster-based parallel GIS», Computers and Geosciences, **37**(4), 426-434, 2011, doi:10.1016/j.cageo.2010.05.024.

[24] P.M. Bartier, C.P. Keller, «Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW)», Computers and Geosciences, **22**(7), 795-799, 1996, doi:10.1016/0098-3004(96)00021-0.

[25] M.P. Armstrong, R.J. Marciano, «Local interpolation using a distributed parallel supercomputer», International Journal of Geographical Information Systems, **10**(6), 713-729, 1996, doi:10.1080/02693799608902106.

[26] Q. Huang, C. Yang, «Optimizing grid computing configuration and scheduling for geospatial analysis: An example with interpolating DEM», Computers and Geosciences, **37**(2), 165-176, 2011, doi:10.1016/j.cageo.2010.05.015.

[27] H. Hu, H. Shu, «An improved coarse-grained parallel algorithm for computational acceleration of ordinary Kriging interpolation», Computers and Geosciences, **78**, 44-52, 2015, doi:10.1016/j.cageo.2015.02.011.

[28] N.C.-N.Y.A.W.-I. Publicatiorq, undefined 1993, «Statistics for spatial data. revised edition».

[29] M.F. Hutchinson, «A new procedure for gridding elevation and stream line data with automatic removal of spurious pits», Journal of Hydrology, **106**(3-4), 211-232, 1989, doi:10.1016/0022-1694(89)90073-5.

[30] R. Franke, «Smooth interpolation of scattered data by local thin plate splines», Computers and Mathematics with Applications, **8**(4), 273-281, 1982, doi:10.1016/0898-1221(82)90009-8.

[31] J.A. Becerra Oviedo, L.F. Sánchez Mazorca, P.M. Acosta Castellano, J.L. Díaz Arévalo, «Regionalización de curvas IDF para el uso de modelos hidrometeorológicos en la Sabana Occidental del departamento de Cundinamarca», Ingeniería y Región, **14**(2), 143, 2016, doi:10.25054/22161325.701.

[32] P.M. Acosta Castellanos, L.X. Sierra APonte, «Evaluación de métodos de construcción de curvas IDF a partir de distribuciones de probabilidad y parámetros de ajuste», REVISTA FACULTAD DE INGENIERÍA, **22**(35), 25, 2013, doi:10.19053/01211129.2512.

[33] Colombia: Diagnóstico departamental de situación de DDHH y DIH - Boyacá - Colombia | ReliefWeb, oct. 2020.

[34] T Izquierdo., Álvaro Márquez, «Estudio comparativo de diferentes métodos de interpolación para la realización de mapas de precipitación en la isla de La Gomera (Islas Canarias): Evaluación de la incorporación de la altura - Dialnet», Geogaceta, 131-134, 2007.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1328-1334 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

SH-CNN: Shearlet Convolutional Neural Network for Gender Classification

Chaymae Ziani*, Abdelalim Sadiq

MISC Laboratory, Faculty of Sciences, Ibn Tofail University, Kenitra, 14000, Morocco

ARTICLE INFO

Article history:
Received: 04 September, 2020
Accepted: 01 December, 2020
Online: 16 December, 2020

Keywords:
CNN
Shearlet transform
Gender prediction
Features extraction

ABSTRACT

Gender detection and age estimation become an active research area and a very important field today, wish has been widely used in various applications including them: biometrics, social network, Targeted advertising, access control, human-computer interaction, electronic customer, etc. The need to further improve the recognition or classification rate keeps increasing day after day. In this paper, we explore how deep learning techniques can help in the classification of gender from human face images and moreover raise the recognition rate. We propose in this contribution an approach called SH-CNN based on Discrete Shearlet Transform (DST) as a first step of feature extraction layer, and Deep Convolutional Neural Network (DCNN) as a second automatic feature extraction layer and also a classification step. The idea behind our contribution is to generate trough DST several features (in different decomposition and orientation) of an image. These features of each image will be the input of the DCNN, to enrich the training step, and so, improve the recognition rate. The obtained results have shown that the proposed approach import a significant enhancement.

1. Introduction

This paper is an extension of work originally presented in The Third International Conference on Intelligent Computing in Data Sciences ICDS2019 [1].

Automatic age and gender classification of humans through their image faces is a fundamental task in computer vision. It is an important research focus this last years, and plays a key role in a large variety of applications. In this last years, The Convolutional Neural Networks (CNNs) model has shown high level classification capacity, and mainly in cases of face images taken in real-world conditions. The performance of this model result in the ensured high level feature extraction and classification, through the used layers.

The Feature Extraction Function (FEF) is the key factor that determines the efficiency of a Face Recognition System and considered as one of the important parts. Several feature extraction function have been proposed in this domain of face feature extraction, such as, wavelets [2], SIFT [3] and other [4], [5]. Until today, the need for more efficient feature extraction function increases, seen the use of face gender recognition system in sensitive and important area. The Shearlet transform as feature extraction function allows generating more sensitive and significant features [6]-[8].

The field of learning has undergone a revolution with the appearance of CNN because the CNN can extract features of different complexities. Unlike the learning technique, convolutional neural networks learn the features of each image automatically, and that's where their strength lies. Today, the most successful models for image classification are those based on CNN. Since the model is based on learning, so we will enrich this phase with features extracting by DST. The Discrete Shearlet Transform (DST) is used in our paper as a first feature extraction layer for our proposed SH-CNN. The idea is to extract more significant features of the input face, to further enhance the gender recognition rate.

2. Related work

There is already a rich literature on the subject of classification by age and gender in an attempt to further improve the recognition rate. We briefly reviewed the age and gender classification studies that use neural network or Convolutional neural network. The authors in [9] present a structure including CNN and Extreme Learning Machine (ELM), where the CNN is used to extract the features from the input images and ELM, to classify the interposed results, and to combine the co-action of the two classifiers, to perform age and gender classification. In [10], the authors

*Corresponding Author: Chaymae Ziani, Email: chema.zin@gmail.com, chaymae.ziani@uit.ac.ma

implement a CNN of 3 convolutional layers and 3 fully connected layers for the tasks of estimating age and gender, it is trained exclusively on Audience. For the forecast, they do several crops around the face in different sizes. The final prediction is the average prediction value for all of these crops. They achieve an 86.8% rate of good classifications on the audience test images. A Deep gender classification is performed in [11], which is based on AdaBoost-fusion of isolated facial features and foggy faces. They applies an approach based on local features of the face by separating it into several parts (mouths, eyes, nose). They also include the discriminating features surrounding the face while blurring it. These images are then used to train several CNNs followed by an Adaboost classifier grouping the classification scores of the CNNs. This method yielded 91.75% on LFW and 83.06% on Adience using Gallagher as training data. In [12] the authors address the estimations problem with a classification based model which considered the age value as an independent etiquette and a separate class, and propose a deep convolutional neural network model learns the pertinent informative age and gender characterizations from the image pixel. A Male/female identification method from very low resolution face images by using neural network have proposed in [13]. In [14], the authors proposed a method named the full component CNN, which holds the full faces and the facial component networks, to obtain human age/gender classification through the facial image data. For age and gender classification, the authors in [15] propose a method applying deep learning based learning, the feature extractor HOG is applying on the input image on the network. The structure of the network used in this study is GoogLeNet. In [16], the authors present a review and a study of the advanced research technique for gender classification, which can be useful for researchers who work on gender classification. In [17], a convolutional neural network is used for the estimation of age and gender of a person from the facial image, and the PCA is applied to reduce the dimensions of the extracted features.

Among the research works countable on fingers that use CNN and shearlet transform, we find a contribution in [18], where the authors use an algorithm based on CNN and Shearlet transform, to classify the patient having Alzheimer's disease and early or late mild cognitive impairment using the florbetapir PET (positron emission tomography) amyloid imaging data. The authors had concatenated shearlet coefficients into the convolution layers on the CNN.

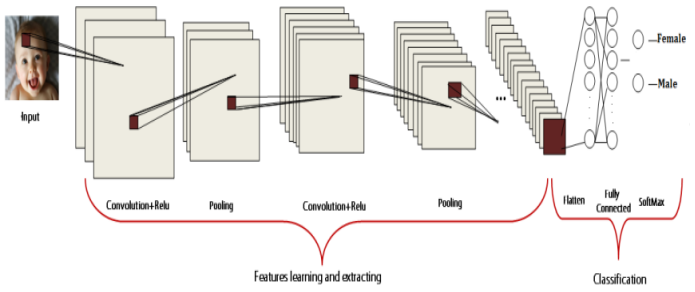


Figure 1: Example of a DCNN architecture for gender prediction

In our knowledge, this is the first time that Shearlets are used as first pretreatment, in order to enrich the training phase of the DCNN. The idea is to add as first treatment a DST, to exploit the calculated features extracted through decompositions and orientations. These operations allow us to generate from one input image, several feature views. Based on these generated features, we enrich the training phase, by injecting to DCNN the image and its feature views (the DST decompositions and orientations). The proposed approach will allow us to enhance the gender recognition, seen that the training phase is improved.

3. Deep Convolutional neural network and shearlet transform

3.1. DCNN

Convolutional Neural Networks [19] was proposed mainly to treat 2D shapes variability. They have been used widely in computer vision and they showed the best results and outperformed all other technique. Now, DCNN has become a major research orientation for researchers in the domain of computer vision, especially for image classification.

DCNN is powerful model, which is capable to capture effective information due to his architecture (Figure 1). DCNN architecture is divided into 2 steps, the first step is learning and extracting features and the second step is classification. There are four types of layers for a CNN: the convolutional, the pooling, the ReLU correction, and the fully-connected.. These components work together to learn a dense feature representation of an input.

- The convolution layer (Figure 2) is the important component of CNN, and always constitutes their first layer. Its goal is to identify features in the images received as input. To do this, convolution filtering is applied, by sliding window that represents feature, and by computing the convolution product of each image part and feature. Then, a feature is viewed as a filter. An activation map is obtained for each pair (image, filter) which indicates where the features are located in the image, when the value elevated, the more the corresponding place in the image looks like the feature. And then, these feature maps are normalized (with an activation function) [20].
- The ReLU correction layer used as an activation function, it is defined by ReLU (x) = max (0, x) [20], to replace all negative values with zeros

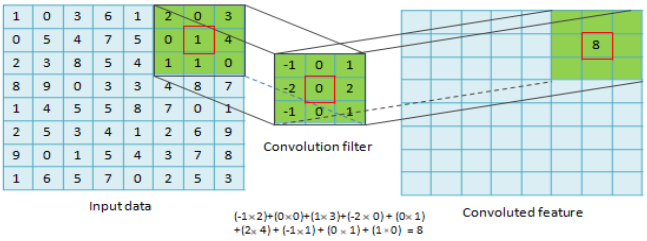


Figure 2: Convolution filter applied on input image

- The pooling layer (Figure 3): feature maps are injected as input to this layer. The principal of pooling is to reduce image size without losing the essential characteristics. In general, the pooling layer is nested in two convolution layers, and aims to enhance network-learing, by minimizing over-calculations [20].
- As last layer of the network. The fully-connected layer [20] has as input a vector, on which it applies linear combination

and activation function, and so, generated output vector. The returned vector by this layer to classify the input image represents the possible output classes of the classification system. Each value of the vector represents the corresponding probability of the input image to a class. The input matrix resulting from the previous layer contains an activation map, which represents the position of feature in the image. According to the position of feature in the image, the fully-connected layer can calculate the corresponding probability to each output classes. In general, for the DCNN classification systems, the softmax method is often used at the end of the network [21].

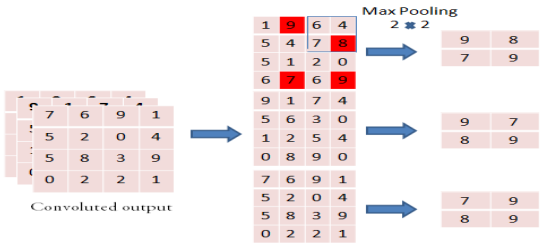


Figure 3: Example of a max pooling operation [22]

3.2. Shearlet transform

A multidimensional version of wavelet transform were invented in 2006 [23] named Shearlet, that encode anisotropic features in multivariate problem classes. The limitation of wavelets is the incapacity of capturing anisotropic features. Therefore, this transform provided optimal sparse representation for images with edges, surpassing the wavelet.

The successful use of shearlets in many image processing applications, especially face detection, feature extraction and denoising [24]-[26], is due to their properties: parabolic scaling, shearing, and translation. Expressly, the Continuous Shearlet Transform (CST) [27] is defined as:

$$f \longmapsto SH_\psi f(a,s,t) = \langle f, \psi_{a,s,t,} \rangle, \quad f \in L^2(\mathbb{R}^2)\ , (a,s,t) \in \mathbb{R}_{>0} \times \mathbb{R} \times \mathbb{R}^2 \tag{1}$$

where,

$$\psi_{a,s,t(x)} = \mid det M_{a,s} \mid^{-\frac{1}{2}} \psi\left(M_{a,s}^{-1}(x-1)\right) \tag{2}$$

The parameter of this system calls as wavelets with composite dilations.

And
$$M_{a,s} = \begin{pmatrix} a & s \\ 0 & \sqrt{a} \end{pmatrix}. \tag{3}$$

The matrix $M_{a,s}$ is factorized as B_sA_a, where $B_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ is a shearing matrix and $A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$ is a parabolic scaling matrix (or anisotropic dilation matrix).

The Shearlets transform function is based on three variables (Figure 4): a is the scale parameter measuring the resolution level, s is the shear parameter measuring the directionality and t is the translation parameter measuring the position.

To get a frame (or Parseval Frame) for $L^2(R)$, a CST sampling is applied on a suitable discrete set of the three parameters: scale, shear, and translation. A discrete shearlets can be generated by sampling those three parameters:

$$a_j = 2^j (j \in \mathbb{z}), \qquad s_{j,k} = k a_j^{\frac{1}{2}} = k 2^{\frac{j}{2}} (k \in \mathbb{z})\ ,$$
$$\text{and} \qquad t_{j,k,m} = D_{a_j,s_{j,k}} (m \in \mathbb{z}^2) \tag{4}$$

For $\psi \in L^2(\mathbb{R}^2)$, the Discrete Shearlet Transform of $f \in L^2(\mathbb{R}^2)$ is the mapping defined by:

$$f \longmapsto SH_\psi f(j,k,m) = \langle f, \psi_{j,k,m} \rangle,$$
$$f \in L^2(\mathbb{R}^2)\ , (j,k,m) \in \mathbb{z} \times \mathbb{z} \times \mathbb{z}^2 \tag{5}$$

Where (6) is the discrete Shearlets function (DST)

$$\psi_{j,k,m,}(x) = 2^{\frac{3j}{2}} \psi(B_k A^j x - m), j \geq 0, -2^j \leq k \leq 2^j - 1, \quad m \in \mathbb{z}^2 \tag{6}$$

and $M_{a,s}$ is discretized as $M_{j,k} = B_k A^j$. Thus, SH_ψ maps the function f to the coefficients $SH_\psi f(j,k,m)$ correlated with the indexes of scale j, the orientation k, and the position m.

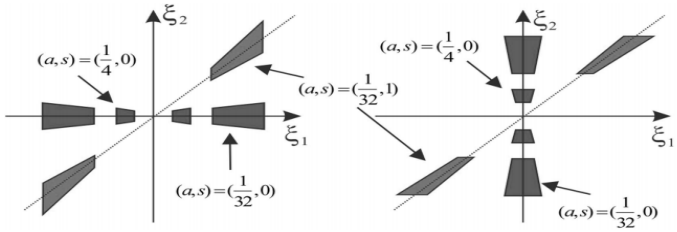
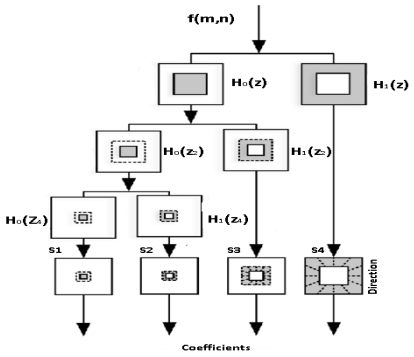


Figure 4: The horizontal and vertical shearlets Frequency support [28]



The shearlet transformation is a decomposition function based on scales and direction set. As shown in Figure 5, the decomposition is done by 4 scales levels denoted S1, S2, S3, and S4 with 6 directional filtered coefficients. At first, repetitively, a decomposition of the input image is done to a low-pass image and a high-pass image, based on a decomposition of two-channel non-subsampled [30], [31]. The used filters are: $H_0(z)$ for low pass filter function, and $H_1(z)$ for high pass filter function. The decomposition steps are: first, a frequency transformation domain applied on the high pass image, using two dimensional Fourier transform. Secondly, a cartesian coordinate system with six directions is used, to generate six directional subbands from the frequency domain. For the low pass image, a decomposition of two-channel non-subsampled is applied again, and the same

decomposition steps are repeated. The shearlet coefficients are generated as result of this transformation [29].

4. Proposed Approach

The idea which motivated us to propose this solution is the objective of enriching the learning phase to improve gender recognition. To enrich the DCNN learning input sample. We proposed to use as the first preprocessing of the CNN, a layer which uses the DST. The latter allows us to generate from a CNN learning input image, several views made up of the features generated from the decompositions and orientations of the DST. At this point, in the learning phase, the input will be represented not just by the images but also by their feature views. In this way, we will increase the input sample from the same number of starting images. Such a technique will be very useful in the case where the image sample in the training phase is small.

In our proposed approach, the training and the testing phase of the system performs essential operations. Each operation contributes to ensure a high recognition level. Figure 6 and Figure 7 showed these operations:

- Face detection
- Preprocessing
- Features extraction using DST
- Automatic features extraction and classification

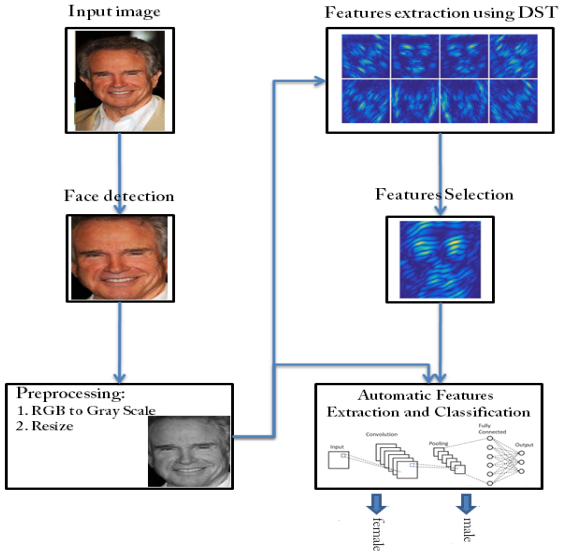


Figure 6: Proposed SH-CNN approach (training phase)

In our approach, haar-cascades [32] are used for face detection in order to eliminate any unusable information in the image. As a result, we will focus just on faces.

4.2. Preprocessing

At this level, two main treatments are applied on each face photo. The first treatment is to convert the photo from a RGB color level to a Gray level. The second treatment is to reduce the size of the faces photos. These steps reduce the calculation time in extracting features, and also in CNN layers.

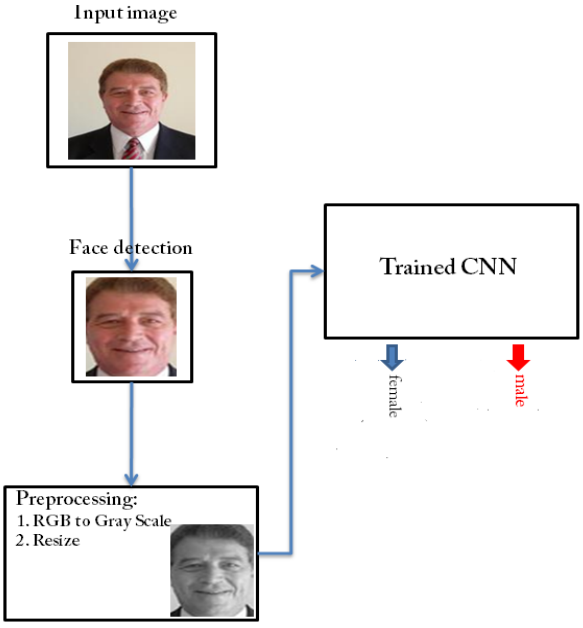


Figure 7: Proposed SH-CNN approach (testing phase)

The Feature Extraction Function (FEF) is a key factor that determines the efficiency of gender Recognition System. DST as a FEF is applied in each face image; the images will decompose along a number of directions and scales to extract important facial features. At the end, for each image we get a set of images in different orientations. That will enrich the CNN's learning step and facilitate the classification phase. And that is the goal for combining it with the CNN. This operation is used only in training phase, and so, we will not increase the processing time in the testing phase.

4.4. Automatic features extraction and classification (CNN)

The idea is very simple. The trainable system consists of a series of modules, each one representing a processing step. Each module is trainable, with adjustable parameters similar to the weights of linear classifiers. The system is trained from end to end: in each instance, all parameters of all modules are adjusted to bring the output produced by the system closer to the desired output.

To enrich the training phase, we inject as an input in our CNN the face and the corresponding coefficients calculated using DST. This operation allows the CNN to adapt more the weights and to increase the learning step.

5. Experiment results

This section presents the experimental results of our proposed approach. Our method is implemented using Python as a high-level programming language, with the Tensorflow [33] (an OSS framework for machine learning on decentralized data) and Keras(is a deep learning API) [34] libraries for CNN learning and classification.

5.1. UTKFace dataset

UTKFace dataset [35] is composed from over 20,000 face images (e.g. Figure 8), which are tagged by age and gender. The

images of this dataset present multiple facial expressions. Certain images are with high and low resolution, things that allow testing our system well. The UTKFace dataset is frequently used in gender/age classification and face detection, etc.

The images size is 200×200 pixels. We have resized all the face with the dimension of 100×100 pixels.

5.2. Operations of extracting features and classification

In the previous section, we presented our approach with successive operations. In this section, we present more details on the proposed SH-CNN approach for automatic gender recognition.

The DST ensures a good representation level of images with edges. At the features extraction step, the pre-treated images were decomposed into 4 levels (Table 1), by varying shear parameter from 2^0 to 2^3, and scale parameter from 0 to 3, and hence we obtained shearlet coefficients on many directions (Figure 9).As we said before these coefficients will be used later as input of a CNN. We repeat this procedure for all faces of the training database.

Table 1: Used DST parameters.

Decomposition levels	Directions	Shearing filters size
Level 1	4	32×32
Level 2	8	32×32
Level 3	16	16×16
Level 4	16	16×16

In our proposed approach, and during all experiments for gender classification, we used a CNN composed from three convolutional layers and one fully-connected layer. Gender classification on the face dataset requires distinguishing between two classes, that the reason why we choose a smaller network.

The input layer is a face image of size 100 × 100 pixels. The three convolutional layers are respectively of 96@7 × 7, 256@5 × 5, and 384@3 × 3 feature maps. After each one of these convolution layer, a ReLU function and a max pooling layer (the max value of 3×3) are applied. To generate the needed 512 neurons for the fully connected layer, a flatten layer is used. The next layer is a fully connected, that maps to the final gender classes. As a last step, a soft-max layer is applied to attribute a corresponding probability for each class. The class having the probability max value represents the closest prediction

The used criteria in this experiment results take into consideration our need in terms of features extraction part and CNN classification part. For the DST feature extraction part, we have used four level of decomposition with directions (4, 8, 16, and 16). In general in the existing works, and as the recommended values, the used decomposition number is approximatively 3, 4 or 5 levels, and for directions we found (4, 8, 16, and 32) [6]-[8], [26]. In our contribution we have used four level with directions (4, 8, 16, 16), which have clearly shown the added value of our contribution. For the CNN classification part, we have used three convolutional layers with a kernel size of 7x7, 5x5 and 3x3, in order to reduce the training time. And for the fully connected layer, we used 512 neurons, since using larger values in case of 100x100 input image size will not improve the results, contrary, it will increase the calculation time.

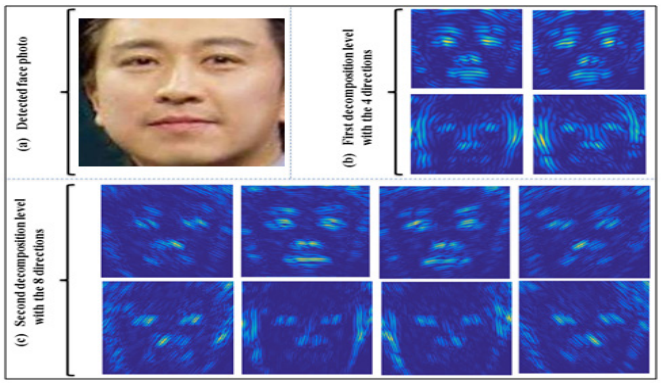


Figure 9: (a) Detected face photo (b) 1st decomposition level with 4 orientations applied on the detected face (c) 2nd decomposition level with 8 orientations applied on the detected face.

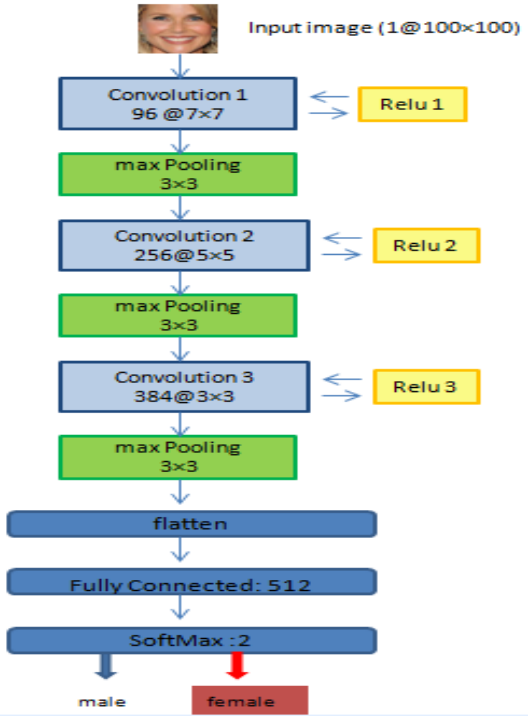


Figure 10: Proposed CNN architecture

5.3. Results and discussion

Regarding the comparison phase of the proposed approach (SH-CNN), we carried out a direct comparison study with the CNN, without seeing the other models that already exist, proposed as part of the improvement of the gender recognition, and based on CNNs. It is seen that our first objective was to test the added value of our SH-CNN compared to CNN. Add to this that the proposed preprocessing layer concatenated with the CNN can be concatenated with all the proposed models based on the CNN as

well. Because the add layer plays the role of an enrichment phase of the sample of the learning part of the CNN. From this explanation, we understand that the comparison must be made with CNN directly.

The obtained results in terms of accuracy rate and loss rate (see Figure 11 and 12), shown clearly the added value of our proposed approach. The generated coefficients for each face using DST, enrich the training phase as expected, and so, increase the recognition rate. In this simulation, we have used just one generated coefficient of DST as input with the associated face. If we use more coefficient as input to enrich the training phase, we may obtain more efficient recognition system, and this will be our coming research work. Also, in our simulation, we have used just 3 epochs, to show the capacity of our approach to increase more the accuracy rate within a few epochs.

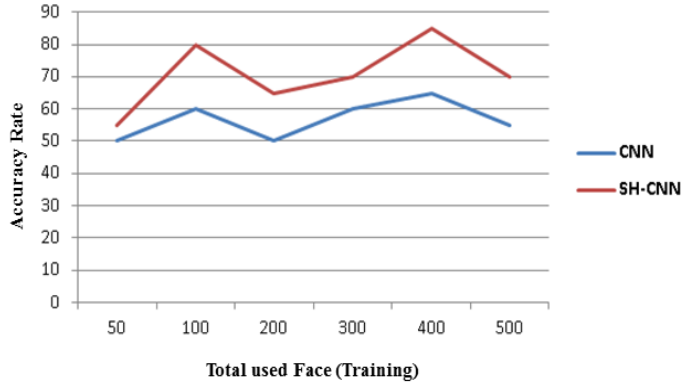


Figure 11: The obtained Accuracy Rate (CNN vs SH-CNN)

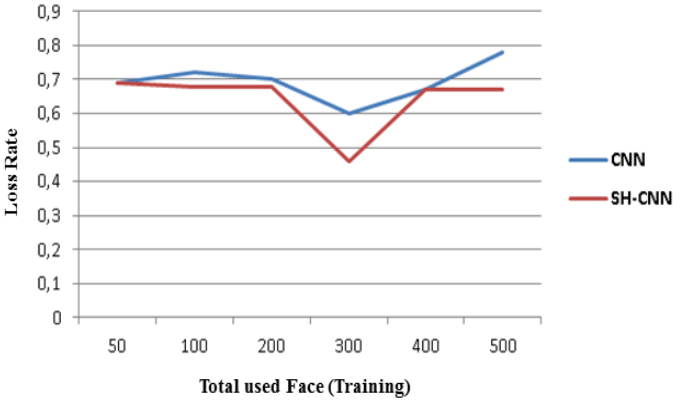


Figure 12: The obtained Loss Rate (CNN vs SH-CNN)

6. Conclusion

The need to increase the gender recognition rate has pushed us to search lower-cost solutions (processing time, dataset size, etc.). In this paper, we have proposed an approach based on CNN and especially on DST as a first layer. The idea behind this solution is to enrich more the training phase, by using the generated coefficients of each image using DST. The obtained results have shown the clear added value of this approach. The ensured accuracy rate of our system is also related to the capacity of DST to extract the significant features of each image.

In our future work, we will add also the age as output of our proposed system. We will also use high and low values of epoch, and explore the capacity of our proposal in others databases (FERET, etc.).

References

[1] C. Ziani, A. Sadiq, "Shearlet Convolutional Neural Network Approach for Age and Gender recognition," in 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS), 1-4, 2019, doi: 10.1109/ICDS47004.2019.8942359.

[2] K. Ramesha, K. B. Raja, "Face Recognition System Using Discrete Wavelet Transform and Fast PCA," in Information Technology and Mobile Communication, **147**, 13-18, 2011, https://doi.org/10.1007/978-3-642-20573-6_3.

[3] P. G. Ambhore , L. Bijole, "Face Sketch to Photo Matching Using LFDA,". International Journal of Science and Research (IJSR), 3(4), 11-14, 2014,

[4] J .Qian, J. Yang, "A Novel Feature Extraction Method for Face Recognition under Different Lighting Conditions," in Chinese Conference on Biometric Recognition (CCBR), **7098**, 17-24 , 2011, https://doi.org/10.1007/978-3-642-25449-9_3.

[5] T. T. Do, T. H. Le, " Facial Feature Extraction Using Geometric Feature and Independent Component Analysis," In Knowledge Acquisition: Approaches, Algorithms and Applications, **5465**, 231-241, 2009, https://doi.org/10.1007/978-3-642-01715-5_20.

[6] Z. Zeng, J. Hu, "Face recognition based on shearlets transform and principle component analysis," in 2013 5th International Conference on Intelligent Networking and Collaborative Systems, 697-701, 2013, doi: 10.1109/INCoS.2013.134.

[7] A. Danti, K. M. Poornima, "Face Reconition using Shearlets," in 2012 IEEE 7th International Conference on Industrial and Information Systems, 1-6, 2012, doi: 10.1109/ICIInfS.2012.6304796..

[8] X. Sun, X. Wu, and Y. Wei, "Face Recognition Based on Shearlet Transform and Fast ICA," in 2014 Fourth International Conference on Instrumentation and Measurement, Computer, Communication and Control, 832–835, 2014, doi: 10.1109/IMCCC.2014.175.

[9] M. Duan, K. Li, C. Yang, K. Li, "A hybrid deep learning cnn–elm for age and gender classification" , Neurocomputing, **275**, 448–461, 2018, https://doi.org/10.1016/j.neucom.2017.08.062

[10] G. Levi , T. Hassncer, "Age and gender classification using convolutional neural networks," in 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 34-42, 2015, doi: 10.1109/CVPRW.2015.7301352.

[11] A. Mahmoud ,A. Abdelhamed, "AFIF4: Deep Gender Classification based on AdaBoost-based Fusion of Isolated Facial Features and Foggy Faces," Journal of Visual Communication and Image Representation, **62**, 77-86, 2019 , https://doi.org/10.1016/j.jvcir.2019.05.001.

[12] O. Agbo-Ajala, S. Viriri, " Age Group and Gender Classification of Unconstrained Faces," In Advances in Visual Computing, **11844**, 418-429 , 2019 https://doi.org/10.1007/978-3-030-33720-9_32.

[14] C. Huang, Y. Chen, R. Lin , C. J. Kuo, "Age/gender classification with whole-component convolutional neural networks (WC-CNN)," 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 1282-1285, 2017, doi: 10.1109/APSIPA.2017.8282221.

[16] C. P. Divate, S. Z. Ali, "Study of Different Bio-Metric Based Gender Classification Systems," in 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 347-353, 2018, doi: 10.1109/ICIRCA.2018.8597340.

[17] B. Agrawal, M. Dixit, "Age Estimation and Gender Prediction Using Convolutional Neural Network," In: Intelligent Computing Applications for Sustainable Real-World Systems ,Proceedings in Adaptation, Learning and Optimization, **13**, 163-175, 2020, https://doi.org/10.1007/978-3-030-44758-8_15.

[18] E. Jabason, M. O. Ahmad, M. N. S. Swamy, "Shearlet based Stacked Convolutional Network for Multiclass Diagnosis of Alzheimer's Disease using the Florbetapir PET Amyloid Imaging Data," in 2018 16th IEEE International New Circuits and Systems Conference (NEWCAS), 344-347, 2018, doi: 10.1109/NEWCAS.2018.8585550.

[19] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning

applied to document recognition," in Proceedings of the IEEE, **86**(11), 2278-2324, 1998, doi: 10.1109/5.726791.

[20] R. Waseem , Z. Wang. "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review," Neural Computation, **29**(9), 2352-2449, 2017, https://doi.org/10.1162/neco_a_00990

[21] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun.ACM, **60**(6), 84–90, 2017, https://doi.org/10.1145/3065386

[22] "Machine Learning, Max Pooling With Color Images", https://dev.to/sandeepbalachandran/machine-learning-max-pooling-with-color-images-2i21, 2020

[23] G. Kanghui, G. Kutyniok, D. Labate, "Sparse multidimensional representations using anisotropic dilation and shear operators," Wavelets and Splines , Nashboro Press, Nashville, 189–201, 2006.

[24] G.R. Easley, D. Labate, "Critically sampled wavelets with composite dilations," IEEE Trans. Image Process, **21**(2), 550-561, 2012, 10.1109/TIP.2011.2164415.

[25] G.R. Easley, D. Labate ,V. Patel, "Directional multiscale processing of images using wavelets with composite dilations," J. Math. Imag. Vision, **48**, 13-44, 2014, https://doi.org/10.1007/s10851-012-0385-4.

[26] S. Yi, D.Labate, G.R. Easley, H. Krim, "A Shearlet approach to Edge Analysis and Detection," IEEE Trans. Image Process, **18**(5), 929–941, 2009, doi: 10.1109/TIP.2009.2013082.

[27] G. Kutyniok , D. Labate, "Resolution of the wavefront set using continuous shearlets," Transactions of the American Mathematical Society, **361**(5), 2719-2754, 2009, http://www.jstor.org/stable/40302874.

[28] C.Liu, D.Wang, J.Sun, T.Wang, "Crossline-direction reconstruction of multi-component seismic data with shearlet sparsity constraint," Journal of Geophysics and Engineering, **15**(5), 1929–1942, 2018, https://doi.org/10.1088/1742-2140/aac097.

[29] F.Yuan, L. Po , M.Liu, X. Xu, W. Jian., K. Wong, K. Cheung , "Shearlet Based Video Fingerprint for Content-Based Copy Detection," Journal of Signal and Information Processing, **7**, 84-97. doi: 10.4236/jsip.2016.72010.

[30] J.P.Zhou, A.L. Cunha, M.N. Do, "Nonsubsampled Contourlet Transform Construction and Application Enhancement," IEEE International Conference on Image Processing, 1-469, 2005, doi: 10.1109/ICIP.2005.1529789.

[31] G.Easley, D.Labate, W.Q.Lim , "Sparse Directional Image Representations Using the Discrete Shearlet Transform," Applied and Computational Harmonic Analysis, **25**, 25-46, 2007, http://dx.doi.org/10.1016/j.acha.2007.09.003

[32] P. Viola, M. J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", IEEE Computer Society, **1**, 511-518, 2001, doi: 10.1109/CVPR.2001.990517.

[33] "keras", https://keras.io/, 2020

[34] "Tenserflow", https://www.tensorflow.org/federated, 2020

[35] "UTKface database", https://susanqq.github.io/UTKFace/, 2020

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1335-1339 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Quality of Nursing Care in Hospitalized Patients of the Carlos Lanfranco La Hoz Hospital, 2019

Amancio Izquierdo-Príncipe[1], Jaqueline Garcia-Núñez[1], Brian Meneses-Claudio[2,*], Hernan Solis-Matta[1], Lourdes Matta-Zamudio[1]

[1]*Faculty of Health Sciences, Universidad de Ciencias y Humanidades, 15314, Lima-Perú*

[2]*Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, 15314, Lima-Perú*

ARTICLE INFO

Article history:
Received: 03 August, 2020
Accepted: 08 December, 2020
Online: 16 December, 2020

Keywords:
Quality of care
Patient Satisfaction
Hospitalized Patients
Nursing Care
Hospital

ABSTRACT

***Background**: The quality of care of the nursing professional in relation to the hospitalized patient refers to the set of knowledge and attitudes that the nursing professional must provide quality care, preventing risks and improving patient satisfaction, for which its objective is to determine the quality of nursing care in hospitalized patients at the LanFranco la Hoz Hospital.* ***Method:** it is a quantitative, non-experimental, descriptive, and cross-sectional study with a study population of 143 patients in total hospitalized at the Carlos LanFranco la Hoz Hospital.* ***Results:** The results obtained in the quality of nursing care in hospitalized patients, it can be seeing that 86 patients representing 60.1%, are moderately satisfied, followed by 36 patients that represent 25.2% are satisfied, 12 patients that represent 8.4% are dissatisfied, 7 patients that represent 4.9% are very satisfied and finally 2 patients that represent 1.4% are very dissatisfied. In the dimension, experience of the nurse in the care of hospitalized patients, it can be observed 69 patients that represents 48.3% are moderately satisfied with the experience of the nurse in the care and in the dimension satisfaction with nursing care in hospitalized patients, it can be see 93 patients that represents 65% are satisfied with the quality of care.* ***Conclusions:** It is concluded that the high demand for patients sustains a decrease in quality nursing care due to the scarcity of human resources in the hospital and it is recommended that the hospital have a permanent evaluation system of quality of care to meet the expectations of hospitalized patients.*

1. Introduction

The World Health Organization (WHO) notes that 421 million people are hospitalized annually, these patients suffer approximately 42.7 million adverse events according to the most accurate evaluations and damage to patients during their hospital care are the cause of global mortality [1].

One of the factors of the quality of care is the perception that the hospitalized patient has during its hospital stay under care of hospital health personnel [2], since they fix their gaze on health personnel, easily judge and criticize the care provided, therefore, the position of the hospitalized patient will be evident when it comes to refusing care and advising that they should not be treated in that health facility [3].

Likewise, the care provided by the nurse towards patients has been perceived as deficient, because technology has influenced the nurse, where patient care is managed in a less humanized way, therefore the nurse-patient relationship loses every time trust [4].

*Corresponding Author: Brian Meneses-Claudio, Mr., +51 1 950159924 & bmeneses@uch.edu.pe

Likewise, the relationship of the family members of the hospitalized patient and the nurse who cares is essential for comprehensive health care, the adverse effects generated by negative environments for not providing good information timely and understandable for the family are negative factors for hospitalized patients [5].

The length of stay of hospitalized patients in the hospital is an impact on the accessibility of new patients entering the hospitalization service for the first time, generating demand both in the emergency area and hospitalization in the health sector [6].

In [7], the authors presented a study carried out in Jordan, it indicates that in the health area a topic of quality and safety is needed, this contributes to all personnel being trained responsibly, providing quality care and safety. The presence and permanence in health services are considered as a fundamental factor to promote the safety of hospitalized patients.

In [8], the authors mentioned a study carried out in Colombia, it was possible to find a total quality of care reached in 62%. The

dimension with the highest quality was " monitoring and follow-up" (73.2%) and the lowest was "anticipates" (56.1%). 42% of the patients said they had not been attended and did not recognize the nurse of the service.

In [9], the author described a study carried out in Russia, it refers that the perception, both of the health professional and the users, have a link in the human resource, which is ethical, the evaluations of hospitalized patients showed that are linked to the changeable, the concern of the hospital, and the obligation of public employees to provide quality care.

In [10], the authors presented a study carried out in Mexico, indicates that the quality of nursing professionals care towards patients highlighted 31patients that represents 44.0% reported that the nursing staff managed their work tools (baumanometer, stethoscope, secretion aspirator, among others) as they expected; 22 patients that represents 31.0%, better than they expected, and 32 patients that represents 30.0%, as they did not expect.

The objective of the study is to determine the quality of nursing care in hospitalized patients of the Carlos Lanfranco la Hoz hospital, in which it will allow to observe the quality of care that nursing professionals have towards hospitalized patients. This study is important since it will give relevant and real data about how the nursing care is provided to patients hospitalized for a disease.

In the study, the Spanish-version quality of care data collection instrument (CUCACE) was applied by Alonso R. and the original version The Newcastle Satisfaction with Nursing Scales (NSNS) by Thomas in Scotland has been demonstrated its usefulness to assess the quality of nursing care.

The data collection was processed through the survey of hospitalized patients of the Carlos LanFranco la Hoz Hospital, the data to be entered was carried out in a data matrix that will be designed in the statistical program IBM SPSS Statistics Base 25.0, it was proceeded to its corresponding analysis, in which it will allow a better data processing to make statistical tables and graphs so that can be described and interpreted in results and discussions, respectively.

The following research work is structured as follows: In section II, the development of the hospitalized patient data collection process will be presented, as well as the guidelines to consider so that they are within the research work. In section III, the results will show the quality of care presented by nursing professionals according to the specified dimensions of the instrument in the measurement of the variable. In section IV, it will present the discussions of the research work, in section V, the conclusions and in section VI the recommendations as well as the future work that is intended to be reached with the research work.

2. Methodology

In this part, the type and design of the research will be developed, as well as the population and the sample that will be carried out in the research work, the inclusion and exclusion criteria in detail and finally the technique and the instrument for collecting data.

2.1. Research type and design

The present study, due to its characteristics, the way of collecting data and measurement of the variables involved is a quantitative approach. Regarding the methodological design, it is a non-experimental, descriptive, cross-sectional study [11].

2.2. Population

The study worked with a total population of 143 hospitalized patients from the Carlos LanFranco la Hoz Hospital, which are distributed in 4 areas (Medicine, Surgery, Gynecology and Emergencies).

2.3. Inclusion criteria

- Adult patients with more than 2 days of hospitalization.
- Patients not receiving sedatives as treatment.
- Patients who have signed the informed consent.

2.4. Exclusion criteria

- Adult patients with less than 2 days of hospitalization.
- Patients with an altered state of consciousness, a language disorder, or a risk complication.
- Patient who has not signed the informed consent.

2.5. Technique and instrument

The data collection technique used in the research work was the survey, which important data will be obtained optimally and effectively. The survey will determine the quality of care in hospitalized patients using the (CUCACE) questionnaire, the Spanish version was validated by Alonso R. and the original version is The Newcastle Satisfaction with Nursing Scales (NSNS), validated by Thomas in Scotland. The CUCACE questionnaire has three sections: The first is the experience with nursing care that it has 26 items, the second is the satisfaction with the nursing care that it has 19 items and the third is the sociodemographic variable such as: age, sex, educational level, days of stay in the service and some general questions [2].

Regarding statistical validity, this was obtained using the Kaiser-Meyer-Olkin sample adequacy tests and Bartlett's sphericity test. The results of the sphericity test are significant while the sample adequacy test obtained a value of 0,897 (KMO> 0.6) and the results of the sphericity test obtained a favorable value (X^2 3702,347; g.l. = 990; p = 0,000). Likewise, Cronbach's alpha index obtained a value of 0.939 ($\alpha > 0,6$). All the above tests confirm an acceptable index of validity and reliability of the instrument.

This research work will consider the safety aspects of human research participants regarding to quality in hospitalized patients. The ethical behavior of deontological and moral health personnel. The bioethical principles: autonomy, justice, beneficence and non-maleficence, which constitute the basis for reasoning and orientation to their actions [12].

Medical ethics deals with the foundations that govern professional conduct in the practice of medicine, both in the doctor-patient and interprofessional relationships [13]. For this reason, the ethical authorization was carried out correctly when

coordinating with the hospital authority and with the patient itself and its family to carry out the research work.

Bioethical principles are interpreted into fundamental values and rights of hospitalized patients [14]. In the present research work, the bioethical principles of beneficence, non-maleficence, justice, and autonomy were mainly taken.

2.6. Place and Application of the Instrument

The questionnaire to measure the quality of nursing care in hospitalized patients of the Carlos Lanfranco la Hoz hospital, Lima - Peru was carried out on a programming date with the hospital, the time used in each service was approximately 25 minutes, in each service a document was requested for the hospitalized patient to participate in the research work (according to the criteria of inclusion and exclusion), concluding with good satisfaction when collecting the questionnaires from hospitalized patients since they provided a lot of support to carry out the research work.

It is important to emphasize the presence of relatives of some hospitalized patients at the time of completing the questionnaire, since while the family member has the decision in turn that the hospitalized patient can participate in the research, in which they provided support to complete the questionnaires. At the same time, the presence of the nursing staff, since the questionnaire is based on patient care, which allows to observe if the nursing staff adequately performs the appropriate and/or necessary care for the patient.

3. Results

Below is a summary table of the surveys carried out following the guidelines corresponding to the research work:

Table 1: Sociodemographic Data of Quality of Nursing Care in Hospitalized Patients of the Carlos Lanfranco La Hoz Hospital, 2019 (N = 143)

Participant Information	**Total**	
	N	**%**
Total	143	100
Gender		
Female	99	69,2
Male	44	30,8
Marital Status		
Single	37	25,9
Married	21	14,7
Separated	5	3,5
Cohabiting	79	55,2
Widow	1	0,7
Type of Family		
Nuclear	100	50,2
Extended	26	6,6
Expanded	7	11,1
Single parent	6	17,0
Reconstituted	3	6,2
Single Person	1	0,7
Study Degree		
No study	4	2,8
Incomplete primary	12	8,4
Complete primary	22	15,4
Incomplete secondary	33	23,1
Complete secondary	46	32,2
Incomplete University Studies	7	4,9
Complete University Studies	19	13,3
Occupation Condition		
Stable	14	9,8
Eventual	41	28,7
Without occupation	75	52,4
Retired	2	1,4
Student	5	3,5
Does not apply	6	4,2
Days of hospitalization		
Three days	51	35,5
Four days	54	37,8
Five or more days	38	26,6

In Table 1, it has the sociodemographic data of the study participants in total, there were 143 hospitalized patients from the Carlos Lanfranco La Hoz Hospital. Regarding gender, the female sex predominated with 99 patients representing 69.2%, followed by the male with 44 patients representing 30.8%.

Regarding the study degree, complete secondary school stands out with 46 patients representing 32.2%, followed by incomplete secondary school with 33 patients representing 23.1%, followed by complete primary school with 22 patients representing 15.4%, followed by complete university studies with 19 patients representing 13.3%, followed by incomplete primary school with 12 patients representing 8.4%, followed by incomplete university studies with 7 patients representing 4.9% and finally no study with 4 patients representing 2.8%.

Regarding the days of hospitalization, four days predominate with 54 patients representing 37.8%, followed by 51 days with 51 patients representing 35.5% and finally five or more days with 38 patients representing 26.6%.

Below are figures that show the results about the total quality of nursing care in hospitalized patients and its dimensions according to the data based on the research instrument:

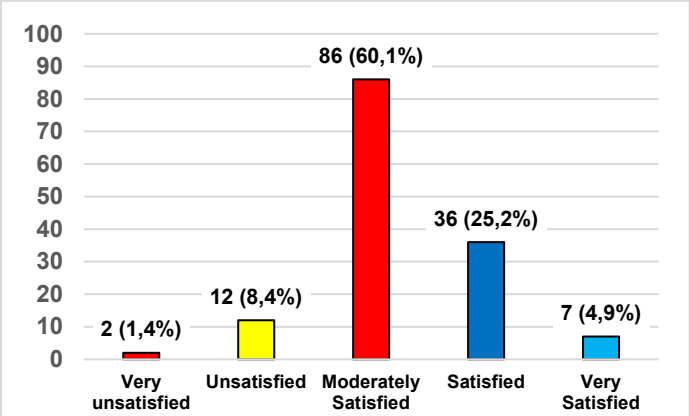


Figure 1: Quality of nursing care in hospitalized patients of the Carlos Lanfranco la Hoz Hospital, 2019 (N = 143)

In Figure 1, the quality of nursing care in hospitalized patients of the Carlos Lanfranco la Hoz Hospital, it can see that 86 patients representing 60.1% are moderately satisfied, followed by 36 patients representing 25.2% are satisfied, 12 patients representing 8.4% are unsatisfied, 7 patients representing 4.9% are very satisfied and finally 2 patients representing 1.4% are very unsatisfied.

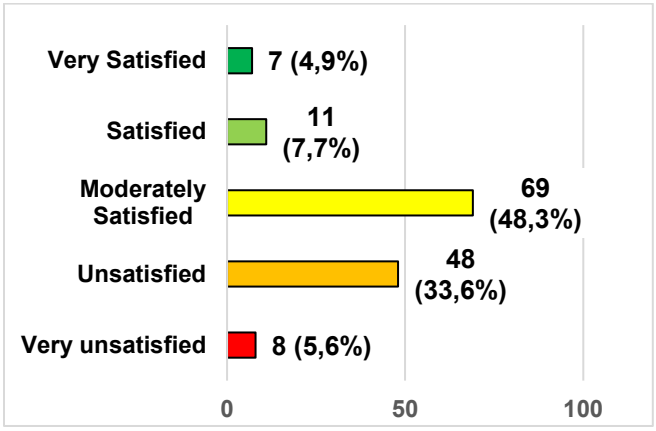


Figure 2: Quality of nursing care according to its dimension experience with nursing care, in inpatients in a public hospital in Lima Norte, 2019 (N = 143)

In Figure 2, the quality of nursing care according to its dimension experience with nursing care, in hospitalized patients of the Carlos Lanfranco La Hoz hospital, it can see that 69 patients representing 48.3% are moderately satisfied, followed by 48 patients representing 33.6% are unsatisfied, 11 patients representing 7.7% are satisfied, 8 patients representing 5.6% are very unsatisfied and finally 7 patients representing 4.9% are very satisfied.

In Figure 3, the quality of nursing care according to its dimension of satisfaction with nursing care in hospitalized patients of the Carlos Lanfranco La Hoz hospital, it can see that 93 patients representing 65% are satisfied, followed by 32 patients representing the 22.4% are moderately satisfied, 13 patients that represent 9.1% are very satisfied, 5 patients that represent 3.5% are unsatisfied and finally 0 patients that represent 0% are very unsatisfied.

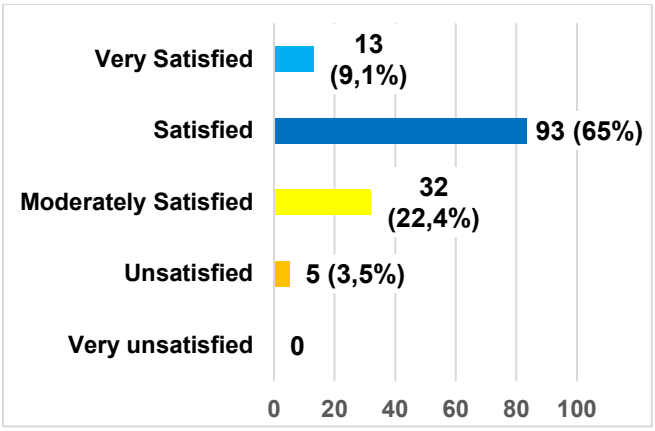


Figure 3. Quality of nursing care according to its dimension of satisfaction with nursing care, in hospitalized patients of the Carlos Lanfranco La Hoz hospital, 2019 (N = 143)

Knowledge of these results is important since it will allow to show if the nursing professional is adequately caring for hospitalized patients, in the same way, it will allow to show if the hospitalized patient in turn collaborates with the staff of nursing for its care during its hospital stay.

4. Discussion

The quality of nursing care constitutes a fundamental tool that values nursing care in hospitalized patients in all aspects and dimensions that are involved in the health of the person and the population.

We agree with what was indicated by Kol E. et al., who affirm that the fundamental objective of measuring the quality of health-related nursing care is to assess the care provided in health services [15].

Regarding the quality of nursing care in hospitalized patients of a public hospital, 60.1% predominantly are moderately satisfied, followed by 25.2% are satisfied, 8.4% are unsatisfied, 4.9% are very satisfied and finally 1.4% are very unsatisfied. These findings can be explained due to the perception that hospitalized patients have, during their stay in the hospital, to the care of health personnel since they fix their gaze and easily judge, criticize the care provided.

Our findings coincide with those of Weldetsadik A. et al., they determine that the satisfaction of hospitalized patients symbolizes the care provided, if it meets their expectations. This magnitude is not necessarily directly related to the degree of scientific-technical quality of the care provided [16].

The quality of nursing care is related to the dimension of experience with patient care mentioned by Amanda G. et al., they highlight the experience that hospitalized patients will have with the health professional, they are linked to the time that is provided during their hospital stay [17].

According to Becker R., he mentions that the wisdom of human resources is safe in medical practice, not only the care of patients is required and the attempt to provide quality care, the gift of expression is needed so that hospitalized patients can recover and reintegration into their family [18].

According to Allen A. et al., they point out the importance of technology as management support, but that it is not essential if there are trained or capable human resources. In many cases, a technological defect or the improper use of the equipment can lead to a failure in clinical analyzes that can lead to severe consequences for the life of hospitalized patients [19].

5. Conclusions

It is concluded that the use of a questionnaire CUCACE is important to evaluate the care received from hospitalized patients by the nursing professional, to evaluate the dimensions of experience and satisfaction with nursing care.

It is concluded that the high demand for patients sustains a decrease in quality nursing care due to the scarcity of human resources in the hospital.

It is concluded that the experience and work capacity of the nursing professional improves the quality and care of hospitalized patients.

It is concluded that it is essential to continue executing studies of this category, to analyze the outcome found and compare with current research, as well as facilitate its validity and reliability of CUCACE to measure the quality of nursing care in hospitalized patients.

6. Recommendations

It is recommended to carry out activities that promote communication with patients, since this interaction improves the nurse-patient relationship.

It is recommended that the hospital have a permanent quality of care evaluation system to meet the expectations of hospitalized patients.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Organización Mundial de la Salud, 10 Datos Sobre Seguridad Del Paciente., OMS, 2018.

[2] C. Lenis, F. Manrique, "Calidad del cuidado de enfermería percibida por pacientes hospitalizados.," Aquichan, **15**(3), 413–425, 2015, doi:10.5294/aqui.2015.15.3.9.

[3] P. Albujar, "Cuando el medico es el Paciente.," Fondo Editorial Comunicational, **32**(3), 1–164, 2015, doi:10.35663/amp.2015.323.103.

[4] P. Ramírez, C. Müggenburg, "Relaciones personales entre la enfermera y el paciente.," Enfermería Universitaria, **12**(3), 134–143, 2015, doi:10.1016/j.reu.2015.07.004.

[5] S. Geravandi, F. Soltani, M. Mohammadi, R. Alizadeh, A. Valipour, A. Hoseini, B. Rastegarimehr, A. Yari, A. Karimyan, A. Ghomeishi, "The effect of education on the nursing care quality of patients who are under mechanical ventilation in ICU ward.," Data in Brief, **16**, 822–827, 2018, doi:10.1016/j.dib.2017.11.090.

[6] C. Wang, M. Zhao, T. Zeng, "Progress in applying patient experience in nursing quality improvement.," Chinese Nursing Research, **4**(1), 1–4, 2017, doi:10.1016/j.cnre.2017.03.001.

[7] K. Al-Awamreh, M. Suliman, "Patients' satisfaction with the quality of nursing care in thalassemia units.," Applied Nursing Research, **47**, 46–51, 2019, doi:10.1016/j.apnr.2019.05.007.

[8] E. Amed, G. Villareal, C. Alvis, "Calidad del cuidado de Enfermería brindado por profesionales en una ciudad colombiana. Revista Ciencia y Cuidado[revista en Internet] 2019 [acceso 01 de julio de 2019]; 16(2): 108-119.," Revista Ciencia y Cuidado, **16**(2), 108–119, 2019, doi:10.22463/17949831.1615.

[9] V. Sarkisova, "Quality of nursing care, availability and accessibility of treatment and prevention of communicable and chronic diseases.," International Journal of Nursing Sciences, **4**(4), 350–351, 2017, doi:10.1016/j.ijnss.2017.10.016.

[10] M. Marín, C. Álvarez, J. Valenzuela, "Percepción de la atención de enfermería y satisfacción laboral en un Hospital de Seguridad Social.," RICSH Revista Iberoamericana de Las Ciencias Sociales y Humanísticas, **6**(11), 287, 2017, doi:10.23913/ricsh.v6i11.119.

[11] C. Fernández, P. Baptista, Metodología de la Investigación. 6ta ed. México: Mc Graw-Hill/Interamericana., 2015.

[12] O. Vera, "Aspectos bioéticos en la atención de los pacientes de las unidades de cuidados intensivos.," Revista Medica La Paz, **21**(1), 61–71, 2015.

[13] J. Martínez, C. Laucirica, E. Llanes, "La ética, la bioética y la investigación científica en salud, complementos de un único proceso.," Revista Médica Electrónica, **37**(4), 310–312, 2018.

[14] L. Mora, "Los principios éticos y bioéticos aplicados a la calidad de la atención en enfermería.," Revista Cubana de Oftalmología, **28**(2), 228–233, 2015.

[15] E. Kol, F. Arıkan, E. İlaslan, M. Akınci, M. Koçak, "A quality indicator for the evaluation of nursing care: determination of patient satisfaction and related factors at a university hospital in the Mediterranean Region in Turkey.," Collegian, **25**(1), 51–56, 2018, doi:10.1016/j.colegn.2017.03.006.

[16] A. Weldetsadik, T. Gishu, T. Mekonnen, A. Mekonnen, L. Girma, T. Demas, "Quality of nursing care and nurses' working environment in Ethiopia: Nurses' and physicians' perception.," International Journal of Africa Nursing Sciences, **10**, 131–135, 2019, doi:10.1016/j.ijans.2019.03.002.

[17] M. Trincado, E. Fernández, "Calidad en enfermería.," Rev Cubana Enfermer, **11**(1), 1–5, 1995.

[18] R. Becker, "Recordando a Sir William Osler a 100 años de su fallecimiento: ¿qué podemos aprender de su legado?.," An Fac Med, **76**(1), 71–6, 2015, doi:10.15381/anales.v76i1.11078.

[19] A. Allen, J. Robinson, M. Stewart, "Healthcare Quality: A Concept Analysis.," Nursing Forum, **52**(4), 377–386, 2017, doi:10.1111/nuf.12207.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1340-1344 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Quality of Life in Patients with Type 2 Diabetes of the Central Hospital of the Peruvian Air Force, 2019

Jared Zavala-Izaguirre[1], Fanny Mego-Llanos[1], Sarita Cornejo-Quispitongo[1], Brian Meneses-Claudio[2,*], Hernan Solis-Matta[1], Lourdes Matta-Zamudio[1]

[1]*Faculty of Health Sciences, Universidad de Ciencias y Humanidades, 15314, Lima-Perú*

[2]*Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, 15314, Lima-Perú*

ARTICLE INFO

Article history:
Received: 03 August, 2020
Accepted: 05 December, 2020
Online: 16 December, 2020

Keywords:
Quality of life
Type 2 diabetes
Diabetes control
Social burden
Sexual functioning

ABSTRACT

This research shows the study carried out on the Health Related to Quality of Life (HRQL) focuses on aspects related to the perception of health experienced and declared by the person, in different dimensions such as physical, mental, social, general perception of health and satisfaction achieved measured at different levels; the objective of the study is to determine the quality of life in patients with type 2 diabetes at the Central Hospital of the Peruvian Air Force, 2019. To determine the validity of the instruments, the Kaiser-Meyer-Olkin (KMO) sample adequacy index was used and the Bartlett sphericity test, which result in a significance value of 0.005, being an acceptable validity of both instruments. Among the most relevant results is that the following dimensions: Energy and Mobility, 147 patients representing 73.5% have low level. Likewise, regarding to Social Burden, 156 patients representing 78.0%, have a low level. Lastly, regarding the Sexual Functioning dimension, 178 patients representing 89.0%, have a low level; patients significantly affect quality of life. "Non-adherence to drug treatment" represents 159 patients, of which 115 who represent 72.3% have a low level with respect to their quality of life, of 44 patients who represent 27.7% have a high level regarding their quality of life; "Adherence to pharmacological treatment" 41 patients representing 92.7% have a low level regarding their quality of life and 3 patients who represent 7.3% have a high level regarding their quality of life.

1. Introduction

According to the International Diabetes Federation, 382 million adults worldwide had diabetes and 5.1 million died from the disease. About half (48%) of those who died were under the 60 years old. Three of four people living with diabetes (352 million) are in an active age (that is, between 20 and 64 years old) this number is projected to increase to 417 million by 2030 and 486 million by 2045 [1].

They also estimate that by 2035, if current trends continue, the global prevalence of diabetes will rise to 592 million, and the World Health Organization predicts that by 2030 it will become the seventh leading cause of death in the world [2].

This disease is considered as one of the main reasons for deaths worldwide, due to the poor quality of life without measuring the complications that it will bring them in the long term [3].

This non-communicable pathology is a worldwide concern, since due to the multiple organic complications it causes deaths, mostly related and complex with the organs of the cardiovascular system [4].

Diabetes is a chronic disease that is released when the pancreas does not produce insulin necessary for the body (a hormone that regulates the level of sugar, or glucose in the blood) or when the body cannot vigorously use the insulin it produces; they have a substantially increased risk of cardiovascular disease. It has also been related to hypertension and hyperlipidemia, it has vascular complications and is considered one of the four non-communicable diseases [5], [6].

Its own symptoms such as thirst, polyuria, blurred vision, weight loss, and sometimes polyphagia is recurrent. Perennially, the symptoms are not severe or may be distant; consequently,

*Corresponding Author: Brian Meneses-Claudio, Mr., +51 1 950159924 & Email: bmeneses@uch.edu.pe

hyperglycemia can provoke functional and pathological changes for a long time before diagnosis [7].

Diabetes is a chronic disease, and the associated stress of long-term symptoms, medical appointments, research, and daily treatments can affect sexual function through a person's mood, anxiety levels, fatigue, and well-being. In general, people with diabetes may have comorbidity of hypertension, hyperlipidemia, cardiovascular disease, or other endocrine dysfunction, and their associated treatments, may affect sexual function [8].

In [9], these problems cannot be addressed without a solid medical and scientific base that allows local solutions to be identified and proposed; unfortunately, the scientific production on diabetes in Peru is limited with only 81 scientific articles published in Web of Science and Scopus between 1996 and 2015.

In [10], diabetes was positioned as the fifth cause of death with 10 528 (7.11%) of the cases, in addition, the numbers of patients with this disease who died from causes of kidney complications were from 2 266 cases (1,191 men and 1,075 women) in 2011 to 3,190 cases (1,638 men and 1,552 women) in 2016.

In [11], this study shows that the mean time of evolution of the disease was 11.37 years old. Quality of life was self-perceived as deficient in 67.57% of older adults at the expense of satisfaction with their health (75.68%).

In [12], it is believed that, in Peru, more than 2 million people suffer from diabetes. The national average is 8% and in Lima, up to 10%. However, it is good to clarify that, in many regions, up to at least 50% of the population does not know if they have diabetes.

According to the [13], 4.0% of the population residing on the Coast reported having diabetes mellitus, being the urban area with the highest incidence of this disease (4.1%) and in the rural area only 2.7%; followed by the jungle regions with 1.9%, the urban area registers 2.7% and rural 1.1%; while the highlands area register 1.6 %, urban area 2.5% and rural with 1.0%.

The objective in this study is to determine the quality of life of patients with diabetes mellitus who come to the consultation of the Central Hospital of the Peruvian Air Force, using the Diabetes questionnaire 39.

2. Methodology

2.1. Research Design Approach

The present study has a quantitative, descriptive, and cross-sectional approach. A survey was carried out as a data collection technique, by applying a questionnaire as a documentary instrument. The study population was made up of patients with diabetes mellitus from the Central Hospital of the Peruvian Air Force [14].

2.2. Inclusion Criteria

All continuous patients over 18 years old with type 2 diabetes mellitus, regardless of the sex treated at the Central Hospital of the Peruvian Air Force.

2.3. Exclusion Criteria

- Patients with type 1 diabetes mellitus or gestational diabetes.
- Patients seen for the first time.
- Resigned with intellectual inability to recognize the question, and those who refuse to report the study.

2.4. Measurement Techniques and Instruments

Data collection technique: The technique used for this study was the survey, it is not the same as the questionnaire, however, the questionnaires are usually used in a survey, which requires a systematic way of collecting the data for the later to be able to analyze them statistically and obtain conclusions about the study population [15].

To measure the quality of life of patients diagnosed with type 2 diabetes mellitus, it will be applied with the diabetes 39 instrument, which includes 39 items that evaluate five domains of the patient's life: energy and mobility (15 items), control of diabetes (12 items), anxiety and worry (4 items), social burden (5 items) and sexual functioning (3 items). Each item is calculated from the evaluation made by the patient itself, in relation to its quality of life, how much was affected during the last month by action or activity that each item expresses- placing an "X" on a scale that is presented as a continuous line, with vertical marks that delimit spaces in which the numbers from 1 to 7 are identified, in which, on the value scales, the number 1 represents the quality of life that was not affected at all, and number 7, extremely affected. Two final items are incorporated into the five domains that rate the patient's perception about their quality of life in general, with a range from 1 (minimum) to 7 (maximum), and the severity of diabetes mellitus with a 1 (no severity) to 7 (extremely serious) [16].

The data collected was entered into a Microsoft Excel 2013 data matrix, to be later exported to the statistical program IBM SPSS Statistics Base 24.0.

Figure 1: Endocrinology area for conducting surveys at the Central Hospital of the Peruvian Air Force.

3. Results

To determine the validity of the instrument, the Kaiser-Meyer-Elkin sample adequacy index (KMO) and the Bartlett's test of sphericity were used. For the Quality-of-Life Test for patients with type 2 diabetes, a sample adequacy coefficient of 0.834 and a significance value of 0.005 (x^2 = 2633,620; gl = 741; p <0.05) were obtained in the test of Bartlett's sphericity.

The reliability of the instrument was determined based on the Cronbach's Alpha statistic. For the instrument Quality of life test

for patients with type 2 diabetes, it was determined that it presents an internal consistency index of 0.741 (α> 0.6); while the Morisky-Gree test instrument of adherence to pharmacological treatment has a coefficient of 0.591 (α <0.6). Therefore, a high level of general reliability of the Quality-of-Life Test is identified for patients with type 2 diabetes.

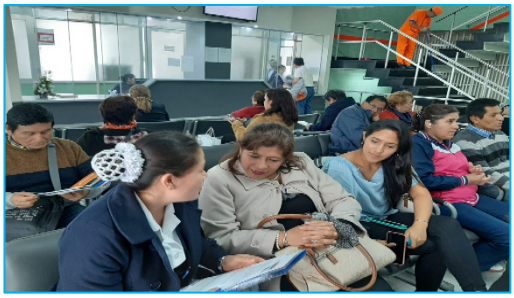


Figure 2: Application of the Survey in the Central Hospital of the Peruvian Air Force

3.1. Sociodemographic characteristics

Regarding the age of the study participants, it was determined that they fluctuate in an interval of 20 to 90 years, with an average of 56.45 (± 13,847) years old. Regarding sex, 60% of the sample is male. Likewise, 72.1% of the sample lives with their parents.

Table 1: Quality of life in patients with type 2 diabetes of the Central Hospital of the Peruvian Air Force, 2019 (N = 200)

Quality of life	N	%
Low Level	153	76,5
High Level	47	23,5
Total	200	100,0

In Table 1, it can observe the quality of life in patients with type 2 diabetes of the Central Hospital of the Peruvian Air Force, where 153 patients representing 76.5% have a low level followed by 47 patients representing 23, 5% have a high level.

In Table 2, it can see the quality of life according to its dimensions, in patients with type 2 diabetes of the Central Hospital of the Peruvian Air Force, where the most affected dimension was sexual functioning with 178 patients representing 89% have low level followed by 22 patients representing 11% have high level.

Table 2: Quality of life in patients with type 2 diabetes according to its dimensions, from the Central Hospital of the Peruvian Air Force, 2019 (N = 200)

Dimensions		N	%
Energy-mobility	**Low Level**	147	73,5
	High Level	53	26,5
Diabetes control	**Low Level**	132	66,0
	High Level	68	34,0
Control of anxiety - concern	**Low Level**	85	42,5
	High Level	115	57,5
Social burden	**Low Level**	156	78,0
	High Level	44	22,0
Sexual Functioning	**Low Level**	178	89,0
	High Level	22	11,0

The importance of obtaining these results in continuing patients with diabetes mellitus is to be able to identify in which of the dimensions are most affected by this disease and to be able to act in a timely manner in the dimension and above all to teach how to be responsible and aware of the disease. So, they could learn that diabetes is a chronic disease that involves the variation of lifestyles. For its treatment, it is essential that people learn to manage it correctly to have a good quality of life and avoid possible complications.

Table 3: Contingency table between the variables Quality of life and Adherence to pharmacological treatment and chi-square tests

			Quality of life		Total
			Low Level	High Level	
Adherence to pharmacological treatment	**Non-adherent**	Count	115	44	159
		% within APT	72,3%	27,7%	100,0 %
	Adherent	Count	38	3	41
		% within APT	92,7%	7,3%	100,0 %
Total		Count	153	47	200
		% within APT	76,5%	23,5%	100,0 %
Pearson's chi-square			Valor X^2	g.l.	Sig. (p)
			7,513	1	0,006

APT = Adherence to Pharmacological Treatment

In Table 3; on adherence, it can observe in patients with type 2 diabetes of the Central Hospital of the Peruvian Air Force, where the "Non-adherence" represents 159 patients, 115 of which is equivalent to 72.3% have low level and 44 patients 27.7% with

high level; "Adherence" 41 patients representing 92.7% have low level and 3 patients representing 7.3% have high level.

It is important that people with diabetes and their families should receive education about adherence and non-adherence; ignorance is one of the reasons for not achieving the necessary benefits that medications can provide in their disease; diabetes is a chronic disease that affects different organs of the human body.

4. Discussions

Quality of Life is a representative variable that allows us to measure the impact it has on a certain pathology (in this case Diabetes Mellitus type II) in the person's life, which has an impact on various biological and psychosocial levels in the individual health.

The results obtained according to the Diabetes 39 questionnaire demonstrate that of a population made up of 200 diagnosed continuing patients who are treated at the Central Hospital of the Peruvian Air Force, the dimensions most affected are determined that, regarding the Energy and mobility dimension, 147 is equivalent to 73.5% have low level. Likewise, regarding to social burden, 156 is equivalent to 78.0%, have low level. Lastly, regarding the Sexual Functioning dimension, 178 is equivalent to 89.0% have low level.

These results coincide and are reinforced by the study carried out in [17]. They show that patients with Diabetes Mellitus regarding their Quality of Life, who were surveyed with 101 diabetics; the most affected domains in order of score were: energy and mobility (60) equal 59.4%, anxiety-concern (62) equals 61.3% and sexual functioning (66) equals 65.3%.

Regarding adherence, it can observe in patients with type 2 diabetes of the Central Hospital of the Peruvian Air Force, where non-adherence represents 159 patients, 115 representing 72.3% have low level and 44 patients representing 27.7% have high level; adherence 41 patients representing 92.7% have low level and 3 patients representing 7.3% have high level. No relationship was found between the percentage of adherence to treatment with sex, age, educational level, and employment status; it is the lack of awareness about its treatment.

According to [18], it was found, the low level was 21 (32.8%) and with high level adherence 24 (37.5%). No relationship was found between the percentage of adherence to treatment with sex, age, educational level, and employment status.

In the analysis of each dimension of the instrument, it was observed that there were significant differences. In the dimension between quality of life and adherence to treatment in patients with treatment, as a strength of the study, it can be highlighted that there was no rejection by the patients, all the questionnaires were correctly filled out, with no missing data or loss in the sample. It is also important to note that a widely used instrument with a good internal consistency index was used.

Among the limitations of the study, the selection of patients is found, since as it is not a probabilistic sampling, there may be the possibility of selection bias. For future studies on the subject in primary care, it would be interesting to carry out a random sampling that provides greater reliability in the results obtained. Another important aspect to control would be trying to match the sample by sex.

5. Conclusions

It is concluded that the dimensions that highly affect patients are Energy and Mobility, Social Burden, Sexual Functioning, having a low level of more than 70% of the sampled cases.

In addition, the results of therapeutic adherence show that 72.3% of patients present a low level with respect to their quality of life, therefore, compliance with treatment, physical activities and a balanced diet help diabetic patients a better recovery and a balanced quality of life.

6. Recommendation

A restructuring of promotional preventive intervention programs targeting such patients with type II diabetes is recommended. This intervention program must involve a multidisciplinary team of professionals, who see aspects related to the disease and the environment of the person and are accelerators for their recovery from physical, biological, and psychosocial health.

It is suggested an awareness program, educational talks for both the patient and family members for their support and control, home health personnel visit programs.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] R. Williams, Guía de incidencia política de la novena edición del atlas de la diabetes de la FID 2019, International Diabetes Federation, **9**, 1–28, 2019.

[2] G. Chen, A. Iezzi, J. McKie, M. Khan, J. Richardson, "Diabetes and quality of life: Comparing results from utility instruments and Diabetes-39.," Diabetes Research and Clinical Practice, **109**(2), 326–333, 2015, doi:10.1016/j.diabres.2015.05.011.

[3] C.F.D.C. and Prevention, Informe nacional de estadísticas de la diabetes, National Center for Chronic Disease, 1–33, 2017.

[4] Organización Panamericana de la Salud, El número de personas con diabetes en las Américas se triplicó desde 1980, 2016.

[5] P. Miguel, Y. Sarmiento, A. Mariño, Y. Llorente, T. Rodriguez, M. Peña, "Prevalencia de enfermedades crónicas no transmisibles y factores de riesgo en adultos mayores de Holguín," Revista Finlay, **7**(3), 155–167, 2017.

[6] R. Wang, P. Zhang, Z. Li, X. Lv, H. Cai, C. Gao, Y. Song, Y. Yu, B. Li, Y. Cui, "The prevalence of pre-diabetes and diabetes and their associated factors in Northeast China: a cross-sectional study," Scientific Reports, **9**(1), 1–8, 2019, doi:10.1038/s41598-019-39221-2.

[7] J. French, A. Godoy, ¿Cuál es la definición de diabetes? Criterios diagnósticos (prueba/s a realizar [glucemia plasmática en ayunas, sobrecarga, etc.]) y puntos de corte, Fundación RedGDPS, 1–3, 2015.

[8] E. Holloway, "Sexual problems in diabetes," Medicine (United Kingdom), **47**(2), 106–109, 2019, doi:10.1016/j.mpmed.2018.11.004.

[9] R. Carrillo, A. Bernabé, "Diabetes mellitus tipo II en Perú: una revisión sistemática sobre la prevalencia e incidencia en población general.," Rev Peru Med Exp Salud Publica, **36**(1), 36–36, 2019, doi:10.17843/rpmesp.2019.361.4027.26.

[10] J. Guerrero, L. Parra, J. Mendoza, "Autoeficacia y calidad de vida en pacientes con diabetes mellitus tipo 2 sometidos a hemodiálisis," Revista Cubana de Salud Publica, **42**(2), 193–203, 2016.

[11] D. Jiménez, P. Casado, R. Santos, D. Jiménez, G. Hernández, "Percepción de la calidad de vida en pacientes adultos mayores con diabetes mellitus tipo II," Revista Electrónica Medimay, **26**(1), 54–62, 2019.

[12] V. Arrunátegui, "Despistaje de diabetes mellitus tipo 2 en una población adulta urbana del distrito de Coishco, Ancash, Perú.," Revista Medica Herediana, **26**(3), 173, 2015, doi:10.20453/rnp.2015.2585.

[13] Instituto Nacional de Estadística e Informática, En el Perú 3 de cada 100

personas de 15 y más años reportan tener diabetes., INEI, 2015.
[14] C. Fernández, P. Baptista, Metodología de la Investigación. 6ta ed. México: Mc Graw-Hill/Interamericana., 2015.
[15] R. Ortega, R. Veloso, O. Hansen, "Percepción y actitudes hacia la investigación científica," Revista de Investigación En Ciencias Sociales y Humanidades, **5**(2), 101–109, 2018, doi:10.30545/academo.2018.jul-dic.2.
[16] J. López, R. Rodríguez, "Adaptación y validación del instrumento de calidad de vida Diabetes 39 en pacientes Mexicanos con diabetes mellitus tipo 2.," Salud Publica de Mexico, **48**(3), 200–211, 2006, doi:10.1590/S0036-36342006000300004.
[17] M. Brítez, E. Torres, "Calidad de vida en pacientes con diabetes mellitus tipo 2.," Rev. Nac (Itauguá), **9**(1), 78–91, 2017, doi:10.18004/rdn2017.0009.01.078-091.
[18] Y. Ramos, R. Morejón, Y. Cabrera, D. Herranz, W. Rodríguez, "Adherencia terapéutica, nivel de conocimientos de la enfermedad y autoestima en pacientes diabéticos tipo 2," Gaceta Médica Espirituana, **20**(3), 13–23, 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1345-1349 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Social skills and Resilience in Adolescents of Secondary Education of the Kumamoto I 3092 Educational Institution, of the Puente Piedra District – Lima 2019

Evelyn Roncal-Cespedes[1], Gloria Castillo-Laban[1], Brian Meneses-Claudio[2,*], Hernan Matta-Solis[1], Lourdes Matta-Zamudio[1], Eduardo Matta-Solis[3]

[1]*Faculty of Health Sciences, Universidad de Ciencias y Humanidades, 15314, Lima-Perú*

[2]*Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, 15314, Lima-Perú*

[3]*Health Sciences, Instituto Peruano de Salud Familiar, 15304, Lima-Perú*

ARTICLE INFO

Article history:
Received: 03 August, 2020
Accepted: 05 December, 2020
Online: 16 December, 2020

Keywords:
Social skills
Psychological resilience
Adolescents
Students

ABSTRACT

Social skills are the behaviors that the individual adopts on a regular basis at specific moments in everyday life. Resilience is the way in which the individual faces the adversity that life imposes on them, as well as manages to overcome it. In adolescence, problems are perceived because important changes are generated in an accelerated way, many adolescents develop psychological disorders or alterations that could compromise their own lives, if not detected in time, therefore, the objective of this study is to determine the social skills and resilience in adolescents of secondary education of the Kumamoto Educational Institution I 3092, of the Puente Piedra district - Lima, 2019. Applying a quantitative, non-experimental, descriptive, and correlational study, with a population of 626 adolescents from the Kumamoto I 3092 educational institution in the Puente Piedra district, who answered a questionnaire with sociodemographic data and the instruments of the social skills scale and the Connor-Davidson questionnaire for resilience. In the results, the predominance of the medium level in both variables, in social skills with 87.5%, in resilience with 58.3%, concluding the need to carry out interventions in adolescent students, to prevent the inadequate development of their personality.

1. Introduction

The social skill or assertive behavior is a set of expressions that is manifested in a verbal and non-verbal way, where the person externalizes in an interpersonal way all his deficiencies, emotions, ideas, favoritism, rights, avoiding anxiety and hostility [1], [2]**.** Resilience is the ability that a subject or a group obtains that projects itself into the future despite difficult situations and on some occasions can generate trauma and serious consequences in its own life [3].

In adolescence, internal problems such as anxiety disorders, depression, post-traumatic disorders, and even obsessive-compulsive disorders can be perceived. They may also perceive external problems associated with abrupt behavior, conflicts with their environment and aggressive behavior [4]. Family problems can have a strong impact on the adolescent's life, which is why it is important that the family manages its problems correctly during this period, since otherwise this problem will affect the adolescent, diminishing his or her abilities in the academic aspect in society and in a personal way [5].

Worldwide, social skills are specifically oriented to prevent risk behavior and promote a healthy lifestyle [6]. Skills education has emerged since the increase of the problems that affects the community, but directly to adolescents as there are several events such as diseases generated through sex, drug addiction, bullying and even suicide [7].

According to the Pan American Health Organization, it promotes healthy lifestyle and the development of the adolescent stage in the Caribbean and Latin America. Regarding resilience, it establishes support for adolescents based on strategies to reduce various risk behaviors. This process will be continuous, where

*Corresponding Author: Brian Meneses-Claudio, Mr., +51 1 950159924 & bmeneses@uch.edu.pe

adolescents will use adaptation skills to seek adequate development within society, the family, school and culture [8].

PAHO establishes that social skills require the adoption of attitudes that are indispensable for optimal behavior to achieve success in the various challenges that life imposes. This favors healthy physical, social and mental behaviors, where adolescents have various possibilities for satisfying their needs, developing their aspirations and coping their environment [9].

In [10], un studio entre 170 adolescents between 15 to 17 years old participated in this study, to whom the social skills questionnaire was applied, where the predominance of the medium level is seen with 50.6%, continuing the low level with 38.2%, then the high level with 11.2%. In the explanation of the dimensions, it is observed that the self-expression factors in social situations predominate the low level with 40.6%, making requests 59.4%, while the high level corresponds to the expression of anger and disagreement with 45.3%. Regarding its study, it mentions the importance of developing social skills in an adequate way to reduce the risk of developing antisocial behaviors.

In [11], it developed the study with a population of 49 adolescents, where the prevalence of medium level resilience was found with 51%, then the low level with 43% and a high level of 6%. In its study, it was shown that adolescents have a medium capacity of resistance to remake and try to escape complex situations.

In [12], the present study had a population of 119 adolescents, where the results show that 71.75% is equivalent to the medium level, in their study, they determined to resort to development programs for teachers in order to include social skills topics in the training of adolescents.

In [13], they indicate that their total population was 160 adolescents, of whom the prevalence of medium-level resilience was obtained with 52.5%, continuing the high level with 46.9% and the low level of 0.6%. Regarding its control and purpose dimensions, they obtained a high level of 61.9%, while control under pressure had a medium level with 59.4% and spirituality a low level of 8.1%. Concluding that the present work will be use as reference to suggest precise interventions where the beneficiary will be the adolescent.

The main objective of this research study is to determine the relationship of social skills and resilience in adolescents because there is a deficit in research on this enigmatic topic in adolescence.

This study presents a quantitative approach, it will be carried out without manipulating the variable, which is why it is a non-experimental and descriptive cross-sectional design. The technique to be used in the research work will be the survey, since it is a measurement instrument which the data will be collected appropriately in a reliable, objective, and valid way.

The Elena Gismero Gonzales, E.H.S questionnaire includes 33 items, 28 of which are written in the sense of lack of assertion or deficit in social skills and 5 of them in the positive sense. The Connor –Davidson Questionnaire (CD-RISC) This test was created in 2003 and consists of two fundamental versions, one of 25 items and the other of 10, in which to carry out the research work the 25-item version was used. All the data collected was entered into the matrix designed by the statistical software SPSS 24.0., This procedure was carried out cautiously to not have errors and avoid values being lost.

This study is important because it will show us the capacity that adolescents have in the face of stimuli of changes in their personality and in their environment, in which it allows the adolescent to improve their capacity for understanding and adaptation both on a social, personal, and family level.

2. Methodology

Starting by making a request and receiving the authorization of the director of the educational institution, it was always established that ethical aspects were applied during the process, respectively, the students voluntarily answered the questionnaires. It is worth mentioning that these questionnaires were provided by the researchers, in the different classrooms, in the presence of the teachers from each grade. The descriptive cross-sectional and correlational methodological design was chosen. Once the database was assembled, the descriptive analysis began, where the percentages achieved, as well as its dimensions, are emphasized by the contingency table that will later be reflected in the bar charts. The stages of data collection are shown in Figure 1 where the processes by which the students were part are shown.

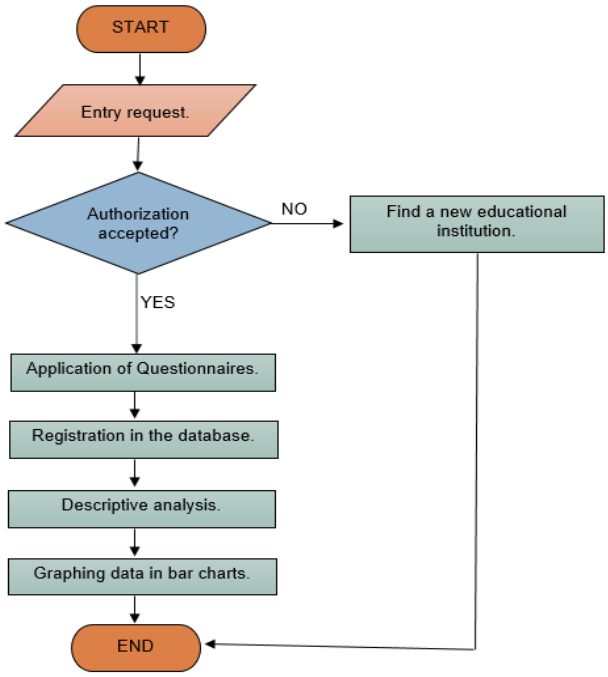


Figure 1: Flowchart for data collection

1.1. Population and sample

For this research, a vulnerable population was selected from a state school in the Puente Piedra district, where 626 students participated, they are the total population, ranging in age from 11 to 18 years old, who are in 1st to 5th grade of secondary education.

1.2. Type of study

In the present research work, due to its characteristics, data collection and the measurement of the variables involved is a quantitative approach. Regarding the methodological design, it is a non-experimental, descriptive, and correlational study.

Table 1: Distribution of classrooms at the secondary level of the Kumamoto I 3092 educational institution

Grade	Shifts	Classrooms	Total Students
1st	Afternoon	5	145
2nd	Afternoon	5	135
3rd	Afternoon	6	133
4th	Afternoon	4	103
5th	Afternoon	4	110
TOTAL		24	626

1.3. Inclusion and exclusion criteria

Inclusion:

- Adolescents who study in 1st to 5th grade of high school.
- Adolescents who have signed the informed consent.
- Parents who have signed the informed consent for the application of the survey in the adolescent.

Exclusion:

- Adolescents and parents who have missed classes and who have not signed the informed consent.
- Parents who have not signed the informed consent.

1.4. Analysis of the variable

The following study has two main variables, social skills, and resilience in adolescents, both variables according to their nature, are qualitative or categorical, and their measurement scale is ordinal.

The technique for data collection in the educational institution was the survey, which is very useful to collect data appropriately in a reliable, objective and valid way [14].

To assess social skills, the instrument of Elena Gismero Gonzales E.H.S was used, which comprises 33 items, 28 of which are written in the sense of lack of assertion or deficit in social skills and 5 of them in the positive sense. The final factor analysis has revealed 6 factors: self-expression in social situations, defense of their own rights as a consumer, expression of anger or disagreement, saying no and cutting interactions, making requests and initiating positive interactions with the opposite sex [15]. This instrument has been valid in different Latin American countries and in other countries, as well as in Peru that has construct validity, the meaning attributed to the measured construct (assertion) is correct. Regarding statistical validity, it was obtained using the Kaiser-Meyer-Olkin (KMO) obtained a sample adequacy coefficient of 0.923 and the Bartlett's Sphericity Test of 0,000 (x^2 = 4036,183; g.l. = 300; $p < 0,05$). Regarding the reliability of the instrument, it was determined based on the Cronbach's Alpha statistic. For the Social Skills Scale instrument, it was determined that it presents an internal consistency index of 0.805 ($\alpha>0.6$).

The Connor –Davidson questionnaire (CD-RISC) was used to assess resilience. This test was created in 2003 and consists of two fundamental versions, one of 25 items and the other of 10 in which the 25-item version was used to carry out the research work. The construct consists of five factors: persistence-tenacity-self-efficacy (items: 10-12, 16, 17, 23-25); control under pressure (6, 7, 14, 15, 18, 19, 20); adaptability and support networks (1, 2, 4, 5, 8); control and purpose (13, 21, 22) and spirituality (3, 9). The scores for each item are added together and it is interpreted that the higher the score in each dimension, the more indicators of resilience the individual shows. Measurement scale 0 - 33 Low resilience, 34 - 66 Medium resilience, 67 - 100 High resilience [16][17]. The Spanish translation was carried out by two members of the research team and supervised by a bilingual person with training in Psychology. Regarding the statistical reliability of the data collection instrument, a coefficient of 0.882 ($\alpha>0.6$) was obtained.

All the collected data will be entered into the matrix designed by the statistical software SPSS 24.0., this procedure was carried out cautiously to not have errors and avoid the values being lost.

1.5. Place and Application of the Instrument

To start the process, the followed steps were taken, making a request document attaching the letter of presentation provided by the nursing professional school of the Universidad de Ciencias y Humanidades, addressed to the director of the educational institution Kumamoto I 3092, located in the department de Lima, Puente Piedra district, in the AA. HH Laderas de Chillón, to access the permit and be able to apply the questionnaires formally, also it coordinated a meeting with the director, in the interview the authors explained in detail about the importance of conducting such research, which was accepted for the director and allowed the study to be carried out.

On July 19, 2019, the informed consent and assent was delivered to all students from the 1st to the 5th grade of secondary education, on July 22 the documents were collected verifying the correct filling, then the authors proceeded with the delivery of the surveys. The time elapsed in each classroom was 25 minutes. They were informed of the importance of their participation, in a brief explanation of the instrument, also it was mentioned that the confidentiality of their responses would be respected, the doubts they presented were also clarified. The students voluntarily answered the questionnaire provided. At the end of the application of the instrument, each of the surveys was verified with the correct completion and they were thanked for their participation in the research.

3. Results

After applying the surveys, the data was processed in a statistical program, resulting in the summary tables presented below.

Table 2 shows the results of the gender of the high school adolescents, where 306 (48.9%) of the adolescents are female and 320 (51.5%) of the adolescents are male.

Table 3 shows the mean age of secondary school adolescents, where the minimum age of adolescents is 11 years and the

maximum age of adolescents is 18 years, therefore the average age of adolescents in primary education is 14 years old.

Table 2. Gender of secondary education adolescents from the Kumamoto I 3092 educational institution, from the Puente Piedra district - Lima, 2019 (N = 626)

Gender of the adolescents			
		Frequency	**Percentage**
Valid	Female	306	48,9
	Male	320	51,1
	Total	626	100,0

Table 3. Age of secondary education adolescents from the Kumamoto I 3092 educational institution, from the Puente Piedra district - Lima, 2019 (N = 626)

Adolescents Age					
	N	**Minimum**	**Maximum**	**Mean**	**SD**
Adolescent Age	626	11	18	14,33	1,485
Valid N (per list)	626				

Table 4: Social Skills in adolescents of Secondary Education of The Kumamoto I 3092 Educational Institution, of The Puente Piedra District - Lima, 2019 (N = 626)

		N	**%**
Social Skills	**Low Level**	76	12,1
	Medium Level	548	87,5
	High Level	2	0,3
Total		626	100,0

In Table 4, the social skills in the adolescents of secondary education of the Kumamoto I 3092Educational Institution, of the Puente Piedra district - Lima, 548 adolescents that represent 87.5% have a medium level, followed by 76 adolescents representing 12.1% have a low level and finally 2 adolescents representing 0.3% have a high level.

Table 5: Resilience in adolescents of Secondary Education of The Kumamoto I 3092 Educational Institution, of The Puente Piedra District - Lima, 2019 (N = 626)

		N	**%**
Resilience	**Low Level**	24	3,8
	Medium Level	365	58,3
	High Level	237	37,9
Total		626	100,0

In Table 5, it can see the resilience of adolescents in secondary education at the Kumamoto I 3092 Educational Institution, in the Puente Piedra district - Lima, where 365 adolescents representing 58.3% have a medium level, 237 adolescents representing 37.9% have a high level and finally 24 adolescents representing 3.8% have a low level.

Table 6: Social Skills and Resilience in adolescents of Secondary Education of The Kumamoto I 3092 Educational Institution, of The Puente Piedra District - Lima, 2019 (N = 626)

			Social Skills			**Total**
			Low Level	**Medium Level**	**High Level**	
Resilience	**Low level**	Count	4	20	0	24
		% Resilience	16,7%	83,3%	0,0%	100,0%
	Medium Level	Count	34	330	1	365
		% Resilience	9,3%	90,4%	0,3%	100,0%
	High Level	Count	38	198	1	237
		% Resilience	16,0%	83,5%	0,4%	100,0%
Total		Count	76	548	2	626
		% Resilience	12,1%	87,5%	0,3%	100,0%
Pearson's chi-square			X^2 Value	g.l.		Sig. (p)
			6,767	4		0,149

In Table 6, the relationship between the variables Social Skills and Resilience was determined based on the Pearson's Chi square (X^2) statistical test. The significance level of the test obtained a value of 0.149 ($p>0.05$) ($X^2 = 6.767$; g.l. = 4). Therefore, the dissociation hypothesis cannot be rejected, for which there is no statistical evidence of a significant relationship between social skills and resilience of a sample of students from the Kumamoto I 3092 Educational Institution, in the Puente Piedra District - Lima. 2019.

4. Discussion

In this research study, the issue of social skills and resilience was raised, promoting adolescent mental health, therefore, it seeks to collaborate with educational institutions that are linked to education as part of the global training of adolescents, where they must incorporate plans or programs that allow the adolescent to strengthen and improve all the capacity to interrelate and be able to face better in daily life and social aspect.

Regarding social skills in adolescents at the secondary level of the Kumamoto I 3092 Educational Institution, in the Puente Piedra district - Lima, the medium level stands out, then the low and finally the high. This study is similar to [10], in which they found medium scores of 50.6% as the predominant average, however, it differs in terms of dimensions, the lowest level most affected is making requests with 59.4% and in our study shows a low score in the expression of anger or disagreement with 34.3%. At the same time, in [12], it presents a percentage of 44.54%, being the predominant medium level, for this reason we agree with the proposal of resorting to development programs for educators and this allows covering topics of social skills within the training of students.

Regarding resilience, it was evident that the medium level predominates, a similar value was found in the study [11], where a greater proportion of the medium level of resilience is evident, represented by 51% followed by 43% of low level and 6% of the high level. At the same time, it shows the average capacity of adolescents to cope with complicated situations. Similarly, in [13], the author determine that the predominant level of resilience was

the medium, and also proposes to carry out interventions with a multidisciplinary vision to comprehensively address adolescents, in the same way, it is evidenced in the study [18], who argue that the resilience in the participants was 17.5% low resilience, 67.1% medium resilience and 15.4% high resilience.

The research study shows that the results obtained indicate the need to carry out an intervention to the adolescent population with social skills and medium-level resilience, otherwise they would present an inadequate development in their personality.

5. Conclusions

Eighty-seven-point five percent of the study participants reported to have medium level of social skills, twelve point one, medium level and cero point three, high level. Regarding social skills according to their dimensions, where the most affected dimension was the expression of anger or disagreement with 215 representing thirty-four-point three percent of adolescents. All students from 1st to 6th grade of secondary level presented medium level resilience with fifty-eight-point three percent, high level with thirty-seven-point nine percent and low level with three-point eight percent.

Regarding resilience according to its dimensions, the most affected dimension was spirituality with 132 representing twenty-one-point one percent of adolescents.

As future work, it is expected that health professionals will have a greater interest in conducting more research, in vulnerable locations, that cover the adolescent stage in a holistic way, physically, emotionally, and socially, where adolescents can develop in an intrapersonal or interpersonal context.

The limitation in the present research work is that some adolescents were not present at the time of conducting the surveys also the interpretation of some questions, so nurse professionals were taking the survey to help school students understanding the questions.

Likewise, educators must assume the responsibility they have regarding the adolescent's physical, affective, psychological-cognitive, social development, guiding them to increase their activities to contribute to the management of good self-esteem, social skills, resilience, among others.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

[1] E. Gismero, Escala de Habilidades Sociales - EHS. Manual. 3a ed. Madrid - España: TEA- Ediciones; 2010. 56 p.

[2] E. Flores, M. Garcia, W. Calsina, A. Yapuchura, "Las Habilidades sociales y la comunicación interpersonal de los estudiantes de la Universidad Nacional del Altiplano – Puno.," Comunicacción, **7**(2), 2016.

[3] J. García del Castillo, Á. García del Castillo, C. López, P. Dias, "Conceptualización teórica de la resiliencia psicosocial y su relación con la salud.," Health and Addictions/Salud y Drogas, **16**(1), 59–68, 2016, doi:10.21134/haaj.v16i1.263.

[4] R. Vera, F. Raquel, L. Tobar, V. María, V. Delgado, "Habilidades Sociales.," Revista Salud y Ciencias, **1**(2), 08–15, 2017.

[5] C. Salavera, P. Usán, P. Teruel, "Contextual problems, emotional intelligence and social skills in Secondary Education students. Gender differences.," Annales Médico-Psychologiques, Revue Psychiatrique., **177**(3), 223–230, 2019.

[6] B. Gunn, E. Feil, J. Seeley, H. Severson, H. Walker, "Promoting School Success: Developing Social Skills and Early Literacy in Head Start Classrooms," NHSA Dialog, **9**(1), 1–11, 2006, doi:10.1207/s19309325nhsa0901_2.

[7] M. Aparicio, Habilidades para la vida. Manual de conceptos básicos para facilitadores y educadores., CEDRO, 1–38, 2013.

[8] Organización Panamerica de la Salud, Manual de identificación y promoción de la resiliencia en niños y adolescentes., 1–90, 1998.

[9] Ministerio de Salud, "Guía técnica de gestión de promoción de la Salud en Instituciones educativas para el Desarrollo Sostenible," 46, 2011.

[10] M. Quispe, "Habilidades sociales y consumo de alcohol en adolescentes de un Colegio Nacional de Lima.," Revista Ciencia y Arte de Enfermería, **2**(2), 43–49, 2017, doi:10.24314/rcae.2017.v2n2.08.

[11] L. Badaracco, "Factores de resiliencia y riesgo suicida en adolescentes en estado de abandono.," Temática Psicológica, **9**(9), 35–42, 2013, doi:10.33539/tematpsicol.2013.n9.838.

[12] A. Gonzales, P. Quispe, "Habilidades sociales y rendimiento Académico de los estudiantes de la Facultad de Ciencias de la Educación - Universidad Nacional del Altiplano - Puno Perú.," Revista de Investigaciones Altoandinas - Journal of High Andean Research, **18**(3), 331–336, 2016, doi:10.18271/ria.2016.222.

[13] E. Choque, H. Matta, "Nivel de resiliencia y funcionalidad familiar en adolescentes de un albergue de Ventanilla , Perú.," Ágora Rev. Cient., **5**(2), 1–7, 2018.

[14] C. Fernández, P. Baptista, Metodología de la Investigación. 6ta ed. México: Mc Graw-Hill/Interamericana., 2015.

[15] H. As, E. Huaire, "Desarrollo de habilidades sociales en contextos universitarios," Horizonte de La Ciencia, **8**(14), 123, 2018, doi:10.26490/uncp.horizonteciencia.2018.14.430.

[16] M. Serrano, M. Garrido, B. Notario, R. Bartolomé, M. Solera, V. Martínez, "Validity of the Connor-Davidson resilience scale (10 items) in a population of elderly | Validez de la escala de resiliencia de Connor-Davidson(10 ítems) en una población de mayores no institucionalizados.," Enfermeria Clinica, **23**(1), 2020, 2013, doi:10.1016/j.enfcli.2012.11.006.

[17] M. Cresp, V. Fernandez, C. Soberon, "Adaptación española de la escala de resiliencia de connor-davidson (cd-risc) en situaciones de estrés crónico," Behavioral Psychology / Psicología Conductual, **22**(2), 219–238, 2014.

[18] R. Singh, S. Mahato, B. Singh, J. Thapa, D. Gartland, "Resilience in nepalese adolescents: Socio- demographic factors associated with low resilience.," Journal of Multidisciplinary Healthcare, **12**, 893–902, 2019, doi:10.2147/JMDH.S226011.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1350-1355 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Social Skills and Resilience in Adolescents of an Educational Institution in North Lima, 2019

Lili Sifuentes-Gomez[1], Doris Vega-Davila[1], Betty Flores-Paz[1], Brian Meneses-Claudio[2,*], Hernan Matta-Solis[1], Eduardo Matta-Solis[3]

[1]*Faculty of Health Sciences, Universidad de Ciencias y Humanidades, 15314, Lima-Perú*

[2]*Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, 15314, Lima-Perú*

[3]*Research and Intellectual Creativity Direction, Universidad María Auxiliadora, 15408, Lima-Perú*

ARTICLE INFO

Article history:
Received: 03 August, 2020
Accepted: 25 November, 2020
Online: 16 December, 2020

Keywords:
Social Skills
Resilience
Educational institution
Daily Life
Coping

ABSTRACT

Social skills are a set of partially independent and situationally specific verbal or nonverbal responses, in terms of resilience, it is the set of qualities, resources or strengths that favor adolescents to face the adversities of life, therefore the objective of the research work is to determine the social skills and resilience in adolescents of an educational institution in Lima - North. It is a non-experimental study, descriptive of a quantitative approach, it is a cross-sectional correlational research, with a population of 706 students of the secondary level, also with demographic data, and the Social Skills Scale (SSS), and the scale of Resilience (CD-RISC) that will measure the level of resilience of adolescents. The relationship between the variables Social Skills and Resilience. It was determined based on the Pearson's Chi square test (X^2). The significance level of the test obtained a value of 0.000 ($p<0.05$) ($X^2 = 28.626$; g.l. = 4). So, the dissociation hypothesis is rejected and it can be affirmed with statistical evidence that there is a significant relationship between social skills and resilience. The predominant level of social skills is medium with 648 (91.8%) participants and in terms of resilience, the predominant level is high with 366 (51.8%) of participants. Social skills and resilience are mutually related since both allow the adolescent to face situations of daily life looking the ability to develop their coping capacity.

1. Introduction

Social skills or assertive behaviors are understood as a set of partially independent and situationally specific verbal and nonverbal responses, through which adolescents express in an interpersonal their needs, feelings, preferences, right opinions without excessive anxiety and in a non-aversive way, expressing itself in self-expression of social situations, defense of its own rights as a consumer, expression of anger or disagreement, saying no and cutting interactions, initiating positive interactions with the opposite sex and making requests. Social skills are involved in many areas of a person's life, such as adaptation to the environment in which they live and health protection, both physical and mental. It helps to solve various situations such as appropriateness in behavior, decision making, facing difficult situations, assigned essential responsibilities in the development of learning [1].

Regarding resilience, it is the set of qualities, resources or strengths that favor adolescents to face life's adversities, overcome them or even be transformed by it, which is characterized by presenting traits of persistence-tenacity-self-efficiency, control under pressure, adaptation and support environments, control and purpose and spirituality [2].

Social skills and resilience can prevent problems that may affect adolescents in their daily life, where risky conditions such as family problems, drug abuse or abandonment make it evident that resilient factors are looking for a way to excel in risky conditions in an adolescent-friendly way [3], [4].

Therefore, social skills and resilience play an important role in the adolescent since it will allow them to develop on a social and emotional level so that they can adapt to their environment and at the same time they can face adverse situations in their daily life [5].

In Peru, according to the report of the Instituto Nacional de Estadística e Informática, 32.9% of households are made up of

*Corresponding Author: Brian Meneses-Claudio, Mr., +51 1 950159924 & bmeneses@uch.edu.pe

girls and boys and/or adolescents, of whom there is at least one child under 18 with a caloric deficit, being higher in rural areas (33.5%) than in urban areas (33.2%) and Metropolitan Lima (32.0%). On the other hand, 41.4% of children under 18 years of age had a problem that affected their health such as a new disease, a relapse or an accident. Health problems occurred more frequently in children and adolescents in the urban area, not including Metropolitan Lima (42.5%), Metropolitan Lima (41.4%) and rural areas (39.8%) [6].

In [7], the authors presented a research work carried out in Colombia, significant differences were shown in emotional intelligence. In the high social acceptance group, low scores were found in the social skills evaluated, being the ability to make requests lower. Likewise, statistically significant correlations were found between several of the factors of the Baron test and those of the Social Skills Scale.

In [5], the authors described a study carried out in Spain, in the sample made up of 844 participants of both sexes, it shows that the most resilient adolescents have better results in all dimensions of quality of life and the effect of resilience is greater in the dimensions related to mental health and in all those that measure social relationships.

In [3], the authors described a study carried out in Colombia, a study was presented regarding the ability to listen, the ability to start a conversation, the ability to hold a conversation and the ability to ask for help, it is evident that most of people obtained scores between high and medium level. On the other hand, in the expression subscale of feelings, it is stated that most men have a medium level, while women are between the medium level and a high level.

The objective of the study is to determine the social skills and resilience in adolescents of an educational institution in North Lima in which it will allow to give important data about the social skills and resilience of adolescent students at the secondary level because nowadays, adolescents tend to have little communication with others in which this behavior cannot help them in a social situation effectively.

The study applied the data collection instrument Social Skills Scale (SSS), it was created by Elena Gismero Gonzales and the Connor-Davidson Resilience Scale (CD-RISC), which has been useful to assess social skills and resilience of adolescents at the secondary level of an educational institution.

The data collection was processed through the survey of adolescents at the secondary level, the data was entered in a data matrix that will be designed in the statistical program SPSS (Statistical Package for the Social Sciences) in its version 24.0 , in which it will allow a better data processing to make statistical tables and graphs so that they can be described and interpreted in results and discussions, respectively.

The following research work is structured as follows: In section II, the development of data collection processing for each adolescent at the secondary level will be presented, as well as the guidelines to consider so that they are included in the research work. In section III, the results will show the social skills and resilience of adolescents at the secondary level of the educational institution according to the specified dimensions of the instruments in the measurement of the variables. In section IV, it presents the discussions of the research work, in section V, the conclusions and in section VI the recommendations as well as the future work that is intended to be reached with the research work.

2. Methodology

In this section, the type and design of the research will be evidenced, as well as the population and sample that will be carried out in the research work, in addition to the inclusion and exclusion criteria and finally the technique and the instrument of data collection.

2.1. Research Type and Design

The present research is non-experimental, descriptive of a quantitative approach, it is a cross-sectional correlational research.

2.2. Population and sample

The population is made up of 706 students from 1st to 5th secondary level of the educational institution 3049 Imperio Del Tahuantinsuyo.

Inclusion Criteria

- Students who normally attend the educational institution.
- Students who have informed consent and assent.
- Students who have the authorization signed by their parents and agree to voluntarily join the study.

Exclusion Criteria

- Students who do not regularly attend to the institution.
- Students who do not have the authorization of their parents.

2.3. Technique and instrument

The technique used is the survey, using the questionnaire or instrument for collecting data on social skills and resilience, which aims to measure social skills and resilience in adolescents of an educational institution in North Lima.

The Social Skills Scale (SSS) was created by Elena Gismero Gonzales, from the Universidad Pontificia de Comillas (Madrid) in 1997, its application is given individually and collectively, with an approximate duration of 10 to 15 minutes, it is focused in adults and adolescents. The social skills scale consists of 33 items grouped into 6 dimensions, self-expression in social situations (8 items), defense of one's rights as a consumer (5 items), expression of anger or disagreement (4 items), saying do not cut off interactions (6 items), make requests (5 items) and initiate positive interactions with the opposite sex (5 items) that are evaluated by a Likert-type scale where “1 = I do not identify myself, in most times it does not happen to me or it does not I would do ”,“ 2 = it has nothing to do with me, even if it happens to me sometime ”,“ 3 = describes me roughly, even if I don't always act that way or feel that way ”and“ 4 = strongly agree, I would feel that way or act that way in most cases". The scale range varies from 33 to 132, the higher the score, the higher the adolescent's social skills [8].

The Connor-Davidson Resilience Scale (CD-RISC) was built by Connor and Davidson in the United States in 2003.

Its application is focus for the adolescent and adult population from 11 to 80 years old, the resilience scale consists of 25 items grouped into 5 dimensions, persistence-tenacity-self-efficacy (8 items), control under pressure (7 items), adaptation and ability to recover (5 items), control and purpose (3 items) and spirituality (2 items) that are evaluated with a Likert-type scale where "0 = never", "1 = rarely", "2 = sometimes", "3 = often" and "4 = almost always". The range of the scale varies from 0 to 100, the higher the score, the higher the resilience in the adolescent [9].

2.4. Place and Application of the Instrument

The survey to measure social skills and resilience in adolescents was carried out at the 3049 educational institution of Imperio del Tahuantinsuyo in North Lima.

The survey was carried out in two consecutive shifts, morning and afternoon according to the schedule of the students of the secondary level, the questionnaire was taken to the secondary level students with a time of 20 minutes to each selected section (1st from high school to 5th grade according to the inclusion criteria), it was concluded with a good satisfaction at the time of collecting the surveys since support from high school students was presented to carry out this research work.

It is very important to emphasize the presence of classroom teachers at the time of completing the surveys, because this helps students to relay on making questions using the sign of upper hand. In addition to the presence of nursing professionals since the mental health of adolescents is very important, because they are the future where the ways of communication tend to be high and that allow them to generate self-confidence and be able to perform effectively to any situation that is happening.

Figure 1: Application of the Survey to the students of the Educational Institution

Figure 2: Nursing staff is supporting in carrying out the survey

In Figure 1, the resolution of survey is shown in a classroom of the educational institution at the secondary level by the students, the collaboration of the students can be seen since this allows interaction with them.

In Figure 2, the reader can see the direct interaction with the adolescent student for the research process, it should be noted that the students gave confidence when explaining about the completion of the surveys.

3. Results

The table 1 is a summary of the surveys carried out following the guidelines corresponding to the research work:

Table 1: Social Skills and Resilience in secondary level students of Independencia Educational Institution – North Lima, 2019 (N = 706)

			Social Skills			Total
			Low level	Mediu m level	High level	
Resilience	Low level	Count	0	16	0	16
		% within Resilienc e	0,0%	100,0%	0,0%	100,0 %
	Mediu m level	Count	6	312	6	324
		% within Resilienc e	1,9%	96,3%	1,9%	100,0 %
	High level	Count	3	320	43	366
		% within Resilienc e	0,8%	87,4%	11,7 %	100,0 %
Total		Count	9	648	49	706
		% within Resilienc e	1,3%	91,8%	6,9%	100,0 %
Pearson's Chi-square			Value X^2	g.l.	Sig. (p)	
			28,626	4	0,000	

The significance level of the test obtained a value of 0.000 ($p < 0.05$) ($X^2 = 28.626$; g.l. = 4). This indicates that there is a significant relationship between social skills and resilience.

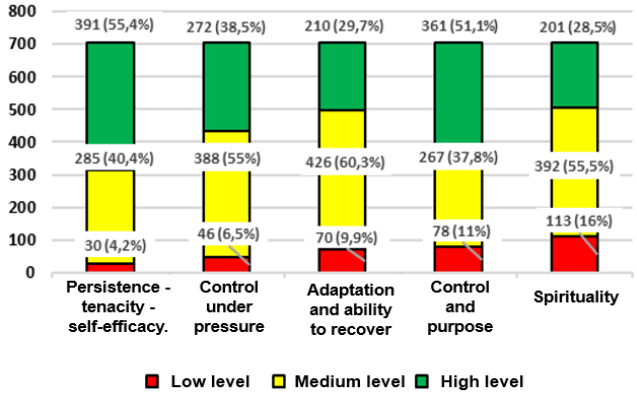


Figure 3: Social skills according to their dimensions in secondary level students in an Educational Institution of Independencia – North Lima, 2019 (N = 706)

In Table 2, it can see the social skills in secondary level students in an Educational Institution of Independencia – North Lima, where the middle level predominates with 648 (91.8%) students.

Table 2: Social Skills in secondary level students in an Educational Institution of Independencia – North Lima, 2019 (N = 706)

		N	%
Social Skills	**Low Level**	9	1,3
	Medium Level	648	91,8
	High Level	49	6,9
Total		706	100,0

In Figure 3, it can observe social skills according to their dimensions, where the most affected dimension was the expression of anger or disagreement with 97 (13.7%) students.

Table 3: Resilience in secondary level students in an Educational Institution of Independencia – North Lima, 2019 (N = 706)

		N	%
Resilience	**Low Level**	16	2,3
	Medium Level	324	45,9
	High Level	366	51,8
Total		706	100,0

In Table 3, it can see the resilience of secondary level students in an Educational Institution of Independence, where the high level predominated with 366 (51.8%) students.

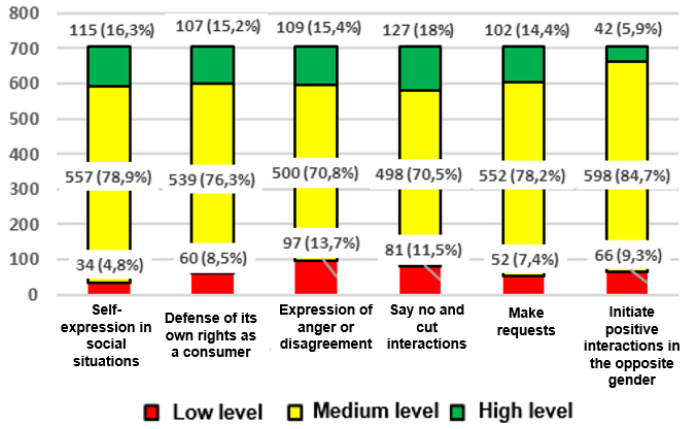


Figure 4: Resilience in secondary level students in an Educational Institution of Independence - North Lima, 2019 (N = 706)

In Figure 4, it can see the resilience according to its dimensions in secondary level students in an Educational Institution of Independence, where the most affected dimension was spirituality with 113 (16%) students.

Both studies can be interpreted that social skills and resilience in adolescents are linked to the promotion of mental and physical health of adolescents, knowing that resilience is relevant because it will allow the person to respond appropriately to negative events or situations. It also focuses on social skills because it is an essential element to achieve the socio-emotional development of the human being and favors school learning. It is vitally important to develop social skills and have a high level of resilience, but unfortunately there is a knowledge gap, reviewing the scientific information related to this topic.

These results will help to the educational institution since it will provide them with information that will allow to evaluate more concisely the adolescents of the secondary level, giving importance to the mental health of the student, in which it is beneficial for them and so they can communicate in an appropriate way that allows to have opportunities in the future.

4. Discussion

In the present study, the issue of social skills and resilience is raised from the point of view of promoting adolescent health, which seeks to contribute to institutions that are in the educational field as part of comprehensive adolescent training, including programs that allow adolescents to enhance their ability to interact, be extrovert, and cope with problems to improve their social life.

Based on the social skills in adolescents of the educational institution 3049 Imperio del Tahuantinsuyo, it prevailed in medium level followed by high and low. These results are based on two purposes: the first, if there are no interventions in the group with a medium level of social skills, these will not have a full development their personality. On the other hand, making interventions can raise the level of social skills, which would be beneficial for adolescent development. In [10], the authors showed that adolescents are normally receptive to new ideas, and are eager to take full advantage of their growing capacity to make decisions; their curiosity and interest lead to a great openness to new possibilities and among them could be considered relating to their peers. Engaging in positive and constructive activities offers opportunities to build relationships with adults and peers. The reason for these findings is due to the fact that in the institution there are no teachers prepared for the tutoring course, they are the teachers of different courses who are in charge of some sections and the participation of parents in the meetings is required.

By implementing well-designed social skills programs, the level of social skills can be raised for the benefit of adolescents and mental health. In these processes of strengthening social skills, the family must also be involved, since this constitutes an important Socialization space and parents play an important role as Trainers and counselors. Conditions must be implemented in schools such as not saturating student classrooms and having the necessary infrastructure for the work of the teacher and tutor; favoring the work of professionals for the benefit of adolescents [11].

Regarding social skills in its self-expression dimension in social situations in adolescents of the educational institution 3049 Imperio del Tahuantinsuyo-Independencia 2019, the medium level prevailed followed by the high and low. It reflects the ability of each individual to express themselves spontaneously in different social and work events. We agree with what was stated by Pérez, who points out in its study that an institutional intervention program can significantly improve social skills in adolescents.

Regarding to social skills in their defense of their rights as a consumer dimension in adolescents of the educational institution 3049 Imperio del Tahuantinsuyo-Independencia 2019, the medium level prevailed followed by the high and the low. Obtaining a high score reflects the expression of assertive behavior towards strangers in defense of their own rights in situations of consumption. Regarding social skills in their dimension defending their own rights as consumers in adolescents from a pre-university center in Lima Norte, the medium level predominated, followed by the low and high. This dimension is related to being an active participant in social life that allows to exercise full citizenship. In [12], the author deduces in his study that the development of social skills and their dimensions, play a protective role for the adolescent, away from risky doings and behaviors such as marijuana use.

Regarding to social skills in its expression of anger or disagreement dimension in adolescents from the educational institution 3049 Imperio del Tahuantinsuyo-Independencia. 2019, the medium level prevailed followed by the high and low. It rates the individual's ability to handle their conflicts with another person in a negative or positive way. In [13], the authors point out from the findings of their study, that the training institution must assume a leading role in promoting social skills; teachers play an important role, so they must be permanently trained to join and contribute to the promotion of social skills in them.

In relation to social skills in it say no and cut interactions dimension in adolescents of the Educational Institution 3049 Imperio del Tahuantinsuyo-Independencia 2019, the medium level prevailed followed by the low and high. This dimension reflects the ability of the individual to cut off interactions that they do not want to maintain, as well as refusing to lend something when dislike doing so, it was found that the vulnerable adolescents who are victims of bullying do not know how to respond to these situations and they have a hard time saying “enough” or “not”.

Respecting of social skills in its making requests dimension in adolescents from the educational institution 3049 Imperio del Tahuantinsuyo-Independencia 2019, the medium level prevailed followed by the low and high. This dimension reflects the ability of the individual to make a request without difficulties, it should be noted that in cooperative learning processes they promote the development of social skills and can enhance the ability to make requests in an appropriate and timely manner.

Regarding social skills in its dimension, initiating positive interactions with the opposite sex in adolescents from the educational institution 3049 Imperio del Tahuantinsuyo-Independencia 2019, the medium level prevailed followed by the low and high [10]. This dimension qualifies the ability of the individual to relate, interact with the opposite sex. This dimension presented a higher percentage in the low level of social skills.

Concerning the level of resilience in adolescents of the educational institution 3049 Imperio del Tahuantinsuyo-Independencia 2019, the high level predominated, followed by a minimal difference from the medium and low level. The results would be explained in different ways, one of the advantages is that the highest percentage lives with their parents, they share family experiences, which favors them as a protective factor so that they can overcome adversities and get out with great ease. The adolescence plays a very important role in resilience since at this stage of life, it goes through various changes such as: biological, psychological and social. This is where most conflicts, rebellions are generated, testing their ability to adapt [12].

In relation to its spiritual dimension, it has a higher percentage in the low level of resilience. The role of spirituality in improving an adolescent's ability to cope with stressful life situations is often under-appreciated, especially in the context of mental health services within a juvenile detention environment.

5. Conclusions

We can conclude that social skills and resilience are mutually related since both allow adolescents to develop functions that allow them to have resources or strengths to cope with situations of daily life.

It is concluded that the social skills are very important for adolescents since it is fundamental for their daily life because they can get success, have proper communication with other people and not only of verbally as well as affectively and emotionally.

It is concluded that resilience is very favorable for adolescents since, within their social environment, they will have the capacity to solve stressful situations and will make them vulnerable to any situation, since resilience will allow adolescents to adapt to these situations and be able to solve them effectively.

6. Recommendations

It is recommended in this institution to have programs that help improve communication between the students themselves so that they can improve their communication with other people.

It is recommended in this institution that it be carried out annually with a control system on social skills and resilience since these are essential for the adolescents so this will allow them to communicate with society and also with their environment.

It is also recommended that health professionals attend educational institutions where they can observe certain deficiencies in social skills and resilience in order to allow them to carry out psycho-affective programs where the adolescent can interact with their social environment and also develop both verbal and non-verbal skills.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] R. Estrada, E. Escoffié, R. García, “Habilidades sociales en adolescentes institucionalizados: Una intervención con juego grupal.,” Psicumex, **6**(2), 27–38, 2016, doi:10.36793/psicumex.v6i2.285.

[2] A. Rodríguez, E. Ramos, I. Ros, A. Fernández, L. Revuelta, “Bienestar subjetivo en la adolescencia: el papel de la resiliencia, el autoconcepto y el apoyo social percibido.,” Suma Psicologica, **23**(1), 60–69, 2016, doi:10.1016/j.sumpsi.2016.02.002.

[3] S. Betancourth, C. Zambrano, A. Ceballos, V. Benavides, N. Villota, “Habilidades sociales relacionadas con el proceso de comunicación en una muestra de adolescentes.,” Psicoespacios, **11**(18), 133, 2017, doi:10.25057/21452776.898.

[4] C. Huang, Y. Chen, S. Cheung, L. Greene, S. Lu, “Resilience, emotional problems, and behavioural problems of adolescents in China: Roles of mindfulness and life skills.,” Health and Social Care in the Community, 1–

9, 2019, doi:10.1111/hsc.12753.
[5] M. Simón, R. Fuentes, M. Garrido, M. Serrano, E. Larrañaga, S. Yubero, "Influencia de la resiliencia sobre la calidad de vida relacionada con la salud en adolescentes.," Enfermería Clínica, **28**(5), 283–291, 2018, doi:10.1016/j.enfcli.2018.06.003.
[6] Instituto Nacional de Estadistica e Informatica, Estado de la Niñez y Adolescencia., 2018.
[7] C. Coronel, M. Levin, S. Mejail, "Las habilidades sociales en adolescentes tempranos de diferentes contextos socioeconómicos.," Electronic Journal of Research in Educational Psychology, **9**(1), 241–261, 2011.
[8] E. Miranda, E. Riquelme, H. Cifuentes, "Revista Latinoamericana de Psicología Análisis factorial confirmatorio de la Escala de habilidades sociales en universitarios chilenos.," Child Psychiatry & Human Development, **41**(2), 73–82, 2010.
[9] G. Guihard, L. Deumier, B. Alliot-Licht, L. Bouton-Kelly, C. Michaut, F. Quilliot, "Psychometric validation of the French version of the Connor-Davidson Resilience Scale.," Encephale, **44**(1), 40–45, 2018, doi:10.1016/j.encep.2017.06.002.
[10] M. Valera, L. Buil, E. Rigo, A. Casero, E. Aguilar, "Habilidades sociales en preadolescentes con trastorno específico del lenguaje.," Revista de Logopedia, Foniatria y Audiologia, **36**(2), 55–63, 2016, doi:10.1016/j.rlfa.2015.03.002.
[11] E. Maguiña, R. Mino, Facultad De Ciencias De La Salud Escuela Profesional De Enfermería., 2018.
[12] F. Morales, "Relaciones entre afrontamiento del estrés cotidiano, autoconcepto, habilidades sociales e inteligencia emocional.," European Journal of Education and Psychology, **10**, 41–48, 2017, doi:10.1016/j.ejeps.2017.04.001.
[13] A. Gonzales, P. Quispe, "Habilidades sociales y rendimiento Académico de los estudiantes de la Facultad de Ciencias de la Educación - Universidad Nacional del Altiplano - Puno Perú.," Revista de Investigaciones Altoandinas - Journal of High Andean Research, **18**(3), 331–336, 2016, doi:10.18271/ria.2016.222.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1356-1365 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Improved Design and Recommendations for Street Lighting in Gitega City

Ntawuhorakomeye Noel[1,*], Ndayiragije Leonidas[2], Belov Mikhail Petrovich[1]

[1]*Department of Robotics and Industrial Automation, Saint- Petersburg Electrotechnical University ETU"LETI", 197227, Russia*

[2]*Department of Electromechanical Engineering, Burundi University, 1550, Burundi*

ARTICLE INFO

Article history:
Received: 07 September, 2020
Accepted: 01 December, 2020
Online: 16 December, 2020

Keywords:
Solar mini power plant
Gitega climate
Public lighting
Smart lamp

ABSTRACT

The article discusses the possibility of using solar energy for street lighting systems in the city of Gitega, in the Republic of Burundi. Analysis of weather and climate conditions of the city was carried out and an effective street lighting system based on solar mini-power plants using an intelligent control system developed. Calculations of technical and economic efficiencies are made and the principle of operation of the developed system is described. A comparison of the methods of transmitting control signals and the performance of different types of lamps was made as well as an assessment of the battery discharge process. The different attachment types of solar panels to their supports were highlighted, the illumination received on a public space according to the installation height of the luminaire and the used LED power as well as the requirements in terms of illumination and brightness depending on the place to be lit were discussed. The system developed consists of two solar panels, connected in parallel, one battery and six LED lamps with adjustable power. At present, the problem of modernizing the lighting of public places and roads of regional and local significance in Gitega city remains relevant. With the aim of promoting clean energy and reducing energy consumption due to global concern about climate change, the use of modern, cost-effective lighting systems based on the use of LED technology and is one of the most dynamic trends in the world at present and in the near future for the well-being of the population.

1. Introduction

Light is a fundamental element for human activity, the day is lit by the sun but when night comes artificial light becomes mandatory. Since solar energy is available all over the globe, the city of Gitega, now the political capital of the Republic of Burundi, located in the center of the country, is privileged to have a strong solar deposit. This work is an extended version of the report [1] presented in 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), where a public Lighting by Mini Solar Power Plants in the Republic of BURUNDI is proposed. With a duration of insolation exceeding 3000 hours per year and a considerable annual average daily irradiation, this region can be a pole of excellence for the use and promotion of photovoltaic solar energy [1, 2]. In this work, we propose a design and dimensioning of autonomous systems for street lighting using solar photovoltaic energy as the primary source. Obtaining information about the intensity of solar radiation in a specific geographic and climatic zone gives an idea of the possibility of using solar energy to generate a certain amount of electricity. The climatic conditions of the city of Gitega make this value almost constant throughout the year [2]. This is an important parameter when calculating any type of solar power plant (SPP). However, in order to improve the energy efficiency of photovoltaic lighting systems, it is necessary to use both high-efficiency photovoltaic modules as well as efficient batteries and charge regulators for energy storage and management; but also it is essential to use luminaires with high energetic efficiency (number of lumens/Watt) [3–7].

Nowadays, artificial lighting has become more than a means of obtaining light; it is an indispensable element of human life in general and in urban areas in particular [8]. The lighting of public places, traffic lanes, monuments and sites in cities and the countryside has now become a major concern for electricity

*Corresponding Author: Ntawuhorakomeye Noel, noentaw1@mail.ru

distributors and administrative and political authorities. Public lighting plays a very important role in various places of attendance both economically and socially [8, 9]. It is above all a measure aimed at increasing the safety of public spaces, whether in traffic areas or in relaxation and recreation areas. The objective is not to provide light only for safety purposes, but also to create a pleasant nighttime atmosphere giving a new vision and approach to the spaces [9]. To produce light, there are three main techniques [1, 8]: incandescent (classical and halogen), fluorescence and LED (Light-emitting diode). Due to the high efficiency and the absence of infrared radiation, LED lamps practically do not heat up during operation. Currently existing LED lamps consist of dozens, and sometimes hundreds of LEDs. The quality of colored light is still an unrivaled achievement of LED technology. The main difference between LED luminaires (lamps) and all others is their unprecedented long service life (up to 100,000 hours) [5, 8]. For electric lighting systems that operate at night, the period of electricity accumulation during daylight hours is a very important. A 120 amp-hours (120 Ah) accumulator battery with two 200-watt solar panels (200 W) were used in the design of this mini SPP to power the six public lighting LEDs. The developed intelligent system is a solution for remote control of street lighting, which has the ability to control the lamps and the lighting level of each street lamp. Increasing the reliability of street lighting networks will ensure energy saving, uninterrupted power supply, and as a result, ensure road safety as the number of accidents and illegal actions is significantly reduced when the city is well lit. The use of an intelligent street lighting control system will ensure safer traffic conditions, pedestrian safety, and significantly improve the city's architectural, tourist, and commercial aspects [10, 11]. All this with the aim of promoting clean energy, and reducing energy consumption due to global concern on climate change.

2. Characteristics and Features of Gitega City Climate

Gitega city is located in the center of the country about 100 km east of the economic capital Bujumbura. Its geographical coordinates are 3° 25' South, 29° 55' East according to Meteonorm 7.3, Sat = 100% of PVSYST. It is part of Africa in the southern hemisphere. This region is characterized by hills composed of metamorphic rocks such as schists, quartzites and phyllites. The Gitega province is located in the Central Plateau area with an altitude ranging from 1600 to 2000 m. It has a tropical climate tempered by altitude. Under normal conditions, the climate of the Gitega city is characterized by two (2) well-marked seasons namely the dry season, which begins in May and continues until October with minimas in June and July as well as the rainy season. The average precipitation is 1200 mm and the average temperature over the year 22.6°C (As shown in Table 1). The main nearby river is the Ruvubu with two main tributaries namely the Ruvyironza and the Mubarazi. Several districts make up the town of Gitega for about 10 km, along and around the RN2 (national road № 2) that crosses it: Magarama, Nyamugari (Swahili district), the commercial district, the administrative district, Musinzira, Shantanya, Rutonde, Nyabiharage, Nyabututsi and Mushasha (religious district). Located on an important relief, there is significant solar irradiation whose average value received annually is estimated at 1800 kWh/m^2/year. The following Table shows the average temperatures, solar radiation and wind velocity, by months according to Meteonorm 7.3, Sat = 100% of PVSYST.

Table 1: Temperature, Average irradiation, Wind velocity and Relative humidity at Gitega

	Average temperature (^{0}C)	Average irradiation (kWh/m^2/mth)	Wind velocity (m/s)	Relative humidity (%)
January	23.3	149.5	3.30	64.0
February	24.3	134.3	3.49	57.5
March	24.4	156.9	3.50	61.1
April	23.2	147.6	3.39	68.6
May	22.7	151	3.70	66.5
June	21.3	151.1	3.80	65.7
July	20.7	158.5	3.80	64.9
August	21.0	156.2	3.80	63.6
September	21.6	149.7	3.99	62.5
October	22.8	154	3.80	62.0
November	22.7	136.0	3.10	68.3
December	23.4	146.5	3.19	65.8
Year	**22.6**	**1791.3**	**3.6**	**64.2**

Based on the abovementioned and presented data in the table, we can conclude that the climatic conditions of the city of Gitega are ideal for the effective use of electricity from solar panels (SP). The Figure1 shows the temperature and irradiation variation according to the months of the year.

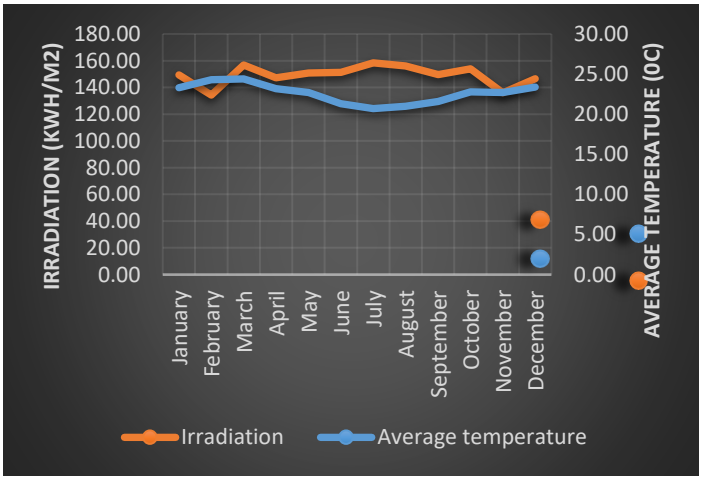


Figure 1: Temperature and Irradiation variation

Figure 2 Shows the Sun Path and Horizon Line respectively at Gitega. It is noted that the city of Gitega sees 9 hours of sunshine per day. Large solar irradiation begins to be observed at 10 o'clock, and begins to decrease at 15 o'clock with a maximum at 12 o'clock [1].

Knowing that there is no public lighting in general in this city and that the only light that the streets and public squares receive comes from the exterior lights of the surrounding houses and buildings, the development of a public lighting system is very important for the well-being of residents and passers-by, especially at night.

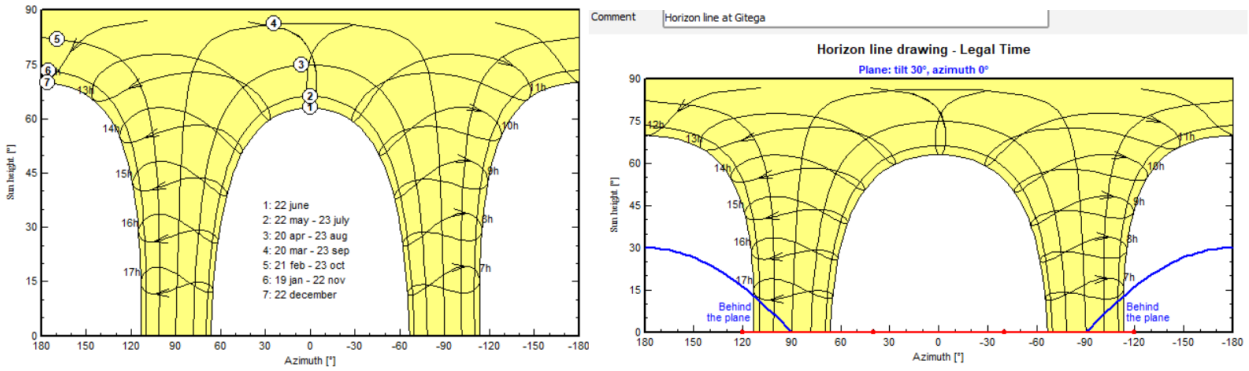


Figure 2: Sun Path Shape Horizon Line

Table 2: Requirements for Roads and Streets Lighting

Lighting Category	Streets and Roads	The Highest Traffic Intensity in Both Directions, units / h	Coating Average Brightness, Cd /m^2	Coating Average Horizontal Illumination, lx
A	Main roads, main streets of citywide significance	More than 3000	1,6	20
		1000 to 3000	1,2	20
		500 to 1000	0,8	15
B	Main streets of district significance	More than 2000	1,0	15
		1000 to 2000	0,8	15
		500 to 1000	0,6	10
		Less than 500	0,4	10
C	Local streets and roads	500 and more	0,4	6
		Less than 500	0,3	4
		Single cars	0,2	4

Table 3: Horizontal Illumination of Some Places

Places to light	Average horizontal illumination, Lx
Main pedestrian streets, impassable parts of squares of categories A and B, and pre-factory squares	10
Pedestrian streets: -within community centers	6
-in other territories	4
Sidewalks separated from the roadway on streets	2-4
Public transport boarding areas on all categories of streets	10
Pedestrian bridges, Children's playgrounds where outdoor games equipment is located	10
Pedestrian tunnel: -in the daytime	100
- evening and night	50
Pedestrian tunnel stairs in the evening and at night	20
Household sites and waste collection sites	2

3. Main Requirements for Public Lighting Systems

When developing solar street lighting, it is possible to achieve maximum efficiency of the system only using, so-called "smart" technical means, which would not only control operation modes, but also control the operation of the entire system in automatic mode. Additional maintenance of street lighting should be carried out in case of violation of the recommended lighting standards (lamps fail, stop shining or Shine dimly). The working hours of

lamps are regulated by municipal services, which approve the schedule for switching on and off lighting.

In cities and rural settlements, the requirements for lighting streets and roads vary significantly. Table 2 shows the standards for artificial lighting of roads and streets in cities and rural settlements. At the same time, the average brightness of a highway's coverage is 1.6 cd/m^2 in cities and at least 1.0 Cd/m^2 outside of cities at the main entrances to airports, river and sea ports, regardless of traffic intensity [8, 12].

The average horizontal illumination at the level of coverage of impassable parts of streets, roads and squares, boulevards and squares, pedestrian streets and territories of micro districts in urban settlements should be taken according to table 3.

At night, lighting level of city streets, roads and squares can be reduced to a standardized average brightness of more than 0.4 cd/m^2 or an average illumination of more than 4 lux by switching on no more than half of the lamps. It is allowed to reduce the lighting level with the regulator in order to obtain additional energy savings in the evening and morning hours of the day [3, 6]: by 30% with a decrease in traffic intensity to 1/3 of the maximum value; and by 50% with a decrease in traffic intensity to 1/5 of the maximum value.

On streets and roads at standardized values of an average brightness of 0.3 cd/m^2 or an average illumination of 4 lux or less, on pedestrian bridges, parking lots, pedestrian walkways and roads, internal, service and fire-fighting passages, as well as on rural streets and roads settlements, partial or complete switching off of lighting at night is not allowed. The following Table (Table 4) shows the variation of the illumination received (lux) on a public space according to the luminaire installation height and the power of the used Led (W) [13].

4. Automatic Control Systems of Public Lighting

Switching street lights on and off as needed is an excellent measure to save energy consumption. However, the various systems are not equally suited to this rapidly evolving technology. It is particularly interesting for public spaces used infrequently, such as public toilets, public waste collection points, and rural roads with little traffic. For areas with high safety sensitivity, it is not recommended (e.g. for busy roads, areas of heavy road or pedestrian traffic, and areas requiring overnight surveillance). On the other hand, for autonomous solar lighting systems, a regular and homogeneous consumption cycle is ideal. Uneven battery discharge due to erratic interruptions leads to rapid degradation and the need to replace them before the end of their normal lifespan. Therefore, measures to interrupt the nighttime lighting duration of such systems are only suitable to a limited extent. For solar systems, regular long-term interruptions are feasible to be managed through programmable astronomical clocks. Motion detectors are only recommended for small isolated spotlights. Automatic street lighting control systems usually operate under the control of a zone controller or server. Depending on the control algorithm, the controller generates a signal, for example, turning on a group of streetlights [10, 13, 14]. To transmit this signal to actuators, the following means are used: -low-current signal lines (twisted pair, RS-485, Ethernet, etc.); -radio channel; -GSM-channel; -signal transmission via power cable.

Table 4: Examples of illumination data for a lighting unit

Installation height of the luminaire	**Illuminated area**	**Maximum illumination (lux) received on a road**					
		LED 18 W	LED 30 W	LED 50 W	LED 65 W	LED 80 W	LED 120 W
6 m	8×20 m	10 Lux	19 Lux	30 Lux	39 Lux	49 Lux	59 Lux
8m	10×25 m	6 Lux	12 Lux	18 Lux	23 Lux	29 Lux	35 Lux
10m	13×30 m	4 Lux	8 Lux	12 Lux	15 Lux	19 Lux	24 Lux
12 m	16×35 m	2 Lux	5 Lux	8 Lux	11 Lux	14 Lux	17 Lux

Table 5: Comparison of control signal transmission methods

	1. Low Current Control	**2. GSM-Channel**	**3. Power Transmission Lines**	**4.Radio Channel**
Addressing (Economically Feasible)	possible Individual lamp controls	Group control only	Group control only	Group control only
Control Method	Digital control Protocol for example based on a calendar schedule	Phone call or SMS to the controller in the control Cabinet	Control via a power cable connected to the controller in the control Cabinet	Radio signal transmission from the control room to the receiver in the control cabinet
Factors Affecting the Reliability	Accumulation of timing error	Dependence on the workload of the public network of the GSM operator.	Upon failure of the required manual switching cable.	Radio interference may cause the control signal to be unable to be received
Labor Costs	High labor costs	Low labor costs through public use network	With individual lamp control, cable laying is labor-intensive	High labor costs when installing transceivers

Territory Coverage	Snap to city / region	Cellular coverage area	The length of the control power cable cannot exceed 1 km	Control is possible only in the area of confident reception of the radio signal.
Territory Size	City district, small settlement	City and nearest suburb	Limited territory	City and suburbs, area along the motorways
Cost Factors	Individual control unit in each lamp	Subscription and connection fees, messaging	The cost of individual power cables installation	Cost of dispatcher's equipment, relay stations and receivers
Factors Affecting Maintenance Cost	Timer adjustment is constantly needed		High repair costs for electrical equipment	A qualified dispatcher is required

Setting up the public lighting network is easier with a computerized centralized management system. Among other functions, this system normally allows: a) Easy and quick re-setting of the whole or parts of the network, for example during extraordinary events, which require a temporary adaptation of the lighting duration in a commune area. b) Accurate detection of faults and malfunctions in light points and electrical cabinets. c) Monitoring of energy consumption by the lighting network. d) Linking of network operation with climatic conditions and traffic volume. e) Management of equipment and human resources related to network maintenance.

Such a system consists of having controllers integrated into the electrical cabinets with a link to the light points or individual controllers integrated into the light points themselves. The controllers communicate with the remote management station (a computer, a tablet or a smartphone) via a wireless network, for example WIFI or radio. Note that for the control of the lighting time of streetlights, programmable time switches preferably of the so-called astronomical, digital or mechanical clock kind which are normally integrated in the control boxes as well as twilight detectors with a photoelectric cell (sensor) which are either integrated in the equipment (for example a regulator), or they are stand-alone units. Lighting controlled by a twilight switch automatically activates the switching off and switching on of public lighting according to daylight. To ensure the reliability and accuracy of the switch, it is essential to correctly position and orient the sensor so that it is not affected by random light sources or disturbed by shadows. It should also be checked and cleaned regularly to prevent dust accumulation. Twilight switches can be associated with clocks in order to cut off the control during part of the nighttime. The programmable time switches control the switching off and switching on of public lighting at specific times. They are installed in public lighting cabinets. If it is not an astronomical clock, it must be adjusted and updated regularly in order to approximate sunrise and sunset times [8, 10].

5. Design Features of the Developed System and its Operation Principle

The functional scheme of the solar-powered street lighting system developed is shown in Figure 3. It is an economical lighting system with intelligent control operating with an intelligent lamp having a step change of the luminous flux.

1. Solar panel; 2. Block of Sensor; 3. Lamp with Progressive Adjustment of the Luminous Flux; 4. Control Block; 5. Storage Battery; 6. Diagnostic Connector.

The photovoltaic solar panel (1) generates an electric current passing through the control block 4, which includes the charge controller of the storage battery. A "smart lamp" (3) is a lamp with stepwise adjustment of the light output power. Depending on its capacity, a battery to be charged by the solar panel ensures the operation of the entire system without recharging for a period corresponding to the duration of the night. The Block (2) consisting of temperature and light sensors sends accurate temperature and illumination data to the Control Unit 4, thereby controlling the operation mode of the lamp. The light sensors in block 2 will turn on the lamp at nightfall, and turn it off as soon as the first rays of the sun appear. Lamp operates in three different modes namely: maximum power mode, average power and economy mode depending on the time of year, the solar radiation intensity and the battery charge level. In addition, if the system notices a long lamp operation in Economy mode, it evaluates the battery charge level and can switch to timer mode in which the lamp turns on in Economy mode for a time T1, after which it turns off and turns on during time T2 before sunrise. Depending on the battery charge level, the T1 and T2 times may vary in the direction of decrease or increase. A computer is connected to the public lighting system via a serial interface according to the USB 2.0 standard from the diagnostic connector 6 to allow the use of software specially designed to control all the parameters of the system, program its operation, determine the dirt level from the solar panel, change the operating modes etc.

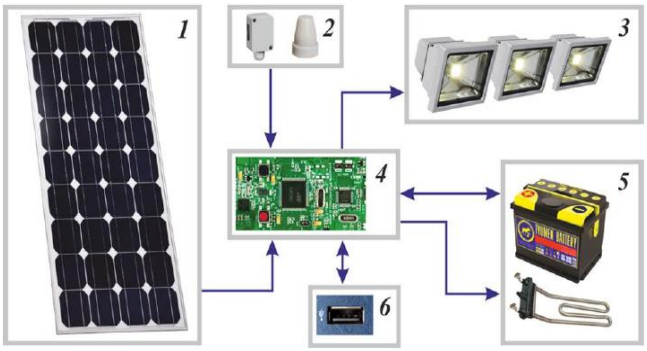


Figure 3: Block Diagram of the Developed Public Lighting

6. Proposed System Description and Operation

The developed street lighting system consists of smart lamps-lamp with a three-stage adjustment of the luminous flux: 10 W-20 W-30 W (the maximum power of each lamp is 30 watts). Each mini-SPP uses two solar panels of 120 W each to power six (6) smart lamps mounted 2 by 2 on three metal poles (As shown in Figure 4). The minimum total power of these lamps will be P_{min} =

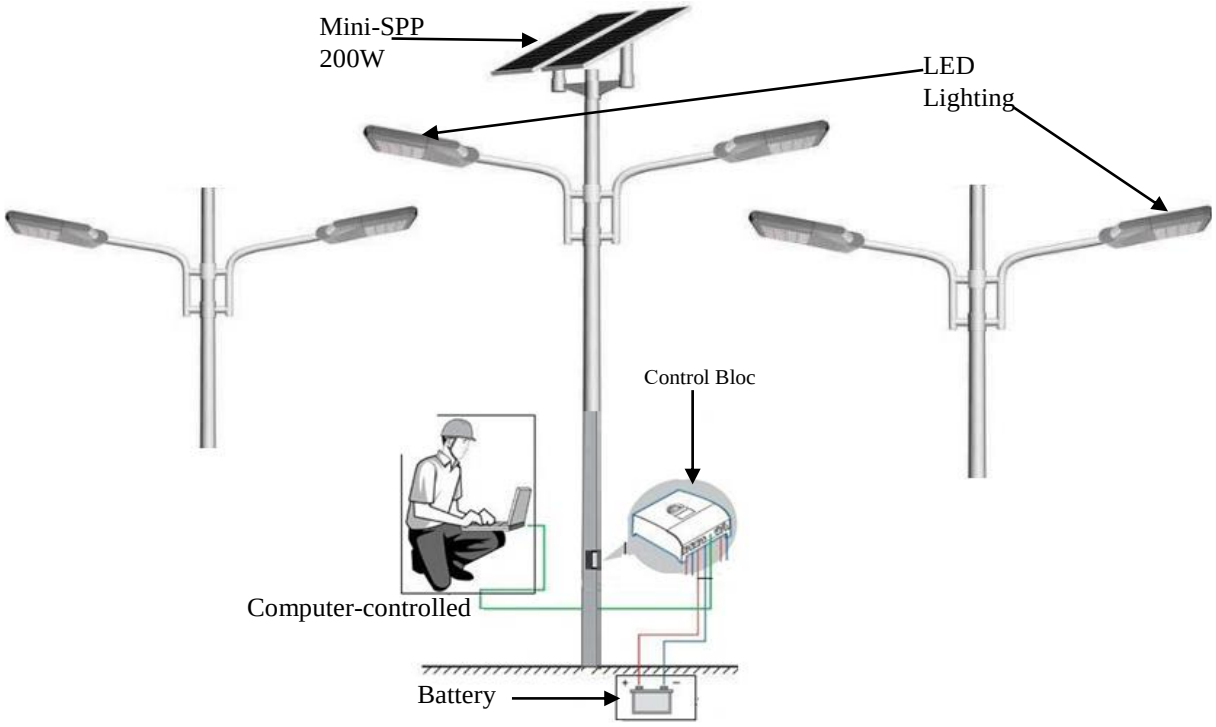


Figure 4: Developed Solar Mini-Power Plant 200 W

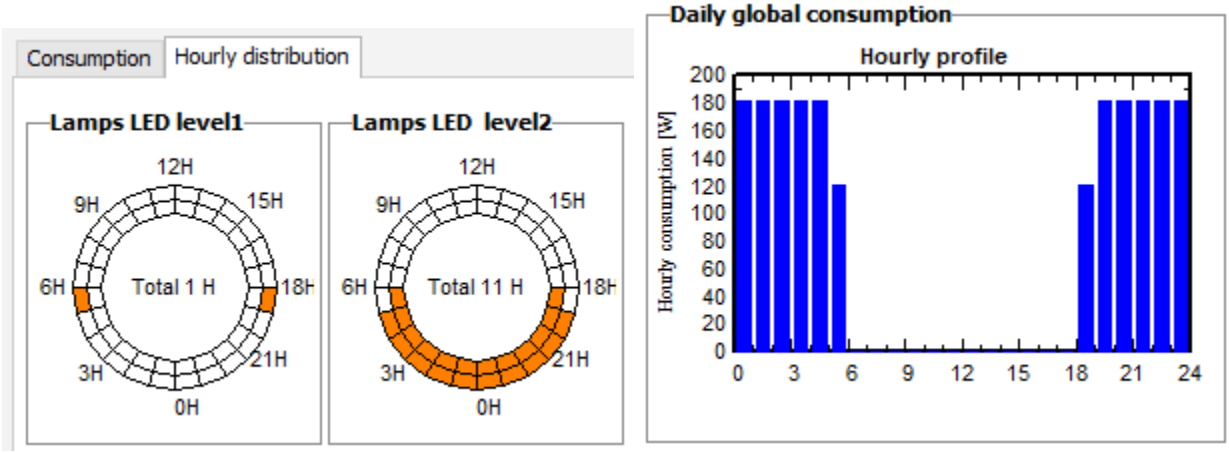


Figure 5: Lighting System hourly Distribution and Daily Consumption

60 W, with a maximum of P_{max} = 180 W. As Shown in the Figure 4, two of the lamps are installed on the central pole supporting the solar panels, the other 4 on the other 2 poles surrounding the first. A voltage of 24 V is transmitted from the battery of the mini solar power plant to the lamps via an underground power line. It should be noted that LED lamps are more advantageous than luminescence or incandescent lamps used in lighting systems in general. The following Table shows the relationship between the power of some lamps (W) and their luminous flux (lumen) [12].

The street lighting system shown in Figure 4 is designed to illuminate boulevards, public areas, etc. The distance between the supports on which the lamps are installed will be about 50 m. In order to reduce the cost of design and installation of the developed system, modernized conventional LED spotlights with systematic adjustment of the luminous flux are used. The following Diagram (As Shown in the Figure 5) shows the use state of this lighting over 24 hours (one day).

The diagram of operation and energy consumption of these lighting lamps shows that from 18: 00 (assumed time of the beginning of the night time),the lamps illuminate under a minimum power of 10 W for 30 minutes after which they illuminate at their maximum power of 30 W each.This minimum power which lasts 30 minutes will also be observed in the morning at the end of the night time so at 05h30 minutes.

7. Solar Panels Orientation and Positioning

In this street lighting system, the electrical energy source is usually located on the same support on which the luminaire is fixed. In this regard, special attention should be paid to the design of attachment of the solar panel to the support, since the amount

of electricity generated will also depend on its orientation. Most often, solar panels are installed at a certain angle (declination) to the sun, in the direction of azimuth to the South or South-East (As Shown in Figure 6) [12, 15].

Table 6: Power Ratio of Various Lamps Types

Incandescent lamp, W	Fluorescent lamp, W	LED lamp, W	Luminous flux, Lm
60	15-16	8-10	700
75	18-20	10-12	1200
100	25-30	12-15	1200
150	40-50	18-20	1800
200	60-80	25-30	2500

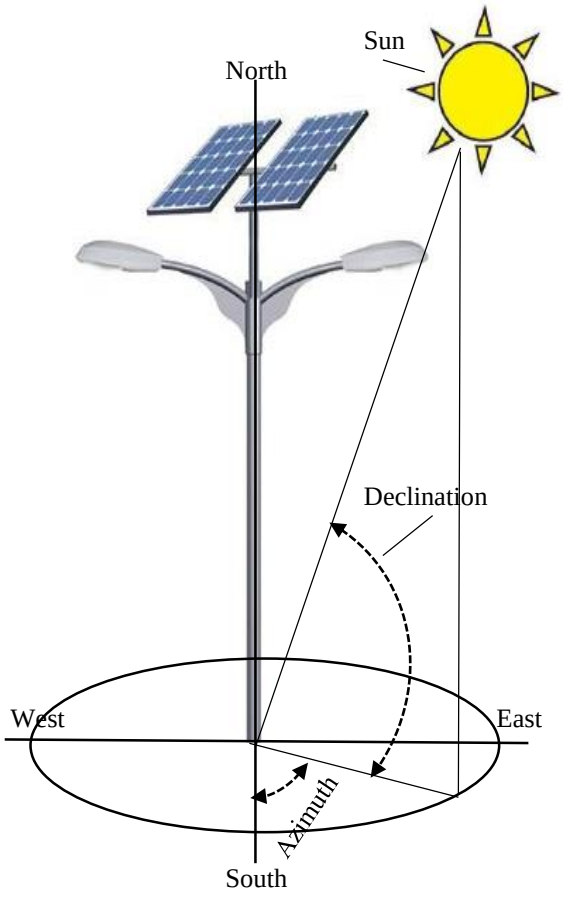


Figure 6: Solar Panel Orientation in Space

It is known that solar panels generate electricity in the absence of direct solar radiation, in cloudy weather, when the sunlight is weak and diffused, the photovoltaic system will produce electricity. The influence of various conditions on the production of solar panels as a percentage of the rated power is shown in Table 7 [12].

The tilt angle of solar panels relative to the earth's surface depends on geographical coordinates. Moreover, to increase the performance of the system, this angle must be changed during the year. As shown in Figure 7, for a fixed orientation throughout the year, the optimal tilt angle for the city of Gitega is 30 degrees.

Table 7: Dependence of Solar Panel Power on Different Conditions

Conditions	**Percentage of the Maximum Radiation (%)**
Bright sun	100
Light cloudy	60-80
Window glass, one layer	91
Overcast weather	20-30
Artificial light indoors	0,2-1,5

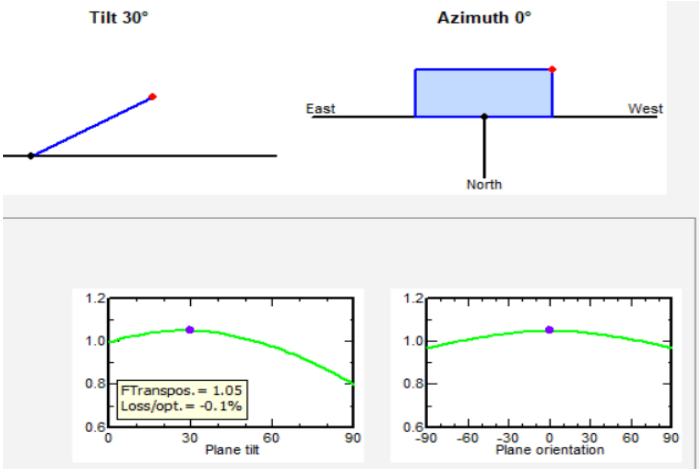


Figure 7: Tilt angle and Azimuth at Gitega

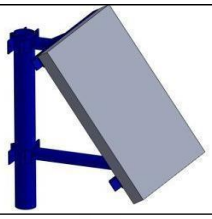
The tilt angle of a solar panel plays an important role, however small changes in the range of 0 to 5 degrees cannot have a serious effect on the system performance. Therefore, the tilt angle for localities (communes) of this region such as Bugendana, Bukirasazi, Buraza, Giheta, Gishubi, Itaba, Makebuko, Mutaho, Nyanrusange and Ryansoro will also be the same.

8. Solar Panels Fixings to Their Supports

Ideally, the luminaire support is fixed to a concrete footing through a metal plate to facilitate its replacement, if necessary. Depending on the function, the aesthetic appearance, the space and the environment, several types of supports can be identified, for example, poles made of steel, aluminum, cast iron, concrete, wooden etc. [12].

At present, metal supports are the most often used. Their manufacturing is more technologically advanced; they have a relatively small mass and retain their operating characteristics for a sufficiently long time. As the metal support is hollow inside, it is possible to run cables through it, connecting the solar panel to the control block and to the storage battery. In general, four (4) fixing points are used to fix a solar panel on a metal support. It is also taken into account that the panel should be oriented vertically, not horizontally. Table 8 shows some standard methods of solar panels fixing.

Table 8: Solar Panels Fixing

Denomination	Solar Module Width	Landing Dimensions	Image
MSU51585 Universal fastening of the solar module 50-150 W to the pipe up to 85 mm	up to 700 mm	Pipe Ø45-85 mm or on the wall	
MSU515220 Universal fastening of solar modules 50150 W to the pipe up to 220mm	up to 700 mm	Pipe Ø80-220 mm	
MSU153085 Universal fastening of solar modules 150-300W on a pipe up to 85mm	up to 1000 mm	Pipe Ø45-85 mm or on the wall	
MSU1530240 Universal fastening of the solar module 150-300 W on the pipe up to 220mm	up to 1000 mm	Pipe Ø80-240 mm	
MSU2x1530280 Universal fastening of two solar modules 150-300W on a pipe (wall) 50-280mm	up to 1000 mm	Pipe Ø 50-280 mm or on the wall	

Note that it is necessary to use a design that facilitates the changeability of the tilt angle of the solar panels according to the time of day or year in order to increase the efficiency of the mini-solar power plant for public lighting.

9. Solar Panels Power Calculation

The energy generated by a solar panel with a power Pw for a selected period is calculated as follows [16, 17]:

$$W = \frac{k \cdot E \cdot P_w}{1000}$$

E – insolation value for the selected period; Pw – solar panel power, k - coefficient equal to 0.5 and 0.7 in summer and winter periods, respectively. This factor takes into account the solar panels heating by the sun, as well as the incidence angle during the day. The insolation value E = 4.91 kWh/day.

The solar panel power to be used is calculated taking into account the daily consumption of all the LED lamps operating during the night hours. Remembering that they are of the adjustable power type and therefore their minimum power $P_{\Sigma lamp}$ will be 60watts, their maximum 180 W. Their maximum daily energy consumption $W_{cons.}$ is given by (1) as follow:

$$W_{cons.} = P_{\Sigma lamp} \cdot t_{\max} \quad (1)$$

где $t_{\max}$ – the operating time of the lamps

The latitude of the city of Gitega and the sun declination on which the average longitude of the day depends, give an interval of the day equal to 12 hours. Accordingly, the maximum daily energy consumption of the system will be 2040 Wh/day according to (1). Since the average number of hours of sunshine is estimated at 9 hours (9h) per day, to determine the daily energy production (generation) of solar panels, it is necessary to multiply their rated power P_n by their operating time per day: $W_{gen} = P_n \cdot t_{op.}$

Where $t_{op.}$ – Operating time (9h), W_{gen} – Daily energy generation. This daily energy produced by the solar panels will be 2016 Wh/day. The 120 W 24 V solar panels that will be chosen from the ranges produced by different manufacturers must have an optimal operating current of at least 6.7 A.

10. Battery Capacity Assessment and Calculation

The developed public lighting systems are intended to work during the night hours, while during this time the solar panels do not generate the electric energy, the electricity stored in the batteries is used. The capacity of the battery to be used, i.e. the amount of energy it can store, plays a big role. Their total capacity depends on the number of storage batteries, and therefore the power of the solar power plant in general. Storage batteries in a solar power plant must meet a number of requirements. They must withstand a large number of charge-discharge cycles. When calculating the parameters of the storage battery, it is necessary to take into account the energy losses during its storage and transformation. Conventional car batteries do not withstand a large number of cycles and have significant self-discharge. Most often, special storage batteries are used for solar power plants. Table 9 shows the states of charge (Battery charge levels) of a 12-volt battery depending on the temperature [12, 18].

Table 9: 12V- Battery Charge Status Assessment

Battery charge level, %	Electromotive force (EMF) , depending on temperature		
	+20…+25°C	+5…–5°C	–10…–15°C
100	12,70 – 2,120	12,80 – 13,00	12,9 – 13,10
75	12,55 – 12,65	12,65 – 12,75	12,75– 12,85
50	12,20 – 12,30	12,30 – 12,40	12,40 – 12,50
25	12,00 – 12,10	12,10 – 12,20	12,20 – 12,30
0	11,70 – 12,00	11,80 – 12,00	11,120 – 12,10

The energy generated and stored in the battery during the day is used at night to power the load (streetlights) of public lighting. This energy shows the capacity of the battery; i.e. how long it can power the load if it is fully charged. The change in battery capacity ΔC during the load supply time $\Delta t_{nig.}$ is determined by the expression (2):

$$\Delta C = \frac{P_n}{U_n} \cdot \Delta t_{nig.} = \frac{P_n}{U_n} \cdot (24 - \Delta t_{dayl.}) \qquad (2)$$

где P_n – rated load power; U_n – rated load voltage; $\Delta t_{nig.}$ – the night time interval (12hours); $\Delta t_{dayl.}$ – day time interval

Battery manufacturers define the final discharge voltage, once reached, the battery must be disconnected from the load and charged, this because deep discharge can damage it. A battery should not be more than 70-80% discharged for it to work for a long time. The battery discharge level is determined by the expression [16]:

$$S_r = \frac{C_n - C_{\min}}{C_n} \cdot 100\% = \frac{\Delta C}{C_n} \cdot 100\% \qquad (3)$$

The required capacity C_n of the accumulator battery is obtained from (3) while taking into account (2):

$$C_n = \frac{100}{S_r} \cdot \frac{P_n}{U_n} \cdot \Delta t_{nig.} \qquad (4)$$

In the city of Gitega, taking into account that the nighttime interval is 12hours, with the 200-Watt solar panels as it was said above, a 120 Ah will be ideal for this mini solar power plant consisting of six LEDs of street lighting. Its energy capacity will be 2880 Wh. The higher the voltage supplied to the load, the lower the capacity can be, since the discharge current $I_r = \frac{P_n}{U_n}$ of the storage battery will be lower. For conventional lead acid batteries, the maximum discharge current is limited to a value that, in amperes, is between 5 and 25 times the battery capacities. The lower the discharge current, the lower the power loss, and therefore the higher the efficiency of the system. It is the increase in voltage and decrease in current that explains the use of voltage 24 V instead of 12 V, when placing "smart lights" at a significant distance from the solar mini-power plant.

11. Conclusions

The task of our research was to develop and study the public lighting system of the city of Gitega using mini solar power plants as a source of renewable energy. This project was designed taking into account the current state of the lighting in Gitega, the current global trends, as well as the great importance of public lighting in general and through photovoltaic solar power plants in particular such as: increasing the citizens' safety, reducing the number of crimes and offenses and increasing their detection potential, reducing cabling costs for transmission systems, reducing the burden on emergency services and electrical technicians. This public lighting system with an automated solar power plant will regulate energy consumption, monitor the integrity of the equipment and quickly send a signal to operating personnel on emergencies on the network. The current state of public lighting in the city of Gitega and its surroundings is poor due to the lack of financial resources needed using traditional energy sources. The development of such solar street lighting systems that would not require large investment costs, promote clean energy, reduce energy consumption and which maintenance would be carried out once or twice a year, is very relevant.

Conflict of Interest

The authors declare no conflict of interest regarding the publication of this paper.

References

[1] N. Ntawuhorakomeye, M.P. Belov, "Development and Analysis of Public Lighting by Mini Solar Power Plants in The Republic of BURUNDI," in 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), IEEE, St. Petersburg, Russia: 785–790, 2020, doi:10.1109/EIConRus49466.2020.9039030.

[2] Ministry of Energy and Mines of the REPUBLIC OF BURUNDI, Opportunités dans le secteur des énergies renouvelables au Burundi Ministère de l'Energie et des Mines REPUBLIQUE DU BURUNDI, Bujumbura,Burundi, 2012.

[3] D. V. Shatilov, Проект «Программы внедрения автономных систем энергоснабжения уличного освещения на основе использования экономичных светодиодных светильников и солнечных батарей для стимулирования применения методов энергосбережения в жилищно-коммунальном хозяйстве, Svetorezerv LLC, Moscow, 2008.

[4] C. Muchunku, K. Ulsrud, D. Palit, W. Jonker-Klunne, "Diffusion of solar PV in East Africa: What can be learned from private sector delivery models?," WIREs Energy and Environment, **7**(3), 2018, doi:10.1002/wene.282.

[5] I.A. Suvorov, Электротехнологические промышленные установки и освещение (2009) — книга автора Суворова И.А. | НЭБ, Kirov, 2009.

[6] R. Ciriminna, F. Meneguzzo, L. Albanese, M. Pagliaro, "Solar street lighting: a key technology en route to sustainability," Wiley Interdisciplinary Reviews: Energy and Environment, **6**(2), e218, 2017, doi:10.1002/wene.218.

[7] V.K. Satyakar Veeramallu, S. Porpandiselvi, B.L. Narasimharaju, "Analysis and implementation of soft-switched bidirectional buck-boost DC-DC converter for solar PV-fed LED street lighting systems," International Journal of Circuit Theory and Applications, **47**(12), 1990–2018, 2019, doi:10.1002/cta.2706.

[8] G.P. Kolesnik, Электрическое освещение Основы проектирования Учебное пособие.pdf, Vladimir, 2006.

[9] Metode penelitian Nursalam, 2016, A.. Fallis, "GUIDE TECHNIQUE MODULE 3 : ECLAIRAGE PUBLIC," Journal of Chemical Information and Modeling, **53**(9), 1689–1699, 2013.

[10] N. Khatavkar, A.A. Naik, B. Kadam, "Energy efficient street light controller for smart cities," in 2017 International Conference on Microelectronic Devices, Circuits and Systems, ICMDCS 2017, Institute of Electrical and Electronics Engineers Inc.: 1–6, 2017, doi:10.1109/ICMDCS.2017.8211714.

[11] R.C. Marques, R.R. Geddes, "THE USE OF PPP ARRANGEMENTS IN STREET LIGHTING: A WIN-WIN OPTION?," in Annals of Public and Cooperative Economics, Wiley-Blackwell Publishing Ltd: 311–327, 2019, doi:10.1111/apce.12229.

[12] A.S. Stolkov, Разработка эффективной системы уличного освещения с питанием от солнечной мини - электростанции, Altai State Technical University, 2018.

[13] Р.Ф.Д.А. Медведева, Автоматизированные системы управления наружным освещением. Опыт Москвы - ЭнергоСовет.ru, 2011.

[14] M. Wadi, A. Shobole, M.R. Tur, M. Baysal, "Smart hybrid wind-solar street lighting system fuzzy based approach: Case study Istanbul-Turkey," in 2018 6th International Istanbul Smart Grids and Cities Congress and Fair (ICSG), IEEE: 71–75, 2018, doi:10.1109/SGCF.2018.8408945.

[15] E. Alsakkaf, "An Automatic Off-Grid Solar Street Lighting System," IJIREEICE, **4**(3), 1–3, 2016, doi:10.17148/IJIREEICE.2016.4301.

[16] M.P.Belov. N. Ntawuhorakomeye, "TECHNICAL METHOD OF CALCULATING POWER OF SOLAR POWER PLANTS IN TROPICAL CLIMATES AS IN THE REPUBLIC OF BURUNDI," Izvestiya SPbGETU "LETI," **8**, 66–74, 2019.

[17] N. Ntawuhorakomeye, M.P. Belov, "Replacement of Liquid Fuel by Solar Electrical Installation for the Production of Electric Energy by Diesel-Generators of the «REGIDESO» Industrial Group in the Republic of Burundi," in 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), IEEE: 1021–1024, 2019, doi:10.1109/EIConRus.2019.8656909.

[18] M.D. Vijay, K. Shah, G. Bhuvaneswari, B. Singh, "LED based street lighting with automatic intensity control using solar PV," in 2015 IEEE IAS Joint Industrial and Commercial Power Systems / Petroleum and Chemical Industry Conference (ICPSPCIC), IEEE: 197–202, 2015, doi:10.1109/CICPS.2015.7974074.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1366-1371 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Investigating Students' Computational Thinking through STEM Robot-based Learning Activities

Sasithorn Chookaew[*,1,2], Suppachai Howimanporn[1,2], Santi Hutamarn[1,2]

[1]*Department of Teacher Training in Mechanical Engineering, Faculty of Technical Education, King Mongkut's University of Technology North Bangkok, 10800, Thailand*

[2]*Industrial Robotics Research and Development Center, King Mongkut's University of Technology North Bangkok, 10800, Thailand*

ARTICLE INFO

Article history:
Received: 26 September, 2020
Accepted: 09 December, 2020
Online: 16 December, 2020

Keywords:
Educational robotics
STEM education
Computational thinking

ABSTRACT

Nowadays, robots are an important learning tool in education and are increasingly used inside and outside the classroom to foster the development of students' knowledge, skills, and attitudes connecting to a real-world situation. In past years, using robots in STEM (science, technology, engineering, and mathematics) education has been proposed as one of the strategies to integrate the different fields in order to construct more effective projects and innovations. Especially, engineering is the combination of math and science in solving a problem. Nevertheless, students could not understand and appreciate how to apply the knowledge of interdisciplinary integration to operate certain tasks by the engineering method. This research article presents STEM robot-based learning activities (STEM-RoLA) with sixty high-school students who were studying science and engineering. The obtained results show that the proposed STEM-RoLA is beneficial for students, especially when compared with high and low robotics performance students. The results found that high robotics performance students have higher computational thinking than low robot performance students, and they have a positive engagement response in the learning activity.

1. Introduction

At present, there is a proposal to increase robot usage in educational establishments undoubtedly because the robots have been integrated into the learning process to encourage students to think and connect to a real-world situation. Hence, using robots for education has gained a lot of attention from researchers and educators. This paper is an extension of work originally presented in the 7th International Congress on Advanced Applied Informatics (IIAI-AAI 2018) [1].

Nowadays, many studies focus on educational robots that help students develop essential skills inside and outside the classrooms that are significant in connecting the concept to real-life. Moreover, the learning activities were designed in consideration of promoting STEM discipline (Science, Technology, Engineering, and Mathematics) with a challenging learning mission in different situations [2]. Computational thinking (CT) is one of the essential core skills for students in the 21st century that is often regarded as the basic skill of computer science. CT is a kind of analytical thinking that uses the general way of thinking mathematically to solve real-life situations with scientific thinking [3]. The robot kit is a learning tool for use in the educational system, it is easy to use for students and enables them to have hands-on experience of activities and allows for a number of challenges for students to give them opportunities of robotics experience [4].

In the education context of Thailand, the students in high school have learned various subjects independently, such as Science (Physics, Chemistry, Biology, Mathematics), and Computer Programming. However, they may not understand how to apply the knowledge of interdisciplinary integration to perform or operate certain tasks in the engineering process [5].

The Thai government has attempted to push STEM education to enhance learning seamlessly for a new generation of students. Also, with the growing interest in STEM, educational robots have several benefits. Many educators have indicated that STEM integration would help enhance engineering students' learning performance with a flexible learning process. Besides, using

[*]Corresponding Author Sasithorn Chookaew, Faculty of Technical Education, King Mongkut's University of Technology North Bangkok, Thailand, Email: sasithorn.c@fte.kmutnb.ac.th

appropriate robotics-based STEM activities can be an effective way to introduce students' thinking to operate tasks systematically [6].

With the swift development of the field of robotics, it was proposed that robotics not only engage in active learning environments but also develop computational thinking knowledge and skills for students [7]. Of course, educational robotics was popular to bring an adaptive instrument for learning. Therefore, the robotics has been to promote STEM content, innovation, and creativity among students through an educational robotics competition [8].

This learning activity has integrated the benefits of robotics with the STEM education framework to promote senior high school students' computational thinking with the learning engagement process of students that was held outside the classroom for three days. This study aims at addressing the following research study:

- Do students who participate in STEM robot-based learning activities have computational thinking?
- What are the students' engagements in the STEM robot-based learning activities?

2. Literature Review Literature Review

2.1. Using Robots in Education

Robotics is a branch of engineering that includes a variety of subject matters that combine with STEM disciplines (science, technology, engineering, and mathematics). A robot is a machine applied in a variety of tasks such as industrial systems, medicinal purposes, and educational tools. Too many researchers state that educational robotics is a core element when STEM is a significant tool to integrate into the curriculum for students in the 21st century. A recent study has found that using a robot in education is increasingly being defined as an excellent instrument of teaching and learning that integrates STEM education, which is crucial for the future success of students [9].

Many studies proposed the benefit of an educational robot able to motivate students' learning with authentic learning activities based on real-world problems [10-11]. In [12], the author proposed robotic cooperative learning to promote students' critical thinking, and STEM interest indicates that students participating in robotic service-learning can benefit from this experience. Besides, robot hands-on activities can be easily adopted in classrooms to connect interdisciplinary core ideas for engineering design and STEM learning [13-14]. A robot in education can be used not only inside and outside of the classroom but also in meaningful learning activities for students as well.

Robotics makes the students able to connect and find relevance in the experience, together with the complexity of the global situation. So, it can be suggested that this is how to apply the concept of a robot for enhancing computational thinking skills.

2.2. Computational Thinking

The computational thinking basis concept of mathematics education research by Seymour Papert [15] proposed the aspects of computation that are: how to use computation to create new knowledge; how to use computers to enhance thinking and how to change patterns of access to knowledge. There is the basic concept that using technology can provide new ways to learn and teach with the logical thinking process. After that, computational thinking was described often by many researchers worldwide. In the last twenty years, computational thinking has become a fundamental skill for everyone in every field, not just for computer scientists, especially for students in the 21st-century who need skills such as analytical ability, reading, writing, and arithmetic to be processed systematically.

Based on the core concepts of computational thinking as a fundamental skill for everyone's analytical ability along with reading, writing and arithmetic, in [16], the author proposed that these should include decomposition, pattern recognition, abstractions, and algorithm design. In [17], the author classified computational thinking core components into three dimensions, including computational concepts which focus on the concepts that students employ when they are learning. Computational practices focus on problem-solving practices that occur in the learning process. Computational perspectives focus on the students' understanding of themselves and can connect learning with the world around them.

Measuring students' computational thinking is complicated, but it is a necessary task for understanding the effectiveness of the STEM-RoLA project. Our investigation focused on the implications of claims about three computational thinking dimensions consisting of computational concepts: the scientific concepts that students employ to learn and understand during activities, computational practices: problem-solving practices that occur in the learning process, and computational perspectives: students' understanding of themselves and the relationships between group members.

3. Methodology

3.1. Participants

The study was implemented in a public school in central Thailand. We selected a curriculum that prepares all students to be a scientist or an engineer. The participants in this study were sixty high school students, grades 10-12 (all males). The students were separated into twenty groups, so that each group had three members, and they learned and worked together in those groups.

3.2. Instruments

- The items in this observation checklist were developed to the three dimensions of computational thinking that cover all concepts of the robot activities including fourteen items: computational concept (4 items), computational practices (5 items), and computational perspectives (4 items). For this observation checklist, the mentors have rated the students on a 5-point Likert scale (1 = Failed, 2 = Passed, 3 = Acceptable, 4 = Good, 5 = Excellent). The scale validity was evaluated by five professors and teachers of educational robotics with more than seven years of experience. The internal consistency for the overall scale was 0.82.

- The items in this engagement questionnaire were adopted from a revised version in [5]. There are three dimensions of students' engagements after attending the robot learning activity that consist of eleven items asking about students' experiences toward STEM-RoLA. Three dimensions included behavioral engagement (3 items), cognitive engagement (4 items), and emotional engagement (4 items). The Likert scale is a five-point scale that allows the individual to express how much they agree or disagree with each item (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = agree, 5 = Strongly Agree). The internal consistency for the overall scale was 0.79.

3.3. Robot-based learning activity

The commercial robotic kits chosen in this study were mBot platform that operates and engages the students' learning experience of the mechanics, electronics, control systems, and computer programming [18]. The mBot is the easiest educational robot kit for student learning in robotics. The robot shape is easy to assemble and contains approximately 45 pieces in total, so students have a sense of achievement when they are able to assemble it quickly.

Normally, the components of mBot included the body, main control board, sensors, communication (Bluetooth), and power supply (battery). Also, it has three pre-set control modes of supporting student learning, consisting of obstacle avoidance mode, line-follow mode, and manual control mode. The students could feel that they had achieved and feel proud when they could program the mBot easily without writing difficult codes and language.

In addition, mBlock is a block-based coding platform for Scratch programming. It is easy-to-use software that makes teaching and learning coding fun and interactive. Scratch 2.0 is a very popular graphical logical programming software for teachers and students' learning. It has been proved to be one of the easiestto use graphical programming tool. Therefore, we have appled mBlock software to use Scratch coding to program Arduino and robots. Also, graphical programming is a visual programming technique where visual block connections are used for a student to code instead of texts, making it easy for non-coders to implement algorithms. It plays a meaningful role in enhancing computational thinking in the learning process.

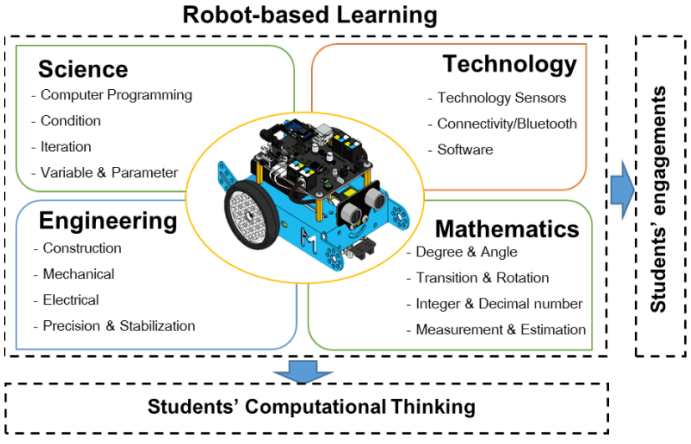


Figure 1: The STEM robot-based learning framework

Figure 1 shows the overall research of STEM-RoLA. We attempted to design the learning activity with STEM integrated. The students have learned about the mBot kits in a stepwise manner that describes its components with scientific understanding.

To achieve the tasks in this learning activity, students are encouraged to employ knowledge, skills, experience, and a teamwork process based on three dimensions of computational thinking:

- Computational thinking concepts, the students employ the knowledge when they code or control the robot for operating the mission and competition.
- Computational thinking practices, the students solve the problems that occur in activities: experimenting, iterating, and testing.
- Computational thinking perspectives, the students' understanding of themselves and their relationships with a group member.

Figure 2: The steps of STEM robot-based learning activity

To be an effective activity, we provide mentors who are pre-service engineering teachers. The mentor will help students while they work in a group during the activity. The mentors can provide

technical consultation and can be a source of support when students need it most. For the time being, each group has to come up with better solutions by integrating the knowledge of STEM and robotics, while their computational thinking is promoted accordingly.

The eight-steps to facilitate their learning process with hands-on experience lasts approximately twenty-one hours as follows:

- *Step 1: Robot introduction*: the first step addresses motivation to student-related robots that are important in daily life. After that, they introduce the functions of the robot and the introduction of the robot's components based on the engineering design process. At this moment, the students have to get acquainted with the robot components through several mini-labs such as the main control board, sensors, communication options (USB cable, Bluetooth, Wireless), and power supply (battery). In addition, the robot has three different modes of functionalities.
  - Mode A – Remote control driving or mobile device: the robot can be controlled by arrow buttons on the device to drive it forwards and backwards as well as to turn left or right.
  - Mode B – Obstacle avoidance: the robot is driven forward on its own until it detects an obstacle. If the robot detects an obstacle, it turns then continues driving forward.
  - Mode C – Line following: the robot drives forward while following a line
- *Step 2: Robot assembly:* The students were separated into twenty groups (3 students per group) then students of each group worked together. They can be learning at their own pace with support from the group mentor. This step requires students to get acquainted with each part and start to join each part together by robot function. Then, they learn three pre-set modes consisting of IR control mode, obstacle avoidance mode, and line-following mode. After that, they get the robot started by pressing on-board and push the button to change the mode.
- *Step 3: Primary programming*: the students can learn an easy programming language to control the robot. They use mBlock based on Scratch 2.0 software that is graphical programming for writing on a laptop or personal computer. The student receives the coding questions for programming practice. For example, how to control the robot to move forward or to turn left or right. They can drag the module block to move the robot area. During this step, the students in each group must analyze the facing proposition and situation in which they require to test-run-revise the code in a stepwise manner.
- *Step 4: Robot in the field:* the students from each group work together on the given tasks, a range of tasks designed to test the robot in the field. Each group has to adapt the physical robot as well as the programming instructions to overcome different challenges. This step is the basis for improving reliable robot working. The robot was assembled by students who programmed it to perform certain tasks and then there is the testing to operate, prove system functionality, and gain confidence that a system will perform as expected. Testing in the field also allows the student to learn the weaknesses, problems, and constraints of the robot and improve upon them.
- *Step 5: Robot mission:* the students learned to analyze and solve the problems in a logical way by taking the knowledge integration of STEM. The adjustment of one point will affect the related points. In this step, each group is faced with different problems in their robot's settings and programming. For example, the robot starts to move at a certain speed. When there is a barrier in front, the robot will turn right; when there is not, the robot keeps moving forward.
- *Step 6: Sharing:* The students have opportunities to share knowledge. What they have learned about each other helps build positive thinking. The representative student presents ideas to organize and control the robot to carry out each task in the missions. The students in each group help to justify the approach to solving the problems and present their ideas.
- *Step 7: Practice before the battle:* the students in each group have a rehearsal before the competition. They have analyzed and discussed opportunities and weaknesses to compete in the competition. This will increase their team's confidence in the final robot, while the final configuration may be needed. They receive the robot tasks for learning line-follow and obstacle avoidance and find a solution and strategy for solving the problem. For example, the students have learned how to program robots to take one step at a time. Meanwhile, students get to know the mechanical structure and understand how to control a robot.
- *Step 8: The robot competition:* the robot competition requires students to control the robot that makes learning robot programming simple and fun. The student members in each group are encouraged to apply knowledge of what they have learned to accomplish the goal effectively on the robot competition. There are two separate modes, including an autonomous robot following a task and a control robot via a mobile application on the Balloon Battle Game. They have to use the Makeblock open-source to control the robot with a mobile device, as shown in Figure 3.

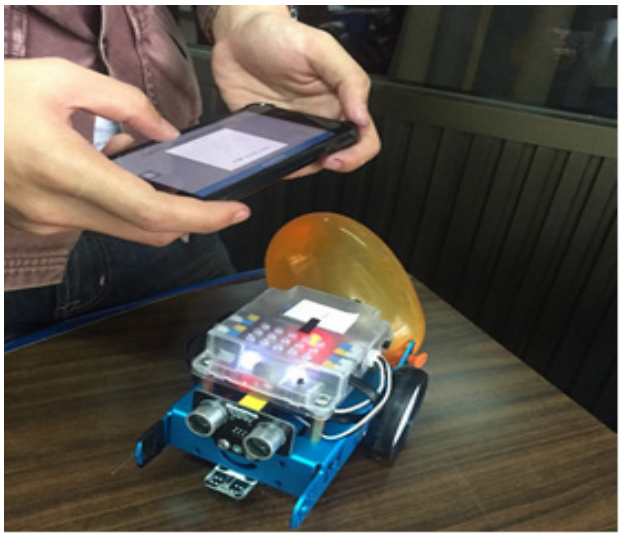


Figure 3: Illustration of a robot controlled by a mobile device

This moment is vital to each group since the robot's performance during rehearsal may be ineffective when competing with others. For this step, each robot has a needle and a balloon attached to the front and back of its body. Then each group's robot tries to stab the balloon's competitor while maintaining their

attached balloon. The loser is eliminated from the competition. For the final of the competition, each qualified group has ten minutes to come up with a better strategy for controlling the robot in a battle while maintaining their balloons. The last group to have a balloon at the end of the competition is the winner.

4. Findings

To understand better the effects of this part of the study, we compared the results with two separate groups of students' data consisting of a tasks score (30 points) and the robot competition score (70 points) total 100 points. Therefore, twenty groups demonstrated the difference between a high robotics performance group (HIRP) for the top ten groups with M=79.00 (SD =12.10) and a low robotics performance group (LORP) for the bottom ten groups. These were elaborated with M= 29.50 (SD =2.69). The findings related to the research questions were demonstrated.

4.1. The students' computational thinking

To analyze the differences between the students' robotics performances in two groups of three computational thinking components included computational thinking concepts (CTC), computational thinking practices (CTP), and computational thinking perspectives (CTPP), and these are shown in Table 1.

Table 1: Means and SDs of the students' computational thinking

CT Component	LORP		HIRP	
	M (SD)	***Remark***	***M (SD)***	***Remark***
CTC	4.01 (0.72)	Good	4.13 (0.22)	Good
CTP	2.88 (0.74)	Acceptable	4.75 (0.43)	Excellent
CTPP	3.25 (0.83)	Acceptable	4.13 (0.54)	Good

As the descriptive statistics in Table 1 show for the proposed STEM-RoLA in computational thinking concepts (CTC), both the HIRP and LORP students have concepts of the robot at a good level with M = 4.01 (SD = 0.72), and M = 4.13 (SD = 0.22), respectively.

Based on the results in computational thinking practices (CTP), the HIRP students performed tasks in each mission at an excellent level (M = 4.75, SD = 0.43), better than the LORP students at an acceptable level (M = 2.88, SD = 0.74).

Furthermore, in computational thinking perspectives (CTPP), the HIRP students tended to have positive perspectives at a good level (M = 4.13, SD = 0.22) while those who were LORP students, it was at an acceptable level (M = 3.25, SD = 0.83).

4.2. The term of students' engagements

The students' engagement results were analyzed based on quantitative and qualitative data for proposing STEM integrated robotics learning. We have presented three different aspects consisting of behavioral, cognitive, and emotional engagement as presented in Table 2.

In the first dimension, the HIRP students revealed behavioral engagement in that they could manage the function of members and have confidence in learning (M=4.63, SD=0.53). The entry indicated that the HIRP students were good at time management and working well together, as they expressed during the interviews, for example:

Table 2: Means and SDs of the students' engagements

Component	LORP		HIRP	
	M (SD)	***Remark***	***M (SD)***	***Remark***
Behavioral	4.32 (0.77)	Agree	4.63 (0.53)	Strongly Agree
Cognitive	4.47 (0.64)	Agree	4.45 (0.67)	Agree
Emotional	4.40 (0.71)	Agree	4.62 (0.51)	Strongly Agree

"Through this activity, we helped in my group about the coding program and explained to each other. I think I have learned a lot about cooperation and teamwork. I attend a robot activity with my friends, it made me have so much fun." (Student A)

"I am so excited about the robot because it is admirable to have this experience that I can construct a robot by myself. Besides, we have an opportunity to cooperate and share the ideas for working." (Student B)

While the LORP students had trouble with cooperating and working together, they did not enjoy learning in a group with members very much, and they had less time because they could not manage time (M=4.32, SD=0.77), for example:

"In my group, we have many members, so somebody does not involve in some activities. We can't organize the tasks and events on time." (Student C)

"I have less time to work on many tasks. It must spend a lot of time to solve the problems. Sometimes, I think I like to work without my friends." (Student D)

For cognitive engagement, the entry indicated that the HIRP students can reflect higher thinking skills on applications in their daily lives (M=4.45, SD=0.67), for example:

"I apply multiple tasking skills, and I can try when not sure. It is important to know how to employ the basic engineering principles for this activity." (Student A)

"For this activity, I think I have a thinking process and much knowledge I play it again when I come back home. Unfortunately, while I have learned a lot of programming, I never took any classes on robotics" (Student B)

While the LORP students just reflected what they had experienced from the activity by putting more effort before the success (M=4.47, SD=0.64), for example:

"I am upset about how to complete the work. However, I think I attempt to find the solution myself." (Student C)

"In the activity, I do not plan, it's not a success. I have attempted, but still failed, again and again, so I consulted the mentor." (Student D)

In the end, the HIRP students have emotional engagement. They enjoyed the learning activities (M=4.62, SD=0.51), especially the students in HIRP who have positive responses such as:

"I think it could be an interesting experience. I like to control a robot to follow my instructions. Each task, I am quite confident that a robot can do some pretty fun stuff." (Student A)

"I very enjoyed the activities in the activity because I think a robot can be a good assistant in the future. It helps me to be more productive in performing daily activities." (Student B)

On the other hand, the LORP students did not concentrate on some activities. However, they exposed their emotions towards the assistance of peer members in the group that could encourage them to proceed with the tasks. So, they have an attitude towards this (M=4.44, SD=0.71), as they expressed during the interviews, for example:

"I think a robot is difficult to use because it was always programmed. I'm running into a lot of infrared sensor issues. I'm attempt to do the typical line follower activity and also go ahead and stop when it finds a black line." (Student C)

"This activity was held during the semester. I have a lot of homework, so I don't concentrate on some activities. I think it should be held in my free time." (Student D)

There was unanimous agreement from the students present that, as a result of this robot activity, the students had significantly developed their computational thinking to solve the problems and challenges, their ability to work as a group, and their resilience. Also, students said they greatly enjoyed participating in the challenges and wanted more opportunities to participate in such STEM-RoLA.

5. Conclusions

Within Thailand's educational context, the integration of interdisciplinary knowledge to enhance crucial skills thinking, and a positive attitude are also important to realize goal. So the robotics STEM was integrated through active learning and the new learning tool has an effective strategy. The robotics challenges allow students to grasp the fundamentals of basic programming and develop their computational thinking with STEM education in an engaging way. The STEM-RoLA was held to motivate the senior high school students in three days (21 hours), consisting of eight steps in total, from robot introduction, assembly, programming, testing in the field, robot mission, sharing, practice before the battle, and final robot competition. The students established their robotics knowledge with hands-on experience based on the STEM strategy. Therefore, the obtained results show that the proposed STEM-RoLA is beneficial for the students, especially when compared with high and low robotics performance students. The results find that high robotics performance students have higher computational thinking than low robotics performance students. Also, they have positive engagement responses to the learning activity.

Nevertheless, the learning activities have shown that there are some limitations due to the availability of the school context, the information, and the educational context students already have before they learn robot information that might affect effectiveness. Also, the number of robot kits are quite expensive and inflexible to learning in some contexts. Future studies may explore whether there are specific learning strategies and learning tools that are particularly useful based on the attitudes of the students. Based on the result, the present research involved all males. So, future studies should clarify the results that can be described by gender differences in learning outcomes and assess the long-term effects on attitudes.

Acknowledgment

This study is supported in part by King Mongkut's University of Technology North Bangkok.

References

[1] S. Chookaew, S. Howimanporn, P. Pratumsuwan, S. Hutamarn, W. Sootkaneung, & C. Wongwatkit, "Enhancing High-School Students' Computational Thinking with Educational Robotics Learning" in 2018 International Congress on Advanced Applied Informatics (IIAI-AAI), 204-208). 2018, doi:10.1109/IIAI-AAI.2018.00047.

[2] T. J. Kopcha, et al. "Developing an integrative STEM curriculum for robotics education through educational design research" Journal of Formative Design in Learning, **1**(1), 31-44, 2017, doi.org/10.1007/s41686-017-0005-1

[3] J. M. Wing, "Computational thinking" Communications of the ACM, 49(3), 33-35, 2006.

[4] B. Zhong, Y. Wang, "Effects of roles assignment and learning styles on pair learning in robotics education" International Journal of Technology and Design Education, 1-19, 2019, doi.org/10.1007/s10798-019-09536-2

[5] S. Hutamarn, S. Chookaew, C. Wongwatkit, S. Howimanporn, T. Tonggeod, S. Panjan, "A stem robotics workshop to promote computational thinking process of pre-engineering students in Thailand: STEMrobot" in 2017 International Conference on Computers in Education, 514-522, 2017.

[6] C. Chalmers "Robotics and computational thinking in primary school" International Journal of Child-Computer Interaction, **17**, 93-100, 2018, doi.org/10.1016/j.ijcci.2018.06.005

[7] E. B. Witherspoon, R. M. Higashi, C. D. Schunn, E. C. Baehr, R. Shoop, "Developing Computational Thinking through a Virtual Robotics Programming Curriculum," ACM Transection on Computing Education, **18**(1), 1–20, 2017, doi.org/10.1145/3104982.

[8] A. Eguchi, "RoboCupJunior for promoting STEM education, 21st century skills, and technological advancement through robotics competition," Robotics and Autonomous Systems, **75**, 692–699, 2016, doi.org/10.1016/j.robot.2015.05.013.

[9] T. Larkin, "Topic Order in Introductory Physics and its Impact on the STEM Curricular Ladder" International Journal of Engineering Pedagogy, **7**(1), 136-150, 2017, https://www.learntechlib.org/p/207332/.

[10] C. Julià, J. Ò. Antolí, "Impact of implementing a long-term STEM-based active learning course on students' motivation" International Journal of Technology and Design Education, **29**(2), 303-327, 2019, doi.org/10.1007/s10798-018-9441-8.

[11] M. Barak, M. Assal, "Robotics and STEM learning: Students' achievements in assignments according to the P3 Task Taxonomy—practice, problem solving, and projects" International Journal of Technology and Design Education, **28**(1), 121-144, 2018, doi.org/10.1007/s10798-016-9385-9.

[12] P. Mosley, G. Ardito, L. Scollins, "Robotic cooperative learning promotes student STEM interest" American Journal of Engineering Education, **7**(2), 117-128. 2016.

[13] S. Ziaeefard, M. H. Miller, M. Rastgaar, N. Mahmoudian, "Co-robotics hands-on activities: A gateway to engineering design and STEM learning" Robotics and Autonomous Systems, **97**, 40-50, 2017, doi.org/10.1016/j.robot.2017.07.013.

[14] S. Chookaew, C. Silawatchananai, S. Hutamarn, S. Howimanporn, W. Sootkaneung, C. Wongwatkit, "An investigation of vocational students' attitude towards STEM robotic activities" in 2019 International Conference on Computers in Education (ICCE) 134–138, 2019.

[15] S. Papert, "An exploration in the space of mathematics educations" International Journal of Computers for Mathematical Learning, **1**(1), 95-123, 1996, doi: https://doi.org/10.1007/BF00191473.

[16] J M. Wing, "Computational thinking and thinking about computing" Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, **366**(1881), 3717-3725, 2008, doi.org/10.1098/rsta.2008.0118.

[17] K. Brennan, M. Resnick, "New frameworks for studying and assessing the development of computational thinking" in 2012 Annual Meeting of the American educational research association, Vancouver, Canada 1, 1-25, 2012.

[18] P. P. Merino, E. S. Ruiz, G. C. Fernandez, and M. C. Gil, "A Wireless robotic educational platform approach," in 2016 International Conference on Remote Engineering and Virtual Instrumentation, 2016, 145–152.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1372-1380 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Effect of Starch Oxidation Degree on the Properties of Hydrogels from Dialdehyde Starch and Polyvinyl Alcohol

Jahel Desire Carrera[1], Daniela Alejandra Viteri Narváez[1], Marco Leon[2], José Francisco Alvarez-Barreto[*,1]

[1]*Universidad San Francisco de Quito, Department of Chemical Engineering, Institute for the Development of Alternative Energies and Materials, IDEMA, Biomaterials Laboratory, 170901, Ecuador*

[2]*Universidad San Francisco de Quito, Department of Mechanical Engineering, 170901, Ecuador*

ARTICLE INFO

Article history:
Received: 30 September, 2020
Accepted: 28 November, 2020
Online: 16 December, 2020

Keywords:
Dialdehyde starch
Hydrogels
Polyvinyl alcohol
Swelling
Controlled drug release

ABSTRACT

Starches have been applied as biomaterials due to their wide availability and biocompatibility. These have also been modified by oxidation, resulting in dialdehyde starch (DAS), to improve their stability in water and mechanical properties. Cassava starch with a low oxidation degree has been introduced into hydrogels based on polyvinyl alcohol (PVA) to improve their properties. However, the behavior of these materials with starch at different oxidation levels has not been previously explored. In the present work, the effect of the oxidation degree of cassava starch on the physical and chemical properties of DAS-PVA hydrogels was evaluated. To modify the degree of oxidation, different concentrations of H_2O_2 were used, and a high degree of oxidation was achieved by incorporating copper sulfate II as a catalyst. Oxidation was confirmed by quantification of carbonyl groups and Fourier Transformed Infrared Spectroscopy. Hydrogels with low and medium oxidation DAS displayed greater swelling, but also lower stability over time. Similarly, scanning electron microscopy confirmed greater porosity in them. On the other hand, hydrogels with high oxidation DAS had lower water absorption capacity, but greater stability over time. Regarding the controlled release of ibuprofen, as a model drug, hydrogels formulated with low and medium oxidation DAS presented a greater and faster release, compared to the formulations with high oxidation DAS. These results showed that the degree of starch oxidation, for the PVA-DAS hydrogel synthesis has a significant effect on the behavior of the polymeric network.

1. Introduction

Hydrogels are three-dimensional polymeric networks that are capable of containing significant amounts of water. They have been used in biomedical applications, such as tissue engineering and controlled drug release [1]. Furthermore, hydrogels have tunable biocompatibility, and respond to external stimuli such as pH, temperature, and concentrations of species, among others [2]. Thus, their porous structures are suitable for the encapsulation of drugs for a subsequent controlled release [3].

Hydrogels can be synthetized from virtually any water-soluble polymer, either natural or synthetic, such as alginic acid, pectin, carrageenan, dextran sulfate, chitosan, poly-lysine, and agarose [4]. Poly(vinyl) alcohol (PVA), is a hydrophilic, synthetic biodegradable polymer that has been widely used, due to its proved biocompatibility [5]. An additional benefit is that this polymer can be crosslinked physically, without the need of potentially toxic chemical agents . This can be achieved through freezing / thawing cycles, where the polymer in aqueous solution depending on its concentration and freeze cycles, form polymer microcrystals [1]. Nevertheless, since it is a petroleum derivative, PVA has limited availability. At the same time, in order to obtain stable structures, PVA concentrations must be high, up to 35% w/v, with molecular weights of 98000 Da or greater [2]. Therefore, its use poses an economical limitation for the creation of new technologies, especially in developing countries.

Starch, on the other hand, is one of the most abundant and cheapest polysaccharides [6]. It is comprised of two main molecules: amylose, a linear macromolecule composed of 250-300

[*]Corresponding Author: José Francisco Alvarez-Barreto, Email: jalvarezb@usfq.edu.ec

α(1→4)-linked-D-glucose, amylopectin, a less hydrophilic and branched macromolecule, and constitutes about 80% of the granular structure. To form starch hydrogels, it is possible to chemically modify its structure in order to have a greater number of functional groups [4,7]. Oxidation, for example, breaks the long glucose chains of the polymer molecules, replacing the hydroxyl groups for carbonyl and carboxyl groups . This results in reduced viscosity [8]. In drug-release applications, integrity and stability in a humid environment are required over a relatively long period of time. It has been shown that the use of dialdehyde starch (DAS), resulting from oxidation, exhibits a substantially lower solubility in water, making it more suitable for this application [9].

In [9], the authors developed protein-based films with dialdehyde starch (DAS), PVA and feather keratin. They found that if DAS was not added, an increase of the PVA concentration in the blended films resulted in a slight decrease of the film's solubility in water. By adding 5% of DAS, the total solubility mass of the film was reduced, which indicates an increase in the degree of cross-linking.

The most commonly used modifier in starch oxidation is hypochlorite, but despite being chemically efficient, it results in the formation of chlorinated products, that could potentially be toxic [10]. Several comparative studies have been carried out on the properties of starches at different degrees of oxidation with hypochlorite[10]; nevertheless, no comparative study is found in the literature on the effect that hydrogen peroxide has on different degrees of oxidation in cassava starch, and thereby in the formulations of hydrogels [11].

In [12], the authors synthesized hydrogels based on PVA and modified cassava starch as carriers for controlled drug delivery. In this research, the oxidation of starch was done with sodium hypochlorite solution containing 5% active chlorine. The hydrogels synthetized with DAS/PVA, showed an enhanced stability in water; nevertheless, the oxidation degree was maintained constant through the entire research. However, it has been demonstrated that, for other biomaterial applications, starch oxidation degree has important implications, such as bioplastics, in material behavior [13,14]. For this reason, the present study proposes to determine the effects of the oxidation degree of cassava starch on the physical and chemical properties of DAS-PVA hydrogels.

2. Materials and Methods

2.1. Chemical materials

Cassava starch was obtained from Ecuador's local stores. Hydroxylamine hydrochloride (ACS Reagent Grade, 5470-11, ACROS Organics), Poly vinyl alcohol (PVA, Mowiol 28-99, MW 145000, 99.0-99.8 mol% hydrolysis, Millipore Sigma, Cat. 10849), Sodium Phosphate, Dibasic, Anhydrous (3828-01, J.T.Baker), Potassium chloride (6858, Mallinckrodt), di-Potassium hydrogen phosphate, anhydrous GR for analysis (1.05104.1000, MERCK), Ibuprofen (≥98% GC, Millipore Sigma, Cat. I4883) were used as received.

2.2. Starch Oxidation

Starch oxidation was performed with hydrogen peroxide (H_2O_2), through adaptation of the methodology proposed by [15]. Briefly, H_2O_2 solutions at 3, 5 and 10% v/v were prepared in distilled water. For the chemical modification, a 2% w/v solution of starch in distilled water was gelatinized at 80ºC for 30 minutes, under moderate stirring. Later, the solution was cooled until room temperature, and its pH was adjusted to 7. For every 2g of gelatinized starch, 6.25mL of H_2O_2 solution were added,, drop wise during 1h at 25ºC, under continuous stirring, maintaining pH at 7.0. When the oxidation process was done, an equal volume of 70% v/v ethanol was added, and starch precipitation was allowed for 1 hour. The supernatant was discarded, and the remaining mixture was centrifuged at 1000g for 5 minutes. Five rinsing cycles were carried out with diH_2O, in a similar fashion. The resulting dialdehyde starch (DAS) was dried at 40ºC until constant weight.

To obtain a high degree of oxidation DAS, copper (II) sulfate was used as a catalyst, as proposed by [16]. Starch was gelatinized with the same methodology described above. Before adding the H_2O_2 solution at 10% v/v, the catalyst was incorporated at 0.1% w/v, and the reaction was allowed to take place for 1 hour at 25 ºC. Subsequently, precipitation and washing were done as previously mentioned.

2.3. Starch characterization

Carbonyl groups were quantified based on the method proposed by [17], with some modifications. Briefly, 1g of DAS was mixed with 25mL of distilled water, gelatinized at 80ºC under constant stirring, and then, cooled at room temperature. The pH was adjusted to 3.2, and, immediately, 3,75mL of hydroxylamine hydrochloride (25g of hydroxylamine hydrochloride in 100 mL of NaOH 0.5M in a 500mL solution) were added to the gelatinized solution, with subsequent incubation at 40ºC for 4h. A potentiometric titration followed, with 0.1M HCl until pH 3.2. The amount of carbonyl groups per 100 glucose units (CO/100GU), was calculated through equation 1:

$$\frac{CO}{100GU} = \frac{(Vb-Vs)\times M\times 0{,}028\times 100}{W} \quad (1)$$

where, Vb and Vs are the volumes (mL) of HCL used to titrate the blank (native starch) and the oxidized starch, respectively; M is the molar concentration of HCl (mol/L), and W is the weight of the sample on a dry basis (g).

On the other hand, carboxyl groups were determined using the protocol proposed by [15]. Briefly, 1g of DAS was suspended in 60mL of distilled water, gelatinized and cooled, as previously mentioned. Later, a potentiometric titration took place until pH 8.2, using NaOH 0.01M. The amount of carboxyl groups per 100 glucose units (CO/100GU), was calculated through equation 2:

$$\frac{COOH}{100GU} = \frac{(Vs-Vb)\times M\times 0{,}045\times 100}{W} \quad (2)$$

where, Vs and Vb are the titration volumes (mL) of the oxidized starch and the blank (native starch), respectively; M is the molarity of NaOH (mol/L), and W is the sample weight (g).

2.4. Hydrogel preparation

Hybrid hydrogels with Polyvinyl Alcohol (PVA) and dialdehyde starch were synthetized through a freezing/thawing method, by adaptation of the protocol reported by [2]. The

common protocol for the preparation was to completely dissolve PVA in distillated water under continuous stirring at 95ºC, at 20% w/v. In parallel, a starch suspension, at a given concentration, was gelatinized as previously mentioned. Both solutions were then mixed at a 1:1 ratio in a vortex homogenizer, with alternation with heating in a water bath at 70 ºC, for a total time of 30 min. Then, 5 mL of the polymeric mixture were poured into 1-inch petri-dishes.

Crosslinking was induced through freezing/thawing cycles; the polymeric mixture in the petri dish was frozen at -20ºC for 12 h, followed by thawing at 25ºC for 3h. Table 1 specifies the variables that were modified for the different conditions of hydrogel preparation in order to identify the formulations that could be suitable for the subsequent ibuprofen encapsulation and release. It was decided to vary the number of freeze/thawing cycles because through every cycle, there is a formation of microcrystals through the polymeric network, allowing a high crosslinking degree. Consequently, after thawing, the mixture was frozen again, and this process was repeated according to the chosen number of cycles. Since one of the objectives of this research is to partially substitute the amount of PVA in hydrogels formulations, the final concentration of the latter was maintained constant at 10% w/v. Hydrogels with native starch were used as controls.

Table 1: Variables for the synthesis of hybrid DAS/PVA hydrogels

Variable	**Values (nomenclature)**		
Oxidation degree ($\%H_2O_2$)	3 (3DAS)	5 (5DAS)	10, 10* (10DAS, 10DASC)
DAS concentration (% w/v)	5	10	15
Crosslinking cycles	1	2	3

*Copper (II) sulfate was added to obtain high oxidation DAS

According to the protocol described by [18], after crosslinking, hydrogels were cut in 5mm diameter discs, and placed in 24-well plates with 2mL of 1X phosphate buffered saline (PBS, pH 7.3), in an incubator, at 37ºC. The hydrogel weight was registered at 0, 5, 15, 60 min, 2, 3, 4, 5, 24, 48 and 72h. Water-uptake was calculated according to equation 3.

$$Water-uptake\ (\%) = \frac{(W_t - W_1)}{W_1} \times 100 \quad (3)$$

where W_1 is the initial weight of the hidrogel, and W_t its weight at a given time.

2.5. FTIR Analysis

Analyses were conducted in a Fourier transformed infrared spectrometer; model Cary 630 FTIR-ATR (Agilent Technologies) for the native and oxidized starches, as well as for the different hydrogel formulations [15]. Spectra were acquired in the region of 4000-500 cm-1.

2.6. Scanning Electron Microscopy (SEM)

Hydrogel morphological analyses were carried out in a JEOL JSM-IT300 Scanning Electron Microscope (Tokyo, Japan) with lyophilized hydrogels. Samples were cut and placed on metallic stubs with carbon tape. Images were obtained at 5kV and 50 Pa, at different magnifications..

2.7. Ibuprofen encapsulation and release

Ibuprofen was incorporated into the hydrogels at a concentration of 10 mg/mL, as proposed by [18], by adding the drug into the PVA solution at 50ºC. The same process described above for hydrogel preparation was followed. The release was performed as previously described [18]. DAS/PVA hydrogels with ibuprofen were cut into 5 mm diameter and placed in 2 mL of PBS at 37 ºC. Later, 400 µL of supernatant were collected and replaced with fresh PBS at 15, 30, 60, 120, 180, 240, 300, 360 min, 24, 48 and 72 h. Four samples of each hydrogel formulation were tested for Ibuprofen release.

Each supernatant was placed in 2.6 mL of PBS, in a quartz cuvette, and the optical density of this final solution was determined at 222 nm in a CECIL CE 2041 UV-VIS spectrophotometer. The sample concentration was calculated with a calibration curve done with known solutions of ibuprofen in PBS. The release kinetics was expressed as the percent cumulative release of ibuprofen over time.

2.8. Statistical analyses

All assays were performed with three samples of each formulation, and four in the case of ibuprofen release. The results are expressed as the average ± standard deviation. A two-way analysis of variance (ANOVA) was also applied, and multiple pairwise comparisons were carried out using Turkey-HSD, with a 95% confidence level ($p<0.05$).

3. Results and discussion

3.1. Oxidized starch characterization

Carbonyl and carboxyl contents of native and oxidized starches (dialdehyde starch, DAS) with low and high degrees of oxidation are shown in Table 2. It can be observed that carbonyl was the primary functional group produced during the oxidation process, whereas, a significantly minor amount of carboxyl groups was formed. It has been reported that the reaction mechanism of hydrogen peroxide with starch proceeds via a radical chain reaction [19] . It is more evident in dialdehyde catalyzed starch with copper sulfate, since, in the presence of the metal catalyst, the H_2O_2 is decomposed into a hydroxyl radical. Then, this radical rapidly reacts with the carbohydrate by eliminating hydrogen from a C-H group on the sugar ring, forming a R*CHOH radical, which further undergoes acid or base catalyzed rearrangement resulting in the cleavage of the glycosidic bond and formation of a carbonyl group [19].

One assumption that the authors in [19] proposed, is that the reaction between starch and hydrogen peroxide is so rapid that the major quantity of reagent was consumed for hydrogen abstraction, affording high amounts of carbonyl groups during the early period of time; therefore, less oxidant was available for the oxidation of

carbonyl to carboxyl groups. At the same time, the reaction time on the current research was stablished at one hour; which is in accordance with the literature since it has been shown that the oxidation of hydroxyl groups in starch molecules by H_2O_2 was almost completed within the first 30 minutes.

Table 2: Carbonyl and carboxyl contents of dialdehyde cassava starch at different oxidation levels

H_2O_2 Concentration (%v/v)	Carbonyl content [CO/100 GU]	Carboxyl content [CO/100 GU]
3	0.255 ± 0.091	0.046±0.010
5	0.419 ± 0.124	0.034±0.009
10	0.209 ± 0.108	0.052±0.010
10 (Catalyzed)	2.875 ± 0.089	0.025±0.003

GU: Glucose units

Oxidation degree is determined by carbonyl and carboxyl content. The relatively bulky carbonyl and carboxyl groups are introduced with a partial depolymerization of the starch chains; for this reason, the determination of these values shows the degree of oxidation degree. A major carbonyl content is an indication of a higher oxidation degree [20].In this study, a low degree of oxidation will be considered as having a carbonyl content of less than 1CO/100GU; thus, all starches oxidized without catalyst will be within this category.

3.2. Hydrogel morphological analyses

It was important to evaluate hydrogel morphology as it can have important implications in swelling behavior and controlled released, for which a porous structure is desired. Lyophilized hydrogels with the same compositions, but different DAS are presented in Figure 1. Natural porous materials have no uniform porosity; rather, it is distributed allowing to maximize the overall performance of the structure [21]. For drug delivery applications, a balance between adequate porosity and mechanical stability is required.. At comparing the different images, it is possible to recognize that hydrogels with low oxidation degree have flatter and larger pores; while, in high oxidation DAS, the pores seem to be smaller and rounded. There is clear indication that pore arrangement is altered by the incorporation of starch, and that the oxidation degree further influences pore morphology and structure. Additionally, PVA only hydrogels (without starch) appeared to less porous.

Hybrid hydrogels composed with hydrophilic-hydrophobic polymers have been reported previously in literature. It has been shown that one of the major issues in obtaining porosity is due to the intrinsic immiscibility of hydrophilic (i.e. PVA) and hydrophobic (i.e. DAS) polymers, which, in turn, affect pore structure, and potentially, distribution..

3.3. FTIR analysis

The Fourier transformed infrared absorption spectra of lyophilized hydrogels composed with 10% w/v PVA and 10% w/v starch is presented in Figure 2. In the case of FTIR spectra for different hydrogel formulations with oxidized starch, there is a decreased intensity. As previously reported in the literature, this lower intensity corresponds to crosslinked polymer structures, probably associated with a major bond vibration restriction [22]. A stretching band can be observed at 2850 cm^{-1} due to C-H broad alkyl. Another lower intensity stretching is because of the strong-OH group band for free unreacted alcohol and hydrogen bonded bands at 3600-3200 cm-1, which are associated with freeze-thawed PVA [23]. As expected, the peak of dialdehyde starch appears at 1630 cm^{-1} - 1710 cm^{-1}, due to C=O and C=C stretching frequency groups that are present in DAS [24].

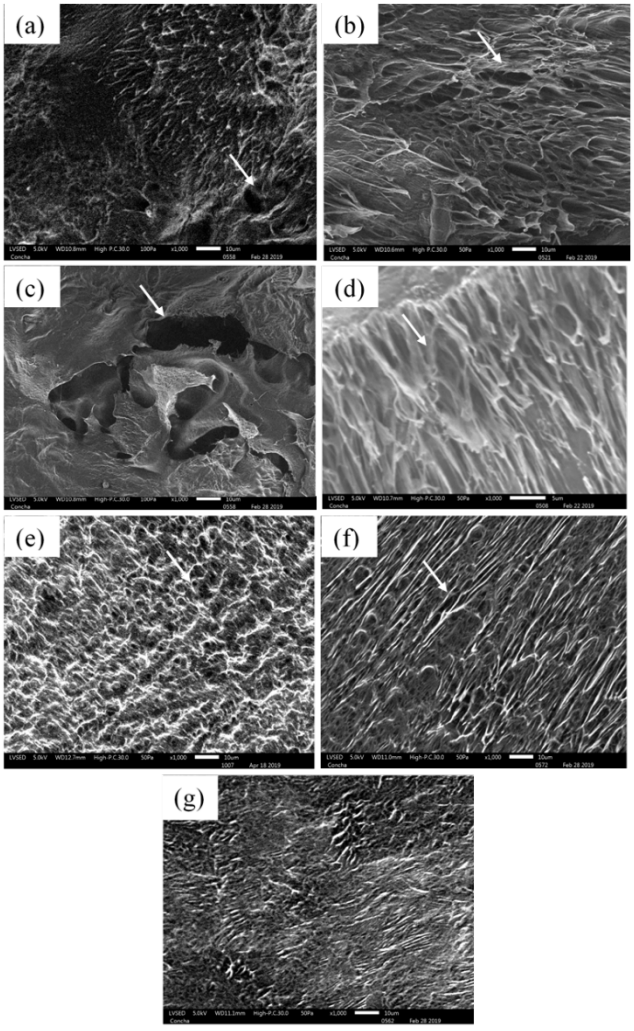


Figure 1: Scanning electron micrographs of lyophilized hydrogels composed of 10% w/v PVA, 10% w/v Starch and 1 crosslinking cycle: (a) NS, (b) 3DAS, (c) 5 DAS, (d) 10 DAS, (e) 10DASC, (f) Solely 20% w/v PVA and (g) Solely 10% w/v PVA. Calibration bar: 10 µm. Figure 1d was reported at 3000X (Calibration bar: 5 µm) in order to better observe the porous structure.

When comparing the spectra of hydrogels composed solely with PVA to hybrid hydrogels, there is lower intensity absorbance in the different bands; nevertheless, those associated with -OH groups, could be related to a strong interaction between both polymers forming a tight network, as suggested by [22]. PVA addition enhances the elasticity of hydrogels, which is corroborated by the increased intensity of the C-H stretching

absorption band at 2031 cm-1, and the C-C stretching vibrational band at 1200-1400 cm^{-1} [25].

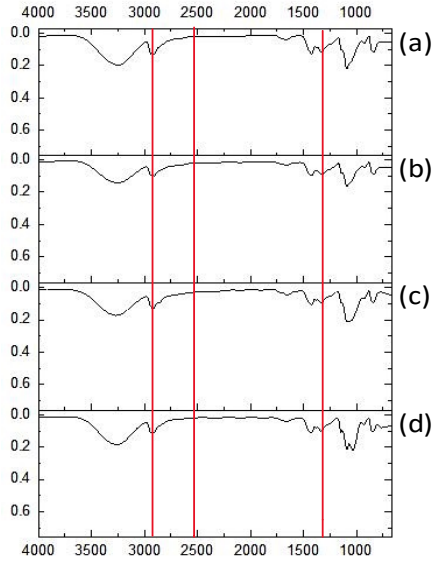


Figure 2: FT-IR spectra of lyophilized hydrogels formulations with 10% w/v PVA, 10% w/v starch and 1 crosslinking cycle: (a) Native starch, (b) 5DAS, (c) 10DAS and (d) 10DASC.

3.4. Hydrogel Water-uptake

Hydrogel water-uptake, or swelling ratio, is one of their most important characteristics, since this parameter is directly related to drug release mechanisms, as well as structure stability in aqueous media [1]. Water uptake capacity (%) of the different hydrogel formulations specified in table 1 are presented in Figure 3 through 5, where PVA composition remained constant at 10%. All assays where performed at 37ºC in PBS to emulate physiologic conditions. Since one of the objectives of this research was to partially substitute the weight/volume percentage of PVA in hydrogels formulations, it was necessary to compare the water-uptake percentage vs. time, according to hydrogels composed solely with PVA. In general, it was observed that 1 crosslinking cycle presented the largest swelling capacity; nevertheless, they showed the lower stability in time, with important swelling reduction toward the end of the assay. Two and three crosslinking cycles exhibited similar behaviors. At the same time, by increasing the quantity of PVA, hydrogels displayed greater stability (data not shown).

According to the analysis of variance, hydrogel formulation (both starch content and oxidation degree), number of crosslinking cycles and time had significant effects on water uptake ($p<0.05$). Figure 3a shows the swelling kinetics of hydrogels with 15% starch. There were no statistical differences between native and low oxidation DAS (3, 5 and 10DAS), which displayed fast water uptake, with important fluctuations over time with one crosslinking cycle. This could be due to degradation effects that, ultimately, become relevant, as the swelling decrease after a maximum at 48h. When increasing the number of cycles (Figure 3a and b), however, these fluctuations become less apparent, and decreases in water uptake are lower. These changes could indicate an improvement in hydrogel stability. In the case of 10DASC, on the other hand, had rapid swelling and seemed to reach equilibrium after 15h. This high oxidation DAS resulted in significantly lower swelling than the other formulations; moreover, this parameter tends to decrease with greater number of crosslinking cycles.

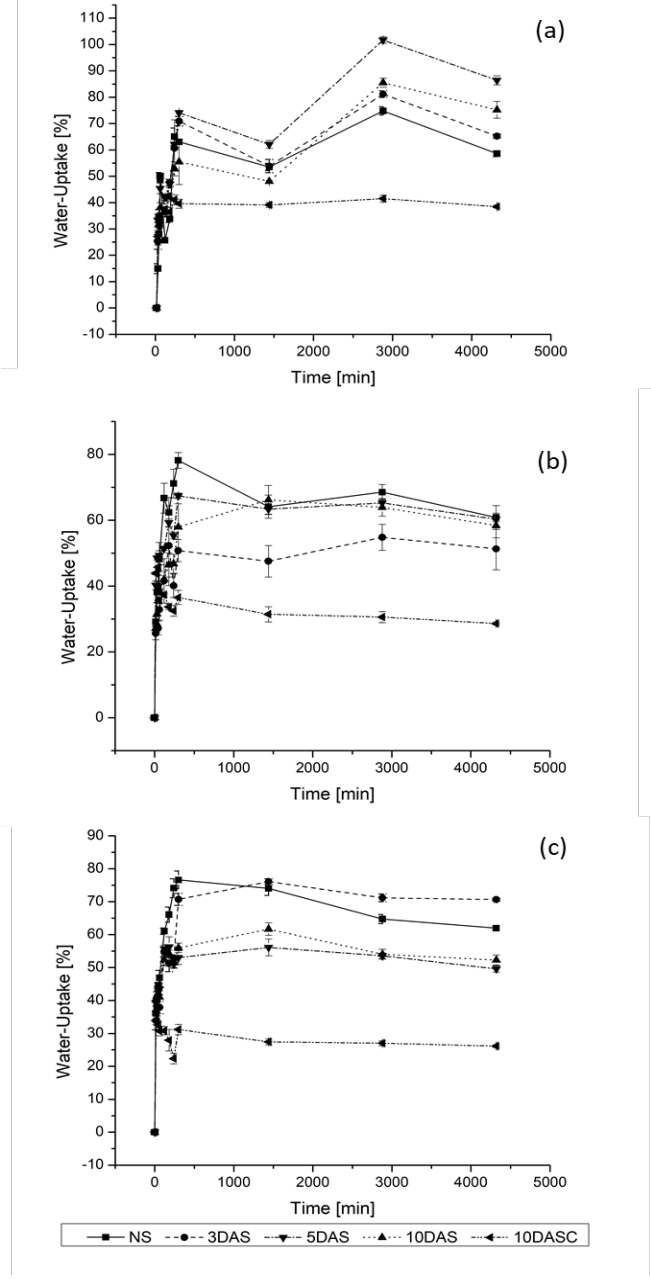


Figure 3: Swelling kinetics, expressed as water uptake percentage in time, for hydrogels composed of 10% w/v PVA and 15% w/v starch, either native (NS) of dialdehyde (3DAS, 5DAS, 10DAS and 10DASC). Formulations with (a) one crosslinking cycle, (b) two crosslinking cycles, and (c) three crosslinking cycles.

Figure 4 shows water-uptake capacity for hydrogels with 10% starch, in which, after an analysis of variance, it is determined that formulation and time affect water-uptake percentage ($p<0.05$). Once again, the major water-uptake is obtained with one freeze-thaw cycle. The stability in this case is significantly reduced in comparison with Figure 3(a), where hydrogel degradation, regardless of formulation, begins at 240 min until 72h. After the

statistical analysis, it was confirmed that oxidation degree had a significant effect on swelling ratio ($p<0.05$); nevertheless, there is not a clear pattern in the way the amount of carbonyl groups affects the swelling degree.

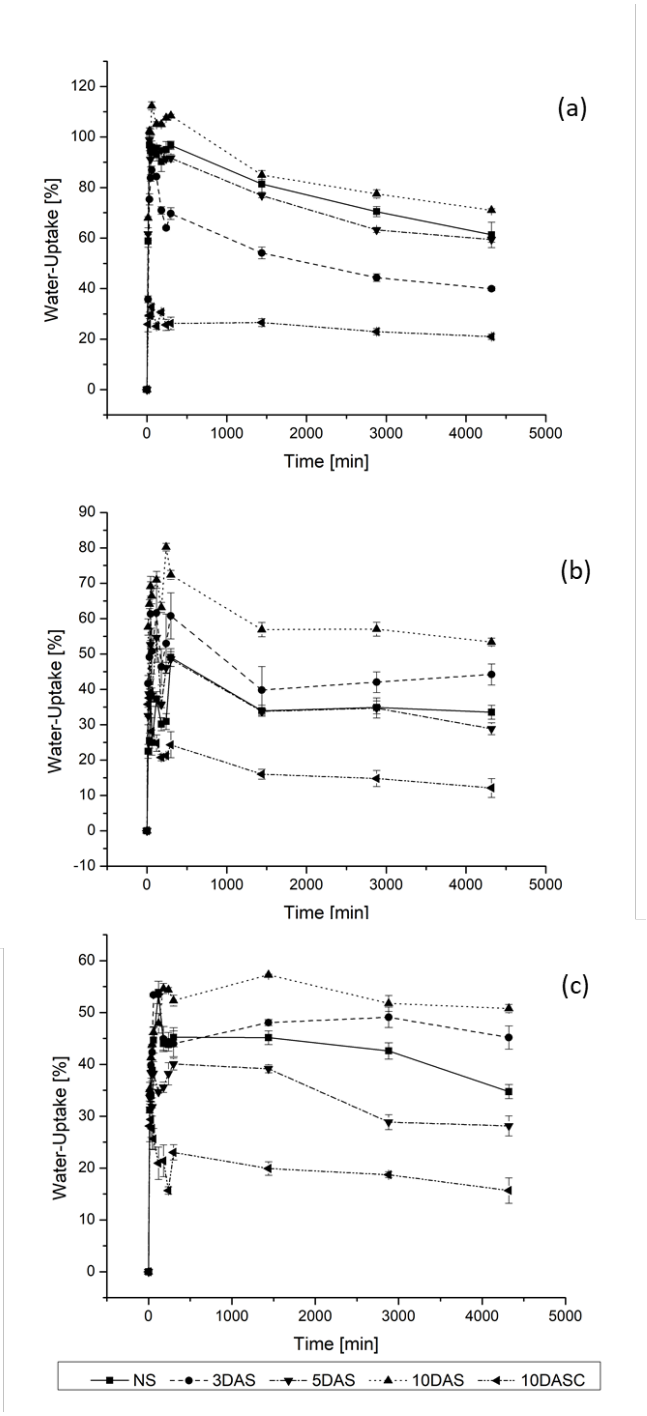


Figure 4: Swelling kinetics, expressed as water uptake percentage in time, for hydrogels composed of 10% w/v PVA and 10% w/v starch, either native (NS) of dialdehyde (3DAS, 5DAS, 10DAS and 10DASC). Formulations with (a) one crosslinking cycle, (b) two crosslinking cycles, and (c) three crosslinking cycles.

Figures 4(b) and (c) show a more accurate pattern within the effect of DAS, where starch with lower carbonyl content presents the greater swelling degree. After a Tuckey comparison on values of Figure 4(b), all the formulations are significantly different between them; nevertheless, the greatest water uptake was obtained in formulations with 3DAS and 10DAS, which agrees with lower carbonyl content. At three freeze-thaw cycles, Figure 4(c), Tukey analysis shows that there is not a difference between formulations with 3% DAS and 10% DAS. In all cases, the way time affects swelling degree, establishes an equilibrium tendency up to 72h. Finally, the number of cycles does affect hydrogel response, being inversely proportional to the swelling degree. However, the swelling degree within a time greater than 24 h, is not statistically different between formulations with 2 and 3 freeze-thaw cycles.

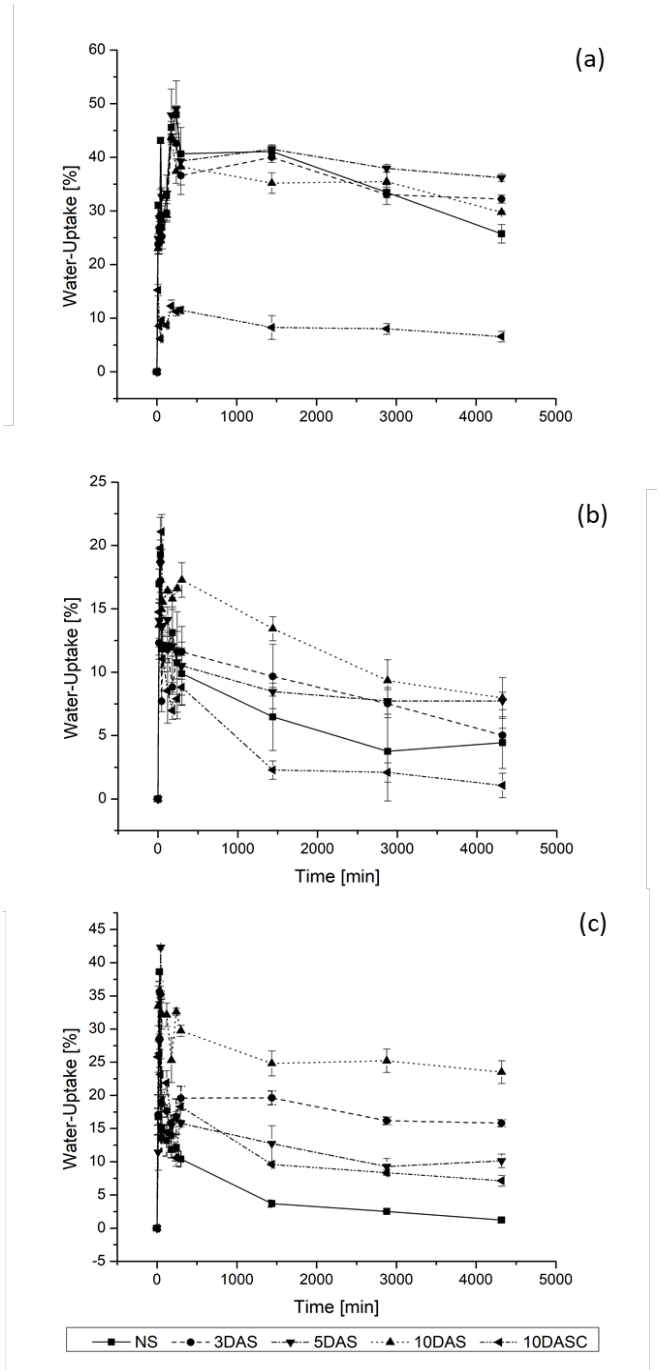


Figure 5: Swelling kinetics, expressed as water uptake percentage in time, for hydrogels composed of 10% w/v PVA and 5% w/v starch, either native (NS) of dialdehyde (3DAS, 5DAS, 10DAS and 10DASC). Formulations with (a) one crosslinking cycle, (b) two crosslinking cycles, and (c) three crosslinking cycles.

By comparing hydrogels with 15% and 10w/v starch (Figures 3 and 4, respectively), there are clear differences. The greater water-uptake in the first case is presented at 300 min; while, for the latter, this value is obtained at 180 min. There is a major hydrogel degradation within the first 24 h at a lower starch content. These results suggest that oxidized starch content affects swelling.

Finally, swelling kinetics for hydrogels formulations with 5% starch are shown in Figure 5. Considering factors such as formulations and number of cycles, swelling degree in all cases was considerably lower, compared with 15 and 10% starch (Figures 3 and 4), due to the decrease of polymer content in the structure. The statistical analysis resolved that there is not a significant variation on swelling, considering the formulation of the hydrogel as a swelling response. At this low concentration of starch, there is not a clear pattern on the effect of the oxidation degree on swelling, which could be due to a predominant effect of PVA, independent of the starch´s characteristics.

Literature shows that starch oxidation affects the surface in the granule by increasing the porosity and presence of holes in the structure [20,26]. This could explain the reason behind a large water-uptake in hydrogels formulations with dialdehyde starch. Even if there is a significant presence of hydrophobic groups in modified starch, this porosity helps hydrogels to increase not only their swelling degree, but also their stability during a period of time.

After the statistical analysis, it was possible to show that there is a major increasing swelling degree during the first 5 hours in hydrogels with a lower carbonyl content; which agrees with previous literature reported by [27]. According to previous works, higher PVA composition would result in hydrogels with lower swelling ratio due to its higher degree of crystallinity [25]. Nonetheless, the opposite effect occurs after the addition of dialdehyde starch. In [28], the authors reported that oxidation occurs mainly in the amorphous lamella of the semi-crystalline growth rings in starch granules. They evaluated the effect of different degrees of oxidation with sodium hypochlorite on bean starch. A higher carbonyl content in dialdehyde bean starch resulted in a decrease of the crystallinity, suggesting that the amylopectin chains were damaged and therefore, there is a degradation of the crystalline region. The conclusion proposed in the current research behind the large water-uptake of hydrogels composed with PVA and dialdehyde starch, is due to the significant presence of less crystalline polymer.

An important fact about dialdehyde starch at a high oxidation degree was evidenced through a Tukey analysis. The way in which time affects swelling degree, in all cases, follows a tendency in which water uptake within the first hour is rather high. Suddenly, there is a significant de-swelling until 240 min. Beyond this time, hydrogels exhibit stability. In [10], the authors reported that at higher carbonyl content, the crystallinity decreases, and at a certain point, starch structure becomes amorphous. It can be assumed that this de-swelling is due to the less crystallinity arrangement, even though, there is stability during times greater than 240 min.

Finally, swelling degree is related with the porosity of hydrogels, evidenced through the scanning electron micrographs of lyophilized hydrogels. As presented in Figure 1, oxidation degree does affect the morphology of hydrogels, by allowing flatter and seemingly larger pores in the case of modified starch at a low oxidation degree. The behavior according to porosity and swelling degree is as expected, since 10DAS presents not only large porosity, but also greater water-uptake; unlike, 10DASC, which exhibits a lower porosity and so was the swelling degree with these formulations.

3.5. Ibuprofen release

Hydrogels have sparked interest in drug delivery applications due to their porous structures, that assists drug loading in the polymeric matrix, for a subsequent sustained release. In the present work, ibuprofen was used as a model drug that can give insights into potential applications for other non-steroidal anti-inflammatory drugs (NSAIDs). In this case, control hydrogels were done at 20% PVA, since those at 10% were rather unstable and disintegrated quickly. According to ANOVA, ibuprofen release was significantly affected by composition, time and number of freeze-thaw, or crosslinking, cycles. Release assays are presented in Figure 6.

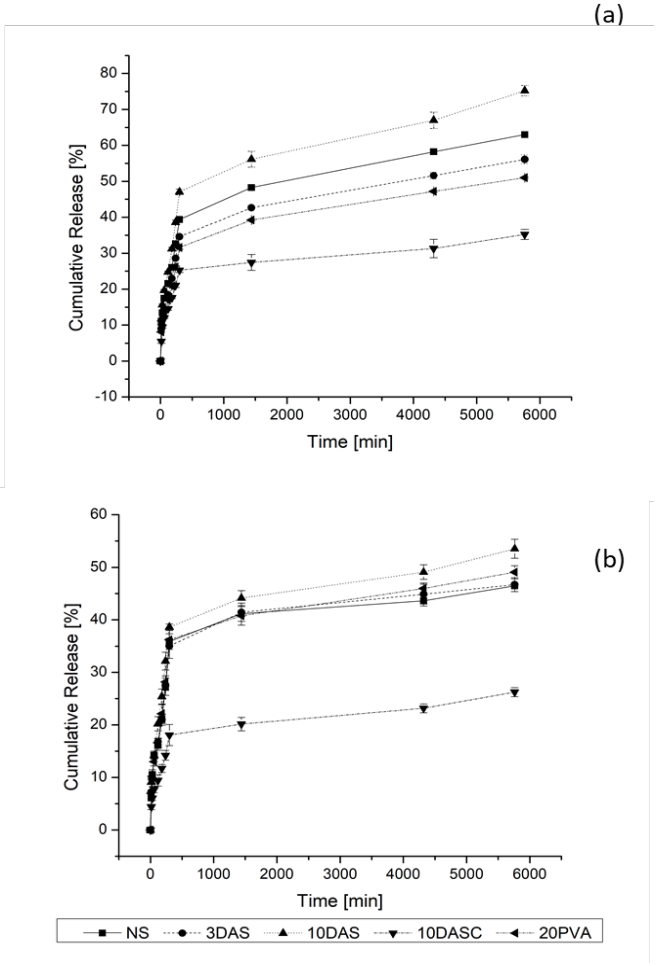


Figure 6: Ibuprofen cumulative release profiles from hydrogels with 10% /v PVA and 10% starch at different oxidation levels. Only PVA at 20% was used as control. (a) One crosslinking cycle and (b) Two crosslinking cycles.

For all formulations, there was a fast release in the first 6 hours, followed by small, not statistically significant, increases, indicating a tendency toward equilibrium. In hydrogels with one crosslinking cycle (Figure 6a), there is a very similar ibuprofen

profile in every hydrogel formulation. After Tukey comparison, it was shown that formulations with native starch, 3DAS, 10DAS and solely 20% PVA have no significant differences between them. Interestingly, hydrogels with native starch and low oxidation DAS displayed greater swelling, and some degradation that played an important role in the release. Hydrogels with 10DASC, on the other hand, did present significantly lower cumulative release at every time point. All these previous formulations exhibit the same behavior; the largest drug release was performed within the first 6 hours of contact with PBS; and, there is an equilibrium tendency at 72h.

Values of ibuprofen release in hydrogels with two crosslinking cycles, Figure 6b, were significantly affected by formulations and time. Again, Tukey comparisons indicated that formulations with 3DAS, 10DAS and solely 20% PVA are similar among them, representing the formulations with largest drug release percentage. Similar to what occurred with one cycle, 10DASC resulted in hydrogels with the lowest levels of release. This time, nonetheless, values of cumulative release were lower than those found with one crosslinking cycle, which relates to the lower swelling capacities found at greater number of freeze-thaw cycles. Moreover, there was a tendency to a more sustained release, possible due to the greater stability provided by more crosslinking

When comparing Figures 6a and b, it is possible to affirm that as presented in literature, the high-water content of hydrogels results in an increased release of drugs from the polymeric matrix [3]. This is confirmed by the statistical analysis. The tendency in both cases is that hydrogels with 10DAS exhibit the largest drug release, and the lowest is attributed to those with 10DASC.

In hydrogels, the rate of drug diffusion through the polymeric structure depends on the extent of cross-linking. Converting the ensemble of individual polymer chains into a macromolecular network allows the water diffusion into the interstices of the network. Comparing Figures 6a and b, there is indication that the extent of cross-linking minimizes the swelling and, thereby, the diffusion of the drug. The reason behind this behavior is due to the fact that ibuprofen is an entrapped drug, and the rate of diffusion depends on the interchain separation and the size of the diffusing drug [29].

4. Conclusions

The oxidation degree of cassava starch has important effects on the properties of hydrogels with PVA, particularly at larger carbonyl contents. At a low oxidation degree obtained with 10% v/v H_2O_2 and 3% v/v H_2O_2, hydrogels showed the major suitable characteristics of hydrogels for drug delivery applications, since the water-uptake percentage was higher, as it was drug release. Considering hydrogels formulations with DAS catalyzed with copper sulfate (II), and the largest carbonyl content, there was a significant reduction of water-uptake through the polymeric structure, due to strong intermolecular bonds, which resulted in a reduced swelling degree, but, as expected, those showed the major stability during time, proving an improved stability in physiological environment. It is possible that, for the latter formulation, the hydrogels could be more appropriate for tissue engineering because, given that they do not swell significantly, their mechanical properties could be superior to those that presented higher swelling ratios. Additionally, being more stable in time, they could potentially meet mechanical demands for longer periods, and we recommend doing this assessment in future studies.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was funded by the PoliGrant program at Universidad San Francisco de Quito. Publication of this article was funded by the Academic Articles Publication Fund of Universidad San Francisco de Quito USFQ

References

[1] A.S. Hoffman, "Hydrogels for biomedical applications," Advanced Drug Delivery Reviews, **64**(SUPPL.), 18–23, 2012, doi:10.1016/j.addr.2012.09.010.

[2] J.L. Holloway, A.M. Lowman, G.R. Palmese, "The role of crystallization and phase separation in the formation of physically cross-linked PVA hydrogels," Soft Matter, **9**(3), 826–833, 2013, doi:10.1039/c2sm26763b.

[3] T.R. Hoare, D.S. Kohane, "Hydrogels in drug delivery: Progress and challenges," Polymer, **49**(8), 1993–2007, 2008, doi:10.1016/j.polymer.2008.01.027.

[4] E.A. Kamoun, "N-succinyl chitosan-dialdehyde starch hybrid hydrogels for biomedical applications," Journal of Advanced Research, **7**(1), 69–77, 2016, doi:10.1016/j.jare.2015.02.002.

[5] S. Jiang, S. Liu, W. Feng, "PVA hydrogel properties for biomedical application," Journal of the Mechanical Behavior of Biomedical Materials, **4**(7), 1228–1233, 2011, doi:10.1016/j.jmbbm.2011.04.005.

[6] Y. Zhu, C. Romain, C.K. Williams, "Sustainable polymers from renewable resources," Nature, **540**(7633), 354–362, 2016, doi:10.1038/nature21001.

[7] H. Hu, F.J. Xu, "Rational design and latest advances of polysaccharide-based hydrogels for wound healing," Biomater Sci., **8**(8), 2084–2101, 2020, doi:10.1039/d0bm00055h.

[8] A.R.G. Dias, E.D.R. Zavareze, E. Helbig, F.A. De Moura, C.G. Vargas, C.F. Ciacco, "Oxidation of fermented cassava starch using hydrogen peroxide," Carbohydrate Polymers, **86**(1), 185–191, 2011, doi:10.1016/j.carbpol.2011.04.026.

[9] Y. Dou, B. Zhang, M. He, G. Yin, Y. Cui, I.N. Savina, "Keratin/polyvinyl alcohol blend films cross-linked by dialdehyde starch and their potential application for drug release," Polymers, **7**(3), 580–591, 2015, doi:10.3390/polym7030580.

[10] Y.R. Zhang, X.L. Wang, G.M. Zhao, Y.Z. Wang, "Preparation and properties of oxidized starch with high degree of oxidation," Carbohydrate Polymers, **87**(4), 2554–2562, 2012, doi:10.1016/j.carbpol.2011.11.036.

[11] V.M. de Oliveira Cardoso, B. Stringhetti Ferreira Cury, R.C. Evangelista, M.P. Daflon Gremião, "Development and characterization of cross-linked gellan gum and retrograded starch blend hydrogels for drug delivery applications," Journal of the Mechanical Behavior of Biomedical Materials, **65**, 317–333, 2017, doi:10.1016/j.jmbbm.2016.08.005.

[12] A.P. Erazo Bastidas, J. Alvarez Barreto, Synthesis and Characterization of Hydrogels based on Poly (vinyl alcohol) and Modified Starch as Carriers for Controlled Drug Delivery, Quito, 2018.

[13] Y.-R. Zhang, X.-L. Wang, G.-M. Zhao, Y.-Z. Wang, "Influence of oxidized starch on the properties of thermoplastic starch," Carbohydrate Polymers, **36**(1), 358–364, 2013, doi:10.1016/j.carbpol.2013.03.093.

[14] Y. Yu, Y. Wang, W. Ding, J. Zhou, B. Shi, "Preparation of highly-oxidized starch using hydrogen peroxide and its application as a novel ligand for zirconium tanning of leather," Carbohydrate Polymers, **174**, 823–829, 2017, doi:10.1016/j.carbpol.2017.06.114.

[15] P. Parovuori, A. Hamunen, P. Forssell, K. Autio, K. Poutanen, "Oxidation of Potato Starch by Hydrogen Peroxide," Starch - Stärke, **47**(1), 19–23, 1995, doi:10.1002/star.19950470106.

[16] Y. Lu, Q.M. Kong, R. Jing, X. Hu, P.X. Zhu, "Solid state oxidation of polyvinyl alcohol by hydrogen peroxide-Cu (II)," Polymer Degradation and Stability, **98**(6), 1103–1109, 2013, doi:10.1016/j.polymdegradstab.2013.03.022.

[17] R.J. Smith, Characterization and analysis of starches, Academic Press, New

York and London: 620–625, 1967.
[18] P.M. Castro, P. Fonte, A. Oliveira, A.R. Madureira, B. Sarmento, M.E. Pintado, "Optimization of two biopolymer-based oral films for the delivery of bioactive molecules," Materials Science and Engineering C, **76**, 171–180, 2017, doi:10.1016/j.msec.2017.02.173.
[19] K. Sangseethong, N. Termvejsayanon, K. Sriroth, "Characterization of physicochemical properties of hypochlorite- and peroxide-oxidized cassava starches," Carbohydrate Polymers, **82**(2), 446–453, 2010, doi:10.1016/j.carbpol.2010.05.003.
[20] L.M. Fonseca, J.R. Gonçalves, S.L.M. El Halal, V.Z. Pinto, A.R.G. Dias, A.C. Jacques, E. da R. Zavareze, "Oxidation of potato starch with different sodium hypochlorite concentrations and its effect on biodegradable films," LWT - Food Science and Technology, **60**(2), 714–720, 2015, doi:10.1016/j.lwt.2014.10.052.
[21] Q.L. Loh, C. Choong, "Three-dimensional scaffolds for tissue engineering applications: Role of porosity and pore size," Tissue Engineering - Part B: Reviews, **19**(6), 485–502, 2013, doi:10.1089/ten.teb.2012.0437.
[22] O. Moreno, J. Cárdenas, L. Atarés, A. Chiralt, "Influence of starch oxidation on the functionality of starch-gelatin based active films," Carbohydrate Polymers, **178**, 147–158, 2017, doi:10.1016/j.carbpol.2017.08.128.
[23] E.R. Kenawy, E.A. Kamoun, M.S. Mohy Eldin, M.A. El-Meligy, "Physically crosslinked poly(vinyl alcohol)-hydroxyethyl starch blend hydrogel membranes: Synthesis and characterization for biomedical applications," Arabian Journal of Chemistry, **7**(3), 372–380, 2014, doi:10.1016/j.arabjc.2013.05.026.
[24] A.L. Da Róz, M.D. Zambon, A.A.S. Curvelo, A.J.F. Carvalho, "Thermoplastic starch modified during melt processing with organic acids: The effect of molar mass on thermal and mechanical properties," Industrial Crops and Products, **33**(1), 152–157, 2011, doi:10.1016/j.indcrop.2010.09.015.
[25] S.C. Pang, S.F. Chin, S.H. Tay, F.M. Tchong, "Starch-maleate-polyvinyl alcohol hydrogels with controllable swelling behaviors," Carbohydrate Polymers, **84**(1), 424–429, 2011, doi:10.1016/j.carbpol.2010.12.002.
[26] D. Kuakpetoon, Y.J. Wang, "Locations of hypochlorite oxidation in corn starches varying in amylose content," Carbohydrate Research, **343**(1), 90–100, 2008, doi:10.1016/j.carres.2007.10.002.
[27] A. V. Reis, M.R. Guilherme, T.A. Moia, L.H.C. Mattoso, E.C. Muniz, E.B. Tambourgi, "Synthesis and characterization of a starch-modified hydrogel as potential carrier for drug delivery system," Journal of Polymer Science Part A: Polymer Chemistry, **46**(7), 2567–2574, 2008, doi:10.1002/pola.22588.
[28] N.L. Vanier, E. Da Rosa Zavareze, V.Z. Pinto, B. Klein, F.T. Botelho, A.R.G. Dias, M.C. Elias, "Physicochemical, crystallinity, pasting and morphological properties of bean starch oxidised by different concentrations of sodium hypochlorite," Food Chemistry, **131**(4), 1255–1262, 2012, doi:10.1016/j.foodchem.2011.09.114.
[29] L. Djekic, M. Martinović, V. Dobričić, B. Čalija, Đ. Medarević, M. Primorac, "Effect of Bioadhesive Polymers on Stability and Drug Release Kinetics of Biocompatible Hydrogels for Topical Application of Ibuprofen.," J Pharm Sci., **108**(3), 1326–1333, 2019, doi:10.1016/j.xphs.2018.10.054.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1381-1386 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Image Tag Recommendation based on Ranked Categorical Nearest Neighbors and Weighted Tag Features

Anupama D. Dondekar*, Balwant A. Sonkamble

Department of Computer Engineering, Pune Institute of Computer Technology, Pune, 411042, India

ARTICLE INFO

Article history:
Received: 16 October, 2020
Accepted: 24 November, 2020
Online: 16 December, 2020

Keywords:
Tag Recommendation
Categorical Nearest neighbor
Weighted Tag Frequency
Flickr

ABSTRACT

The huge number of images on the image sharing websites poses challenges for classification and retrieval of the images. On many image sharing websites, tags can be assigned by the users to an image that describes the contextual and visual description of an image. However, ambiguous or incorrect tags have appeared in frequent tags that affect the performance of an image retrieval system. Thus, assigning appropriate tags to the images plays a very important role in image retrieval and classification. In this paper, the ITR-WTF image tag recommendation method is proposed which explores tags from ranked nearest neighbors of each category. For a given input image, the method first determines the neighbors from training images of each category and ranks the neighbors according to the distance from the input image. In the second step, the weight is assigned to each tag based on the vote from each neighbor. Finally, the weighted tag frequency is determined to recommend appropriate tags to a given image. The experimentation is done on two datasets self-generated and NUS-WIDE. The results obtained using the proposed method ITR-WTF gives good results as compared with the existing methods of tag recommendation.

1. Introduction

The rapid development of advanced technology and the high usage of social media have created a large repository of images which poses many challenges for an image retrieval system. On many image sharing social websites, the users are uploading images for faster communication or to find the people with the same interest. The images on the image sharing social websites are associated with the tags which are assigned by the users. The tags describe the visual content of the images along with the context information such as the location and time the images are captured. These tags are used indexing during an image search. Thus, tags play a very important role in tag-based image retrieval systems and classification.

Many times, users assign tags that are imprecise and irrelevant to the image. According to the survey, only 50% of the Flickr tags describe the content of the images [1]. The presence of irrelevant tags affects the performance of the classification and image retrieval system. Hence, it is necessary to design an algorithm that assigns correct tags to the images. It improves the accuracy of the image retrieval system by suggesting proper tags for the images. The advantages of tag suggestions are: it reduces the cost of manual annotation of the images, spelling mistakes. Several studies have been done on tag recommendation using visual contents, tags and metadata. Still, the performance of tag recommendation is not satisfactory due to personalized tags as shown in Figure 1. The tags may not describe the content of the images and may be assigned as per the user perspective. The tags Denmark, 2011 does not describe visual content of an image as shown in Figure 1 (a).

(a)

Tags:

Denmark, 2011, nature, D60

(b)

Tags:

Fish, Canon 50mm, Canon EOS 5D Mark III

Figure 1: Tag based Images

*Corresponding Author: Anupama D. Dondekar, Email: agphakatkar@pict.edu

In this paper, the ITR-WTF image tag recommendation method is proposed which contributes in the following way:

- Obtained ranked visual neighbor of each category for each feature
- The tag score is calculated by combining weighted tf-idf value and vote from rank categorical nearest neighbor to recommend top k tags

The paper is organized as follows: existing methods for tag recommendation are described in section 2. Section 3 explains the ITR-WTF proposed method for image tag recommendation. Section 4 and 5 describe the metric used for evaluating the performance of the proposed method and dataset used for experimentation. Section 6 describes the experimental results. Finally, in section 7 the conclusion and future work is given.

2. Work Done

The methods for tag recommendation are categorized into three methods: classification, semantic and nearest neighbor based.

In classification-based methods, the features of the images are extracted and classified into a category using a classification algorithm. The classifier trained for multiple class result in a multi-class problem. Finally, the tags are recommended based on the category of an input image. In [2] the method was proposed labeling of the images. The images were segmented into regions and identified salient region. Based on particle swarm optimization, the SVDD trained to assign labels to the images by assigning more weights to the salient regions. In [3] the method was proposed using probability and weight based SVM classifier for annotation of images. Given an image with tag, the method identifies the related and unrelated images using majority of voting from SVM in [4]. The system named 'SheepDog' was developed to which identifies the suitable group for the inclusion of photos and suggest appropriate tags to the users on Flickr dataset in [5].

In the semantic method, the tags are recommended based on the joint distribution of image and tag features. The method was proposed in [6] for annotation of images using KCCA framework by constructing semantic space in which the correlation was built between image feature and tag features. In [7] an approach for retagging of social images with diverse semantics was presented. Both the relevance of a tag and the semantic compensation to the already determined tags was fused to determine the final tag list for a given image. The method was proposed in [8] to recommend tags for geotagged images using unified subspace which correlates the textual and visual features. The hyper-graph-based method was proposed for tag-based image retrieval using image features and tag features simultaneously [9].

The nearest neighbor-based tag recommendation methods are very popular due to its effectiveness. The model is based on the assumption that feature based similar images tend to have the same tags. Given an input image, the method determines k nearest neighbors by combining various features either by using early or late fusion and calculates the tag relevance score by collecting votes of a tag from its nearest neighbors. The advantages of the method are: it is scalable and model building is not needed as it makes an assumption based on training data. The tag recommendation method was proposed in [10] using random walk on bipartite graph constructed based on weighted user and image nearest neighbor. In [11], the Bayesian based image annotation model was proposed based on semantic nearest neighbors. The method was proposed in [12] for annotation of images using a variation of traditional kNN algorithm by defining matrix which shows the relationship between labels and images. In [13] the method was proposed in which the given an image the similar images were determined using k nearest neighbor and tag graph was created from tags of neighbors and clustered to assign label to an image. The personalized image tag recommendation method was proposed based on neighbor voting scheme by building tripartite graph to show relationship between user, tags and images in [14]. The VS-KNN method was proposed in [15] for image labeling by exploring image features and label features simultaneously as a maximum posterior estimation. Given an image with label l, the method was proposed which identify the images labeled with l using kNN and denote it by set S [16]. Finally, the labels were assigned to the images by calculating the similarity between the input image and set S. The tag relevance method was proposed by assigning weight to each neighbor based on distance using kNN [17]. The method was proposed in [18] to suggest tags for an image based on visual features and tag correlations using neighbor voting scheme. Photo tagging method was proposed in [19] using history of the users. The method finds the geographical, visual and time neighbors for a given image and recommends tags by counting accumulating votes for each tag from three types of neighbors. In [20], the method was proposed to suggest tags to the images with and without labels. The method first identifies a set of k images using feature-based k nearest neighbor. Finally, assign or recommends the tags by counting the difference between tag frequency count from the entire database and k neighbor. The image annotation algorithm in [21] identifies rank based and weight based nearest neighbor and suggests the label for an input image using a probabilistic model.

In this paper, the proposed method recommends tags based on the nearest neighbor method. Compared with the classification and semantic based method, the nearest neighbor method is popular due to its effectiveness and scalability. Also, it does not require any training.

However, the performance of the existing nearest neighbor-based tag recommendation methods depends on the number of k neighbors and may be affected due to the equal weight and voting irrespective of the class. This observation motivated us to develop a method which improves the performance of tag recommendation. The method first determines the weighted categorical nearest neighbors and suggest the tags based on the ranked categorical neighbor and weighted tag frequency.

3. Research Methodology

In this section, ITR-WTF proposed tag recommendation algorithm is described. The main objective of the proposed algorithm is to improve the accuracy of tag recommendation by identifying distance based nearest neighbor and rank the tags by combining image score and tag score.

3.1. ITR-WTF Tag Recommendation Method

The block diagram of the proposed method for tag recommendation is shown in Figure 2. The proposed method consists of three main modules: feature representation, classification and tag recommendation.

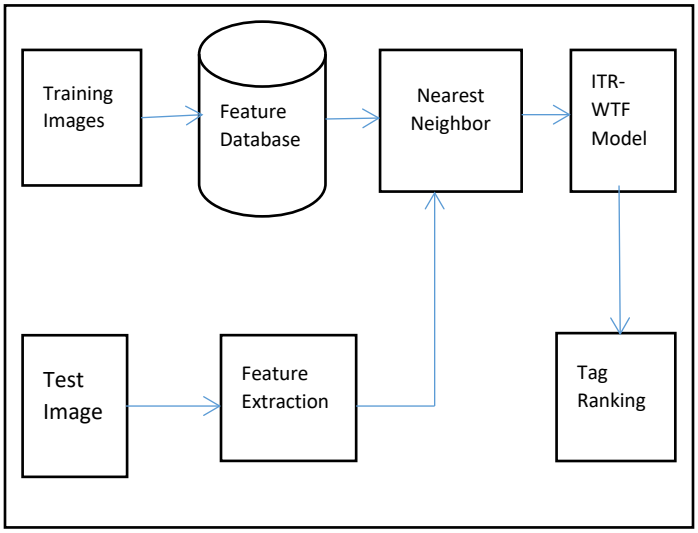


Figure 2: System Diagram of Proposed Method

3.1.1. Feature Representation:

Features play a very important role in image representation. Several researchers have worked on feature extraction methods for image retrieval using color, texture and shape features. Extracting an effective feature and represent them efficiently is very important.

Color is the most used feature in an image retrieval system. Using color features, a human can recognize most images and objects included in the images. Also, the color features are invariant to scaling, translation and rotation of an image. Another important feature is texture. Texture measures look for visual patterns in images and how they are spatially defined.

During the training phase, the features are extracted using color moment and wavelet packet transform [22, 23]. The first, second and third moments are extracted as color features in L*a*b* color space by segmenting an image into two by two sub-block along with a centralized sub-block of the same size resulting into 9 features for each sub-block. For texture features are extraction using wavelet packet transform, an image is decomposed into sub-bands up to level three using daubechies wavelet. The energy and standard deviation of each band are determined as texture features using eq.1 and 2.

$$M_i = \frac{1}{HT*WT} \sum_{h=1}^{HT} \sum_{w=1}^{WT} |Coeff_i(h,w)| \quad (1)$$

$$SD_i = \sqrt{\frac{1}{HT*WT} \sum_{h=1}^{HT} \sum_{w=1}^{WT} (|Coeff_i(h,w)| - M_i)^2} \quad (2)$$

where i = 1 to 4^L and L=3, $Coeff_i(h,w)$ represents the coefficient values of ith decomposed sub-band image at level L, WT and HT are the width and height of the decomposed sub-band image $Coeff_i$.

Each feature has a different range of values. To avoid the influence of one feature due to variation in a range of values, the features are normalized between 0 and 1 range using min-max normalization method.

There are two methods used for combining features: early and late fusion. In early fusion, the individual features are combined before image similarity score calculation. In late fusion, the image similarity score determines for each feature and combine the individual score to calculate the final score. The late fusion method needs more cost of computation [24]. For this reason, the early fusion technique is used to determine the image similarity score.

3.1.2. Nearest neighbor:

After feature extraction, the nearest neighbor classifier determines the neighbors of a test image. In the nearest neighbor classifier, the neighbors for each class/category are determined as shown in Figure 3 and combined to form the final neighbors. The harmonic mean of the neighbors of each category is determined and finally predicts the category with the smallest mean [25].

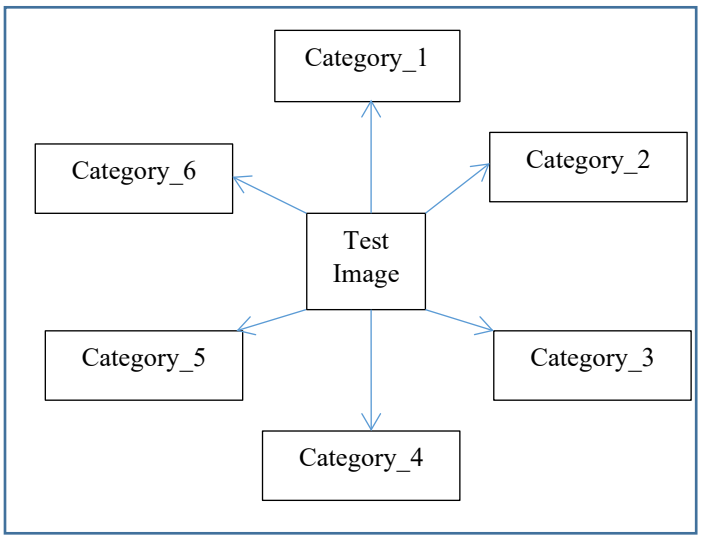


Figure 3: Nearest Neighbor from Each Category

To improve the effectiveness of feature-based neighbors, the weight is assigned to each neighbor so that nearby training image will get more weight and the training images which are farther away will be assigned less weight.

3.1.3. ITR-WTF Model:

Once the feature based neighbors are determined, the method assign an importance to each tag based on image similarity score as follow:

$$P(t) = \sum_{x_i \in NN} sim(I_q, x_i).\delta(x_i, t) \quad (3)$$

where $sim(I_q, x_i)$ denotes the importance of image x_i in predicting tag t according to their visual similarity; $\delta(x_i, t)$ is equal to one if tag is associated with an image x_i, otherwise $\delta(x_i, t)$ is equal to zero. Finally, the score of tag t is calculated as follow:

$$w - tfidf(t) = \left[1 + \log(N_t) * \log\left(\frac{NN}{N_t}\right) * P(t)\right] \quad (4)$$

where N_t represents the number of images associated with t and NN represents the total number of nearest neighbors.

3.2. Algorithm

Algorithm 1: The proposed ITR-WTF Algorithm

Input:

CV: Color feature vector of an input image
TV: Texture feature vector of an input image
$TFV_j = (y_i^j, C_i^j)_1^{NF_j}$: j[th] class texture feature vector of training images
$CFV_j = (x_i^j, C_i^j)_1^{NF_j}$: j[th] class color feature vector of training images
M = $C_1, C_2, \ldots\ldots C_M$: the number of class labels
$T_{mn} = \{T_{1n}, T_{2n}, T_{3n}, \ldots\ldots\ldots T_{mn}\}$: Tag associated with n[th] image

Output: Ranking of Tags

Step 1: Calculate the distance between CV and CFV_j using eq. (3). The set of m number of nearest neighbor for each class C_j is denoted as $CD_{mj}^{NF} = [x_{1j}^{NF}, x_{2j}^{NF}, \ldots\ldots, x_{mj}^{NF}]$

$$dist(CV, x_i^j) = \sqrt{\sum_{l=1}^{m}(CV_l - x_{il}^j)^2} \quad (5)$$

where $x_i^j \epsilon\, CFV_j$

Step 2: Calculate the distance between TV and TFV_j using eq. (4). The set of n number of nearest neighbor for each class C_j is denoted as $TD_{nj}^{NF} = [y_{1j}^{NF}, y_{2j}^{NF}, \ldots\ldots, y_{nj}^{NF}]$

$$dist(TV, y_i^j) = \sqrt{\sum_{l=1}^{n}(TV_l - y_{il}^j)^2} \quad (6)$$

where $y_i^j \epsilon\, TVF_j$

Step 3: Merge the nearest neighbors obtained using color and texture features

$NN = \{CD_{ij}^{NF} \cup TD_{ij}^{NF}\}$ where i = 1……M

Step 4: Assign an importance to each tag based on image similarity score as follows:

$$P(t) = \sum_{x_i \in NN} sim(I_q, x_i).\delta(x_i, t)$$

where

$$sim(I_q, x_i) = dist(I_q, x_i) * wt_j^i$$

$$wt_j^i = \frac{1}{j} \qquad i = 1, \ldots\ldots, m$$

Step 5: Calculate the w-tfidf of each tag as follows

$$w - tfidf(t) = \left[1 + \log(N_t) * \log\left(\frac{NN}{N_t}\right) * P(t)\right]$$

Step 6: Rank the tags according to the w- tfidf values and select top k values

4. Performance Metric

To evaluate the performance of proposed method, we use NDGC, Precision, Recall and F1-score.

Given an image with ranked tag list $T_1, T_2, \ldots, T_n$, the NDCG is computed as where r(i) is the relevance level of the ith tag and Z_n is a normalization constant that is chosen so that the optimal ranking's NDCG score is 1. After computing the NDCG measures of each image's tag list, we can average them to obtain an overall performance evaluation of the tag ranking method.

$$NDCG@k = \frac{1}{z}\sum_{i=1}^{k}\frac{2^{rel(i)} - 1}{\log(1+i)} \quad (7)$$

rel(i) is a binary indicator, which is equal to one if the ith tag in the ranking list is relevant to an input image, and zero otherwise.

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1 - Score = \frac{2 * precision * recall}{precision + recall} \quad (10)$$

5. Dataset

Two datasets are used: self-generated and NUS-WIDE. The self-generated dataset consists of images collected from Flickr image sharing website belonging different categories such as fish,

actor, aeroplane, butterfly, autumn etc. and each category consists of 300 images. The size of the images is fixed with maximum width or height to 320 pixels. The eight categories of the images from NUS-WIDE dataset are used for experimentation. The images are divided into two groups: training and testing images after 10-fold cross validation.

For self-generated dataset, the tags associated with the images are collected from Flickr using public API. However, some tags are do not describe the image content. Therefore, tags related to year, camera and brands are excluded from the tag list.

6. Experimental Results

The performance of traditional K nearest neighbor and nearest neighbor is shown in Table 1 and 2 respectively.

Table 1: Performance of traditional K nearest neighbor

Method	Self-generated dataset	NUS-WIDE dataset
Precision	84.33 %	65.87 %
Recall	84.50%	60.75 %
F1-Score	84.00%	61.37 %

Table 2: Performance of Nearest Neighbor

Method	Self-generated dataset	NUS-WIDE dataset
Precision	87.73 %	69.88 %
Recall	88.66 %	67.00 %
F1-Score	88.00 %	67.13 %

From Table 1 and Table 2 it is observed that the performance of nearest neighbor based on each category achieves good performance as compared to traditional k nearest neighbor.

The Figure 4 shows that the NDGC value is high for tag recommendation result when no of the tags recommended is 5 for self-generated dataset and 15 for NUS-WIDE dataset.

Table 3 shows the performance of the existing and proposed tag recommendation algorithm. The performance of proposed tag recommendation algorithm is better as it can recommend/suggest tags with higher NDGC score.

Tagvoting method [18]: In this method, the feature based similar images are determined and the tags are recommended to an input image based on the frequency of tags that appeared in k visually similar images. The method assigns a uniform weight to each neighbor.

TagProp method [20]: In TagProp method, the weights are assigned each visual neighbor of a query image. The weights are assigned using rank based and distance-based method.

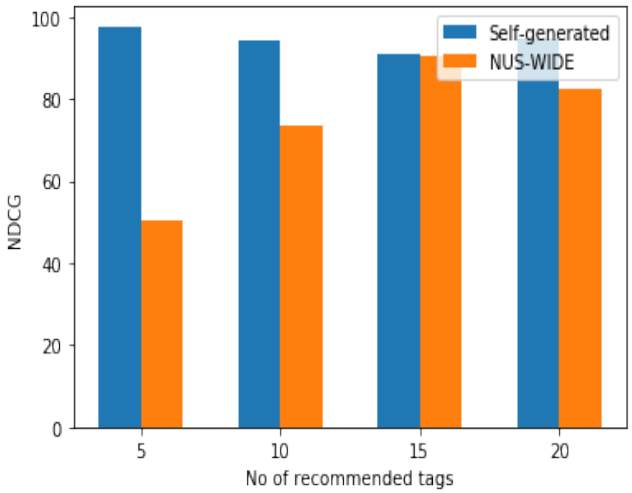


Figure 4: The NDCG values for different number of tags

NVote method [19]: In Nvote method, the tags are recommended based on the difference between global and local tag frequency by assigning equal weight to each neighbor.

Table 3: Performance of different tag recommendation algorithm

Method	Self-generated dataset	NUS-WIDE dataset
Tagvoting	84.45 %	85.20 %
TagProp	89.67 %	68.50 %
NVote	92.07 %	68.96 %
ITR-WTF	95.43 %	88.90%

From Table 3, it is observed that the methods for tag recommendation using the nearest neighbor depend on the value of k which indicates the count of neighbors. The existing methods first identify the uniform/weighted neighbors and then consider only tag information for tag voting which affects the accuracy of tag recommendation. The proposed method improves the accuracy of tag recommendation by combining the image score and tag score.

Table 4: Result of Tag Recommendation

Image	Initial Tags	Recommended Tags
	Brainedge Background Celebration Closeup Clover	Clover Green Nature Leaves Macro

	Bracom (Bram) Bracom Forest Bos Autumn Herfst	Autumn Trees Leaves Path Forest
	Mikepaws London Airport Aircraft Aeroplane	Flying Aircraft Aeroplane Plane Air
	Demerarah Fishalive Tropical-Fish Fish Tropical Fish-Tank Underwater	Fish Aquarium Tropical-Fish Underwater Fishalive

Table 4 shows the result of tag recommendation obtained using the proposed algorithm on self-generated dataset. The initial tag does not describe the entire image content. The proposed algorithm recommends relevant tags to the images. For the second image in table 4 initial tags do not cover the trees, forest and path between the trees which are added by the proposed algorithm.

Therefore, the effectiveness of the proposed method ITR-WTF for tag recommendation is demonstrated using examples in Table 4.

7. Conclusion and future work

In the paper, a method is proposed for tag recommendation of the images by identifying rank neighbors from each category. The method improves the accuracy of tag recommendation by combining the tag frequency score and weighted similarity score of the nearest neighbor images of each category. The experimentation is done two datasets: self-generated and NUS-WIDE dataset. The effectiveness of the proposed method is demonstrated in the experimental results.

The future work will focus on: i) exploring the relationship between tags obtained using from nearest neighbor ii) developing a more optimized approach which works on large dataset iii) exploring metadata associated with the images.

References

[1] E. Spyrou, P. Mylonas, "An overview of flickr challenges and research opportunities," in Proceedings - 9th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP-2014, 2014, doi:10.1109/SMAP.2014.19.

[2] Z. Hao, H. Ge, L. Wang, "Visual attention mechanism and support vector machine based automatic image annotation," PLoS ONE, 2018, doi:10.1371/journal.pone.0206971.

[3] Wu, Wei, Jianyun Nie, and Guanglai Gao, "An improved SVM-based multiple features fusion method for image annotation," Journal of Information & Computational Science **11**(14), 4987-4997, 2014.

[4] X. Li, C.G.M. Snoek, "Classifying tag relevance with relevant positive and negative examples," in MM 2013 - Proceedings of the 2013 ACM Multimedia Conference, 2013, doi:10.1145/2502081.2502129.

[5] H.M. Chen, M.H. Chang, P.C. Chang, M.C. Tien, W.H. Hsu, J.L. Wu, "Sheepdog - group and tag recommendation for flickr photos by automatic search-based learning," in MM'08 - Proceedings of the 2008 ACM International Conference on Multimedia, with co-located Symposium and Workshops, 2008, doi:10.1145/1459359.1459473.

[6] T. Uricchio, L. Ballan, L. Seidenari, A. Del Bimbo, "Automatic image annotation via label transfer in the semantic space," Pattern Recognition, 2017, doi:10.1016/j.patcog.2017.05.019.

[7] X. Qian, X.S. Hua, Y.Y. Tang, T. Mei, "Social image tagging with diverse semantics," IEEE Transactions on Cybernetics, 2014, doi:10.1109/TCYB.2014.2309593.

[8] J. Liu, Z. Li, J. Tang, Y. Jiang, H. Lu, "Personalized geo-specific tag recommendation for photos on social websites," IEEE Transactions on Multimedia, 2014, doi:10.1109/TMM.2014.2302732.

[9] Y. Gao, M. Wang, Z.J. Zha, J. Shen, X. Li, X. Wu, "Visual-textual joint relevance learning for tag-based social image search," IEEE Transactions on Image Processing, 2013, doi:10.1109/TIP.2012.2202676.

[10] L. Zheng, Z. Tianlong, H. Huijian, Z. Caiming, "Personalized tag recommendation based on convolution feature and weighted random walk," International Journal of Computational Intelligence Systems, 2020, doi:10.2991/ijcis.d.200114.001..

[11] Y. Ma, Y. Liu, Q. Xie, L. Li, "CNN-feature based automatic image annotation method," Multimedia Tools and Applications, 2019, doi:10.1007/s11042-018-6038-x.

[12] Q. Ji, L. Zhang, X. Shu, J. Tang, "Image annotation refinement via 2P-KNN based group sparse reconstruction," Multimedia Tools and Applications, 2019, doi:10.1007/s11042-018-5925-5.

[13] V. Maihami, F. Yaghmaee, "Automatic image annotation using community detection in neighbor images," Physica A: Statistical Mechanics and Its Applications, 2018, doi:10.1016/j.physa.2018.05.028.

[14] J. Zhang, Y. Yang, Q. Tian, L. Zhuo, X. Liu, "Personalized Social Image Recommendation Method Based on User-Image-Tag Model," IEEE Transactions on Multimedia, 2017, doi:10.1109/TMM.2017.2701641.

[15] Q. Ji, L. Zhang, Z. Li, "KNN-based image annotation by collectively mining visual and semantic similarities," KSII Transactions on Internet and Information Systems, 2017, doi:10.3837/tiis.2017.09.016.

[16] Y. Verma, C. V. Jawahar, "Image Annotation by Propagating Labels from Semantic Neighbourhoods," International Journal of Computer Vision, 2017, doi:10.1007/s11263-016-0927-0.

[17] S. Lee, W. De Neve, Y.M. Ro, "Visually weighted neighbor voting for image tag relevance learning," Multimedia Tools and Applications, 2014, doi:10.1007/s11042-013-1439-3.

[18] C. Chaoran, Jialie Shen, Jun Ma, and Tao Lian, "Social tag relevance learning via ranking-oriented neighbor voting," Multimedia Tools and Applications **76**(6) 8831-8857, 2017.

[19] X. Qian, X. Liu, C. Zheng, Y. Du, X. Hou, "Tagging photos using users' vocabularies," Neurocomputing, 2013, doi:10.1016/j.neucom.2012.12.021.

[20] L. Xirong, Cees GM Snoek, and Marcel Worring, "Learning social tag relevance by neighbor voting," IEEE Transactions on Multimedia **11**(7), 1310-1322, 2009.

[21] M. Guillaumin, T. Mensink, J. Verbeek, C. Schmid, "TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation," in Proceedings of the IEEE International Conference on Computer Vision, 2009, doi:10.1109/ICCV.2009.5459266.

[22] D.A. Dondekar and Balwant A. Sonkamble, "Analysis of Flickr Images using Feature Extraction Techniques," 4th IEEE International Conference on Computer and Communication Systems (ICCCS 2019), Singapore, 278-282 2019.

[23] D.A. Dondekar and Balwant A. Sonkamble,"Tag-based Image Retrieval using Hybrid Visual-Tag Feature Extraction Method", International Journal of Advanced Science and Technology, **9**(4), 5931 – 5940, 2020.

[24] X. Li, "Tag relevance fusion for social image retrieval," Multimedia Systems, 2017, doi:10.1007/s00530-014-0430-9.

[25] D.A. Dondekar and Balwant A. Sonkamble, "Harmonic Mean based Classification of Images using Weighted Nearest Neighbor for Tagging" International Journal of Advanced Computer Science and Applications, **11**(11), 2020, doi:/10.14569/IJACSA.2020. 0111131

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1387-1393 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Analysis of Physical and Mechanical Properties of Galvanic-Plasma Wear-Resistant Coatings

Mahmood Shaker Albdeiri*, Sergey Sergeev, Vladimir V. Krasilnikov

Belgorod state national research university, Pobedy St, 85, Belgorod 308015, Russia

ARTICLE INFO

Article history:
Received: 11 October, 2020
Accepted: 07 December, 2020
Online: 16 December, 2020

Keywords:
Plasma electrolytic oxidation
Coatings
Microhardness
Porosity
Phases
Wear resistance
Heat-resistance
Pin-on-disk
Tribological properties

ABSTRACT

This article shows the possibility of creating high wear-resistant and corrosion-resistant coating by plasma electrolytic oxidation (PEO) on the AlSi12Cu aluminum alloy. The combination of influencing parameters during processing allows to influencing of the set of quality indicators of coating, such as the thickness of the coating layer, microhardness, porosity, phase and elemental composition. The influence of technological modes and electrolyte composition on the structure and composition of the resulting coating has been established. The structural features of the modified layers was investigated, the effect of the capacitor capacitance on the installation and the duration of treatment on the structure and microhardness of the resulting coating was established. The tribological properties of the different PEO surfaces had been evaluated by usage of the 'pin-on-disk' method. Different current densities were chosen to represent the arrangement component of the PEO layer. Wear test results were used to evaluate the wear resistance ability of the various PEO layers and for establishing a wear model. The surface morphology, porosity distribution on cross-section, and worn surface of the PEO layers were characterized using a scanning electron microscope (SEM). Different material counterparts and contact conditions were applied to simulate a variety of practical states. The results show that the hardness of the PEO layer formed on sample M5 was the highest among the five kinds of samples. The PEO layer formed at sample M5 has the best wear resistance. This can be ascribed to the different size and appropriation of the pores and the diverse proportion of α to the γ stage in the PEO layers when exposed to various limitations during the PEO cycle.

1. Introduction

During the operation of the internal combustion engine, the parts of the cylinder-piston group wear out most quickly. The wear of the cylinder-piston group of the engine takes place in conditions of high temperature, significant cyclic mechanical loads and an aggressive gas environment. One of the effective ways to increase the service life and reliability of the rotor group of an internal combustion engine is *PEO*, which leads to the formation of multifunctional wear- resistant coating based on ceramics on the working surface, which are characterized by high adhesion to the base [1].

PEO is a modern promising technology for electrochemical oxidation of valve group metals, including aluminum, titanium, zirconium, magnesium, niobium, beryllium, and tantalum, as well as their alloys in order to form a uniform ceramic coating with a thickness of 80 - 300 (μm) with high performance and protective properties. The environmentally friendly *PEO* method is optimal for processing and hardening the surfaces of products with complex-shaped parts and limited units of the electrical, aviation, oil and gas processing industries, instrument, and mechanical engineering, as well as shipbuilding, medicine, and utilities [2, 3]. The *PEO* method makes it possible to obtain a coating on the surface of aluminum alloys, which includes the γ- and α-phases of aluminum oxide, mullite (in the case of alloying the alloy with silicon). Corundum (α-phase) has a low coefficient of friction (0.04 and less) and low thermal conductivity, which increases the wear resistance of the friction pair, reduces heat transfer from the working chamber, and provides high corrosion resistance. Moreover, it has a higher hardness compared to the γ-phase of aluminum oxide. On the other hand, the presence of the γ-phase makes it possible to somewhat increase the coating resistance for cyclic mechanical and thermal loads. Mullite, in turn, has a low density, high thermal stability and corrosion resistance, low thermal conductivity, acceptable values of strength characteristics

*Corresponding Author: Mahmood Albdeiri, Belgorod- Russia, engmahmood86@gmail.com

[4, 5]. Many previous studies that have discussed the effect of coating thickness [6], solution concentrations [7], and porosity size [8] on corrosion performance and hardness. In this work, suitable thicknesses between 80 -100(μm) were chosen to obtain high hardness. In an attempt to get the lowest possible porosity, but impossible to obtain a coating that does not contain pores, due to the nature of the process. If necessary, the porosity can be reduced by saturating the surface layer with the application of a polymer. The coating produced had the best results for the lowest possible wear rate and thus reduce the friction coefficient to a minimum.

Favorable factors for the widespread of the PEO process in the industry are a set of such advantages as the possibility of obtaining high-quality wear-resistant, heat-protective coatings in combination with sufficiently high productivity, simplicity of hardware design, environmental friendliness, and cost-effectiveness of process automation [9]. The aim of the study is the influence of the parameters of the PEO process on the structure, mechanical and physical properties of the coating applied to the aluminum alloy AlSi12Cu to get high wear resistance.

Also, the goal of this study was formulated: to assess the possibility of applying GPM to the aluminum alloy AK12MMGH, Modification with the help of GPM makes it possible to increase the resource and reliability of parts of the Internal Combustion Engine (ICE), also protect them from high-temperature gas erosion and reduce the temperature of the base metal by 1.5 times [10]. To work out the modes and to study the structure and properties of the obtained coatings

2. Materials and Research Methods

Prepared AlSi12Cu Samples for coating were made in the form of a ring with a surface area of (3.6) dm^2. The external diameter size of such rings was 200 mm and the internal diameter was 164 mm and the thickness were 10 mm. An ELS device MDO-50-AKT-001 was used to form the coating. Two regimes were used to start the process, the anode regime was used, which made it possible to accelerate the formation of the primary oxide layer on the surface of the samples. After two minutes of anodic treatment, the capacitor-state regime was started, as the most suitable for the formation of a coating with the reduction of micro stresses, which controls energy dissipation due to the coupling of mechanical and thermal fields.

Cross-sections were prepared by mechanical grinding and polishing on a LaboPol - 5 (Struers) machine using abrasive paper and a silicon oxide slurry. The microstructure of the cross-sections of the coating obtained was investigated using Quanta 200 scanning electron microscope (*SEM*) in the backscattered electron mode, with an accelerating voltage of 30 and 15 kV. The thickness and porosity of the coating were determined on a cross-sectional image using the Digimizer program.

The electrolyte of *PEO* consists of Potassium Hydroxide (*KOH*) and Sodium Silicate (Na_2SiO_3). In preliminary experiments various composition of the electrolyte was tested in terms of *KOH* content in the range from 4 to 20 g/L, and in terms of Sodium Silicate content in the range from 4 to 20 g/L. It was found that coatings with less porosity and maximal thickness are obtained in the range 6.4 - 9 g/L for *KOH*, and 5 - 10 g/L for Sodium Silicate. So, this range was chosen for carrying out more detailed experiments to study the structure, elemental and phase composition, etc.

The study was carried out on five samples, the obtaining technological regimes of which are indicated in Table 1.

The duration of the processing was determined by a decrease in the intensity of the process, which means a possible transition of the process from the micro-arc stage to the arc stage. The last limitation is associated with the onset of etching and the destruction of the coating. The samples were cleaned with water for 3 min after treatment to prevent the deposition of electrolyte components during drying.

Table 1: The value of the thickness formed coating under different processing conditions

Sample	A/ dm^2	τ, min	Electrolyte, g/L	
			KOH	Na_2SiO_3
M0	($AlSi_{12}Cu$) Without coating			
M1	30	120	9	10
M2	30	120	9	5
M3	22	120	6.4	10
M4	22	120	9	10
M5	22	120	6.4	5

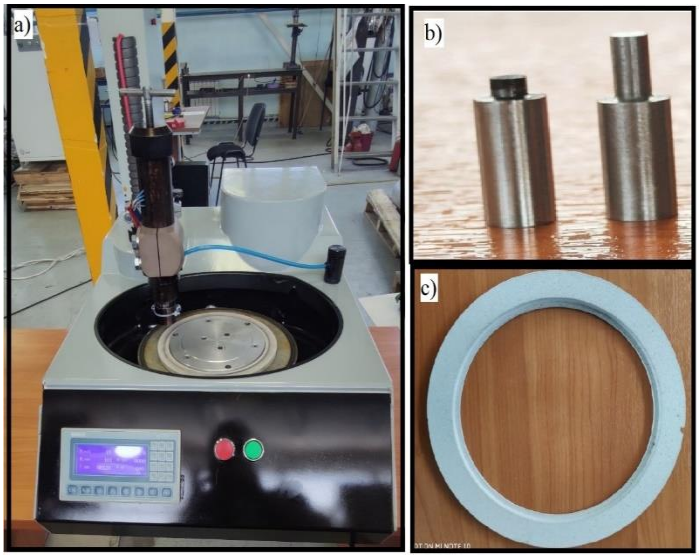


Figure 1: a) Image of the experimental setup for determining wear resistance; b) pin after testing (left), original (right) and c) coated substrate.

The 'pin on disc' wear testing machine used to study the mechanisms, wear rate and for determination of friction coefficients between the composite coating on a ring and the pin acting as a counterbody. The image of this machine is represented in Figure 1. Standard technical oil for a diesel engine was used as a lubricant for such wear tests. An aluminum alloy ring with a coating which deposited on its surface was fixed to the annular disc of the tribometer, and a 6 mm or 10 mm diameter cast iron pin was brought in contact. The load applied by the pin on the coating, as well as the rotation speed of the coated ring, were set and controlled during the experiment. The time of the experiment, the load on the pin, and the rotational speed of the coated ring

were 180 min, 90 N, and 1500 rpm, respectively fixed for all experiments. All tests were carried out at room temperature and atmospheric pressure. The wear rate was determined by evaluating the cross-sectional area of the wear grooves [11]. This data was obtained by analyzing the surface profile across the wear track using a SURTRONIC 25 profilometer. Wear rate was defined as volume loss under a unit load at a unit rotating distance.

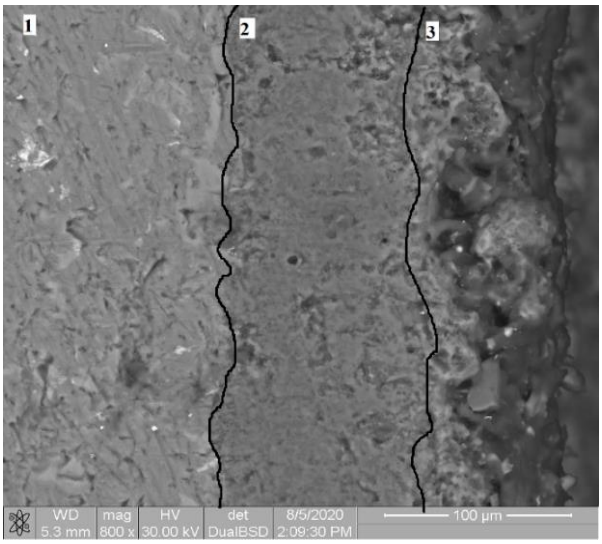


Figure 2: The coatings structure layers on the example of sample M1, obtained by the PEO method: 1-transition layer; 2- working layer; 3- technological layer.

3. Results and Discussion

SEM image of a cross-section of a sample *M1* with a 210 µm thick coating on the AlSi12Cu aluminum alloy is shown in Fig. 2. The ceramic coating contains three layers: a 5 µm thin transition layer, a 120 µm thin working layer and a Technological layer with a thickness of 70 µm. The coatings obtained are characterized by good interfacial adhesion between the aluminum substrate and the ceramic coating. The same result was shown in previous work [12, 13]. Analyzing the cross-section image of the coatings show that the growth of the coating occurs in two directions. The first direction from the surface of the substrate to its depth (working layer) and this take place in the direction of the aluminum substrate, this forms a dense layer with good mechanical properties. The second one is the growth of the coating in the direction of the outer surface accompanied by the formation of a loose structure (Technological layer). The wear resistance of the coating obtained by the *PEO* method will be almost completely determined by the properties of the inner working layer. Therefore, the upper porous layer of the coatings was grinded to the working layer for tribological measurements, as well as for X-ray phase studies.

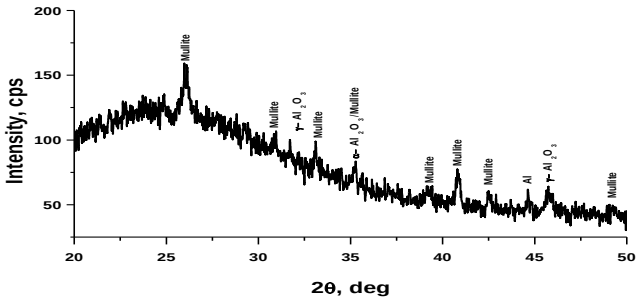


Figure 3: X-ray diffraction pattern of the technological layer of the coating applied according to mode sample *M5*

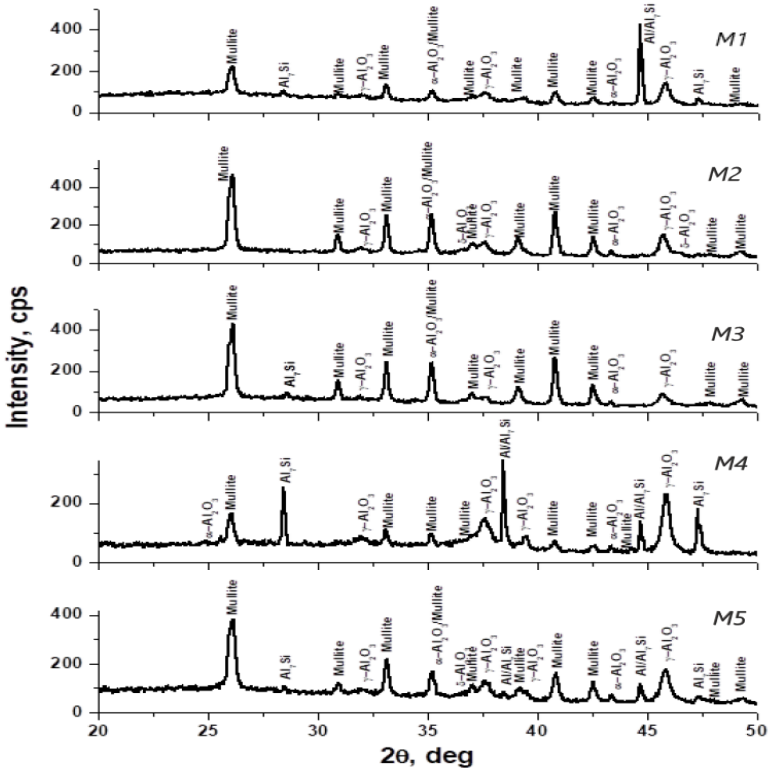


Figure 4: X-ray diffraction patterns of the working layer for applied coatings for the samples represented in table 1.

In (Figure 4) X-ray phase studies have shown the absence of pure aluminum in the composition of the coatings. It is seen that the coatings obtained by the *PEO* method on the AlSi12Cu alloy consist mainly of 3 phases: α- Al_2O_3, γ- Al_2O_3 and mullite. Diffraction reflections (105) α, (200) γ, and (400) mullite with high reflection intensity are identified as characteristic diffraction peaks of α- Al_2O_3 and γ- Al_2O_3, respectively. The working layer of the coating mainly consists of α- Al_2O_3 phase the content of which is much higher than that of γ- Al_2O_3 (see Fig. 4 *M2*). It can be seen that the content of the α- Al_2O_3 phase correspondingly increases from the working layer to the technological layer of the coating comparing the ratio of the intensities of reflection (105) α to reflection (200) γ at different depths of the coating. We assumed that this is due to a change in the cooling rate of molten aluminum oxide in the microarc discharge channel [14].

The presence of porosity is a positive moment when the coating is operating under oil starvation conditions, the lubricant enters the pores of the coating and provides a permanent oil film. The value of volumetric porosity and average pore size for different coatings are represented in Figure 5.

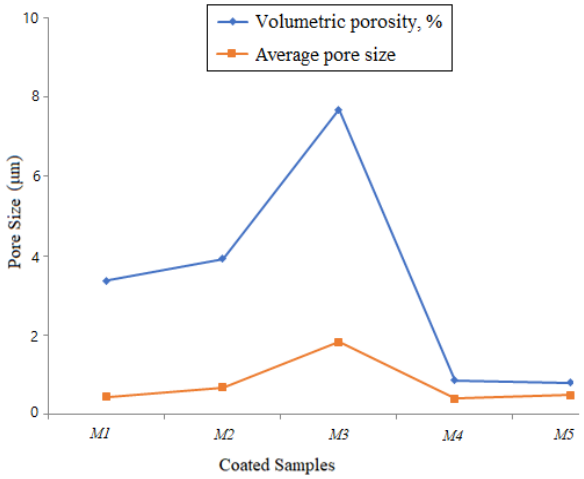


Figure 5: Volumetric porosity and average pore size of various processing modes.

The microhardness of the working layer is 1000-1500 HV (Fig. 6). A piston with a coating based on oxyceramic with a hardness of 6.3–11.5 GPa was tested in [15, 16]. It was also found in previous works that the wear rate of coated pistons is 3.86 times lower than that of uncoated ones [17,18].

Measurement of the microhardness of the applied coatings was carried out on cross-sections surfaces. This was done on the one hand, because it was impossible to make a correct measurement on the surface with the irregularity height comparable to diamond pyramid size and on the other hand it makes it possible to measure the change in microhardness over the thickness of the coating.

Figure 6 shows a graph of the distribution of the microhardness of the coating along with its thickness for example sample *M3*. Zero on the abscissa in this graph corresponds to the surface of the aluminum substrate on which the coating was applied. The microhardness of the coating increases first relatively slowly in the transition layer and then rapidly during the transition to the working layer until it reaches its maximum value at a depth of 40 µm from the substrate-coating interface is 1600 HV. After reaching the maximum value, it begins to gradually decrease with thickness increasing. The hardness of the coating does not exceed the value of 1600 HV in the distances range of 25–100 µm from the Al / Al_2O_3 interface. The microhardness of the aluminum alloy substrate near the Al / Al_2O_3 interface is only 210 HV [19] which is close to the hardness of 185 HV in the substrate, measured far from the interface.

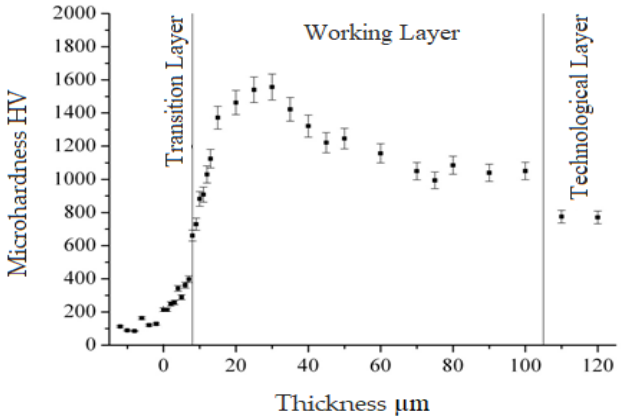


Figure 6: Typical distribution of microhardness over the coating thickness (shown for sample *M3*).

Figure 7 shows the cross-sectional profiles of the surface across the wear track of the a) initial substrate made of the AlSi12Cu alloy sample *M0* and c) polished coating of the sample *M5*. The wear track of the substrate made of AlSi12Cu alloy as shown in Figure 7b has a depth of 61 µm and a height of 23.2 µm, which is much deeper than that of a coated ring whose depth is 30.9 µm and a height of 6.9 µm as shown in Figure 7d. This means that the surface wear rate decreases significantly after the coating deposition on the AlSi12Cu alloy by the *PEO* method. Figure 7b shows the parabolic shaped profile of the wear groove with superficial changes in depth. This profile is formed with uniform abrasive wear of the material. Figure 7d, shows that there is no wear of the coating material over the entire width of the track, but instead there are narrow furrows that could occur if a very small fragment of the coating breaks off during testing and therefore forms a scratch in the area of the friction track.

Roughness profile parameters are showing in Figure 7. The roughness profile parameters for all samples were measured. It was observed that the maximum profile valley depth (*Rv*) and the average roughness (*Ra*) are more than two times higher for the uncoated sample in comparison to the coated one.

As it was mentioned above, the coatings were ground to remove the transition loose layer, which has poor mechanical properties. Since the thickness of the transition and the working layers were different for different samples and it was necessary to grind different amounts of materials. As a result, the remaining coating thickness was different for different samples. So, the microhardness distribution over the thickness (the thickness was determined from the interface between the metallic substrate and the coating) was carried out for different sample thicknesses. Table 2 shows the value of the coating wear rate for different samples obtained at a fixed operating time and applied load. It was

found that the smallest wear rate for coating samples is only 3.4295×10^{-7} mm^3 / Nm, which is a thousand times less than for the uncoated Al substrate. This means that the wear resistance surface of an aluminum alloy by *PEO* can be improved by three orders of magnitude by oxide coating deposition. The loss of coating thickness during the wear test is much less than the thickness of their working layer, so table 2 describes the results of wear of the working layers of coatings.

As shown for coating samples *M2* - *M5*, the wear rate gradually decreases as the remaining coating thickness decreases. However, the difference in the wear rate of samples *M4* and *M5* is small, which indicates a similar wear resistance at different thicknesses of the working layer of the coating. It should be noted that in the process of wear, some of the metal chips can be clogged in the pores of the ceramic coating, which leads to the effect of lubrication, which reduces the coefficient of friction. This leads to the fact that the wear of the coating is greater at the initial stage, however, when the friction couples reach stable wear, the tribocontact zone of the coating and the pin becomes very smooth, and the aluminum oxide coating practically does not wear out.

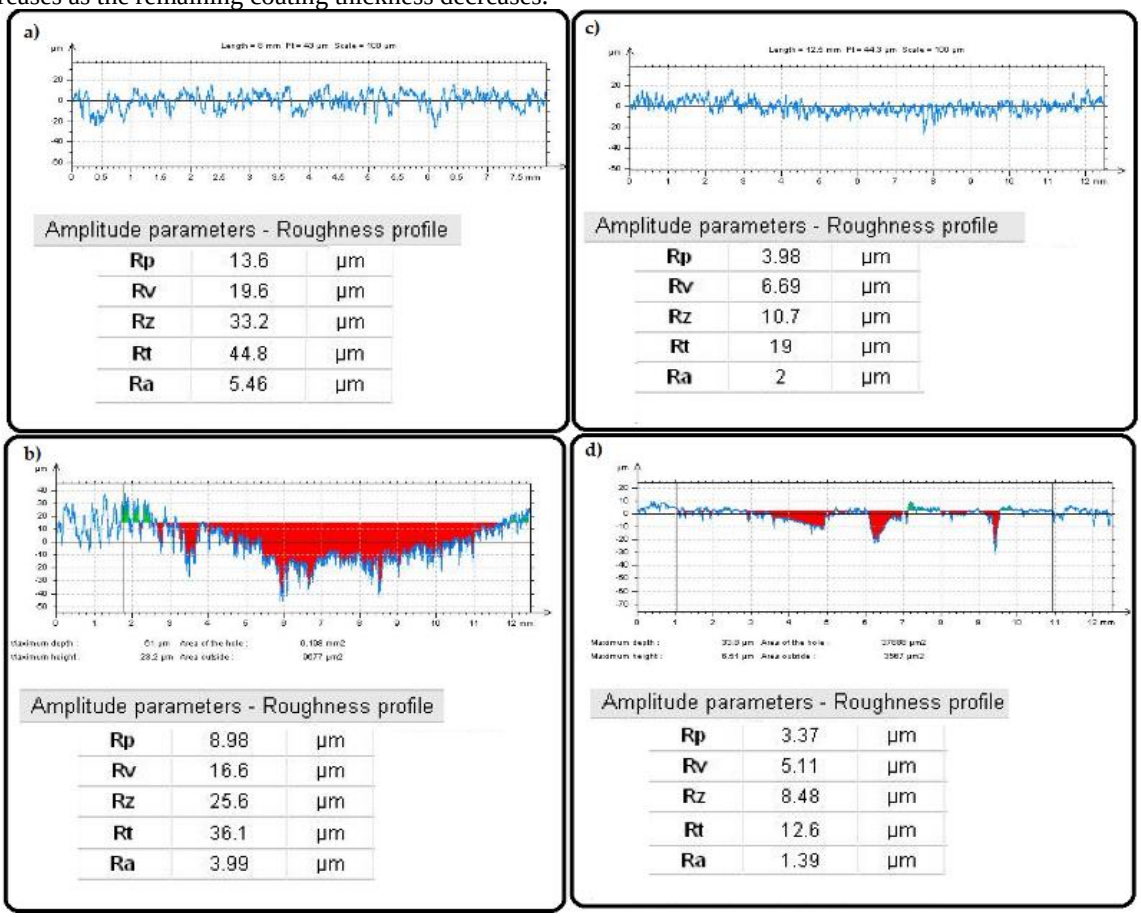


Figure 7: Typical cross-sectional profiles of the surface on an aluminum substrate (sample *M0)* before wear track (a); across the wear track (b); and on the polished coating with a remaining thickness of 180 µm (sample *M2*) before wear track (c); across the wear track (d).

Table 2: Wear rate of a substrate without coating and polished coating by tribological parameters.

Sample	Coating thickness (µm)	Loss of thickness in friction track (µm)	Wearing time (min)	Wear load (N)	Wear rate (mm^3/Nm)
M0	0	50	180	90	8.73016×10^{-4}
M1	210	38	180	90	1.38973×10^{-6}
M2	180	29	180	90	2.54766×10^{-7}
M3	160	27	180	90	3.4295×10^{-7}
M4	80	14	180	90	1.6183×10^{-6}
M5	90	10	180	90	1.4610×10^{-6}

The change of the friction coefficient for sample *M2* is paired with a cast iron pin depending on the time of wear is shown in Fig. 8. In the first minutes of the experiment, the lapping of a pair of rubbing bodies occurs. At this time, the contact area changes (increases), primary chips are formed, which in total leads to an increase of the friction coefficient up to 0.85. During 15 minutes of wear, the coefficient of friction (μ) is still high and remains at the level of about 0.75. At this stage, there is an intensive formation of α-Al_2O_3 and γ-Al_2O_3 microparticles, which cause intensive wear of both the coating and the counter body. This leads to the expansion and deepening of the wear mark on the coating. At the same time, there is an intensive formation of metal chips of the counter body material. In addition, the ceramic grains of α-Al_2O_3 and γ-Al_2O_3 in the coating become smoother, and part of the metal chips is driven into the pores of the coating. Thus, at this stage, the coefficient of friction remains at a high level. As we continue the experiment, the metal particles from the pin become clogged in the pores of the ceramic coating, which leads to the fact that the ceramic-to-metal contact turns into a metal-to-metal contact, as a result of which the coefficient of friction decreases. This contact looks like a thin lubricating layer of cast iron, which creates a lubricating effect between the pin and disc. In this case, the wear loss is very small, which results in a very low wear rate of coatings with longer test times.

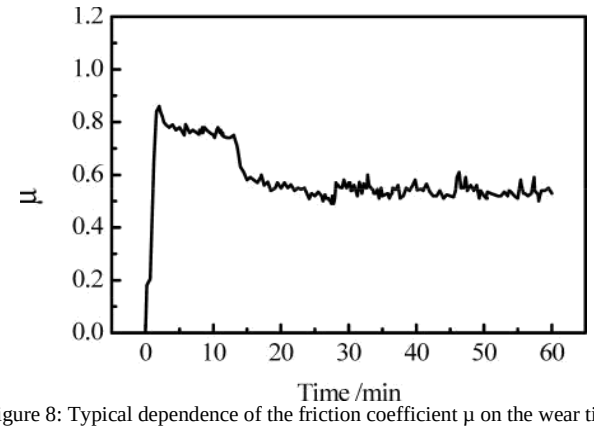


Figure 8: Typical dependence of the friction coefficient μ on the wear time (measured at *M1*).

Figure 9 and 10 shows images of signs of wear on the coating and the counterbody. As shown in figure 9, a wide wear path is formed in the tribological contact zone. The wear mark shows many grooves, which corresponds to the profile in Fig. 7. These grooves are the result of abrasive wear by small particles that have broken off from the coating. The wear trace of the counterbody as shown in figure 10 have more cracks than the coated ring because the hardness of a cast iron rod is less than oxide coating.

4. Conclusions

Wear tests were performed using standard technical oil for a diesel engine as a lubricant. The wear rates of the coatings and counterbody, as well as the friction coefficients for each condition at a normal load of 90 N for 180 min, were obtained. *SEM* and *XRD* were used to analyze the effect of process parameters and we founded that the best range of electrolyte composition for *KOH* (6.4-9 g/L) and Na_2SiO_3 (5-10 g/L) in order to get volumetric porosity less than 1% and hardness more than 1600 HV. Depend on in the intensity of the *PEO* process the selected time was 120 min and current density was in range (22 – 30 A/ dm2). When time or current density of Poe process increased this led to destruction of the coatings. A comparative analysis of the structure of experimental samples allows us to conclude that it is possible to provide different properties of coatings on the same material. In particular, the experimental sample *M5* has low porosity and a working layer thickness of 90 μm; therefore, it has higher mechanical strength and wears resistance. This makes the coating most suitable for resizing and increasing the wear resistance of the piston skirt.

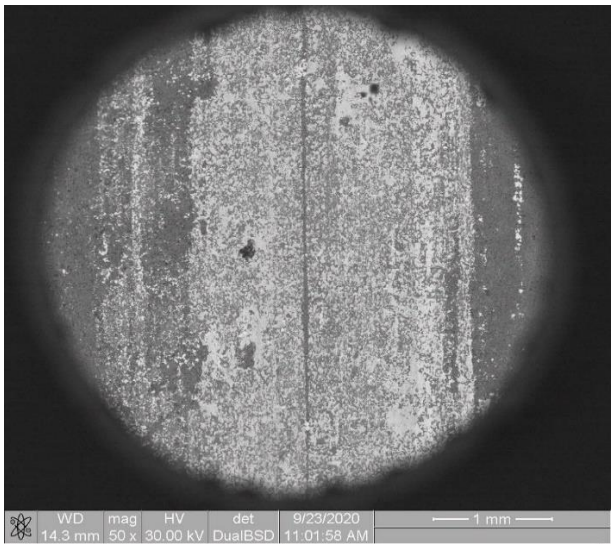


Figure 9: Electron microscopic images of the surface of the wear track for the aluminum coating sample *M4*

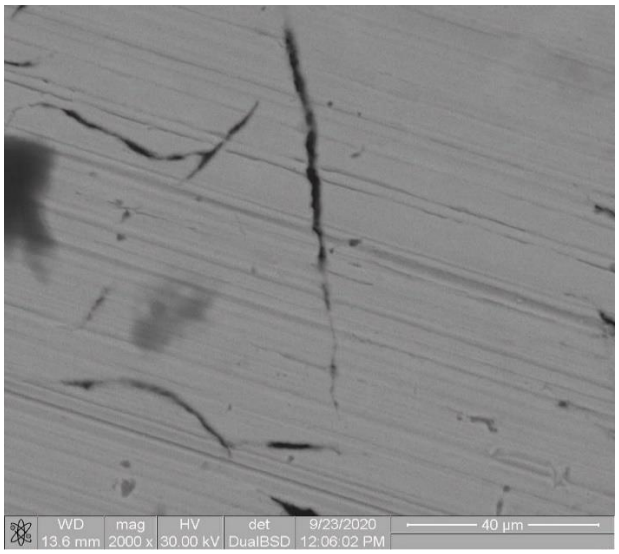


Figure 10: the wear track surface for the counterbody

It was found that the maximum profile valley depth (*Rv*) and the average roughness (*Ra*) for the uncoated sample are more than two times higher in comparison to other samples. Reducing roughness should have a positive effect on engine performance. The wear rate of the samples with oxide coatings is three orders of magnitude less than for uncoated one. Also, coefficient friction of coated samples was less than 0.04.

The *M5* sample has the best thermal insulation characteristics, which combines a low porosity about 10% depending on the concentration of Na_2SiO_3. In addition, the *M5*

sample compared to the other samples is more cost-effective in terms of the electrolyte composition. It is also noted that the wear loss of this sample is less than that of others, which is a consequence of the higher hardness.

Thus, it has been established that by the method of plasma electrolytic oxidation it is possible to obtain high-quality wear-resistant and heat-protective coatings on aluminum alloys with the high silicon content. By varying the technological modes and composition of the electrolyte, it is possible to change the structure, composition and thicknesses of coatings in a wide range from 80 to 300 (μm).

References

[1] M.S.H. Al-Bdeiri, "Review of electroplating-plasma modification methods for the production of anodized coatings on aluminum alloys: microstructure, properties and applications," Bulletin PNRPU. Mechanical engineering, materials science, **22**(3), 51-59, doi: 10.15593/2224-9877/2020.3.07

[2] G. Lv, W. Gu, H. Chen, W. Feng, M.L. Khosa, L. Li, E. Niu, G. Zhang, S.-Z. Yang, "Characteristic of ceramic coatings on aluminum by plasma electrolytic oxidation in silicate and phosphate electrolyte," Applied Surface Science, **253**(5), 2947–2952, 2006. doi:10.1016/j.apsusc.2006.06.036

[3] S. Sergeev, M.S. Albieri, V. Yatsenko, D. Natalya, "Theoretical and practical study of possibility to decrease thermal stress in pistons of internal combustion diesel engine by using galvanic plasma modification," International Journal of Advanced Science and Technology, **28**(8), 2019. doi:10.13140/RG.2.2.32284.44162

[4] Y. Zong, R.G. Song, T.S. Hua, S.W. Cai, C. Wang, H. Li, "Effects of current frequency on the MAO coatings on AA7050," Surface Engineering, **36**(8), 809–816, 2020. doi:10.1080/02670844.2019.1596577

[5] S. Xin, L. Song, R. Zhao, X. Hu, "Influence of cathodic current on composition, structure and properties of Al2O3 coatings on aluminum alloy prepared by micro-arc oxidation process," Thin Solid Films, **515**(1), 326–332, 2006. doi:10.1016/j.tsf.2005.12.087

[6] A. Polat, M. Makaraci, M. Usta, "Influence of sodium silicate concentration on structural and tribological properties of microarc oxidation coatings on 2017A aluminum alloy substrate," Journal of Alloys and Compounds, **504**(2), 519–526, 2010, doi:10.1016/j.jallcom.2010.06.008

[7] V. Dehnavi, B.L. Luan, X.Y. Liu, D.W. Shoesmith, S. Rohani, "Correlation between plasma electrolytic oxidation treatment stages and coating microstructure on aluminum under unipolar pulsed DC mode," Surface and Coatings Technology, **269**, 91–99, 2015. doi: 10.1016/j.surfcoat.2014.11.007

[8] Y.-C. Lin, Y.-H. Chen, "Study on Tribological Behavior of Surface Micro-arc Oxidation 6061 Aluminum Alloy," in 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), 883–887, 2019. doi:10.1109/IEA.2019.8714909

[9] A.L. Yerokhin, V. V Lyubimov, R. V Ashitkov, "Phase formation in ceramic coatings during plasma electrolytic oxidation of aluminium alloys," Ceramics International, **24**(1), 1–6, 1998, doi: 10.1016/S0272-8842(96)00067-3

[10] L. Agureev, S. Savushkina, A. Ashmarin, A. Borisov, A. Apelfeld, K. Anikin, N. Tkachenko, M. Gerasimov, A. Shcherbakov, V. Ignatenko, others, "Study of plasma electrolytic oxidation coatings on aluminum composites," Metals, **8**(6), 459, 2018, doi:10.3390/met8060459

[11] S. SV, M.S. Al-Bdieri, N.A. Dubrovina, "Surface modification of the AK12MMGH aluminum alloy by micro-oxidation technique to improve operating characteristics," Metallurgy, **27**(1), 217–223, 2020, doi: 10.34759/vst-2020-1-217-223.

[12] S. V Sergeyev, M.S. Al-Bdeiri, O.B. Beshevli, "Development of a technique for surface treatment of a rotary machine." International Journal of Engineering Trends and Technology, **68**(9), 31-35, 2020, doi: 10.14445/22315381/IJETT-V68I1P201

[13] J.A. Curran, T.W. Clyne, "Porosity in plasma electrolytic oxide coatings," Acta Materialia, **54**(7), 1985–1993, 2006. doi: 10.1016/j.surfcoat.2004.09.037

[14] P.L. Menezes, "Surface texturing to control friction and wear for energy efficiency and sustainability," The International Journal of Advanced Manufacturing Technology, **85**(5–8), 1385–1394, 2016. doi: 10.1007/s00170-015-8058-2

[15] W. Wang, C. Han, "Microstructure and Wear Resistance of Ti6Al4V Coating Fabricated by Electro-Spark Deposition," Metals, **9**(1), 23, 2019, doi:10.3390/met9010023

[16] S.K. Singh, S. Chattopadhyaya, A. Pramanik, S. Kumar, A.K. Basak, "Effect of lubrication on the wear behaviour of CrN coating deposited by PVD process," International Journal of Surface Science and Engineering, **13**(1), 60–78, 2019, doi: 10.1504/IJSURFSE.2019.097938

[17] B. Haghighat-Shishavan, R. Azari-Khosrowshahi, S. Haghighat-Shishavan, M. Nazarian-Samani, N. Parvini-Ahmadi, "Improving wear and corrosion properties of alumina coating on AA7075 aluminum by plasma electrolytic oxidation: Effects of graphite absorption," Applied Surface Science, **481**, 108–119, 2019, doi: 10.1016/j.apsusc.2019.03.069

[18] M.Sh. Al-Bdeiri, V.V. Krasilnikov, S.V. Sergeev, "A modified quasi-stationary method for studying the change in transition temperatures of diesel engine pistons coated with heat-shielding materials," **24**(5), 954–965, 2020, doi:10.21285/1814-3520-2020-5-954-965

[19] Q. Li, J. Liang, B. Liu, Z. Peng, Q. Wang, "Effects of cathodic voltages on structure and wear resistance of plasma electrolytic oxidation coatings formed on aluminium alloy," Applied Surface Science, **297**, 176–181, 2014, doi: org/10.1016/j.apsusc.2014.01.120

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1394-1401 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

A prediction of Cutting Force, System Vibration, and Productivity in Five-Axis Milling Process of the Spiral Bevel Gear

Nguyen Van Thien, Dung Hoang Tien, Nhu-Tung Nguyen*, Nguyen Van Que, Do Duc Trung, Pham Thi Thieu Thoa

Mechanical Engineering Faculty, Hanoi University of Industry, Hanoi, 100000, Vietnam

ARTICLE INFO

Article history:
Received: 03 November, 2020
Accepted: 08 December, 2020
Online: 16 December, 2020

Keywords:
Five Axial Milling
Spiral Bevel Gear
Taguchi Method
Cutting Forces
Vibrations
Productivity

ABSTRACT

In this study, Taguchi method and ANOVA that are applied to determine the influence of several input parametrs on cutting force amplitude (F), vibration acceleration amplitude (a), and machining productivity (P) in the five-axis finish milling process of the spiral bevel gear. Taguchi orthogonal L_{27} array is used to design the experimental matrix with 27 experiments. During finish milling process of the spiral bevel gear (X40 steel) with a ball-end mill cutter (HSLB-2030), the parameter that has most influence on F was spindle speed (37.51 %). The parameter that has the second influence degree on F was feed rate (22.97 %). The parameter that has the third influence degree on F was maximum scallop height (16.45 %). Toolpath strategy has negligible influence on F (less than 1 %). Feed rate was the most influence factor on a (22.48 %). The parameter that has the second influence degree on a was the maximum scallop height (21.73 %). The parameter that has the third influence degree on a was the toolpath strategy (3.65 %). Spindle speed has negligible influence on a (less than 1 %). The parameter that has the most influence on the P was the maximum scallop height (54.90 %). The parameter that has the second influence degree on P was the feed rate (18.92 %). The spindle speed and toolpath strategy have negligible influence on P (less than 1 %). Quadratic regression that is determined as the most suitable regression of F, a, and P in the finish milling process of the spiral bevel gear with the determination coefficients (R^2) 86.72 % for cutting force amplitude, 93.46 % for vibration acceleration amplitude, and 99.85 % for machining productivity.

1. Introduction

In machining processes, milling processes are ones of the most common machining methods. Especial, the application of CNC technology with high precision and high machinability, milling processes have been used popularly to machine the surfaces with high precision and high productivity. Therefore, more and more studies on the application of milling method have been performed. These studies were performed to improve the quality of machining quality, reduce cutting heat, reduce cutting forces, reduce vibrations, increasing the material removal rate, etc. in milling processes. In application of machines such as for thermal power generation system, marine propulsion elements, mines and quarries, wind's kinetic energy system, automation, etc. the spiral bevel gears are the most important components of power transmission system [1, 2]

In the manufacturing of spiral bevel gears, many different methods were applied to cut the gear teeth of spiral bevel gears. In which, most of the machining of spiral bevel gears were conducted by the specialized machines. Some of machining systems for manufacturing of the spiral bevel gears consist of the Klingelnberg's Palliod system [3], Klingelnberg's and Oerlikon's Cyclo Palliod systems [4], and the Gleason's gearing [5] Several studies were conducted to machining the spiral bevel gears using five-axis CNC milling machines [6-8]

In machining processes, cutting forces, vibrations, material removal rate (MRR) are the most important factors. Cutting forces, vibrations, and MRR influence on the machining quality and machining productivity. So, many studies were performed to predict the cutting forces, vibrations, and MRR, and to determine the influence of cutting parameters on the cutting forces, vibrations, MRR. In milling processes, cutting forces can be modeled by theoretical method for ball end mill [9] or flat-end mill [10] and by experimental method [11] The vibrations also can

*Corresponding Author: Nhu-Tung Nguyen, Faculty of Mechanical Engineering, Hanoi University of Industry, Vietnam, 84988480490 & tungnn@haui.edu.vn

be modeled by theoretical methods [12, 13] and experimental method [14] Normally, the MRR is calculated based on the relationship of cutting parameters such as cutting time, feed rate, cutting velocity, etc.

In the manufacturing of spiral bevel gears using five-axis milling processes, many research directions were performed to improve the quality and productivity of machining processes. Using universal multitasking machines or five-axis milling centers, a large sized spiral bevel gears were manufacture. The different machining strategies, toolpath patterns, and parameters were conducted for both roughing and finishing operations. The machining time, tool wear, and gear surface were analyzed to determine which are the best strategies and parameters for large modulus gear manufacturing on universal machines. Cutting forces, vibrations, and MRR have not been mentioned in this study [6].

A unity transformation model (UTM) was proposed to machine the spiral bevel gears and hypoid gears. This model can be applied in machining process of Gleason spiral bevel gears and hypoid gears, including generation machining and formation machining for wheel or pinion on a universal five-axis machining center. However, this study only focused on the generation of NC program. So, the machining quality and machining productivity have not been mentioned [15].

By combined with the general five-axis NC machine tool, the general expressions of trajectory for five-axis NC machining for the circle bevel gears were proposed. The results of cutter positions and states were verified by 3D modeling. And then, the final tool trajectory was applied in experimental processes. In the milling process of this kind of gear, the cutting forces, vibrations, and MRR have also not been mentioned in this study [16] Besides, several studies were performed using four or five-axis CNC machines to analysis, plan, and manufacture the bevel gears to reduce the cost and time and to improvement of the machining quality in machining processes [15-17].

Figure 1: Spiral bevel gear before using in experimental

2. Material and Experimental Method

2.1. Material

In this study, 40X steel was used to perform the experiment. The equivalent signs of this steel are described in Table 1 according several standard. The chemical compositions of 40X steel are listed in Table 2. The properties of the 40X were the listed in Table 3. The geometry of workpiece (a gear was rough machined) is described in Figure 1.

40X steel is a structural alloy steel intended for use in heavy-duty components. Products made from this steel conform to DSTU 7806 and GOST 4543 standards. 40X structural alloy steel is used in heavy-duty components (shafts, axles, gear shafts, plungers, bars, crank and cam shafts, rings, spindles, drift pins, battens, cogs, bolts, half-shafts, bushings, screws etc.). 40X chromic steel is also used in forged, cold- and hot-stamped components, and the steel product is used in pipe, tank and other production.

Table 1: Equivalent symbol of 40X steel of according several Standard

1	US	5135, 5140, 5140H, 5140RH, G51350, G51400, H51350, H51400
2	Japan	SCr435, SCr435H, SCr440, SCr440H
3	EU	37Cr4, 37Cr4KD, 41Cr4, 41Cr4KD, 41CrS4
4	China	35Cr, 38CrA, 40Cr, 40CrA, 40CrH, 45Cr, 45CrH, ML38CrA, ML40Cr
5	Sweden	2245
6	Poland	38HA, 40H
7	Romania	40Cr10, 40Cr10q
8	Czech Republic	14140
9	Australia	5132H, 5140
10	South Korea	SCr435, SCr435H, SCr440, SCr440H

Table 2: Chemical composition of 40X steel

Element	C	Si	Mn	NI	Cr	Cu	P	S	Fe
%	0.36-0.44	0.17-0.37	0.5-0.8	≤0.3	0.8-1.1	≤0.3	≤0.035	≤0.035	Balance

Table 3: Properties of 40X steel

Youngs module (GPa)	Density (kg/m^3)	Yield stress (GPa)	Brinell hardness (MPa)
214	7820	78.5	179
Specific heat capacity 50/100°C (J/kg°K)	Average CTE 20-300°C (µm/m°K)	Thermal conductivity Ambient temperature (W/m°K)	Electrical resistivity Ambient temperature (µΩm)
466	11.9	46	0.210 - 0.285

2.2. Experimental Machine, Tool, and Measurement systems

A five-axis vertical machining center (DMU 50 - 5 Axis Milling) was used to perform the experiments as shown in Figure 1. A ball-end mill cutter (HSLB-2030) was used to finish milling the Spiral Bevel Gear with number of flutes of 2, helix angle of 30^0, rake angle of 5^0, and the diameter of 6 mm. In this study, the cutting forces were measured using a cutting force measurement system including a dynamometer (Kistler type 9139AA: force ranges: (-3 KN÷3 KN), a data processing box, and a PC with DynoWare software. Vibration measurement system including the acceleration sensor (type 4525-B-001), the data processing box, and the PLUSE software was used to measure the system vibration acceleration. The detail of measurement system was illustrated in Figure 2. The machining time of each gear teeth was determines based on the time machining from the control system of five-axis CNC machine. And then, the productivity was calculated by (1).

a. Workpiece b. Tool c. Dynamometer d. Acceleration sensore. Cutting force signal processing box
f. PC and DynoWare software g. Acceleration signal processing box h. PLUSE software
Figure 2: Experimental machine and measurement system

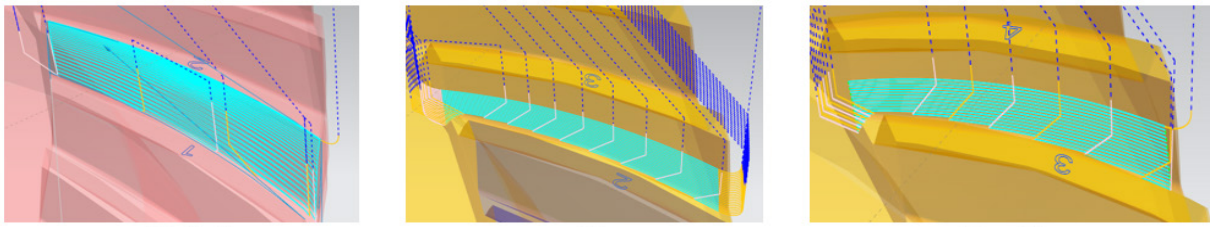


Figure 3: The toolpath strategy in five-axis milling process

$$P = \frac{1}{T} \text{ (tooth/min)} \quad (1)$$

where T is the finish machining time of one gear teeth.

2.3. Experimental design

In this research, four factors that were used as the input parameters were spindle speed, feed rate, tool path strategy, and maximum scallop height, and their levels were expressed in Table 4. The toolpath strategies in five-axis milling process were described in Figure 3, and the scallop height in five-axis milling process was calculated by (2), (3), and presented in Figure 4 [18] In the experimental layout plan, with four factors and three levels, the most suitable Taguchi orthogonal array (L_{27}) was selected to design the experimental matrix as listed in Table 5.

$$h = R(1 - cos\beta) \quad (2)$$

$$\beta = sin^{-1}(\frac{S}{2R}) \quad (3)$$

where R is the tool radius.
S is the step over (radial depth of cut).

Table 4: The factors and their levels

Parameter	Unit	Coded sign	Levels		
			1	2	3
Spindle speed (S)	rpm	A	5000	8000	11000
Feed rate (F)	mm/min	B	400	520	640

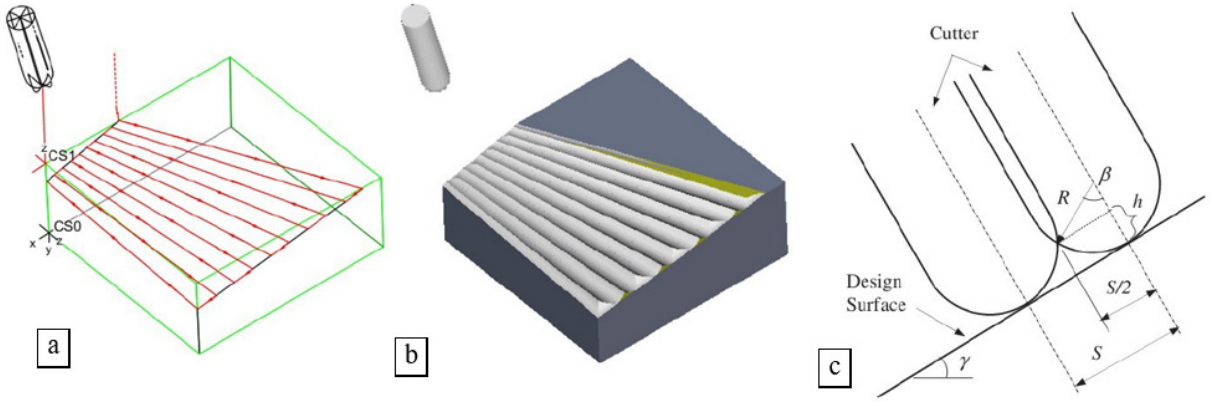
Strategy		C	Helical	Zig	Zig - Zig
Maximum Scallop Height (µm)	µm	D	1	3	5

3. Experimental Results and Discussion

3.1. Evaluation of the Cutting Force

ANOVA for Cutting Force Amplitude

The experimental results were listed in Table 5. The measured results of cutting force were listed in column 10 of this Table. Analysis results of variance of cutting force were listed in Table 6 and the main effect of cutting parameters, toolpath strategy, and the maximum scallop height on F was described in Figure 5. The ANOVA results that were used to evaluate the effect of input parameters on the output parameters with 95% confidence level and 5% significance level. The analysis results in Table 6 and Fig 3 showed that: During five-axis finish milling process of the spiral bevel gear, the parameter that has most influence on F was the spindle speed (37.51 %). The parameter that has the second influence degree on F was the feed rate (22.97 %). The parameter that has the third influence degree on F was the maximum scallop height (16.45 %). The toolpath strategy has negligible influence on the F (less than 1 %).
In addition, the influence of each input parameter on F was shown in Figure 5, the results from this figure showed that the spindle speed, feed rate, and maximum scallop height were the parameters that have significant effect on F And the influence degree of toolpath strategy on the F was quite small.



a. Toolpath b. Machining simulation c. Calculation of scallop height

Figure 4: The calculation of scallop height in five-axis milling process

Table 5: Taguchi L_{27} experimental matrix and results

No.	A	B	C	D	S (rpm)	f (mm/min)	Strategy	Scallop Height (µm)	F (N)	a (m/s^2)	Productivity (tooth/min)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	1	1	1	1	5000	400	Helical	1	63.388	7.081	0.235
2	1	1	2	2	5000	400	Zig	3	81.056	9.119	0.291
3	1	1	3	3	5000	400	Zig-Zag	5	167.432	10.495	0.566
4	1	2	1	2	5000	520	Helical	3	202.366	10.566	0.531
5	1	2	2	3	5000	520	Zig	5	171.882	9.072	0.472
6	1	2	3	1	5000	520	Zig-Zag	1	129.055	9.433	0.323
7	1	3	1	3	5000	640	Helical	5	219.478	10.513	0.667
8	1	3	2	1	5000	640	Zig	1	81.921	8.906	0.243
9	1	3	3	2	5000	640	Zig-Zag	3	134.358	11.486	0.690
10	2	1	1	1	8000	400	Helical	1	88.621	9.637	0.240
11	2	1	2	2	8000	400	Zig	3	115.231	10.735	0.290
12	2	1	3	3	8000	400	Zig-Zag	5	196.270	12.710	0.566
13	2	2	1	2	8000	520	Helical	3	191.313	10.935	0.455
14	2	2	2	3	8000	520	Zig	5	177.027	12.154	0.472
15	2	2	3	1	8000	520	Zig-Zag	1	232.821	11.496	0.280
16	2	3	1	3	8000	640	Helical	5	264.235	13.993	0.789
17	2	3	2	1	8000	640	Zig	1	211.871	10.841	0.243
18	2	3	3	2	8000	640	Zig-Zag	3	285.355	14.711	0.690
19	3	1	1	1	11000	400	Helical	1	163.627	7.188	0.240
20	3	1	2	2	11000	400	Zig	3	214.324	7.798	0.291
21	3	1	3	3	11000	400	Zig-Zag	5	243.956	10.046	0.566
22	3	2	1	2	11000	520	Helical	3	260.916	10.738	0.536
23	3	2	2	3	11000	520	Zig	5	237.999	10.120	0.472
24	3	2	3	1	11000	520	Zig-Zag	1	190.603	9.656	0.323
25	3	3	1	3	11000	640	Helical	5	320.031	12.794	0.789
26	3	3	2	1	11000	640	Zig	1	236.119	9.300	0.242
27	3	3	3	2	11000	640	Zig-Zag	3	289.907	10.939	0.690

Table 6: ANOVA results for cutting force

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	13	105562	86.72%	105562	8120.2	6.53	0.001
Linear	4	94142	77.34%	80109	20027.2	16.10	0.000
A	1	45657	37.51%	45657	45657.0	36.71	0.000
B	1	27956	22.97%	27956	27955.9	22.48	0.000
C	1	510	0.42%	138	137.9	0.11	0.744
D	1	20019	16.45%	5516	5515.9	4.44	0.055
Square	4	7934	6.52%	8405	2101.2	1.69	0.212
A*A	1	254	0.21%	254	253.8	0.20	0.659
B*B	1	823	0.68%	823	822.8	0.66	0.431
C*C	1	6422	5.28%	5239	5239.5	4.21	0.061
D*D	1	435	0.36%	1317	1317.5	1.06	0.322
2-Way Interaction	5	3487	2.86%	3487	697.4	0.56	0.728
A*B	1	838	0.69%	838	837.8	0.67	0.427

A*C	1	98	0.08%	98	97.9	0.08	0.783
A*D	1	442	0.36%	442	441.5	0.36	0.561
B*C	1	2074	1.70%	1329	1329.1	1.07	0.320
B*D	1	36	0.03%	36	35.6	0.03	0.868
Error	13	16166	13.28%	16166	1243.6		
Total	26	121729	100.00%				

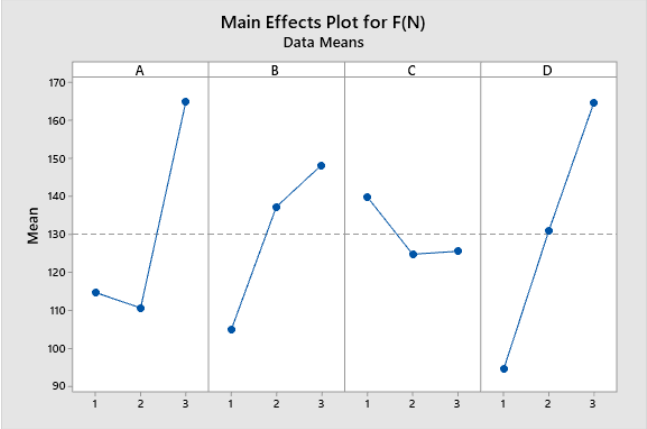


Figure 5: Main Effects Plot for Cutting Force

3.2. Regression of Cutting Force Amplitude Model

The measured data were used to model the cutting force amplitude. The regression model of cutting force amplitude is described by (4). The analysis results showed that the most suitable regression model for cutting force amplitude was a quadratic polynomial function of the spindle speed, feed rate, tool path strategy, and maximum scallop height. The regression model of cutting force amplitude was analyzed and built with high determination coefficients R^2 (86.72 %).

Cutting Force Regression Equation in coded variables

$$F = -109 + 66.1\,A + 109.5\,B - 103.9\,C + 111.0\,D - 6.5\,A*A - 11.7\,B*B + 34.1\,C*C - 17.1\,D*D + 8.4\,A*B + 2.9\,A*C - 6.1\,A*D - 17.2\,B*C - 2.8\,B*D \quad (4)$$

$$R^2 = 86.72\ \%, \quad R^2_{Ajd} = 73.44\ \%$$

3.3. Verification of Cutting Force Amplitude Model

The cutting force amplitude model was successfully verified by comparison of predicted and measured results of cutting force amplitude as described in Figure 6. The compared results from this figure showed that the predicted cutting force amplitude was quite close to the measured cutting force amplitude. It can be concluded that the quadratic regression model can be used to model the cutting force amplitude in in the finish milling process of the spiral bevel gear. The average difference between predicted and measured of cutting force amplitude is about 11.863 %.

3.4. Evaluation of the Vibration Acceleration

3.4.1. ANOVA for Vibration Acceleration Amplitude

The measured results of cutting force were listed in column 11 of Table 5. Analysis results of variance of vibration acceleration were listed in Table 7 and the main effect of cutting parameters, toolpath strategy, and the maximum scallop height on the vibration acceleration was described in Figure 7. In five-axis finish milling process of the spiral bevel gear, the parameter that has most influence on the vibration acceleration was the feed rate (22.48 %). The parameter that has the second influence degree on the vibration acceleration was the maximum scallop height (21.73 %). The parameter that has the third influence degree on the vibration acceleration was the toolpath strategy (3.65 %). The spindle speed has negligible influence on the vibration acceleration (less than 1 %).

In addition, the influence of each input parameter on vibration acceleration was shown in Figure 7, the results from this figure showed that the feed rate and maximum scallop height were the parameters that have significant effect on the vibration acceleration. And the influences degree of spindle speed and toolpath strategy on the vibration acceleration were quite small.

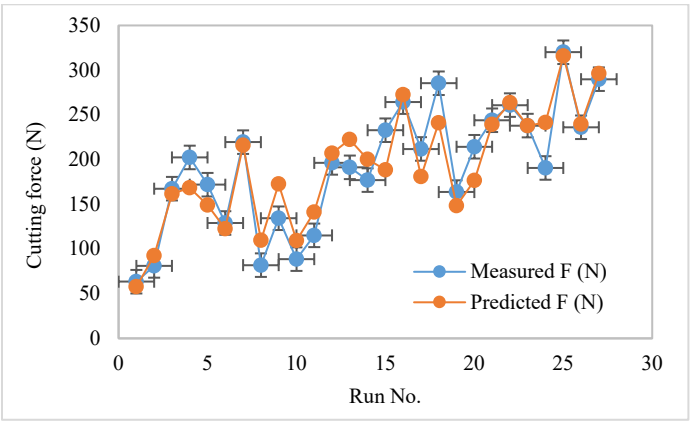


Figure 6: Comparison of predicted and measured cutting force

Table 7: ANOVA results for vibration acceleration

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	13	80.5415	93.46%	80.5415	6.1955	14.29	0.000
Linear	4	41.4468	48.09%	30.5043	7.6261	17.59	0.000
A	1	0.2023	0.23%	0.2023	0.2023	0.47	0.507
B	1	19.3716	22.48%	19.3716	19.3716	44.67	0.000
C	1	3.1463	3.65%	2.0195	2.0195	4.66	0.050
D	1	18.7266	21.73%	6.4265	6.4265	14.82	0.002
Square	4	36.0108	41.79%	35.3239	8.8310	20.36	0.000
A*A	1	28.4142	32.97%	28.4142	28.4142	65.52	0.000
B*B	1	0.0000	0.00%	0.0000	0.0000	0.00	0.992
C*C	1	6.2205	7.22%	4.1697	4.1697	9.62	0.008
D*D	1	1.3761	1.60%	1.9240	1.9240	4.44	0.055
2-Way Interaction	5	3.0839	3.58%	3.0839	0.6168	1.42	0.280
A*B	1	1.1963	1.39%	1.1963	1.1963	2.76	0.121
A*C	1	0.9248	1.07%	0.9248	0.9248	2.13	0.168
A*D	1	0.3875	0.45%	0.3875	0.3875	0.89	0.362
B*C	1	0.5197	0.60%	0.5510	0.5510	1.27	0.280
B*D	1	0.0557	0.06%	0.0557	0.0557	0.13	0.726
Error	13	5.6373	6.54%	5.6373	0.4336		
Total	26	86.1788	100.00%				

The measured data of vibration acceleration were used to model the vibration acceleration amplitude. The regression model of vibration acceleration amplitude is described by (5). The analysis results showed that the most suitable regression model

for vibration acceleration amplitude was also a quadratic polynomial function of the spindle speed, feed rate, tool path strategy, and maximum scallop height. The regression model of vibration acceleration amplitude was analyzed and built with high determination coefficients R^2 (93.46 %).

Figure 7: Main Effects Plot for vibration acceleration (a)

Vibration Acceleration Regression Equation in coded variables

$$\begin{aligned} a = {} & -0.78 + 8.38\,A + 0.89\,B - 2.12\,C + 2.88\,D \\ & - 2.176\,A*A - 0.003\,B*B + 0.963\,C \\ & *C - 0.654\,D*D + 0.316\,A*B \\ & - 0.278\,A*C + 0.180\,A*D - 0.350\,B \\ & *C + 0.111\,B*D \end{aligned} \quad (5)$$

$$R^2 = 93.46\,\%, \quad R^2_{Ajd} = 86.92\,\%$$

3.6. Verification of Vibration Acceleration Amplitude Model

The vibration acceleration amplitude model was also successfully verified by comparison of predicted and measured results of vibration acceleration amplitude as described in Figure 8. The compared results from this figure showed that the predicted vibration acceleration amplitude was quite close to the that one of measured value. It can be also concluded that the quadratic regression model can be used to model the vibration acceleration amplitude in in the finish milling process of the spiral bevel gear. The average difference between predicted and measured of vibration acceleration amplitude is about 3.371 %.

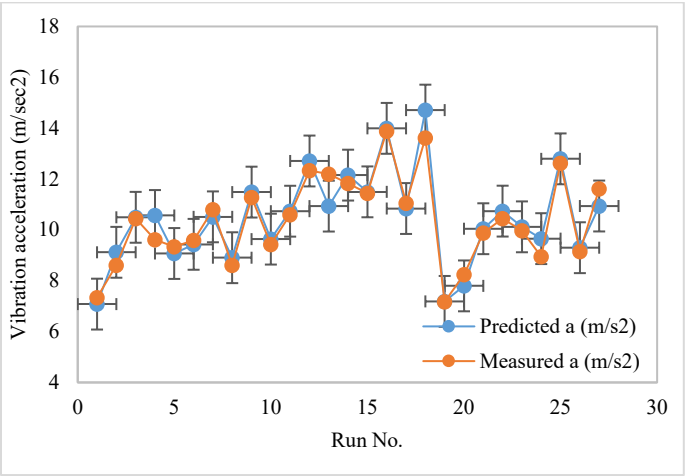


Figure 8: Comparison of predicted and measured vibration acceleration

3.7. Evaluation of the Machining Productivity

3.7.1. ANOVA for Machining Productivity

The measured results of cutting force were listed in column 12 of Table 5. Analysis results of variance of machining productivity were listed in Table 8 and the main effect of cutting parameters, toolpath strategy, and the maximum scallop height on the machining productivity was described in Figure 9. In five-axis finish milling process of the spiral bevel gear, the parameter that has the most influence on the machining productivity was the maximum scallop height (54.90 %). The parameter that has the second influence degree on the machining productivity was the feed rate (18.92 %). The spindle speed and toolpath strategy have negligible influence on machining productivity (less than 1 %). The influence of each input parameter on the machining productivity was shown in Figure 9, the results from this figure showed that the feed rate and maximum scallop height were the parameters that have significant effect on the machining productivity. And the influence degree of spindle speed and toolpath strategy on the machining productivity were quite small.

Table 8: ANOVA results for productivity

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	13	0.895721	98.85%	0.895721	0.068902	85.97	0.000
Linear	4	0.672323	74.20%	0.483583	0.120896	150.84	0.000
A	1	0.000957	0.11%	0.000957	0.000957	1.19	0.294
B	1	0.171417	18.92%	0.171417	0.171417	213.88	0.000
C	1	0.002461	0.27%	0.002616	0.002616	3.26	0.094
D	1	0.497488	54.90%	0.275714	0.275714	344.01	0.000
Square	4	0.215948	23.83%	0.163925	0.040981	51.13	0.000
A*A	1	0.000251	0.03%	0.000251	0.000251	0.31	0.586
B*B	1	0.006666	0.74%	0.006666	0.006666	8.32	0.013
C*C	1	0.182564	20.15%	0.128643	0.128643	160.51	0.000
D*D	1	0.026468	2.92%	0.014949	0.014949	18.65	0.001
2-Way Interaction	5	0.007450	0.82%	0.007450	0.001490	1.86	0.170
A*B	1	0.001143	0.13%	0.001143	0.001143	1.43	0.254
A*C	1	0.001458	0.16%	0.001458	0.001458	1.82	0.200
A*D	1	0.001182	0.13%	0.001182	0.001182	1.47	0.246
B*C	1	0.003151	0.35%	0.001388	0.001388	1.73	0.211
B*D	1	0.000516	0.06%	0.000516	0.000516	0.64	0.437
Error	13	0.010419	1.15%	0.010419	0.000801		
Total	26	0.906140	100.00%				

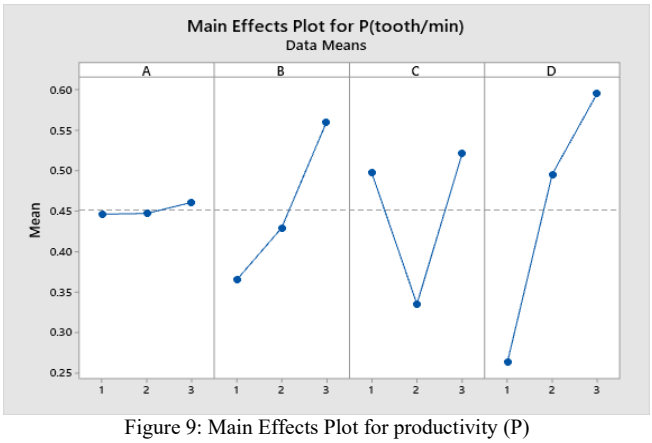


Figure 9: Main Effects Plot for productivity (P)

3.8. Regression of Machining Productivity Model

The measured data of machining productivity were used to model the machining productivity. The regression model of machining productivity is described by (6). The analysis results showed that the most suitable regression model for the machining productivity was also a quadratic polynomial function of the spindle speed, feed rate, tool path strategy, and maximum scallop height. The regression model of machining productivity was analyzed and built with very high determination coefficients R^2 (99.85 %).

Productivity Regression Equation in coded variables

$$\begin{aligned} P = {} & 0.510 - 0.0359\,A - 0.1118\,B - 0.6723\,C + 0.3643\,D \\ & + 0.0065\,A*A + 0.0333\,B*B \\ & + 0.1691\,C*C - 0.0576\,D*D \\ & + 0.00976\,A*B - 0.01102\,A*C \\ & + 0.00992\,A*D + 0.0176\,B*C \\ & + 0.0107\,B*D \end{aligned} \quad (6)$$

$$R^2 = 99.85\ \%, \quad R^2_{Ajd} = 97.70\ \%$$

3.9. Verification of Machining Productivity Model

The machining productivity model was also successfully verified by comparison of predicted and measured results of machining productivity as described in Figure 10. The compared results from this figure showed that the predicted machining productivity was quite close to the that one of measured value. It can be also concluded that the quadratic regression model can be used to model the machining productivity in in the finish milling process of the spiral bevel gear. The average difference between predicted and measured of machining productivity is about 2.867 %.

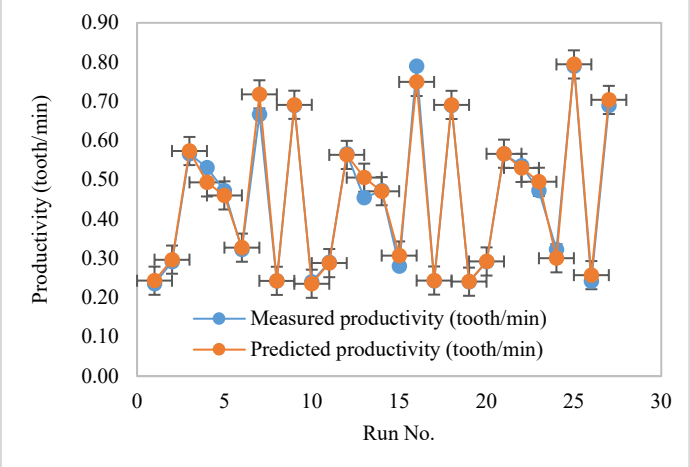


Figure 10: Comparison of predicted and measured productivity

4. Conclusion

In this study, the Taguchi method and ANOVA that were successfully applied to determine the influence of the spindle speed, feed rate, tool path strategy, and maximum scallop height on cutting forces, vibrations, productivity in the finish milling process of the spiral bevel gear. From the achieved results, the conclusions of this study were drawn as follows:

During the finish milling process of the spiral bevel gear with a ball-end mill cutter (HSLB-2030), the parameter that has most influence on the cutting force amplitude was the spindle speed (37.51 %). The parameter that has the second influence degree on the cutting force amplitude was the feed rate (22.97 %). The parameter that has the third influence degree on the cutting force was the maximum scallop height (16.45 %). The toolpath strategy has negligible influence on the cutting force (less than 1 %)

The feed rate was the most influence factor on the vibration acceleration (22.48 %). The parameter that has the second influence degree on the vibration acceleration was the maximum scallop height (21.73 %). The parameter that has the third influence degree on the vibration acceleration was the toolpath strategy (3.65 %). The spindle speed has negligible influence on the vibration acceleration (less than 1 %).

The parameter that has the most influence on the machining productivity was the maximum scallop height (54.90 %). The parameter that has the second influence degree on the machining productivity was the feed rate (18.92 %). The spindle speed and toolpath strategy have negligible influence on machining productivity (less than 1 %).

Quadratic regression that was determined as the most suitable regression of cutting force amplitude and vibration acceleration amplitude, and productivity in the finish milling process of the spiral bevel gear with the determination coefficients (R^2) 86.72 % for cutting force amplitude, 93.46 % for vibration acceleration amplitude, and 99.85 % for machining productivity. These regression models were successfully verified by comparison of predicted and experimental results of cutting force amplitude and vibration acceleration amplitude, and productivity.

Taguchi method, ANOVA, and these regression models in this study can be applied to improve the quality of the machining process by reducing the cutting force components, vibration components, and improve the productivity of machining process. There are the research directions in the next issues of this study.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors thank for the support of Hanoi University of Industry (HaUI) during the implementation of this study.

References

[1] G.Xu, D. Hua, W. Dai, and X. Zhang, "Design and performance analysis of a coal bed gas drainage machine based on incomplete non-circular gears," Energies, **10**(12), 1933, 2017, doi: https://doi.org/10.3390/en10121933.
[2] Z. Qin, Y. Luo, K. Li, H. Peng, "Optimal design of a novel hybrid Electric powertrain for tracked vehicles," Energies, **10**(12), 2141, 2017, doi:https://doi.org/10.3390/en10122141.
[3] H. J. Watson, Modern gear production, Elsevier, 2013.
[4] G. Lechner, Klingelnberg spiral bevel gears, W. Krumme Springer, Berlin, 1968.
[5] H. J. Stadtfeld, Gleason Bevel Gear Technology: The Science of Gear Engineering and Modern Manufacturing Methods for Angular Transmissions, Gleason Works, 2014.
[6] Á. Álvarez, A. Calleja, N. Ortega, L. N. De Lacalle, "Five-axis milling of large spiral bevel gears: toolpath definition, finishing, and shape errors," Metals, **8**(5), 353, 2018, doi:https://doi.org/10.3390/met8050353.
[7] T. Xiang , J. Yi, W. Li, "Five-Axis Numerical Control Machining of the Tooth Flank of a Logarithmic Spiral Bevel Gear Pinion," Transactions of

FAMENA, **42**(1), 73-84, 2018, doi: https://doi.org/10.21278/TOF.42107.
[8] I. Tsiafis, P. Mamouri, S. Kompogiannis, "Design and manufacturing of spiral bevel gears using CNC milling machines," InIOP Conference Series: Materials Science and Engineering, **393**(1), 0120662018, 2018.
[9] Y. C. Kao, N. T. Nguyen, M. S. Chen, S. C. Huang, "A combination method of the theory and experiment in determination of cutting force coefficients in ball-end mill processes," Journal of Computational Design and Engineering, **2**(4), 233-47, 2015, doi:https://doi.org/10.1016/j.jcde.2015.06.005.
[10] Y. C. Kao, N. T. Nguyen, M. S. Chen, S. T. Su, "A prediction method of cutting force coefficients with helix angle of flat-end cutter and its application in a virtual three-axis milling simulation system," The International Journal of Advanced Manufacturing Technology, **77**(9-12), 1793-809, 2015, doi: https://doi.org/10.1007/s00170-014-6550-8.
[11] T. D. Hoang, N. T. Nguyen, D. Q. Tran, V. T. Nguyen, "Cutting Forces and Surface Roughness in Face-Milling of SKD61 Hard Steel," Strojniski Vestnik/Journal of Mechanical Engineering, **65**(6), 375-385, 2019, doi: https://doi.org/10.5545/sv-jme.2019.6057.
[12] N. T. Nguyen, Y. C. Kao, H. T. Dung, "A Prediction Method of Dynamic Cutting Forces and Machine-Tool Vibrations When Milling by Using Ball-End Mill Cutter," In International Conference on Engineering Research and Applications, 47-54. Springer, Cham, 2019, doi:https://doi.org/10.1007/978-3-030-37497-6_5.
[13] K. Weinert, P. Kersting, T. Surmann, D. Biermann, "Modeling regenerative workpiece vibrations in five-axis milling," Production Engineering, **2**(3), 255, 2008, doi:https://doi.org/10.1007/s11740-008-0113-5.
[14] P. S. Sivasakthivel, V. Velmurugan, R. Sudhakaran, "Prediction of vibration amplitude from machining parameters by response surface methodology in end milling," The International Journal of Advanced Manufacturing Technology, **53**(5-8), 453-61, 2011, doi:https://doi.org/10.1007/s00170-010-2872-3.
[15] T. Wang, Y. Xing, L. Zhao, Q. Li, "NC machining of spiral bevel gear and hypoid gear based on unity transformation model," Transactions of Tianjin University, **17**(4), 264-9, 2011, doi:https://doi.org/10.1007/s12209-011-1641-1.
[16] B. Wang, M. Fan, X. Sun, J. Li, A. Xu, G. Li, J. Yang, "Cutter position calculation of machining equal base circle bevel gears with a pot-shaped milling cutter," The International Journal of Advanced Manufacturing Technology, **87**(9-12), 2625-37, 2016, doi:https://doi.org/10.1007/s00170-016-8657-6.
[17] J. Tang, X. Yang, "Research on manufacturing method of planing for spur face-gear with 4-axis CNC planer," The International Journal of Advanced Manufacturing Technology, **82**(5-8), 847-858, 2016, doi:https://doi.org/10.1007/s00170-015-7417-3.
[18] K. H, Product manufacturing and cost estimating using CAD/CAE: the computer aided engineering design series, Academic Press, 2013.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1402-1407 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

A Study on the Tool Wear in Milling Process of the Gleason Spiral Bevel Gear

Hoang Xuan Thinh[*,1], Pham Van Dong[2], Tran Ve Quoc[1]

[1]*Mechanical Engineering Center, Hanoi University of Industry, Hanoi City, 100000, Vietnam*

[2]*Department of Science and Technology, Hanoi University of Industry, Hanoi City, 100000, Vietnam*

ARTICLE INFO

Article history:
Received: 08 September, 2020
Accepted: 02 December, 2020
Online: 16 December, 2020

Keywords:
Gleason spiral bevel gear
Tool wear
20XM steel
Hard alloy coated with CVD Ti (C, N)-Al_2O_3-TiN
Optimization
Cutting parameters

ABSTRACT

In this paper, the investigated results on tool wear in milling process of a Gleason spiral bevel gear was investigated. The CVD Ti (C, N)-Al_2O_3-TiN coated carbide inserts were used in the experimental process. The cutting inserts clamped on the tool head of Gleason tool head system. The workpiece material that was used in this study was 20XM steel. In this experimental research, the experiments were carried out according to the Box-Behnken matrix. In this study, the cutting velocity, the feed rate, and the depth of cut were selected as the input parameters for the experimental processes. By analyzing the experimental results, the influence of cutting velocity, feed rate, and the depth of cut on the tool wear in machining process was investigated. Besides, the tool wear was successfully modelled as a quadratic polynomial function of the cutting velocity, the feed rate, and the depth of cut. Moreover, the optimization process was also conducted to determine the optimal values of input parameters. In this case, the minimum value of the tool wear that was 32.3679 µm was achieved at the cutting velocity of 93.0 m/min, at feed rate of 59.3939 s/tooth, and at the cutting depth of 0.553 mm. The obtained optimal values that were successfully verified by cutting test process. The difference between predicted and experimental values of tool wear was 2.9 %. This experimental method could be used to reduce the tool wear in the milling process of a Gleason spiral bevel gear.

1. Introduction

In recent years, the effect of cutting parameters on the tool wear and other machining characteristics was investigated to improve the quality and efficiency of machining processes many studies were performed to investigate. This research direction was performed in the different machining processes such as milling [1-3], turning [4-6], drilling [7], etc. to machine the different products. Milling is a popular machining method, providing high performance and widely used in mechanical machining. The milling method can be performed to machine many different types of surfaces, with many different materials.

In milling processes, the studies about tool wear and other machining characteristics that were carried out following two directions. In the first direction, the tool wear and other machining characteristics were modeled depending on the physical, chemical, and geometrical phenomena such as friction, temperature, cutting, Etc. [8, 9]. This approach is quite difficult to perform because many factors that influence on the wear of tool and other output parameters. In the second direction, the tool wear and surface roughness were modeled depending on the experimental data. This approach can be applied for specific cases in which only several factors were considered in the investigation of their influence on the tool wear and other machining characteristics [10-12].

In the experimental modelling method, several approaches were applied to model the tool wear and other machining characteristics depending on the cutting time or cutting conditions. The tool wear and other machining characteristics were investigated in milling process depending on the cutting time. This study was carried out to verify the change of machining surface roughness due to increasing tool wear. [13, 14]. The neural approach was applied to investigate the influence of cutting condition of tool wear and average surface roughness in turning process under minimum quantity lubrication (MQL) environment. [15]. The influence of the cutting parameters on the tool wear and the surface finish criteria have been determined through the

[*]Corresponding Author: Hoang Xuan Thinh, Center for Mechanical Engineering, Hanoi University of Industry, Hanoi city, Vietnam, +84988304223 & hoangxuanthinh@haui.edu.vn

response surface methodology (RSM) prediction model. The prediction models of tool wear and surface roughness are also used to determine the combined effect of machining parameters on the tool wear and surface roughness [12, 15]. Taguchi experimental design was used to determine the influence of machining conditions on the tool wear and surface roughness. Besides, the investigation of relationship between tool wear and surface roughness was investigated in in dry pocket milling processes of aluminum alloy Al7075 [16].

In the milling process of a Gleason spiral bevel gear, the tool wear that is one of the important parameters to be selected as a criterion to evaluate the efficiency and quality of the cutting process. The tool wear not only influences on the quality of machined surfaces and the tool life, but also influences on the energy consumption (through cutting power). The studies on tool wear in the milling processes of the Gleason spiral bevel gear have been done by several authors. In order to determine the optimum values of the geometry parameters of cutting tools, the 3D simulation method was applied to ensure the minimum of tool wear value [17, 18]. The influence of the cutting velocity and feed rate on the tool wear was investigated when machining 6MnCr5 material using cutter insert coated (Al, Cr) N [19]. This study showed that cutting velocity and feed rate have significant influence on the tool wear. Tool wear was also was investigated based on the heat simulation in cutting processes [20].

In this study, the experimental of milling a Gleason spiral bevel gear (20XM steel) was performed using the CVD Ti(C, N) -Al2O3-TiN coated carbide inserts. This study was conducted to determine the effect of the cutting velocity, the feed rate, and cutting depth on the tool wear, and determine the optimum values of the cutting parameters to obtain the minimum of the tool wear.

2. Experimental method when milling a Gleason spiral bevel gear

2.1. Experimental machine

The semi-automatic gear milling machine with symbol 525 manufactured by the Russian Federation (figure 1) was used to perform the experiments.

Cool lubricant: Industrial oil 32, flow of 15 lit/min, lubricating directly.

Figure 1: Testing machine

2.2. Cutting insert and cutter head

The cutting inserts that were used in this study were the CVD Ti (C, N) - Al_2O_3 - TiN coated carbide inserts (SANDVIK). This type of cutting insert has many advantages in industry machining such as: heat resistance stability, mechanical stability, impact resistance, and suitable for final cutting conditions, etc. This material is widely used to manufacture cutting tools for machining process and for machining process the of the gears. The properties of this material in comparison to other cutter materials were presented in Table 1. The geometry of the cutting insert is shown in Figure 2.

Table 1: The properties of several cutter material [21]

No.	Cutter material	Cutting velocity V_c (m/ph)	Cutting limit temperature (°C)	Hardness (HRC)
1	High Speed Steel	20 ÷ 30	500 ÷ 650	60 ÷ 64
2	WC coated carbide	200	1000 ÷ 1200	91
3	WC + TiC coated carbide	300	1000 ÷ 1200	91 ÷ 92
4	TiC coated carbide	300	1000	92 ÷ 94
5	CVD Ti (C, N) -Al2O3-TiN) coated	370 ÷ 480	1000 ÷ 1100	90 ÷ 92

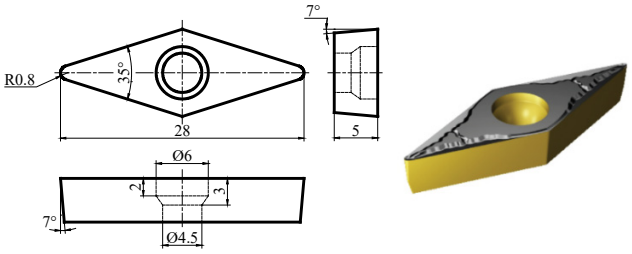


Figure 2: CVD Ti(C, N)-Al_2O_3-TiN coated carbide insert

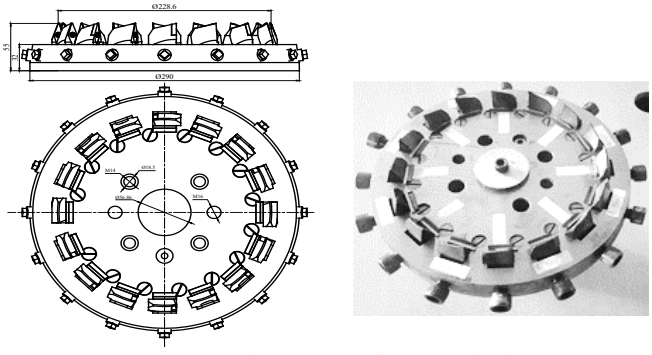


Figure 3: Gleason fine milling cutter head

The tool head for machining gears that was used in this study was a 9-inch Gleason final milling cutter head. This cutter head consists of 16 cutting flutes with 8 external flutes and 8 internal flutes. The tool can be used to machine the gears with module

from the module 4 to module 9 as shown in Fig. 3. The geometrical parameters of a flute were shown in Fig. 4 and Fig. 5.

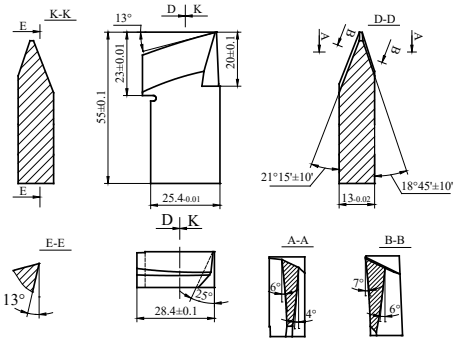


Figure 4: Geometrical parameters of the internal cutting flute of the Gleason cutter head

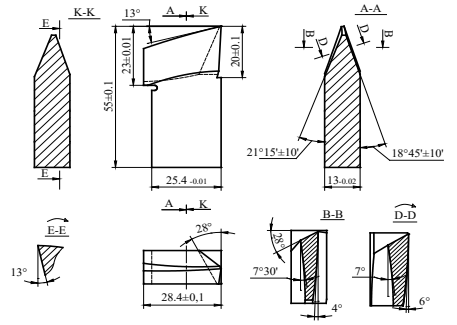


Figure 5: Geometrical parameters of the external cutting flute of the Gleason cutter head

2.3. Experimental workpieces

The workpiece material that was used in this study was 20XM steel (ГОСТ 4543-71). The workpiece chemical compositions are analyzed according to ASTM 415-99A-2005 standard was listed in Table 2. Before each experiment, the workpiece was machined according to the detailed drawing of a gear with the following basic parameters: Module: m_s = 4.5mm; helix angle $\beta_s = 35^0$; number of flutes Z = 27 as shown in Fig. 6 and Fig.7.

Table 2: Main chemical composition of workpiece material %

C	Si	Mn	Cr	Ni	Mo	Cu	S	P	Fe
0.23	0.19	0.68	0.93	0.18	0.24	0.15	0.03	0.03	Balanced

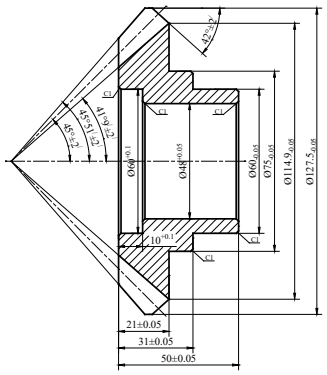


Figure 6: Detailed drawing of a gear workpiece

Figure 7: Experimental workpiece

2.4. Measuring system of tool wear

For each experiment, the tool wear was measured using the VHX-6000 digital microscope of Keyence Japan. Fig. 8 indicated the setup of the microscope for measurement of the tool wear. Fig. 9 indicates the measurement position of the flank tool wear (VB).

Figure 8: VHX-6000 digital microscope

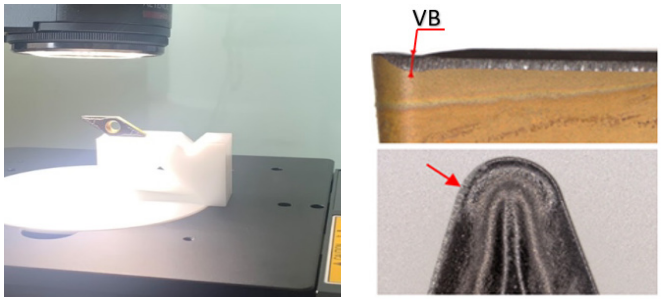


Figure 9: The tool wear of the back in a test

2.5. Experimental matrix

In this study, the experimental plan was designed using Box-Behnken matrix. In this form of planning, each input parameter had three value levels. The input parameters and their levels are presented in Table 3. The experimental matrix was presented in Table 4.

Table 3: Input parameters and their levels

Parameters	Symbol	Unit	Levels and values		
			1	2	3
Velocity of cut	V	m/min	93	117.5	142
Feed rate	S	s/tooth	40	50	60
Cutting depth	t	mm	0.25	0.50	0.75

Table 4: Test matrix and results

No.	Code value			Actual value			VB (µm)
	V	S	t	V(m/min)	S (s/tooth)	t (mm)	
1	-1	-1	0	93	40	0.5	57.22
2	1	-1	0	142	40	0.5	37.42
3	-1	1	0	93	60	0.5	32.95
4	1	1	0	142	60	0.5	82.98
5	-1	0	-1	93	50	0.25	47.73
6	1	0	-1	142	50	0.25	69.42
7	-1	0	1	93	50	0.75	29.69
8	1	0	1	142	50	0.75	66.98
9	0	-1	-1	117.5	40	0.25	37.42
10	0	1	-1	117.5	60	0.25	51.25
11	0	-1	1	117.5	40	0.75	26.85
12	0	1	1	117.5	60	0.75	59.93
13	0	0	0	117.5	50	0.5	30.10
14	0	0	0	117.5	50	0.5	43.39
15	0	0	0	117.5	50	0.5	30.24

3. Analyzed results and discussions

The experimental process was performed according to the experimental matrix as listed in Table 4. The values of tool wear (VB) that were obtained from the experimental process were also listed in this table. The results of ANOVA analysis for the tool wear were presented in Table 5. The influence degree of input parameters on the tool wear, and the influence of the interaction between parameters on the tool wear were shown in Fig. 10 and Fig. 11, respectively.

The analyzed results in table 5, Fig. 10, and Fig. 11 showed that:

The cutting velocity was a parameter with a great influence on the tool wear. When the cutting velocity increased from 93 to 117.5 (m/min), the tool wear decreases slowly, but when cutting velocity increased from 117.5 to 142 (m/min), the tool wear increased quickly.

Other parameter that has a significant influence on the tool wear was feed rate. This conclusion is also consistent with the statement of Fritz et al. [19]. When the feed rate increased, the wear of tool increased.

The cutting depth has little influence on the tool wear. Initially, when increasing the depth of cut, the tool wear decreased slowly, but further increasing the depth of cut, the tool wear increased slowly.

The interaction influence of V, the S, and t on VB were very complex influence. The in Fig. 11 showed clearly this statement.

From the experimental results in Table 4, a tool wear regression model was developed and expressed by Eq. (1). This model was built with a determination coefficient of 0.9169 (R^2 = 0.9169). This value was very close to one, which affirmed that this model had high compatibility with the testing data.

Table 5: ANOVA analysis for the tool wear

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	892.416	191.2756	4.6656	0.0055	400.7266	1384.1059	400.7266	1384.1059
V (m/min)	-8.84014	1.9269	-4.5877	0.0059	-13.7935	-3.8868	-13.7935	-3.8868
S (s/tooth)	-12.7191	4.8412	-2.6273	0.0467	-25.1639	-0.2744	-25.1639	-0.2744
t (mm)	-263.038	134.8144	-1.9511	0.1085	-609.5894	83.5134	-609.5894	83.5134
V * V	0.0230390	0.0072	3.1986	0.0240	0.0045	0.0416	0.0045	0.0416
S *S	0.0423667	0.0432	0.9799	0.3721	-0.0688	0.1535	-0.0688	0.1535
t*t	80.7867	69.1754	1.1679	0.2955	-97.0343	258.6076	-97.0343	258.6076
V*S	0.0712551	0.0170	4.2027	0.0085	0.0277	0.1148	0.0277	0.1148
V*t	0.636735	0.6782	0.9389	0.3909	-1.1066	2.3800	-1.1066	2.3800
S*t	1.92500	1.6615	1.1586	0.2990	-2.3461	6.1961	-2.3461	6.1961

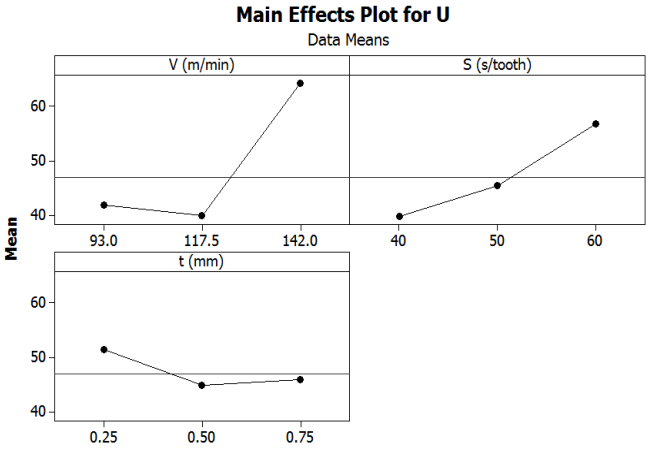
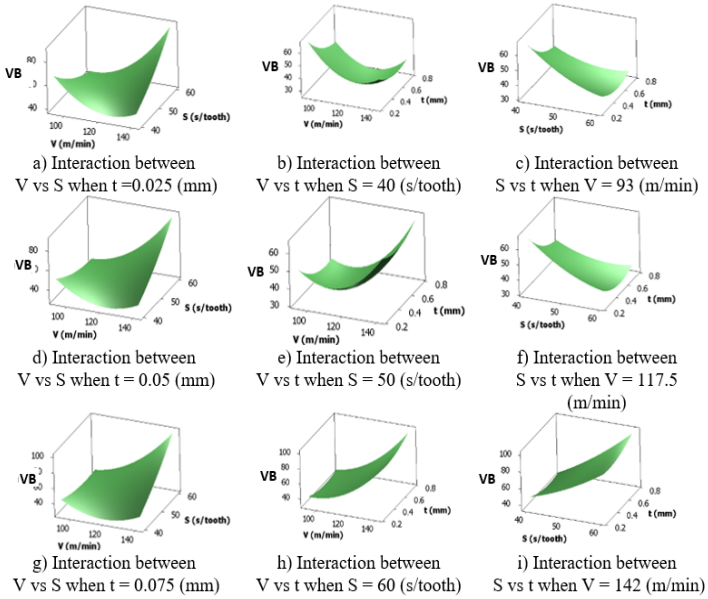


Figure 10: The influence of parameters on the tool wear



a) Interaction between V vs S when t =0.025 (mm)
b) Interaction between V vs t when S = 40 (s/tooth)
c) Interaction between S vs t when V = 93 (m/min)
d) Interaction between V vs S when t = 0.05 (mm)
e) Interaction between V vs t when S = 50 (s/tooth)
f) Interaction between S vs t when V = 117.5 (m/min)
g) Interaction between V vs S when t = 0.075 (mm)
h) Interaction between V vs t when S = 60 (s/tooth)
i) Interaction between S vs t when V = 142 (m/min)

Figure 11: The interaction influence between input parameters on the tool wear

$$VB = 892.416 - 8.84014 * V - 12.7191 * S - 263.038 * t + 0.0230390 * V^2 + 0.042366 * S^2 + 80.7867 * t^2 + 0.0712551 * V * S + 0.636735 * V * t + 1.92500 * S * t \quad (1)$$

The tool wear model presented in Eq. (1) was the basis for choosing the value of the input parameters in the investigated area to obtain the minimum value of the tool wear. In addition, this equation may be used to predict the tool wear corresponding to specific values of the cutting velocity, feed rate and depth of cut. The comparing the tool wear in experimental and predicted processes was shown in Fig.12. It was shown that the predicted results of the tool wear were very close to the that one of experimental results. This showed that equation (1) can completely be used to predict the tool wear corresponding to specific values of the input parameters in the investigated area.

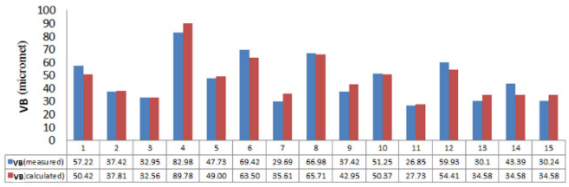


Figure 12: The tool wear when measuring and calculating

4. Optimization process

In order to obtain the optimal values of the cutting velocity, feed rate, and depth of cut to ensure the tool wear with minimum value (tools with maximum life), the constraints for input parameters and the constraints for evaluation criteria should be defined in solving the optimization problem.

The constraints values of the variables are the largest and lowest values of the Box-Behnken matrix. In particular, the constraints of experimental variables were expressed as Eq. (2).

$$\begin{cases} 93\ \left(\frac{m}{min}\right) \le V \le 142\ \left(\frac{m}{min}\right) \\ 40\ \left(\frac{s}{tooth}\right) \le S \le 60\ \left(\frac{s}{tooth}\right) \\ 0.25\ (mm) \le t \le 0.75\ (mm) \end{cases} \quad (2)$$

According to references [21, 22] and recommendation by the cutting tool manufacturer, and practical experience in the manufacturing processes, when the tool wear VB exceeds the value 250 (μm), the accuracy of each gear is not guaranteed. This value of VB is chosen as the constraint of VB when performing the optimization process in this study. Thus, the constraint for output parameter was selected by Eq. (3).

$$0 < VB \le 250(\mu m) \quad (3)$$

Using the Minitab 16 statistical software to solve the equation (1) to ensure the minimum of tool wear with the constraints as in equations (2) and (3), the optimization graph for the tool wear was obtained and shown in Fig. 13.

Fig. 13 showed that, the optimal value of the cutting velocity, feed rate, and depth of cut are 93.0 (m/min), 59.3939 (s/tooth), and 0.553 (mm), respectively. With the expectation function d = 1.000, it means the probability of this case is up to 100%, and then the tool wear has a minimum value of 32,3679 (μm).

The verified experiments were performed by using these optimized values of the cutting parameters. Based on the adjustment ability of the experimental machine, the value of cutting parameters were only selected in the percentage accuracy (two decimal places). The experiments were carried out with five tests in five gear samples with the workpiece velocity of 93.00 m/min, feed rate 59.40 s/tooth, and cutting depth of 0.55 mm. During the experimental process, other parameters were selected the same when performing the test according to the Box-Behnken matrix. The verified results were stored and listed in Table 6.

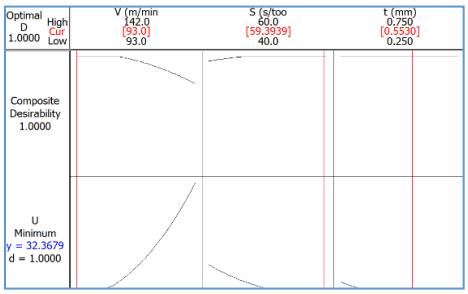


Figure 13: Optimization graph of objective functions for the tool wear

Table 6: Test results verifying the optimal value

The tool wear when testing, VB (μm)						Calculated VB (μm)	**% deviation**
Test 1	Test 2	Test 3	Test 4	Test 5	Average		
31.12	31.83	29.68	30.85	33.66	31.43	32.37	**2.9**

Table 6 indicates that in all 5 tests, the values of the tool wear during the tests are very close to the calculated values of tool wear, the average deviation between the test values and the calculated values was 0.94 μm (about 2.9%). The evaluated results showed that the optimal value of cutting parameters as well as the value of achieved tool wear when performing the optimization process ensured the high accuracy in compared to the actual value.

5. Conclusions

This study was performed to investigate the effect of milling parameters on the tool wear in milling process of the Gleason spiral bevel gear. The conclusion of this study was drawn as following.

(1) The cutting velocity has a great influence on the tool wear, the feed rate was the second factor that influence on the tool wear. The depth of cut has a negligible effect on the tool wear.

(2) The interaction between the velocity of cut, the feed rate, and cutting depth on the wear of tool is very complex.

(3) A regression model showing the relationship between the tool wear and the cutting velocity, feed rate, and the depth of cut was proposed in this study. This model was applied to determine the optimum values of cutting parameters to obtain the minimum value of tool wear. In addition, this model was also used to predict the tool wear in each specific case depending on the cutting parameters. The value of tool wear when calculating was very close to that one when performing experimental.

(4) In this study, by solving the optimization problem, the optimized results of the cutting parameters were determined as

following: The cutting velocity was 93.00 m/min, the feed rate was 59.40 s/tooth, and cutting depth was 0.55 mm. When machining with this set of cutting mode, the minimum tool wear that was obtained was about 31.43 (μm).

(5) The experimental method of this study can be used to improve the cost of the machining process through extending the tool life or reducing the tool wear.

Acknowledgement

The authors thank Hanoi University of Industry (HaUI) for the support in the performing the experimental research of this study.

References

[1] E. Kilickap, A. Yardimeden, Y. H. Çelik, "Mathematical modelling and optimization of cutting force, tool wear and surface roughness by using artificial neural network and response surface methodology in milling of Ti-6242S". Applied Sciences, **7**(1064), 1-12, 2017, doi:https://doi.org/10.3390/app7101064.

[2] Y. Li, G. Zheng, X. Zhang, X. Cheng, X. Yang, and R. Xu, "Cutting force, tool wear and surface roughness in high-speed milling of high-strength steel with coated tools", Journal of Mechanical Science and Technology, **33**(11), 5393-5398, 2019, doi:https://doi.org/10.1007/s12206-019-1033-3.

[3] N. M. Ramezani, A. Rasti, M. H. Sadeghi, B. Jabbaripour, and M. Rezaei Hajideh, "Experimental study of tool wear and surface roughness on high speed helical milling in D2 steel", Modares Mechanical Engineering, **15**(20), 198-202, 2016, doi:http://mme.modares.ac.ir/article-15-8858-en.html.

[4] G. Selim, T. Dinçer, C. K. Melih, "An Investigation on Surface Roughness and Tool Wear in Turning Operation of Inconel 718", Journal of Aerospace Technology and Management, **10**, e1319, 2019, doi:https://doi.org/10.5028/jatm.v11.1030.

[5] B. C. Behera, H. Alemayehu, S. Ghosh, P. V. Rao, "A comparative study of recent lubri-coolant strategies for turning of Ni-based superalloy", Journal of Manufacturing Processes, **30**, 541-552, 2017, doi:https://doi.org/10.1016/j.jmapro.2017.10.027.

[6] R. A. Laghari, J. Li, Z. Xie, S. Q. Wang, "Modeling and Optimization of Tool Wear and Surface Roughness in Turning of Al/SiCp Using Response Surface Methodology". 3D Research, **9**(46), 1-13, 2018, doi:https://doi.org/10.1007/s13319-018-0199-2.

[7] M. Eckstein, M. Vrabeľ, I. Maňková, "Tool wear and surface roughness evolution in hole making process of Inconel 718", In Materials Science Forum, **862**, 11-17, 2016, doi:https://doi.org/10.4028/www.scientific.net/MSF.862.11.

[8] T. Nguyen, K. H. Park, X. Wang, J. Olortegui-Yume, T. Wong, D. Schrock, B. Kramer, "The Genesis of Tool Wear in Machining" International Mechanical Engineering Congress and Exposition, 2016, doi:https://doi.org/10.1115/IMECE2015-52531.

[9] R. Wang, B. Wang, G. C. Barber, J. Gu, J. D. Schall, "Models for prediction of surface roughness in a face milling process using triangular inserts", Lubricants, **7**(9), 1-14, 2019, doi:https://doi.org/10.3390/lubricants7010009.

[10] N. L. Coppini, A. E. Diniz, F. S. Lacerda, M. Bonandi, and E. A. Baptista, "Internal turning of sintered carbide parts: tool wear and surface roughness evaluation", Journal of the Brazilian Society of Mechanical Sciences and Engineering, **40**(216), 1-7, 2018, doi:https://doi.org/10.1007/s40430-018-1139-z.

[11] T. D. Hoang, N. T. Nguyen, Đ. Q. Tran, V. T. Nguyen, "Cutting Forces and Surface Roughness in Face-Milling of SKD61 Hard Steel", Strojniski Vestnik/Journal of Mechanical Engineering, **65**(6), 375-385, 2019, doi:https://doi.org/10.5545/sv-jme.2019.6057.

[12] S. Jeyakumar, K. Marimuthu, and T. Ramachandran, "Prediction of cutting force, tool wear and surface roughness of Al6061/SiC composite for end milling operations using RSM", Journal of Mechanical Science and Technology, **27**(9), 2813-2822, 2013, doi:https://doi.org/10.1007/s12206-013-0729-z.

[13] N. F. Kundor, N. W. Awang, N. Berahim, "Tool wear and surface roughness in machining AISI D2 tool steel", Indian Journal of Science and Technology, **9**(18), 20-25, 2016, doi:https://doi.org/10.17485/ijst/2016/v9i18/88731.

[14] M. S. Said, J. A. Ghani, C. H. Haron, S. Yusoff, M. A. Selamat, R. Othman, "Tool wear and surface roughness when machining AlSi/AlN metal matrix composite using uncoated carbide cutting tool", Materials Science Forum, **773**, 409-413, 2014, doi:https://doi.org/10.4028/www.scientific.net/MSF.773-774.409.

[15] S. M. Ali, N. R. Dhar, "Modeling of tool wear and surface roughness under MQL condition-a neural approach", Canadian Journal on Artificial Intelligence, Machine Learning & Pattern Recognition, **1**(2), 7-25, 2010. doi:https://www.researchgate.net/publication/228744934.

[16] L. N. Trang, T. X. Thai, N. T. Hai, N. Nhu-Tung Nguyen, "An investigation and analysis of surface roughness and tool wear in dry pocket milling of Aluminum alloy Al7075", International Journal of Mechanical and Production Engineering Research and Development, **10**(2), 1307–1320, 2020, doi:https://www.researchgate.net/publication/340858898.

[17] B. Christian, K. Fritz, B. Markus, H. Ario, "Analysis and Optimization of Bevel Gear Cutting Processes by Means of Manufacturing Simulation", Simulation & Modeling Methodologies, Technologies & Appl, AISC, **197**, 271–284, 2013, doi:https://doi.org/10.1007/978-3-642-34336-0_18.

[18] B. Christian, K. Fritz, B. Markus, H. Ario, "Simulation based model for tool life prediction in bevel gear Cutting", Prod. Eng. Res. Devel, **7**, 223–231, 2013, doi:https://doi.org/10.1007/s11740-012-0439-x.

[19] K. Fritz, K. Alexander, "Tool Life and Productivity Improvement - Through Cutting Parameter Setting and Tool Design in Dry High-Speed Bevel Gear Tooth Cutting", GEAR TECHNOLOGY, 41-48, 2006.

[20] K. Fritz, B. Markus, H. Stefan, "Influence of Gear Design on Tool Load in Bevel Gear Cutting", Procedia CIRP, **1**, 66-71, 2012, doi:https://doi.org/10.1016/j.procir.2012.04.010.

[21] B. T. Long, T. T. Luc, T. S. Tuy, Material processing principles, Science and Technics Publishing House, Hanoi, 2013.

[22] T. V. Dich, N. T. Binh, N. T. Dat, N. V. Tiep, T. X. Viet, Machinery manufacturing technology, Hanoi Science and Technics Publishing House, 2006.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1408-1413 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Recommendation System for SmartMart-A Virtual Supermarket

Poonam Ghuli*, Manoj Kartik R, Mohammed Amaan, Mridul Mohta, N Kruthik Bhushan, Poonam Ghuli, Shobha G

Department of Computer Science and Engineering, R V College of Engineering, Bengaluru,560059, India

ARTICLE INFO

Article history:
Received: 15 October, 2020
Accepted: 26 November, 2020
Online: 16 December, 2020

Keywords:
3D environment
Unity Environment
Collaborative Filtering Algorithm
Recommendation system
Online shopping

ABSTRACT

The current online shopping system does not provide the actual shopping experience to the users where they scroll through the web pages and select an item to purchase, which becomes monotonous soon and lacks the shopping experience. This paper describes the development of the SmartMart, a 3D shopping mart where the user can navigate around the mart with the help of a virtual cart and interact with the products as they would do in real life. Further, this paper explains the integration of the recommendation system with the SmartMart that provides efficient and seamless customer experience. Here, the customers are recommended with various products based on their previous purchase history. This gives a more personalized touch to each user adding on to their shopping experience. This paper describes the implementation of recommendation system using three types of algorithms namely item-based collaborative filtering, popularity model and user-based collaborative filtering. These algorithms are integrated with Unity SmartMart application and tested with the open source dataset. This dataset is organized in three ways- original data, data with a dummy field and normalized data. Distance metrics such as Pearson co-relation co-efficient, Jaccard similarity, cosine similarity metrics are used to compute the similarity among customers. This paper presents a comparative study of these different similarity metrics for user-based recommendation algorithm based on precision, recall, F-value and accuracy. From results obtained it can be seen that the Pearson co-efficient metric gives the highest accuracy value (86.6%) but the F-value, mean precision and mean recall values are very less compared to Jaccard Similarity Metric. Hence, Jaccard similarity metric is preferred over Pearson co-relation co-efficient. The design and development of SmartMart along with the recommendation system adds a new dimension to both the existing online shopping system and provides better customer satisfaction.

1. Introduction

3D simulation software provides a rich training experience to the user for various end-user applications, without using any physical asset and loss to the company. This is a key factor of 3D simulation that is expected to boost market growth. The simulation software market was valued at USD 8.24 billion in 2019, and it is expected to reach USD 19.22 billion, by 2025. The 3D simulation enables users to immerse in a computer-generated environment where they can interact with various simulated objects in a way replicating the real-life process. These 3D objects can be generated automatically or manually created by deformation of the mesh structure, or otherwise manipulating vertices. Using 3D simulation for online shopping primarily enhances the customer experience and takes online shopping to a new level of sublimity.

This adds a realistic window which is absent with the existing online shopping websites where the user does not get to perceive the actual depiction of an item which they are inclined to purchase. Though the current websites try to minimize the hassle of the actual shopping process but it cuts away many privileges of a customer which would be helpful in selecting the appropriate item they want to purchase.

Currently, the e-commerce websites display 2D images of the products and facilitates the text-based search to seek out desired items to be purchased. But this does not allow customers to associate themselves with the real-world object experience. So, to replicate the real-world shopping experience for customers, this paper describes the design and development of a 3D shopping mart

*Corresponding Author: Poonam Ghuli, Department of Computer Science and Engineering, RV College of Engineering, 9986421230 & Email: poonamghuli@rvce.edu.in

called SmartMart. Here, each object is modeled in three dimensions to recreate the real-world object and also includes the other aspects of physics such as rotation, rolling, dragging of objects that makes the whole process more engaging. Due to this, the customer's experience of selecting an item from the SmartMart becomes more impeccable.

The SmartMart is designed using Unity along with its copious features to establish a smooth user interface. Also, the products to be bought are modeled and rendered in the gaming engine in way that provides the customers with a flawless online shopping experience. MySQL database used stores the information about the products, customers and their transaction details. MySQL database further enables to scale the number of products or customers without any inconvenience.

To broaden the scope of this proposed idea, a recommendation system has been implemented. To recommend appropriate products to customers the purchase history of multiple users is considered. A few factors that measures E-commerce success and customer satisfaction are as follows:

- What are the most trending products for sale?
- What are most suitable products for a particular customer based on his purchase history?
- What items are more often bought together?

Considering all the above-mentioned key factors, this paper describes implementation of recommendation system using three types of algorithms namely item-based collaborative filtering, popularity model and user-based collaborative filtering. Further, these algorithms are designed and integrated with Unity SmartMart application. For this, an open source dataset is used and organized in three ways- original data, data with a dummy field and normalized data. Further, Pearson co-relation co-efficient, Jaccard similarity, cosine similarity matrices are used to compute the similarity between purchase pattern usages among customers. Jaccard similarity metric provided the best result.

The admin of the SmartMart application is authorized to append the data of the ongoing transactions and periodically train the model with latest dataset. This ensures that the recommendation system takes into account the changing trends of customer's shopping patterns with proposed SmartMart application.

2. Literature Review

In [1], the author described a personalized Recommender System based on data clustering techniques to deal with the scalability problem. User-clusters are formed to categorize them based on their shopping behavior. The proposed work use DBSCAN clustering algorithm for clustering the users, and then implement voting algorithms to recommend items to the user depending on the cluster into which it belongs. Their idea was to partition the users of the Recommendation system using clustering algorithm and apply the Recommendation Algorithm separately to each partition. The system recommends item to a user in a specific cluster only using the rating statistics of the other users of that cluster. They have tested the algorithm on the Netflix prize dataset.

The work in [2], proposed two recommendation models for offline grocery/retail stores. The accumulated PoS (Point of Sales) data is used to understand the shopping behavior of different customers and recommend better products. The proposed recommendation models are-Direct recommendation of products and two step recommendation systems. These recommendation algorithms are analyzed based on precision and accuracy of the model.

In [3], the author presented a recommendation system in their 3D environment where they consider the pre-purchase ratings of users to suggest products. The pre-purchase ratings include the interaction of user with the 3D objects. The user's emotions are captured in the form of electroencephalogram (EEG) signals. This makes the recommendations more users specific and accurate. A combination of this and the already existing recommendation system makes it dynamic.

The study in [4] presents a Recommendation system which offers real-time personalized suggestions on offers to its customers. Using iBeacons personalized notifications telling about the offers which are pushed to the smart phones of the customer in real-time as they navigate through the shopping mart. This approach uses a Bayesian Inference and a Hard k-modes clustering approaches. Bayesian Inference approach uses Entropy-based algorithm to give accurate and user specific recommendations on the offers provided to each user.

In [5], the author has proposed a virtual shopping mart which simulates a real working of shopping mart. The virtual shopping mart enables a cost efficient and high flexible way to store data. Researchers can use this data to improve the shopping experience of the customers. The data can be used to accurately predict the reactions of the customers to price change and the reactions to availability of the products that the customers need. These reactions can be used to optimize the shopping mart in a way to efficiently serve the customer's requirements.

In virtual reality the customers have the exact 3D representation of the items unlike the other 2D shopping websites this helps in analyzing the exact reactions the customers have towards a product in the shopping mart which enables researchers to capture an accurate consumer behavior.

The architecture that is proposed in this paper stands out from the other existing prototypes in a way that it integrates the recommendation system with the 3D environment (Unity software) which makes it possible for the user to have a sublime shopping experience. This entire system is based on client server model that allows concurrent access to the SmartMart by the potential customers. The use of recommendation system which recommends the most popular items and the items that the user would purchase helps the user to manage his/her shopping time more efficiently. The interaction with the 3D environment using keyboard and mouse makes the customer's shopping experience realistic.

3. Proposed Design of Smart-Mart

The system architecture for the proposed SmartMart is illustrated in Figure 1. It manifests a computing model in which a server hosts, delivers and manages most of the resources and services to be utilized by the clients. The project aims at building a Client-Server architecture where the client application entails a 3D UI (User Interface) platform that would grant customers/users a real-time shopping experience at the click of a button.

The customer interacts with the environment through keyboard and mouse input controls to bring in various changes ranging from the state of a product to updating the fields in database in the server. The MySQL database stores all the information about products, customers, transaction details and billing information. A virtual shopping cart designed in unity platform can freely move around the SmartMart.

Also, customers are allowed to add products to shopping cart, thereby facilitating a user-friendly experience to the users, unlike the existing shopping platforms. As the user proceeds towards the billing section, the inventory is updated which would help administrators to track and maintain the stock efficiently.

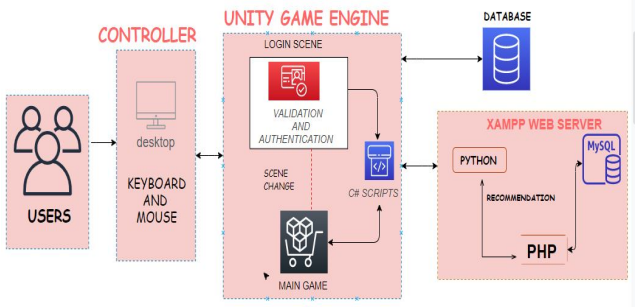


Figure 1: System Design of User interaction with SmartMart.

The Figure 1 describes about the Unity gaming engine that is used to host the entire shopping environment application. The Unity APIs used in the proposed work handle various user interactions and allow communication with third party scripts, namely, C# and PHP (Hypertext Preprocessor). The user interactions are translated into corresponding events by the interface. The application processes these events and triggers other events or updates the SQL database connected to it. A PHP script acts as an intermediary to communicate between the SQL database and the application.

The figure 2 describes the 3D view of the proposed SmartMart. This figure shows a virtual cart an products placed on shelves in the SmartMart. The user can move around virtual around the mart along with the Virtual-Cart.

The uniqueness of this paper is integration of SmartMart with the recommendation system. This recommendation system is deployed using the Python script and connected to the SmartMart application via PHP.

Figure 2. The 3D view of the SmartMart

4. Workflow of Proposed Work

Unity is a cross-platform game engine with a built-in IDE (Integrated Development Environment) used to develop video games for web plug-ins, desktop platforms, consoles and mobile devices. Pro-Builder is a tool that gives Unity the ability to edit the shape of primitive objects and model these 3D objects as per the requirements.

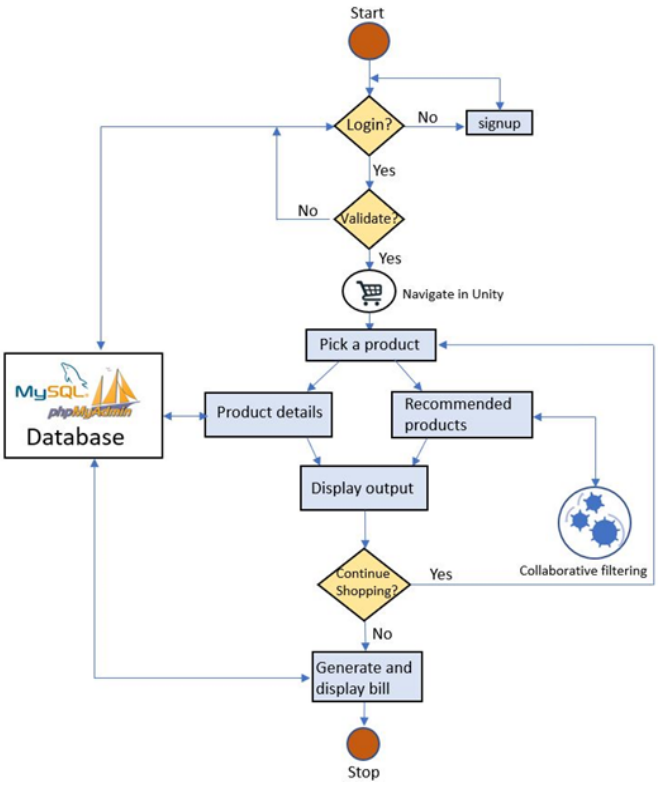


Figure 3: Workflow of Smart-Mart

The work flow of the Virtual Supermarket is depicted in the Figure 3. The application starts with the initial scene where user credentials are validated with username and password. If the credentials are correct, the user gets into the game scene, which is the main application. The user is placed at the entrance of the mart, with a virtual cart placed there. The cart is designed using Pro-Builder. The cart is scripted such that the customer can move it around the mart with the help of the mouse control.

Shelves at the virtual mart are rendered using unity platform and various products are placed on it. The products at the mart are built as 3D objects using Pro-builder platform in order to recreate the real-life products such as baseball bats or celery. These virtual products created in the Unity platform are nothing but Game Objects. Game Objects are the fundamental objects in Unity that represent characters, props and scenery. Every Game Object is associated with a tag which is used to identify, retrieve and store the information about a particular product. Each product/Game Object is associated with a script which allows the user to rotate, drag and drop it into the virtual cart. Whether the customer puts or removes the item to or from the cart, the inventory database is updated accordingly in real time. As soon as the user selects the product, the details of the product along with the recommendations are displayed on the monitor installed at the billing counter. This is accomplished using PHP server that communicates with MySQL and Unity.

Once the user completes the adding of products to the virtual cart, the cart can be dragged to the billing counter where appropriate billing information is displayed. This billing information includes the items purchased, per piece price, total quantity, total price, time of billing, customer id and customer details. These transaction details are stored in the database in real time. Once the billing details are generated, the user signs out and the session is over.

Thus, database manages all the transactions, user and product information efficiently. The details of the products are displayed to customers as required without any latency. The interactions, ranging from dragging and dropping of the items into virtual cart to its movement in 3D environment in the first-person view is flawless.

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

5. Recommendation System

Al the proposed virtual Super market is further integrated with a Recommendation System. In the field of e-commerce, recommendation systems are an attempt to bridge the gap between the online and in-store shopping experience reiterating the very aspect that made this project distinctive. Recommendation systems streamline our online shopping experience by suggesting products that customers might be interested in.

With the growing technological advancements in the fields of business intelligence and data science, recommendation systems have become increasingly popular. They help the customers to know and decide what to buy based on their previous purchase history, their feedbacks and features of the products to be purchased.

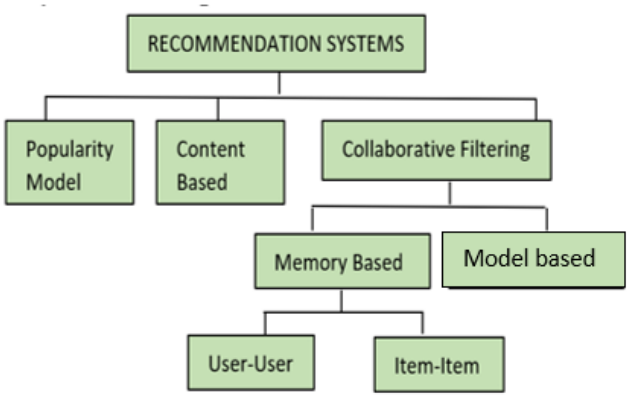


Figure 4: Classification of recommendation system

The two most popular approaches for building a recommendation system are Collaborative Filtering and Content-based algorithms as shown in Figure 4. While the former approach requires only the user's historical preferences over the existing inventory, the latter requires a fair amount of knowledge on the item's characteristics such as ingredients, trademark and features. Although Collaborative Filtering depends on only purchase history of users one assumption that this model makes is that the users who have purchased a certain set of products in the past also tend to agree with the same kind of products in the future.

Collaborative filtering algorithms [6] are mainly classified as model based and memory based approaches. Further there are two types of Memory based Collaborative filtering algorithms namely Item-based collaborative filtering and User-based collaborative filtering . Item-based collaborative filtering that depends on items and their similarity to other items. User-based collaborative filtering algorithm is more resource-intensive as it looks at customer behaviors and preferences. In memory based approach the similarities with the closest users or items are calculated by using Cosine similarity or Pearson correlation coefficients, which depend primarily on arithmetic operations. Model based collaborative filtering algorithms provide recommendations by developing a machine learning model on existing user ratings and then to predict user's rating of unrated items [7]. As building a model is both time and resource consuming. It is difficult to add data to model based systems. The quality of prediction for memory based algorithms is good as it uses the entire database every time it makes a prediction. Thus, we go with memory based algorithms for the proposed work.

The similarity metrics [8] used for recommending products in the SmartMart application are Cosine, Pearson Correlation Coefficient and Jaccard similarity methods.

- By measuring cosine of the angle between any two vectors, Cosine similarity technique calculates how these are related to each other. The Cosine similarity metric is computed as shown in equation 1.

$$\boldsymbol{CS(i,j)} = \frac{\sum_{u \in U_{ij}} \boldsymbol{r_{ui} r_{uj}}}{\left(\sqrt{\sum_{u \in U_{ij}} \boldsymbol{r_{ui}^2}}\right)\left(\sqrt{\sum_{u \in U_{ij}} \boldsymbol{r_{uj}^2}}\right)} \quad (1)$$

where Ui represents users, who have purchased only item i & Uij represents users who have purchased both i & j items.

- Pearson coefficient similarity technique calculates the similarity between two vectors by the linear correlation between them. The values resulting from PCC range from -1, representing negative correlation to +1, representing a high positive correlation. Zero order correlation is when the value is 0 and the vectors do not show any relation. The Pearson Correlation Coefficient similarity metric is computed as shown in equation 2.

$$\boldsymbol{PS(i,j)} = \frac{\sum_{u \in U_{ij}} \left(\boldsymbol{r_{ui}} - \overrightarrow{\boldsymbol{r_i}}\right)\left(\boldsymbol{r_{uj}} - \overrightarrow{\boldsymbol{r_j}}\right)}{\left(\sqrt{\sum_{u \in U_{ij}} \left(\boldsymbol{r_{ui}} - \overrightarrow{\boldsymbol{r_i}}\right)^2}\right)\left(\sqrt{\sum_{u \in U_{ij}} \left(\boldsymbol{r_{uj}} - \overrightarrow{\boldsymbol{r_j}}\right)^2}\right)} \quad (2)$$

where Ui represents users, who have purchased only item i & Uij represents users who have purchased both i & j item.

- Jaccard similarity technique calculates the extent of similarity between two users by taking number of preferences common between them and is shown in equation 3.

$$JS(i,j) = \frac{|U_i \cap U_i|}{|U_i \cup U_i|} \quad (3)$$

where Ui represents users, who have purchased only item i.

As mentioned in the previous section this paper implements recommendation system using three types of algorithms namely item-based collaborative filtering, popularity model and user-based collaborative filtering. Popularity based recommendation systems learn data in a certain way, after which, suggest items that are in trend at the moment. For instance, if there exists a product that is being purchased by many users then it is likely that it will be recommended to every user.

6. Details of Dataset and Data Pre-processing

The proposed work makes use of two datasets. First one is MySQL database created to store details pertaining to the customers, their transactions and product information. The other one is an open-source dataset [9].

The MySQL database being employed has the following tables that stores vital information – the validation details of the customers; the customer table contains the details of all the existing customers; the inventory table contains details like product ID, product name and their respective quantities that directly correlate to the stock available with the seller; the billing table which stores the bill generated during shopping and holds information like user ID, items bought, time of purchase, etc and the orders table that contains all the bills generated along with respective product information.

The open-source dataset that is used to build a recommendation system is available in .csv format. It has information regarding the purchase history of the customers. There are 3 approaches in terms of how data is being used.

Taking customer ID and product ID as they are.

Creating a dummy field for marking whether a customer bought an item or not.

Normalize purchase frequency of each item across users by first creating a user-item matrix.

Since normalization is always used as a basic component for the predictor models and most often produces the best results when it comes to recommender systems, it was decided to proceed with normalized data. We have normalized the purchase history of users on a scale of 0 to 1. 1 depicts the greatest number of purchases for an item and 0 depicts 0 purchase count for that item.

7. Results and Discussions

The proposed work evaluates the different Similarity Metrics for user-item recommendation algorithm based on recall, precision, and F-value with the number of recommended items for each user as 10 [10]. The results presented in Table 1 evaluate the different similarity matrices for user-item recommendation algorithm by using the average value of recall, precision, and F-value. Table 1 also provides the recall, precision, and F-value for popularity model variant of recommendation system. Cut-off 'K' is a user definable integer that is set by the user to match the top-N recommendations. The value of cut-off is kept as 10 (K=10).

Table 1: Performance comparison of Similarity Matrices

Recommendation System Models	Similarity Metrics	Precision	Recall	F-Value
Popularity Model	-	0.002685	0.014068	0.004509
User-based Collaborative Filtering Algorithm	Cosine similarity	0.026191	0.137444	0.043999
	Pearson similarity	0.002706	0.014176	0.004545
	Jaccard similarity	0.027089	0.141439	0.045469

Precision: The number of products bought by the customer out of all the recommended products. Precision at k is the proportion of recommended items in the top-k set that are relevant. Suppose if 10 products were recommended to this customer out of which he buys 5 products, then precision is 0.5. The precision at k is computed using equation 4.

$$\text{Precision} = \frac{\left(\begin{matrix}\text{Number of recommended items that}\\ \text{are relevant in the top k set}\end{matrix}\right)}{\text{Number of recommended items in the top k set}} \quad (4)$$

Recall: The number of products that are recommended out of all the products that a user buys. Recall at k is the proportion of relevant items found in the top-k recommendations. Suppose if a customer buys 10 products and the recommendation decided to show 4 of them, then recall is 0.4. Recall at k and F - value is computed as shown in equation 5 and equation 6 respectively.

$$\text{Recall at k} = \frac{\left(\begin{matrix}\text{Number of recommended items}\\ \text{that are relevant in the top k set}\end{matrix}\right)}{\text{Total number of relevant items}} \quad (5)$$

$$\text{F} - \text{value} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

From Table-1 it can be inferred that the Jaccard similarity has the highest F-Value (0.04546) followed by cosine similarity (0.04399). The mean precision and mean recall are also the highest for Jaccard similarity (0.0271 and 0.1414 respectively). By the results it can be seen that Jaccard similarity is the best metrics to recommend products for the dataset used in this paper.

Table 2 presents the performance comparison of various similarity metrics based on RMSE (Root Mean Square Error) and accuracy. It also provides RMSE and accuracy values for the

popularity model variant of recommendation system. The Pearson metric gives the highest value (86.6%) but since the F-value, mean precision and mean recall values are very less, Jaccard metric is preferred. Note that the popularity model just recommends the most popular products and is similar for every user, so we don't consider it in user-item recommendation.

Table 2: Performance comparison of various similarity metrics based on RMSE and Accuracy

Recommendation System Models	Similarity Metrics	RMSE	Accuracy
Popularity model	-	13.3580%	86.6420%
User-based Collaborative Filtering Algorithm	Cosine similarity	16.1986%	83.8014%
	Pearson similarity	13.3915%	86.6085%
	Jaccard similarity	16.0924%	83.9076%

8. Conclusion

The Virtual SmartMart application is an innovative platform, which evolves the way people shop online when compared to different E-commerce websites. The above deployed recommendation models make sure that the customers never miss out on any current trending products (Popularity model), the most liked products by other users (user-item model) and best products related to the items present in the cart (item-based model). Therefore, the recommended products will provide customers a sense of satisfaction.

This paper presents a comparative study of different similarity metrics for user-based recommendation algorithm based on precision, recall, F-value and accuracy. From results obtained it can be seen that the Pearson co-efficient metric gives the highest accuracy value (86.6%) but since the F-value, mean precision and mean recall values are very less, Jaccard similarity metric is preferred.

From the dealer point of view, Unity Analytics (integrated with Unity game engine) with SmartMart will help him in monitoring the most selling products, the most active time during a day, the number of shoppers active at any point of time and the countries with leading number of shoppers. This helps dealer make appropriate decision at right time. Integration of Unity Analytics with the SmartMart can be considered as future work for this work proposed here.

For any company retaining customers and up scaling of business are two very important aspects. SmartMart Unity application successfully helps a company in doing so. The limitations of the virtual supermarket include cold start problem, data sparsity problem and scalability. This can be overcome by developing a hybrid recommendation system; a combination of collaborative and content-based filtering. This can be taken up as future work.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We would like to express our special thanks to SCII (System Consultant Information India) for funding this work. We are grateful that they provided us an opportunity carry out this wonderful project which encouraged us to learn and explore unity platform to build virtual Smart-Mart application and integrate it with Recommendation system. Their constant guidance and suggestions helped us strive for sublimity and to expand our scope of learning.

References

[1] J. Das, P. Mukherjee, S. Majumder, P. Gupta, "Clustering-based recommender system using principles of voting theory," in Proceedings of 2014 International Conference on Contemporary Computing and Informatics, IC3I 2014, 2014, doi:10.1109/IC3I.2014.7019655.

[2] N. Sano, N. Machino, K. Yada, T. Suzuki, "Recommendation system for grocery store considering data sparsity," in Procedia Computer Science, 2015, doi:10.1016/j.procs.2015.08.216.

[3] G. Guo, M. Elgendi, "A New Recommender System for 3D E-Commerce: An EEG Based Approach," Journal of Advanced Management Science, **1**(1), 61-65, 2013, doi:10.12720/joams.1.1.61-65.

[4] P. Christodoulou, K. Christodoulou, A.S. Andreou, "A real-Time targeted recommender system for supermarkets," in ICEIS 2017 - Proceedings of the 19th International Conference on Enterprise Information Systems, 2017, doi:10.5220/0006309907030712.

[5] W.E. Waterlander, M. Scarpa, D. Lentz, I.H.M. Steenhuis, "The virtual supermarket: An innovative research tool to study consumer food purchasing behaviour," BMC Public Health, **11**, Article No. 589, 2011, doi:10.1186/1471-2458-11-589.

[6] G. Suganeshwari, S.P. Syed Ibrahim, "A survey on collaborative filtering based recommendation system," in Smart Innovation, Systems and Technologies, 503-518, 2016, doi:10.1007/978-3-319-30348-2_42.

[7] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "Item-based collaborative filtering recommendation algorithms," in Proceedings of the 10th International Conference on World Wide Web, WWW 2001, 2001, doi:10.1145/371920.372071.

[8] A. Agarwal, M. Chauhan, "Similarity Measures used in Recommender Systems: A Study," Minakshi Chauhan International Journal of Engineering Technology Science and Research IJETSR Www.Ijetsr.Com ISSN, **4**(6), 619-626, 2017.

[9] S. Mohtaj, F. Tavakkoli, "Maglet: A Persian Journal Recommender System," in 9th International Symposium on Telecommunication: With Emphasis on Information and Communication Technology, IST 2018, 2019, doi:10.1109/ISTEL.2018.8660987.

[10] A. Gunawardana, G. Shani, "A survey of accuracy evaluation metrics of recommendation tasks," Journal of Machine Learning Research, **10**, 2935-2962, 2009.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 1414-1421 (2020)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

Solution of the Semiconductor-Device Equations by the Numerov Process

Nicolò Speciale[1], Rossella Brunetti[2], Massimo Rudan[1*]

[1]*DEI Department, University of Bologna, Viale del Risorgimento 2, I-40136 Bologna, Italy*

[2]*FIM Department, University of Modena and Reggio Emilia, Via Campi 213/A, I-41125 Modena, Italy*

ARTICLE INFO

Article history:
Received: 30 June, 2020
Accepted: 12 October, 2020
Online: 16 December, 2020

Keywords:
Numerov Process
Modeling Charge Transport
Solid-State Devices

ABSTRACT

The one-dimensional form of some types of differential equations that appear in the modeling of solid-state devices, like, e.g., the Poisson and Schrödinger equations, can be solved numerically using the Numerov Process (NP). The accuracy of NP is superior by at least two orders of magnitude to that of commonly-used solution methods like, e.g., the finite-difference one. Another advantage of NP is that the increase in computational cost is modest. This paper investigates the possibility of extending the applicability of NP to the whole set of equations that constitute the mathematical model of semiconductor devices. It shows that the charge-continuity and charge-transport equations can be recast in a form suitable for NP. The performance of the method is tested on an n^+-n-n^+ *device, and the extension of the approach to the two-dimensional transport problem is worked out.*

1 Introduction

A number of physical problems are described by second-order differential equations of the form

$$-z'' = F(z, x) \tag{1}$$

(primes indicate derivatives with respect to x). When a uniform grid with grid size h is used, such a form can be tackled numerically, thanks to the absence of the first derivative, with an $O(h^4)$-accurate discretization method generally known as Numerov Process (NP); this in contrast to $O(h^2)$ of the finite-difference method (the derivation of the scheme is shown in the Appendix). Linearization of (1) yields

$$-z'' = Q(x)\,z + P(x)\,. \tag{2}$$

Scope of this paper is to extend the applicability of NP to the system of equations that describe the transport of electric charge in solid-state devices. This paper is an extension of work originally presented at the SISPAD 2019 Conference [1]; besides the Appendix, more details about NP are given, e.g., in [2] and references therein.

In some cases, (2) can be reduced to equations that appear in the model for semiconductors; for instance, letting $Q = 0$, and naming $z = u$ the electrostatic potential normalized to the thermal voltage $k_B T/q$, and $P = q/(\varepsilon k_B T)\,\varrho$, with ϱ the charge per unit volume, transforms (2) into the Poisson equation, which is one of the equations of the standard semiconductor-device model (in the above definitions, symbols q, ε, k_B, T denote the elementary charge, material permittivity, Boltzmann constant, and lattice temperature, respectively). If, instead, one lets $P = 0$, the form of the time-independent Schrödinger equation is recovered, which enters the semiconductor-device model when nanometer-scale devices are considered; in this case, $z = w$ is the time-independent wave function, whereas the form of Q reads $Q = 2\,m\,(E - V)/\hbar^2$, where $E = \text{const}$ is the total energy of the electron and $V(x)$ the potential energy; in turn, m is the electron effective mass and $\hbar$ the reduced Planck constant.

The elimination of the odd-order derivatives in the Taylor expansion (see Appendix), as required by NP, prescribes the use of a uniform grid; the better accuracy of the method largely compensates for this aspect. An application of NP to the analysis of a semiconductor device was shown in [2], where the coupled solution of the Schrödinger-Poisson system was dealt with; in that case, no manipulation of the equations was necessary because both have already the form (2).

Object of this paper is to show that, by a suitable choice of the unknowns, NP can be extended to the whole set of equations constituting the mathematical model of semiconductor devices. To this purpose it is convenient to briefly outline the recent progresses in modeling, following the semiconductor-device roadmap (now "International Roadmap for Devices and Systems — IRDS™" [3]) and the corresponding evolution of the mathematical models.

The standard model of semiconductor devices (*drift-diffusion* model) is made of the Poisson equation and, for each type of mobile charges (electrons and holes), of the continuity and transport

*Corresponding author:M. Rudan, Dept. DEI, Univ. of Bologna Viale del Risorgimento 2, I-40136 Bologna, Italy +39-051-209-3016, massimo.rudan@unibo.it

equations for the mobile-charge density; the transport equation has two terms: the first one is of the drift type, the second one of the diffusive type. For submicron devices the model is extended by adding another pair of continuity and transport equations, whose unknown is the energy density of the carriers (in this case, the model is termed *hydrodynamic*).

When the lateral extension of the device channel shrinks to the nanometer scale, the device can be described as a cylindrical structure. This makes it possible to separate the longitudinal and transversal directions; then, the calculation of the charge density is accomplished by solving the Schrödinger equation over each cross section, whereas the analysis along the longitudinal direction is still carried out in classical terms. An example is found in the solution of a cylindrical nanowire shown in [2]. Other examples of nanometer-size structures made of compound materials are illustrated in [4].

The n^+-n-n^+ structure considered in this paper is suitable for checking the extension of NP to the solution of the whole transport model of semiconductors. The structure is one of the benchmarks used in the literature (see, e.g., [5] and references therein): besides being one-dimensional, like a nanowire or a gate-all-around transistor, it has the advantage of being unipolar, namely, for its description it suffices to use only one type of carriers (electrons in this case), thus halving the number of continuity-transport pairs of equations to be solved.

The investigations on numerical methods for solving the semiconductor equations have gone in parallel with the strong development of the integrated-circuit technology of the last 50 years. At present, the most successful solution techniques for the drift-diffusion or hydrodynamic models are incorporated in huge device-analysis systems, whose update and maintenance is a concern of the semiconductor Companies in cooperation with specialized Software Houses. The advantage of NP is that it is applicable to the existing solution techniques without excessive changes in the organization of the software, and with a limited increase in the computational cost; for this reason, also considering the improvement in accuracy, the application of NP to the semiconductor equation is worth considering. The present investigation about the applicability of NP to the numerical solution of the semiconductor transport equations considers the drift-diffusion model for the sake of simplicity, since the hydrodynamic model and the other models of higher order, derived from the Boltzmann Transport Equation, have the same structure.

The paper is organized as follows: the model equations are shown in Section 2 along with the transformation that gives them an NP-suitable form; the NP-based discretization in the one-dimensional case is carried out in Section 3. A method for calculating the electric potential at each node independently of the coupling with the transport equation is outlined in Section 4, followed by a stability analysis of the iterative solution (Section 5). The results are shown in Section 6, and the extension of the method to the two-dimensional case, still over a uniform grid, is given in Section 7. Finally, the conclusions are drawn in Section 8, while a summary of the derivation of NP is provided in the Appendix.

2 Model Equations

The semiclassical, drift-diffusion equations describing the transport of carriers in solids is made of the Poisson equation and, for each type of charges (namely, electrons and holes), of a pair of equations made of the continuity and of the transport equations. Here a one-dimensional device of the unipolar type is analyzed, whose carriers are electrons. This situation is found, e.g., in n^+-n-n^+ structures like that shown in Figures 1 and 2, or in memory devices of the phase-change type (e.g., [6, 7]); in these devices, the Poisson equation reads

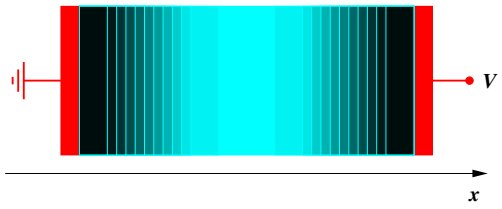


Figure 1: Lateral view of the simulated n^+-n-n^+ structure. The red regions indicate the contacts, while the shades of green mimic the non-uniform dopant concentration.

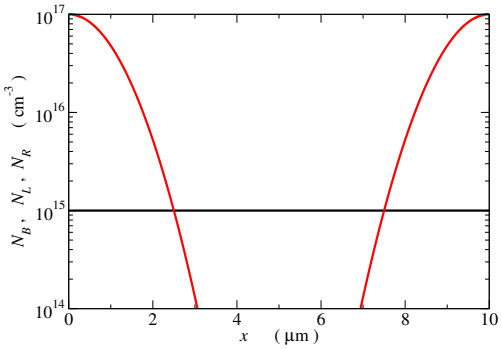


Figure 2: Dopant profiles of the simulated n^+-n-n^+ structure.

$$-u'' = P = \frac{q}{\varepsilon k_B T}\, \varrho(n, N_D)\,, \tag{3}$$

with n the electron concentration and $N_D(x)$ the concentration of ionized donors within the n^+-n-n^+ structure (when phase-change materials are considered, N_D is replaced with the concentration n_T of empty traps). The continuity equation of the electrons, and the corresponding drift-diffusion equation for the electron-current density read, respectively,

$$J_n' = q\,U\,, \qquad J_n = q\,D_n\,(n' - u'\,n)\,, \tag{4}$$

with $U = U(n)$, D_n being the net-recombination rate and the electron-diffusion coefficient, respectively (here a constant value of D_n is considered). Combining (4) yields

$$n'' - u'\,n' - u''\,n = \frac{U}{D_n}\,. \tag{5}$$

The Wronskian determinant of the homogeneous equation associated to (5) reads, by Abel's identity, $W = \text{const} \times \exp(u) \neq 0$.

The system of equations to be solved is then made of (3) and (5), whose boundary conditions are the values of u and n at the two ends of the (finite) integration interval. If the discretization of the interval

produces N internal nodes, the boundary conditions are u_0, u_{N+1} and n_0, n_{N+1}, respectively; in the analysis of solid-state devices it is assumed that the contacts force a condition of electric neutrality, so that n_0, n_{N+1} are such that $P_0 = P_{N+1} = 0$ (compare with (3)).

In a decoupled solution scheme, one assumes that n is given and solves (3) for u, then solves (5) for n using the form of u thus found; the procedure is iterated until convergence is reached. One notes that NP is not applicable to (5) as it stands, due to the presence of the first derivative; to give the equation a more suitable form one defines

$$s = -\frac{U}{D_n}\,\exp(-u/2)\,, \qquad g = n\,\exp(-u/2)\,, \tag{6}$$

this transforming (5) into

$$-\,g'' = c\,g + s\,, \qquad c = \frac{P}{2} - \frac{(u')^2}{4}\,. \tag{7}$$

Transformation (6) does not consist in a reduction to a self-adjoint form; in fact, the expression of the latter would be

$$[n'\,\exp(-u)]' = (U/D_n - P\,n)\,\exp(-u)\,; \tag{8}$$

transformation (6) is not either an "exponential fitting" of the type commonly adopted for solving the semiconductor equations: the exponential fitting would in fact yield

$$q\,D_n\,[n\,\exp(-u)]' = J_n\,\exp(-u)\,, \tag{9}$$

where the current density J_n is approximated with a different constant along each interval between two nodes (in the field of semiconductor-device modeling the scheme based on (9) is also known as *Scharfetter-Gummel method* [8]).

The homogeneous equation corresponding to (7) reads $g''+c\,g = 0$; in the iterative procedure by which the system made of (3) and (7) is solved within the finite integration interval, coefficient c is obtained from the electron concentration n and the normalized electrostatic potential u calculated at the previous iteration. As neither n nor u have poles, all points within the integration interval are ordinary.

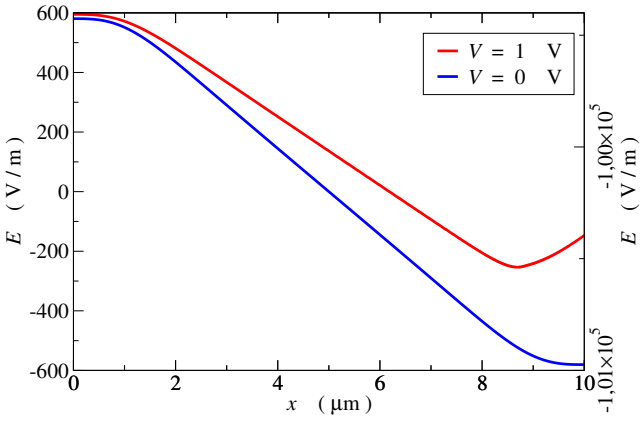


Figure 3: Electric field in the simulated n^+-n-n^+ structure, at different values of the applied bias; namely, 0 V (blue curve, left scale) and 1 V (red curve, right scale).

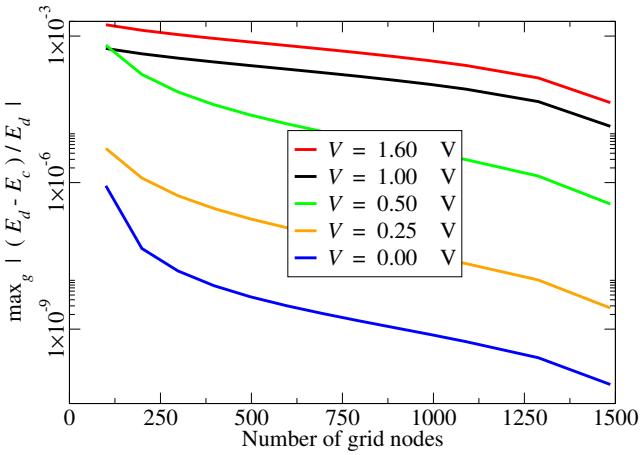


Figure 4: Maximum modulus of the relative variation $\max_g|(E_d - E_c)/E_d|$ of the electric field with respect to the densest grid (2000 nodes). The horizontal axis shows the number of nodes of each grid used in the comparison (more details are given in Section 6). The analysis has been repeated for different values of the applied voltage V, as shown in the inset.

3 Application of NP

In this Section, the Numerov Process is applied to (3) and (7) over a grid made of N, uniformly-spaced internal nodes with element size h. The Poisson equation (3) is transformed into the $N \times N$ algebraic system

$$-\,u_{i-1} + 2\,u_i - u_{i+1} = \frac{h^2}{12}\,(P_{i-1} + 10\,P_i + P_{i+1})\,, \tag{10}$$

$i = 1, 2, \ldots, N$, while (7) transforms into the $N \times N$ system

$$-\left(1 + \frac{h^2}{12}\,c_{i-1}\right) g_{i-1} + \left(2 - 10\,\frac{h^2}{12}\,c_i\right) g_i \tag{11}$$

$$-\left(1 + \frac{h^2}{12}\,c_{i+1}\right) g_{i+1} = \frac{h^2}{12}\,(s_{i-1} + 10\,s_i + s_{i+1})\,.$$

The matrix of the algebraic system (10) is symmetric, whereas that of (11) is non symmetric due to terms c_{i-1} and c_{i+1} at the left hand side. Although in the latter system the unknowns are $g_i = n_i\,\exp(-u_i/2)$, terms c_i and s_i still depend on the original unknown n; also, each g_i depends, through $\exp(-u_i/2)$, on the (arbitrary) zero of the electric potential; finally, due the presence of the exponentials of the electric potential, the matrix of the coefficients of (11) may become stiff. For the above reasons it is convenient to multiply both sides of (11) by $\exp(u_i/2)$. In this way, the following replacements are carried out in (11): $g_{i-1} \leftarrow n_{i-1}\,\exp[(u_i - u_{i-1})/2]$, $g_i \leftarrow n_i$, and $g_{i+1} \leftarrow n_{i+1}\,\exp[(u_i - u_{i+1})/2]$; by the same token, $s_{i-1} \leftarrow -(U_{i-1}/D_n)\,\exp[(u_i - u_{i-1})/2]$, $s_i \leftarrow -(U_i/D_n)$, and $s_{i+1} \leftarrow -(U_{i+1}/D_n)\,\exp[(u_i - u_{i+1})/2]$.

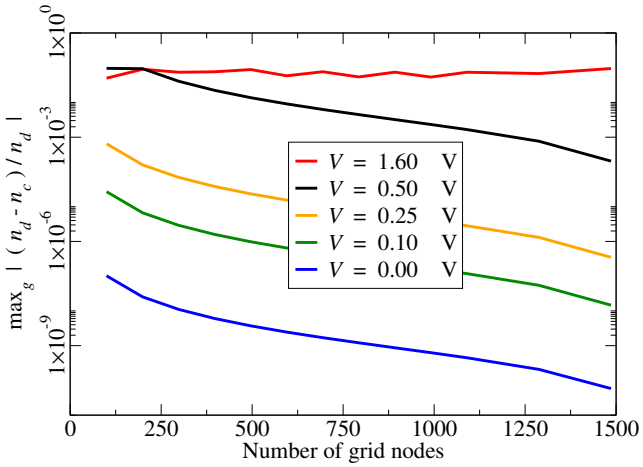


Figure 5: Maximum modulus of the relative variation $\max_g|(n_d - n_c)/n_d|$ of the electron concentration with respect to the densest grid (2000 nodes). Compare with Figure 4.

The new form of (11), where the original unknown appears, is preferable because the exponents bearing differences like, e.g., $u_i - u_{i-1}$, instead of a nodal potential alone, are expected to be relatively small. To complete the calculation it is necessary to express the derivative u' that appears in (7); this is also done using NP, so that the accuracy of the scheme is preserved, and yields

$$u'_i = \frac{u_{i+1} - u_{i-1}}{2\,h} + h\,\frac{P_{i+1} - P_{i-1}}{12}\,. \tag{12}$$

When $i = 2, \ldots N-1$, only the indices of internal nodes appear in (12); if, in contrast, $i = 1$ or $i = N$, then one boundary condition is present (u_0 and $P_0 = 0$ or, respectively, u_{N+1} and $P_{N+1} = 0$); the last possible case occurs when $i = 0$ or $i = N+1$, in which the values of the electric potential and charge in the interior of the contacts appear in (12); remembering that the contacts are equipotential and neutral, it is $u_{-1} = u_0$, $P_{-1} = 0$ and $u_{N+2} = u_{N+1}$, $P_{N+2} = 0$.

4 Decoupling the Model Equations

The model equations are solved by iterating between (10) and (11). In general, equations like these are tackled with algebraic solvers able to provide the values of the unknowns in a given sequence; for instance, considering the solution of (10) obtained through the $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{U}$ decomposition, the ith nodal value of the normalized electric potential u is obtained only after calculating it at nodes 1 through $i-1$ (or at nodes N through $i+1$). Here a different approach is adopted, in which u_i is obtained as soon as necessary, without the need of calculating its other nodal values. Such a result is achievable for the discretized form (10) of the Poisson equation by solving it with the scheme of [9, p. 769], that provides each nodal value u_i independently of the others; using such a scheme after indicating with C_i the right hand side of (10) yields

$$u_1 = u_0 + R\,, \qquad u_i = u_0 + i\,R - Y_{i-1}\,, \qquad i = 2\,\ldots N\,, \tag{13}$$

where

$$Z_j = h^2 \sum_{k=1}^{j} C_k\,, \quad Y_i = \sum_{j=1}^{i} Z_j\,, \quad i = 1\,\ldots\,N\,, \tag{14}$$

$$R = \frac{u_{N+1} - u_0 + Y_N}{N+1}\,. \tag{15}$$

The differences that appear in (11) after carrying out the replacements outlined above, are easily evaluated from (13–15):

$$u_i - u_{i-1} = R - Z_{i-1}\,, \quad u_{i+1} - u_{i-1} = 2\,R - (Z_{i-1} + Z_i)\,.$$

In practice, this scheme decouples (10) from (11); it may appear that this result is due to the discretization: in fact, one can also eliminate the Poisson equation prior to the discretization, by recasting it in integral form [9, p. 781–784]. The calculation of the overall number of operations also shows that the method based on (13–15) requires fewer operations than the $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{U}$ decomposition; specifically, for a matrix of order N the number of operations required by the $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{U}$ decomposition is $6\,(N-1)$ multiplications and $3\,(N-1)$ additions, whereas the method based on (13–15) requires $N-1$ multiplications and $4\,(N-1)$ additions.

Another difference between the standard discretization schemes like, e.g., finite differences, and the present one is the following: here, the discretized functions and their derivatives of all orders "belong" to the nodes; as a consequence, no hypothesis is necessary about the form of the discretized functions over each element. In the finite-difference scheme the functions and their derivatives of even order belong to the nodes, whereas the derivatives of odd order belong to the elements.

5 Stability

Observing that the equations of the model are non linear, it is necessary to implement an iterative solution. Specifically, considering for instance the equilibrium case, the normalized charge concentrations P_{i-1}, P_i, P_{i+1} at the right hand side of (10) depend on the electric potential u; the dependence is exponential in a non-degenerate semiconductor, or via a Fermi integral in a degenerate one [9]. In both cases, the derivatives $\mathrm{d}P_i/\mathrm{d}u$ are negative irrespective of the type of carriers (electrons or holes) that is considered: in fact, electrons (holes) contribute negatively (positively) to the charge density, and the electron (hole) concentration increases (decreases) when u increases. This reasoning also applies in the non-equilibrium case, because the functional dependence of the concentration on the electric potential is the same. In conclusion, linearizing (10) with respect to u adds weight to the main diagonal of the system matrix, resulting in an improved convergence rate; this behavior, typical of the linearized Poisson equation in semiconductors, is due to the dependence of the carrier concentrations on the electrostatic potential and is common to a number of discretization schemes. In contrast, as shown below, the properties of the algebraic system obtained by discretizing the continuity and transport equations with the scheme shown in this paper are quite different from those obtained with the standard scheme.

Coming now to (11), when the expressions of c_{i-1}, c_i, c_{i+1} that appear in (7) are replaced in (11), one finds an algebraic system whose right hand side, in the ith row, is made of three summands: the structure of the first summand, $A_i = -g_{i-1} + 2\,g_i - g_{i+1}$, is the same as that of Poisson's equation (10). The form of the other two terms is

$$B_i = \frac{(u'_{i-1})^2\,g_{i-1} + 10\,(u'_i)^2\,g_i + (u'_{i+1})^2\,g_{i+1}}{48/h^2}\,, \tag{16}$$

$$C_i = -\frac{P_{i-1}\, g_{i-1} + 10\, P_i\, g_i + P_{i+1}\, g_{i+1}}{24/h^2}\,. \qquad (17)$$

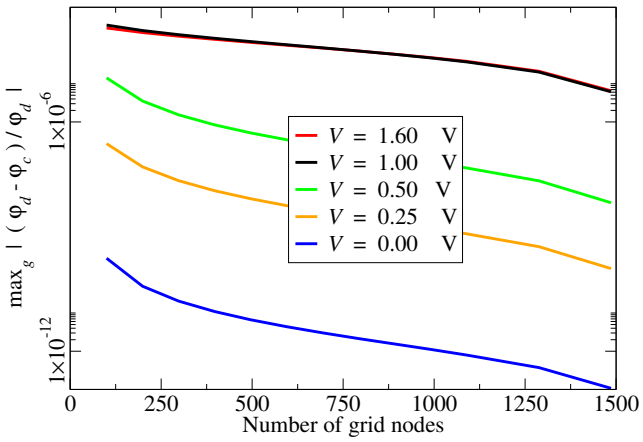


Figure 6: Maximum modulus of the relative variation $\max_g|(\varphi_d - \varphi_c)/\varphi_d|$ of the electric potential with respect to the densest grid (2000 nodes). Compare with Figures 4 and 5. The curves corresponding to $V = 1.0$ V and $V = 1.6$ V practically coincide.

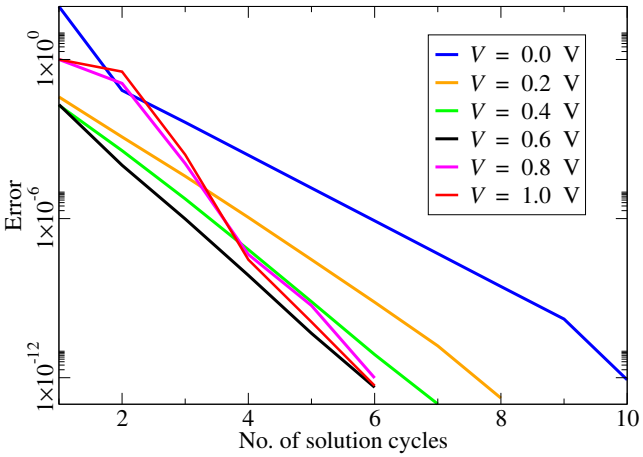


Figure 7: Number of calls to the solver that are necessary to bring the error below a prescribed value, for a given number of grid nodes. Here a 1500-node grid was used. The curves refer to the bias values shown in the inset.

The coefficients defined by (16) are non negative, therefore they add weight to all diagonals of the system matrix (11); due to factor 10, the weight added to the main diagonal is generally dominant, unless node i is near an extremum of u. As for (17), the presence in it of the normalized charge density P, which may have either sign, makes the analysis more complicated; on the other hand, (17) is of order 2 in h, whereas, due to a cancellation that takes place when (12) is replaced in (16), B_i is of the same order as A_i, namely, order 0 in h. In essence, the structure of the algebraic system (11) deriving from the present discretization scheme of the continuity and transport equations is similar to that deriving from the Poisson equation. In contrast, the standard exponential-fitting scheme for the continuity and transport equations, based on (9), approximates U as a piecewise-constant function at each node, and linearizes u between each pair of nodes; with these premises, the scheme yields

$$-K^i_{i-1}\, n_{i-1} + \left(K^{i-1}_i + K^{i+1}_i\right) n_i - K^i_{i+1}\, n_{i+1} = -h^2\, \frac{U_i}{D_n}\,, \qquad (18)$$

whose coefficients K^k_j depend on the normalized electric potential through the Bernoulli function B:

$$K^k_j = B(u_j - u_k)\,, \qquad B(\lambda) = \frac{\lambda}{\exp(\lambda) - 1}\,. \qquad (19)$$

Besides the approximations mentioned above, it is easily found that the main diagonal of the algebraic system (18) is not necessarily dominant [9, p. 781].

6 Results

As an example of application, the solution scheme based on NP is applied here to the n^+-n-n^+ structure whose lateral section and dopant profiles are shown in Figures 1 and 2, respectively. The device length is $L = 10$ μm, the concentration of the light, constant dopant (black line in Figure 2) is $N_B = 10^{15}$ cm^{-3}, while the two profiles (red lines) are given by $N_L = N_0 \exp(-x^2/x_0^2)$ and $N_R = N_0 \exp[-(L - x)^2/x_0^2]$, with $N_0 = 10^{17}$ cm^{-3} and $x_0 = 1.165$ μm, respectively. The total dopant distribution to be used in (3) is $N_D(x) = N_B + N_L + N_R$. The device is uniform in the direction normal to the field; the equations of the models have been solved at different applied voltages V, ranging from equilibrium (0 V) to 1.6 V. The form of the electric field E is shown in Figure 3 for $V = 0$ V and $V = 1$ V.

Each solution was repeated on different grids, starting with a reference grid having $N = 2000$ nodes and successively decreasing it down to $N = 100$. Figure 4 compares the electric field E calculated using the 2000-node grid with that calculated using coarser grids. As all grids are obtained by refining the 100-node one, for each node of the latter there exists a node of the denser grids having the same abscissa. Consider, e.g., the $V = 0$ curve of Figure 4, for which the 500-node grid yields about 4×10^{-9}; to obtain this value one considers two vectors, namely, the electric field E_c at each node of the 500-node grid and the electric field E_d at the corresponding abscissae of the 2000-node grid; from this vector one then extracts the maximum relative difference $\max_g|(E_d - E_c)/E_d|$, which is the strictest metric for the problem in hand. Figures 5 and 6 show the same type of comparison carried out for the electron concentration n and the electric potential φ, respectively. The worst case (about 9%) occurs for the electron concentration at the largest bias; this was to be expected, given the exponential dependence of the concentration on the electric potential.

Another part of the investigation aimed at determining the number of calls to the solver (i.e., iterations) that are necessary to bring the error below a prescribed value, given the number of grid nodes and the applied bias. The error is defined as

$$\text{Error} = \max_i |q_i^{(k+1)} - q_i^{(k)}|\,, \qquad (20)$$

where i is the node index, k is the iteration index, and q stands for φ or n. The number of calls is plotted in Figure 7 for a 1500-node grid, in Figure 8 for a 500-node grid and, finally, in Figure 9 for a 100-node grid, respectively; the bias values are the same in all cases. The curves show that the prescribed error is reached smoothly for all values of the applied bias, with no substantial difference from one grid to another; thus, the method provides a significant gain as far as the computational cost is concerned.

In [1], a different electronic device has been used to compare the error of the method presented here with that of the exponential-fitting one (the *Scharfetter-Gummel method*), in which the discretization scheme for the carrier concentration is based on (9). Like in the example above, after obtaining the reference solution over a dense grid ($N = 2000$), more solutions were calculated by making N to progressively decrease; after each solution, the maximum difference was calculated with respect to the reference solution. The improvement of the present method with respect to the standard one was found to be about one order of magnitude in all cases.

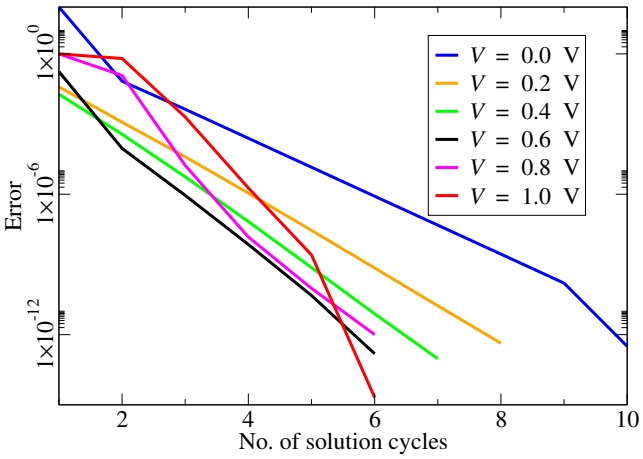


Figure 8: Same as Figure 7, using a 500-node grid.

7 Extension to Two Dimensions

It may be argued that a method for solving one-dimensional problems has a limited range of applicability. This is not so in the modeling of solid-state devices; in fact, when dealing with transport problems in structures like, e.g., carbon-nanotube transistors or nanowires, that are essentially cylindrical objects whose diameter is in the nanometer range, the restriction to the one-dimensional case is not too severe a constraint. In fact, for such devices the problem is typically separated by decoupling the longitudinal coordinate (the channel) from the transversal ones [11]; as no charge transport takes place in the directions transversal to the channel, separating the problem yields, for each transversal section, a system made of the Poisson and Schrödinger equations, which are further separated by exploiting the radial symmetry of the device. As shown in Section 1, the one-dimensional Poisson and Schrödinger equations are solvable by NP. In turn, the transport along the channel is described by equations of the form discussed here, also amenable to the NP scheme.

Clearly the NP method would gain in flexibility from the extension to a non-uniform grid. This issue is outside the scope of this paper; the interested reader may refer to a method for extending NP to a variable stepsize, still in one dimension, based on the k-step Cowell method, which has been tested on the Schrödinger equation (see [10] and references therein).

Another direction for evolving the NP scheme is the extension to more than one dimension using uniform grids of the tensor-product type. Such an extension, again applied to the case of the Schrödinger equation, is described in [12]. It is of interest to ascertain whether the approach of [12] can be extended to the class of equations investigated here. To this purpose, one must first determine the multi-dimensional form of (7); the latter is obtained as follows,

$$g'' \leftarrow \nabla^2 g\,, \qquad u' \leftarrow |\nabla u|\,, \tag{21}$$

this yielding

$$-\nabla^2 g = c\,g + s\,, \qquad c = \frac{P}{2} - \frac{|\nabla u|^2}{4}\,. \tag{22}$$

The multi-dimensional form (22) shows that the equations describing charge transport in a semiconductor can be reduced to a single second-order equation of the elliptic type.

The notation becomes rather awkward even if one limits the extension to the two-dimensional case; taking a uniform, tensor-product grid, and using the matrix notation of [12], namely,

$$a_{11}\,g_{i+1}^{j-1} + a_{12}\,g_{i+1}^{j} + a_{13}\,g_{i+1}^{j+1} + a_{21}\,g_{i}^{j-1} + \cdots \tag{23}$$

$$+\,a_{23}\,g_{i}^{j+1} + \cdots + a_{33}\,g_{i-1}^{j+1} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}_g\,,$$

in which the lower indices of $g_i^j, \ldots$ refer to the x axis, and the upper ones refer to the y axis, the two-dimensional expansion yields an interpolation of the form

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & -4 & 0 \\ 1 & 0 & 1 \end{bmatrix}_g = 2\,h^2\,(\nabla^2 g)_i^j + \tag{24}$$

$$+\,\frac{h^4}{6}\,[\nabla^2(\nabla^2 g)]_i^j + \frac{2}{3}\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}_g\,.$$

Calculating the Laplacian of both sides of (24), eliminating $[\nabla^2(\nabla^2 g)]_i^j$ after neglecting the derivatives of the 6th-order and, finally, replacing $(\nabla^2 g)_i^j$ from (22) and defining the matrices

$$\boldsymbol{M} = -\begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix}, \qquad \boldsymbol{J} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{25}$$

eventually yields the two-dimensional generalization of (11):

$$\boldsymbol{M}_g - h^2\left(6\,\boldsymbol{J}_{cg} - \frac{\boldsymbol{M}_{cg}}{12}\right) = h^2\left(6\,\boldsymbol{J}_s - \frac{\boldsymbol{M}_s}{12}\right). \tag{26}$$

The result expressed by (26) shows that the approach of [12] can indeed be extended to the second-order equation of the general form; like in the one-dimensional case, no special assumption is necessary on the form of the discretized functions inside each elements or along its edges.

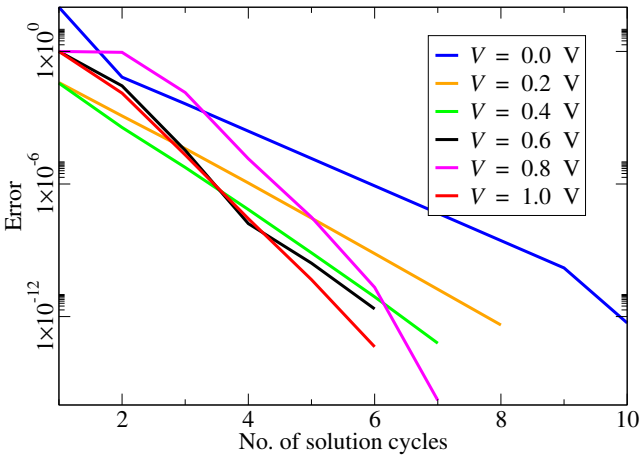


Figure 9: Same as Figures 7 and 8, using a 100-node grid.

8 Conclusions

It has been shown in this paper that the standard charge-transport model in solid-state devices, made of the continuity and transport equations for the charge, can be given a form suitable for the application of the NP scheme, using a uniform grid in one or two dimensions. With respect to the standard discretization schemes used in the analysis of semiconductor devices, the increase in the computational cost is modest; like in the case examined in [2], where the solution of the Schrödinger equation was takled, the NP approach has provided a simple tool for improving the solution of the equations.

In the field of semiconductor-device modeling, the results presented here are important from two points of view, already mentioned in Section 1: first, the NP approach, besides being able to significantly improve the solution, is applicable without excessive changes in the organization of the existing software; second, the drift-diffusion model on which NP has been tested is in fact the simplest level of description of solid-state devices; more sophisticated models exist, made of higher-order moments of the Boltzmann transport equation. On the other hand, such moments can always be grouped in pairs, each pair having the same structure: specifically, one equation of the pair is an even-order moment (order $2\,k$, with $k = 0, 1, \ldots$) whose general form, with a suitable meaning of symbols, reads [9]

$$-\operatorname{div} \boldsymbol{S} = C\,. \tag{27}$$

The next equation of the pair is the odd-order moment of order $2\,k + 1$, whose form is

$$\boldsymbol{S} = a\ \operatorname{grad}\sigma + \sigma\,\nabla b\,. \tag{28}$$

Comparing (27, 28) with (4) shows that all pairs of moments of the Boltzmann transport equation have the same structure. It follows that the method worked out in this paper applies to any order of transport models; also, considering that in the dynamic case the term C in (27) embeds the time derivative $\partial\sigma/\partial t$, the method is applicable also in the dynamic operating mode of solid-state devices.

Appendix

Letting $z_i = z(x_i)$, $z_{i+1} = z(x_i + h)$, $z_{i-1} = z(x_i - h)$, the series expansions around x_i yield

$$z_{i+1} = z_i + h\,z_i' + \frac{h^2}{2}\,z_i'' + \frac{h^3}{6}\,z_i''' + \frac{h^4}{24}\,z_i'''' + \cdots \tag{29}$$

$$z_{i-1} = z_i - h\,z_i' + \frac{h^2}{2}\,z_i'' - \frac{h^3}{6}\,z_i''' + \frac{h^4}{24}\,z_i'''' + \cdots \tag{30}$$

Adding up (29) and (30) and leaving out the terms of the sixth order or higher,

$$z_{i-1} + z_{i+1} \simeq 2\,z_i + h^2\,z_i'' + \frac{h^4}{12}\,z_i''''\,. \tag{31}$$

Multiplying (31) by $h^2/12$, taking the second derivative of both sides, and leaving out again the sixth-order term,

$$\frac{h^4}{12}\,z_i'''' \simeq \frac{h^2}{12}\left(z_{i-1}'' + z_{i+1}'' - 2\,z_i''\right). \tag{32}$$

Eliminating the fourth derivative between (31) and (32),

$$z_{i-1} + z_{i+1} = 2\,z_i + h^2\,z_i'' + \frac{h^2}{12}\left(z_{i-1}'' + z_{i+1}'' - 2\,z_i''\right), \tag{33}$$

or

$$z_{i-1} - 2\,z_i + z_{i+1} = \frac{h^2}{12}\left(z_{i-1}'' + 10\,z_i'' + z_{i+1}''\right). \tag{34}$$

Now one replaces the second derivatives in (34) using (2), to find the three-node interpolation

$$-\left(1 + \frac{h^2}{12}\,Q_{i-1}\right) z_{i-1} + \left(2 - \frac{h^2}{12}\,10\,Q_i\right) z_i - \tag{35}$$

$$-\left(1 + \frac{h^2}{12}\,Q_{i+1}\right) z_{i+1} = \frac{h^2}{12}(P_{i-1} + 10\,P_i + P_{i+1})\,.$$

The method originates from [13] (cited, e.g., in [14]). It provides a three-point interpolation where the terms of the expansion that are left out are those of order six or larger; for this reason, it is superior to the standard finite-difference interpolation, although the computational load is comparable.

The method can be extended to higher orders. For instance, leaving out of the expansion the terms of order eight or larger, and using the short-hand notation $-z'' = F(z, x)$, one finds the five-point interpolation

$$-\,z_{i-2} - 2\,z_{i-1} + 6\,z_i - 2\,z_{i+1} - z_{i+2} = \tag{36}$$

$$= \frac{h^2}{30}\,(F_{i-2} + 56\,F_{i-1} + 246\,F_i + 56\,F_{i+1} + F_{i+2})\,.$$

References

[1] N. Speciale, R. Brunetti and M. Rudan, "Extending the Numerov process to the semiconductor transport equations", in 2019 IEEE International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Udine, Italy, 2019, 1–4, https://doi.org/10.1109/SISPAD.2019.8870513.

[2] F. Buscemi, E. Piccinini, R. Brunetti, and M. Rudan, "High-order solution scheme for transport in low-D devices", in 2010 IEEE International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Yokohama, September 2014, 161–164. https://doi.org/10.1109/SISPAD.2014.6931588.

[3] International Roadmap for Devices and Systems (IRDS^{TM}), 2020 Edition. https://irds.ieee.org/

[4] C. R. A. John Chelliah and R. Swaminathan, "Current trends in changing the channel in MOSFETs by III-V semiconducting nanostructures", Nanotechnol. Rev., **6**(6), 613–623, 2017. https://doi.org/10.1515/ntrev-2017-0155.

[5] O. Muscato, "Monte Carlo evaluation of the transport coefficients in an n^+-n-n^+ silicon diode", COMPEL, **19**(3), 812–828, 2000. https://doi.org/10.1108/03321640010334613.

[6] E. Piccinini, R. Brunetti, and M. Rudan, "Self-heating phase-change memory-array demonstrator for true random number generation", IEEE Transactions on Electron Devices, **64**(5), 2185–2192, 2017. http://dx.doi.org/10.1109/TED.2017.2673867.

[7] R. Brunetti, C. Jacoboni, E. Piccinini and M. Rudan, "Band transport and localised states in modelling the electric switching of chalcogenide materials", Journal of Computational Electronics, **19**(1), 128–136 (2020). https://doi.org/10.1007/s10825-019-01415-2

[8] D. L. Scharfetter and H. Gummel, "Large-signal analysis of a silicon Read diode oscillator", IEEE Trans. Electron Dev., **ED-16**(1), 64–77, 1969. https://doi.org/10.1109/T-ED.1969.16566.

[9] M. Rudan, Physics of semiconductor devices. Springer, 2018. https://doi.org/10.1007/978-3-319-63154-7.

[10] J. Vigo-Aguiar and H. Ramos, "A variable-step Numerov method for the numerical solution of the Schrödinger equation", Journal of Mathematical Chemistry, **37**(3), 255–262, 2005. https://doi.org/10.1007/s10910-004-1467-3.

[11] M. Rudan, A. Gnudi, E. Gnani, S. Reggiani, and G. Baccarani, "Improving the accuracy of the Schrödinger-Poisson solution in CNWs and CNTs," in IEEE International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Bologna, September 2010, pp. 307–310. 10.1109/SISPAD.2010.5604497.

[12] T. Graen and H. Grubmüller, "NuSol–Numerical solver for the 3D stationary nuclear Schrödinger equation", Computer Physics Communications, **198**, 169–178, 2016. https://doi.org/10.1016/j.cpc.2015.08.023.

[13] B. Numerov. Publ. Obs. Central Astrophys., 2:188, 1933 (in Russian).

[14] P. C. Chow. Computer Solutions to the Schrödinger Equation. American J. of Physics, 40:730–734, May 1972. https://doi.org/10.1119/1.1986627

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1422-1427 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Handling Priority Data in Smart Transportation System by using Support Vector Machine Algorithm

Sara Ftaimi*, Tomader Mazri

Laboratory of Advanced Systems Engineering, Ibn Tofail science university, Kenitra, 14020, Morocco

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 04 December, 2020
Online: 21 December, 2020

Keywords:
Vehicular Adhoc network
Autonomous vehicle
Intelligent transportation system
Smart city
Big Data
Machine learning algorithms
Support Vector Machine

ABSTRACT

In an intelligent transportation system (ITS), time is a big challenge since processing a huge amount of data in a short time is very difficult, especially when the processed data is heterogeneous, consisting of a mixture of emergency data, normal data, and noise. In an ITS, an ambulance is one of the priority vehicles, and the data sent by the ambulance to the infrastructure and other vehicles must be treated first because if the ambulance does not receive a status of road from the infrastructure in time, it could take the wrong road or takes a road where there is a traffic jam has a high chance of arriving late, which could put the patient's life in danger.
Prioritizing treatment of this type of data has become paramount and vital in such cases. This paper proposes modifying the big data process to include handling the intelligent transport system's urgent data type. We will add a classification step that allows us to classify the data according to the priority degree. We use the SVM algorithm of machine learning because it has given good results concerning data classification.

1. Introduction

Nowadays, it seems that every city is trying to become a smart city. To make a city smarter, life necessities (water management, waste management, transport, and security) and technology need to be transformed into smart services. A simple definition of a smart city is an environment in which citizens can live smoothly and easily by developing their various facilities and making them smart [1].

In smart cities, sensors will be placed all around the city, and they will generate continuous data. The analysis of such data will be very helpful for the city with the potential to be smarter. As we will have many sensors installed in the city, we will have a very large amount of data generated called Big DATA, and big data is a very large amount of data and data that can come from different sources in different shapes and forms. In the intelligent transportation that will be our case study, data will be generated from sensors placed in the city and sensors placed in the car (traffic light, Lidar camera, etc.). Those sensors will generate a mass of data that will be analyzed to help the car to become autonomous on the one hand and on the other hand to make the ride a joyful experience in addition to reducing the rate of accidents on the road.

As mentioned before, the data will be generated from different sensors located in the city and the sensors placed in the vehicle. In addition to these data, we will also have another type of data circulating in the network, namely the data exchanged between the Road-Side Units and between the latter and the vehicles and between the vehicles and the network data. So the challenge will be which data should be processed first, which data should have a higher priority; we will answer these questions in the present article, which is an extended version of the work presented in [1] where we propose to make a change on big data analysis process to handle priority data.

In the first section of this paper, we will briefly review the intelligent transportation system we will define the autonomous vehicle, explore the sensors embedded in it. We will see the Vehicular Adhoc Network (VANET) and the type of communication in VANET network. In section three, we will define big data in smart transportation. We will apply the principle of classification of transportation data to see different types of data collected by sensors. In the fourth section, we will explain our contribution by explaining the case study and the different steps of the proposed process. We will also define the machine learning algorithm used. In the final section, we will present the experimental result, and we conclude our work.

2. Autonomous vehicle

An autonomous vehicle is defined as a vehicle with the capacity to detect the surrounding environment to drive without human intervention. The autonomous vehicle depends on sensors, a Cloud platform, advanced algorithms, Machine learning systems,

*Corresponding Author: Sara Ftaimi, Email: saraftaimi@gmail.com

and efficient CPUs to operate algorithms. Thanks to the sensors integrated into the vehicle as shown in figure 1, it can generate a map of its environment. The camera and the LiDar are the eyes of the vehicle. They detect traffic lights, obstacles, traffic signs; all this data is processed by an algorithm which allows reading the traffic lights and the traffic signs, to avoid obstacles and to trace the path of the vehicle, and finally to send instructions to the accelerators to accelerate or to brake depending on the existence of obstacles in the path.

2.1. Sensors

Car manufacturers have integrated several sensors into the vehicle, as shown in figure 1 so that it becomes autonomous and can move around safely by detecting obstacles and reading signal lights. For a vehicle to be able to see and understand its environment, at least eight cameras are placed in the vehicle in addition to the LiDar to detect obstacles as well as RADAR, GPS to determine the location of vehicles over and above authors sensors.

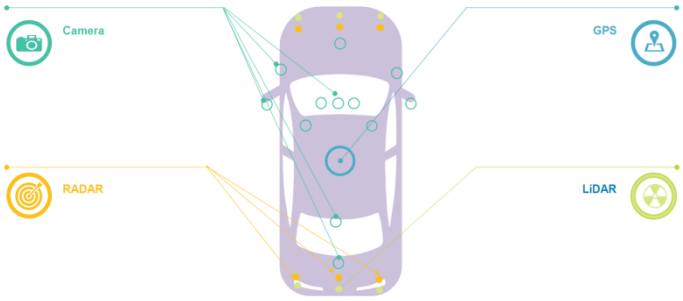


Figure 1: sensors in autonomous vehicle

2.2. VANET network

Vehicle Ad-hoc Networks (VANETs) emerged as one of the most relevant scientific research areas as they focused on supporting intelligent navigation and safe driving. It represents a specific type of MANET network that allows inter-node communication and organizes its communication system without any dependency on any other infrastructure. VANET is an intelligent building block of transport systems since vehicles interact with each other and infrastructure [2]. The high mobility of the nodes is one of the most relevant features of the VANET network vehicles are moving continuously in many directions and at varying speeds. The topology of the vanet network changes rapidly due to the fact that the vehicle can join or leave a cluster of vehicles in a very short time [3].Owing to the two above-mentioned features, combined with climate change and traffic congestion, vehicles can frequently become disconnected from the network. VANET has no energy or storage problems [4–6], unlike MANET networks, where energy constraint is a challenge. Besides having an unlimited network size since it has no geographical limits.The vehicles are referred to be communication nodes in VANET. To make communication possible between vehicles, the vehicle must be equipped with an On-Board Unit (OBU), and along the road, we must install the Roadside Unit (RSU).

The RSU is installed along the road based on the IEEE 802.11p radio technology to establish dedicated short-range communication (DSRC). It allows establishing local connectivity to the moving vehicle. It can also be used to communicate with other network devices in other infrastructure networks [7].

OBU is a device installed onboard the vehicle it allows to make a connection with the RSU, and with other OBUs via a wireless link IEEE 802.11p [8] The OBU gathers data from the sensors in the car, namely the Global Positioning System (GPS), the Event Data Recorder (EDR) and obtains the input power from the vehicle's battery [7].

2.3. Communication in Vanet

In the vanet network, Vehicles communicate with each other and with the infrastructure via three modes of communication

V2V: In V2V communication mode, a vehicle can directly communicate with other vehicles within the same range; for this mode, the architecture is decentralized. It is very efficient to transmit information on road safety services, but due to the changing speed and mobility of the nodes [7, 9] permanent connections between vehicles cannot be guaranteed.

V2I: In this mode of communication, the exchanges of information are between the vehicle and the infrastructure (RSU, antenna, satellite). The V2I mode offers a better use of the common resources,besides the connection status, which is more stable in this mode than V2V communication [7, 9, 10].

V2X means communication between vehicle and X where X can be a vehicle, a pedestrian, infrastructure, network, etc... In this diverse ecosystem, vehicles will communicate with other vehicles, with infrastructures such as traffic lights, pedestrians using smartphones, and data centers via cellular networks using short-range wireless signals [9].

3. Big data in smart transportation

In a smart city, an enormous amount of data will be generated by an intelligent transportation system (ITS). An autonomous vehicle driving around the city generates more than 25 gigabytes of data per hour. The sensors integrated into the vehicle (camera, GPS, Lidar, etc.) continuously generate information that is processed and sometimes sent to other vehicles or the infrastructure. Besides this data, there is also the data generated by the road infrastructure: RSUs (Road-Side Unit), traffic signs, traffic lights, and parking spaces.

All transport components are nowadays equipped with intelligent sensors that generate and send data continuously, e.g. a sensor is installed in a parking space and periodically sends the status of the space either it is empty or not. In addition to this data, there is also the data from the VANET network used for communication. [11, 12].

Figure2 resumes the data in a smart transportation system (ITS). Cars have to manage much larger amounts of data than ever before. Processing this data quickly and transparently is now and will continue to be one of the main challenges for car manufacturers in the future.

4. Taxonomy of data in a smart transportation system

Before classifying the data, we must know the sources of the data. In fact, in an intelligent transport system, the data generated and collected may come from a variety of sources:

- Data shared between vehicles.
- Data shared between the vehicle and the infrastructure
- The environmental data (climate)
- The data generated by the sensors placed on board the vehicles

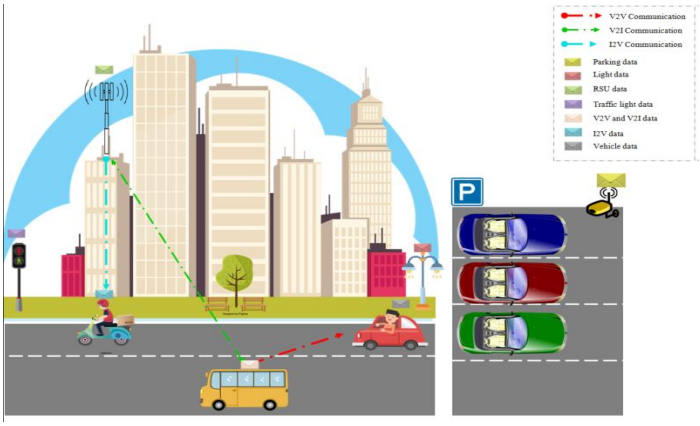


Figure 2: Communication and data type in ITS

Based on data collected from an intelligent transportation system and the taxonomy carried out in [1, 13], the data in an intelligent transportation system can be classified into three categories: the APP-level data, network-level data, and environmental data as shown in figure 3. APP-level data is the data generated by sensors installed in the ITS equipment (traffic light, etc) and the sensors integrated into the vehicles. The Network-level data is the data generated by the network that manages the communication between vehicles and road infrastructure besides the vehicles themselves (VANET network). The last type of data is environmental data.

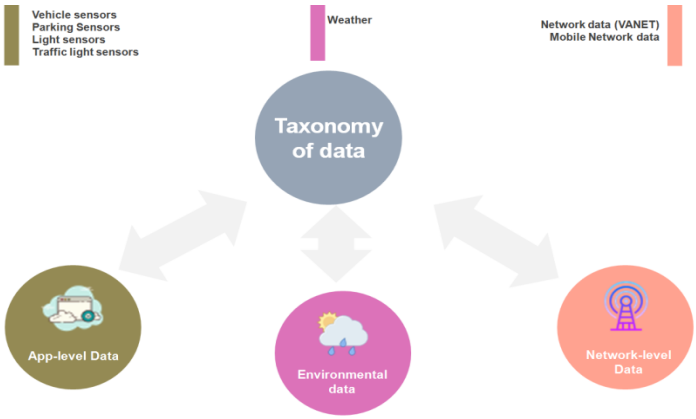


Figure 3: Taxonomy of data in ITS

5. Proposed Method

As the volume of data generated by the various intelligent transport equipment is large, processing has become vital. For a vehicle to move safely, it is necessary to collect, analyze, and transmit the information in short-time. When processing this data, the problem is to determine which data has the highest priority and needs to be processed first. To solve this problem, we propose to make a change in the Big data analysis process. Usually, the process of big data consists of the following steps: data collection, cleaning, analysis, and, finally comes the step of data interpretation[14]. However, to identify the data with the highest priority, we have added a step before the cleaning step, which will be the ordering and classifying step, and we will detail in the next sections.

5.1. Case study

In the intelligent transport system, we have several types of data that are assigned a priority. We are interested in the data shared between vehicles and RSUs, and our case study is as follows:

These RSUs will receive a large amount of data; amongst this data are the critical data that is of a higher priority than others. For instance, in the case of an accident on the road, or a fire engine trying to reach the scene of the incident, the messages sent by these vehicles have a higher priority than those sent by other vehicles, as shown in figure 4. In an intelligent transportation system, it must be able to broadcast information messages to alert road users about the accident on the road and advise them to change or turn back, as well as be able to alert the hospital through notifications, so it will send an ambulance if there is injury. The challenge here is the large number of messages received by the RSUs. The question we ask ourselves in this situation is how can we process this data and come out with the ordered data from the most critical to the least important. Figure 4 represents our case study;

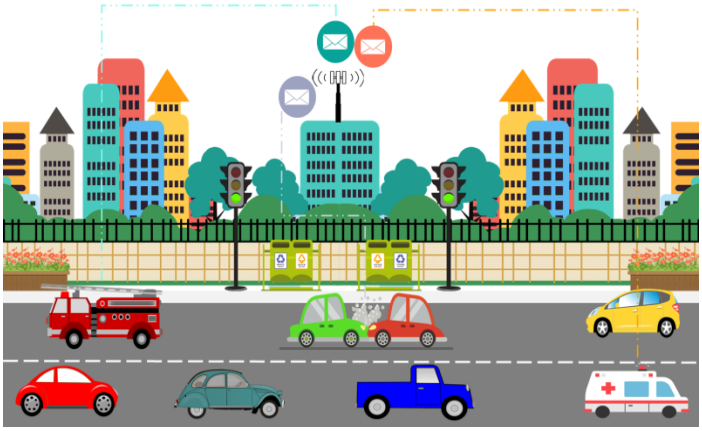


Figure 4: Case Study emergency data

We have three vehicles that have priority over the other vehicles, which are the fire engine, the accident vehicle and finally the ambulance. The messages received by these vehicles must be processed first to avoid causing traffic problems and other more serious consequences, such as loss of life.

5.2. Critical Data

To process the priority data, we must first determine the type of vehicles given priority [15]. In our case study, we have already listed three of them, and we add to them the police vehicles. Therefore, to process the priority data, we propose the following process.

First of all, it is necessary to collect the data received by the different vehicles and then apply a classification algorithm that allows us to determine the classes to which they belong. For this purpose, we have two levels of priority that are shown in figure five.

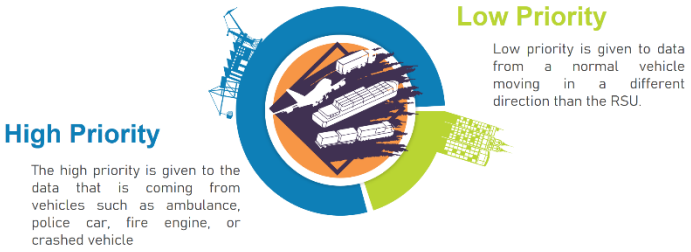


Figure 5: Levels of data priority in ITS

After the data classification step, we will have the possibility to know the data with priority, to process them first. The same process can be applied to the vehicles that will receive the data from other vehicles and the RSU data, and the data generated by the different sensors inside the vehicle. In the case of a vehicle traveling through the city at a given speed, if an accident occurs on its way, the priority must be given to the information of the change of path and the information of the change of the geolocalization map, whereas in the case of the ambulance, the alert must be given so that it passes first and requests the other vehicles in its area to clear the way.

5.3. Data packet

The communication that takes place between the infrastructure and the vehicle is communication managed by the RSU and the OBU. In an intelligent transportation system, the RSU is installed at the roadside to provide a local connection to the vehicle in motion, although it can also be used to communicate with other network devices in other infrastructure networks [7]. RSU uses short-range communication (DSRC), which is based on IEEE 802.11p radio technology. The RSU communicates to the OBU continuously. The main function of the OBU is to connect to the RSU or another OBU via the IEEE 802.11p wireless link [8]. Besides, the OBU gathers information from sensors embedded in the vehicle (e.g. Global Positioning System (GPS), Event Data Recorder (EDR)) and gets input power from the battery vehicles [16]. The packet from OBU to RSU contains the following fields OBU_ID, TIME_STAMP, POSITION, SPEED, DIR, ACC, VEHICLE_TYPE, DUTY_FLAG [15]. The table 1 represents an explanation of the fields.

Table 1: fields of the packet send between the OBU and the RSU

Fields	Explanation
OBU_ID	Contains the vehicle ID
TIME_STAMP	It is the time when the frame was sent.
POSITION	Is the position of the vehicle the value f position is gathered from the GPS sensor
SPEED	Is the speed of the vehicle
DIR	It is the direction of the vehicle.
ACC	Acceleration
VEHICLE_TYPE	Is the type of vehicle (Ambulance, fire engine, etc ...)
DUTY_FLAG	Contains information if there is an accident or the vehicle is on duty

The values of the fields above are calculated from GPS sensors that allow them to obtain directly the position, the speed, and the time. We can deduce the value of DIR from the GPS, the other values that identify the vehicle such as the id and the vehicle type are gathered from the latter.

5.4. Proposed algorithm

5.4.1. Support Vector Machine (SVM)

The SVM is a machine learning algorithm that belongs to the family of supervised learning algorithms. SVM can solve regression and classification problems besides outlier detection. The SVM algorithm can be used to classify linear and non-linear data. Therefore, it can carry out classification in a non-linear way using the kernel trick [3, 17].The SVM can be used to detect network intrusion [18] and spoofing attacks [19]. In [20], the author have used security messages to identify classes and types of attacks using SVM and KNN algorithms. In [20], the author noticed that KNN works better than SVM in some cases. When SVM has a reduction in performance, it is when the training base of both algorithms contains attacks that had extreme levels of misconduct and only a few attacks that were common errors. Also, judging the accuracy of recall, they found that SVM and KNN yield better results in some types of attacks, and SVM outperforms KNN in some attacks. In [21], the author have conducted an experiment on Support Vector Machine (SVM) and Feed Forward Neural Network (FFNN) to design intelligent IDSs that can detect two types of attacks, which are grey hole attacks and Rushing Attacks [21]. It is noted that FFNN-based IDSs could better detect malicious vehicles with lower false-negative alarm rates than SVM-based IDSs. However, the performance of SVM is much faster than FFNN. The choice of SVM algorithms was based on the comparative study conducted in [3] found that SVM is an algorithm that can handle multidimensional data and does not need much memory, and can detect outliers because the outlier detection is very important since the outliers can make false the classification model. Also, SVM can be used for both the separable and non-linearly separable data and work very well with a higher dimension. Besides, by using SVM we can replace the feature engineering, which is a very hard step for exploring data by kernelization.

5.5. Kernels

The SVM algorithm works by using a kernel technique to classify non-linear data, it maps data from the lower dimension data to the higher one and separates them with an optimal hyperplane. Therefore the kernel trick can help to transform non-linear problems into linear separable problems[17].

There are three types of kernel functions that are commonly used: linear kernel, Polynomial kernel, and Radial Basis Function Kernel (RBF). To choose which one to use, the SVM algorithm uses cross-validation to test several combinations of the kernel function and picks the one that gives the best results.

$$K(\bar{x}) = \begin{matrix} 1 & \text{if} \|\bar{x}\| \leq 1 \\ 0 & \text{otherwise} \end{matrix} \quad (1)$$

- Linear kernel: The linear kernel is commonly used to classify text and is also utilized in situations where the data is linearly separable. This means that the data may be separated with only one line. A linear kernel is generally faster to train due to the only optimization of the C regularization parameter.
- Polynomial Kernel(poly): The polynomial kernel is widely used for processing natural language (NLP). The standard degree is d = 2 (quadratic), since higher degrees over-adjust NLP problems.

- Radial Basis Function Kernel (RBF): The radial basis function kernel is the most used in SVM. This is because it has a localized and finite response along the whole x-axis. as well as being used when we have no previous knowledge of the data.

6. Experimental results

Our research results and their explanations, are presented in this section. First, we explain the methodology of our experiment and discuss the experimental result of using the SVM algorithm.

6.1. Methodology

Before starting the cleaning step, we will apply a classification algorithm to determine the type of data e.g, whether it is high-priority data or not.

The proposed process is schematically depicted in figuresix. Firstly, vehicles send every second packet, which is described in section 5.3. Each packet is read by the Machine Learning algorithm, which is the support vector machine (SVM) algorithm that we have chosen for this purpose. It has to classify data into two types see section 5.2. The classification is based on three criteria: the type of vehicle (e.g. ambulance, fire engine, police vehicle, accident vehicle, normal vehicle), the vehicle's direction, and finally, the duty flag field. If the type of vehicle is normal and the value of the duty flag mentions it as not crashed, then the algorithm will classify the data as not priority; otherwise, if the type of vehicle is an ambulance, fire engine, or police car, the algorithm must check the direction of the vehicle either it is the same as the one of the RSU, then the SVM classifies the data as a high priority. Once classified, the data can follow the standard BIG data process described in the figure six.

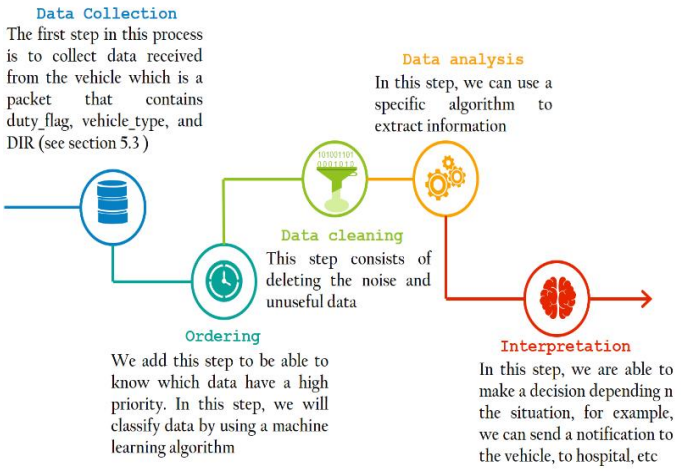


Figure 6: Proposed process

The urgent and priority data is a very challenging type of data in every domain, especially in the transportation system. If we do not treat them urgently, we can cause big damage and sometimes lose a life if the treatment was delayed. Our paper aims to handle this type of data. We propose an SVM algorithm that will read the packet sent to the vehicles or the RSUs and detect which packet has the priority to be treated first. Before starting the classification, we show in figure 7 the distribution of some vehicles based on their speed captured by the GPS sensor. We capture 100 packets sent from vehicles to RSU, and we use it to test our SVM algorithms. The SVM algorithm was applied to this database to classify the data into two types: priority and non-priority. We focused on three types of kernel poly, linear et RBF. The SVM algorithm uses two parameters to build the model, which are the gamma and the C value. In fact, the two variables, gamma, and C could change the classification model. So if the gamma value is very high, the algorithm will consider the closest points to build the limits and aims at it. We can fall into overfitting when C's value is very high and into underfitting if it is small.For this purpose, we applied a function that takes several input values for the C and Gamma parameters to find the optimal values that give the best classification results and high accuracy. The result was resumed in Table 2, where we represent C and Gamma's value that gives an accuracy that is equal to 0.90.

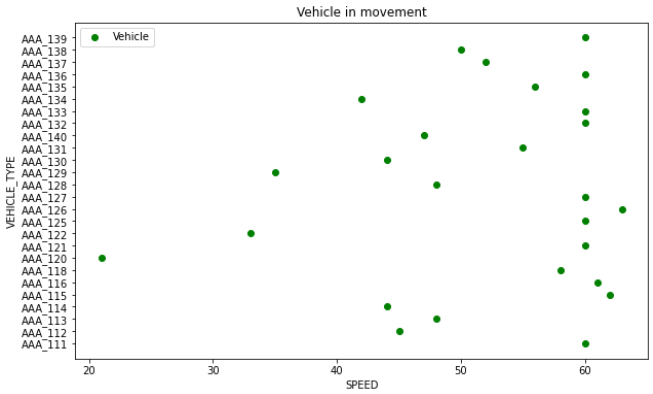


Figure 7: Distribution of vehicle based on speed

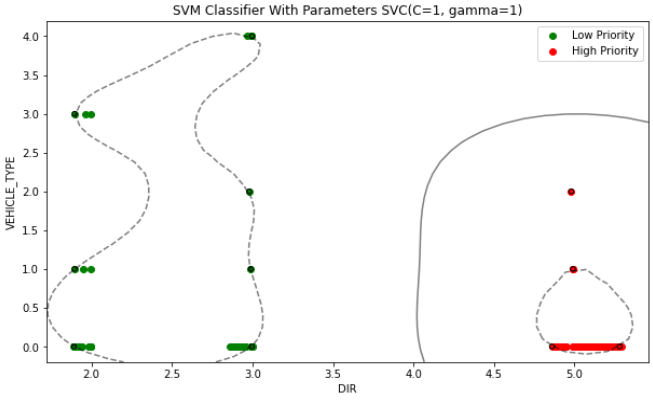


Figure 8 : SVM classifier with RBF kernel

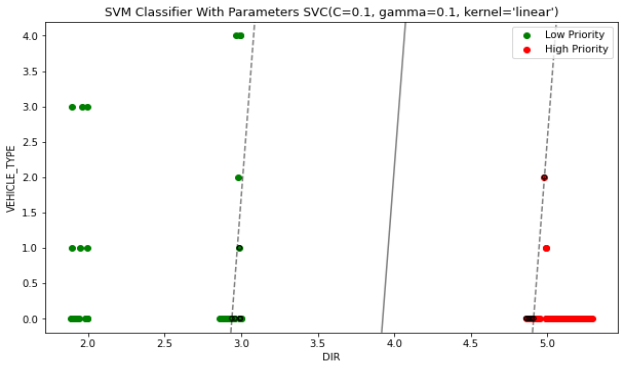


Figure 9: SVM classifier with Linear kernel

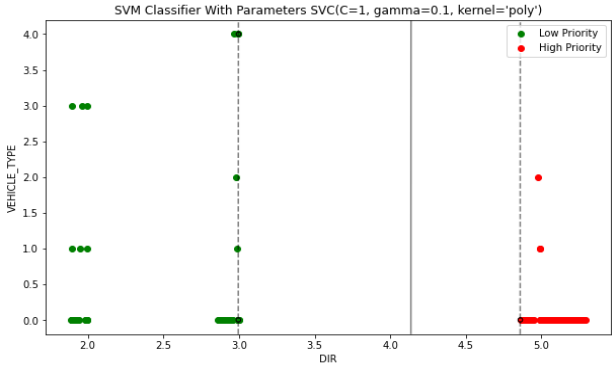


Figure 10: SVM classifier with polynomial kernel

Table 2: The best value of gamma and C

Gamma	C	accuracy	kernel
0.1	1	0.90	poly
0.1	0.1	0.9	linear
1	1	0.9	RBF

The figure 8,9 and 10 represents the classification results with each kernel. Figure 8 represents a two-dimensional classification problem. We plot the priority and the normal packet using two colors, red and green. There is a linear decision boundary for this dataset, but in figure 8, we can see how an RBF kernel can automatically decide a non-linear classification. In figure 9, we have three hyperplanes, and all of them separate the classes well. We can see that the margin between the continuous line and the data point is high compared to the dotted line. Therefore, we name the hyperplane that is represented by a continuous line as the good hyperplane. The reason for choosing the hyperplane with a higher margin is robustness, and if we choose the hyperplane with a low margin, then there is a high risk of wrong classification. The polynomial kernel is used with degree two, and he also gave a result close to the one with a linear kernel.

7. Conclusion

This paper proposes an SVM algorithm for classification problems in the context of the Vanet network. The contributions of this paper are as follows. First, we propose to add an additional step to the big data processing, which is the ordering step, and we have explained its role in identifying the data with high priority. This additional step can help to improve the cleaning step because by ordering data, we can remove the ones with low priority. It can also be employed to eliminate unused data namely the data of vehicle that is sent to the RSU and that possess a different direction from the latter. In the second part of this paper, we apply a machine learning algorithm to classify data. We select the SVM algorithm because of its high performance as a classifier, and we choose three kernels, which are the poly, linear, and the RBF. Since C and gamma's value can make a big change in the construction of the classifier, we apply a function that takes a different value of C and gamma for each kernel and return the value that gives the best accuracy. The paper aims to propose a process that can handle priority and urgent data and treat them at the right time. We propose adding a classification step that will allow us to classify data into two types: normal data and priority data. By using SVM, we achieve a good result. However, In future work, we will use the same proposed process, and we will apply the neural network to find the priority data. We will compare the SVM results with the neural network results. we chose to compare the neural network with SVM because, based on the comparative study conducted in [3], the neural network also gives good results in handling outliers, and also it is a very fast algorithm and can also handle multidimensional data and does not need much memory.

References

[1] S. Ftaimi, T. Mazeri, "The role of big data in a connected environment Case study: smart city," in 2019 International Conference of Computer Science and Renewable Energies (ICCSRE), IEEE, Agadir, Morocco: 1–7, 2019, doi:10.1109/ICCSRE.2019.8807501.

[2] J. Lloret Mauri, ed., Cognitive networks: applications and deployments, CRC Press/Taylor & Francis Group, Boca Raton, 2015.

[3] S. Ftaimi, T. Mazri, "A comparative study of Machine learning algorithms for VANET networks," in Proceedings of the 3rd International Conference on Networking, Information Systems & Security, ACM, Marrakech Morocco: 1–8, 2020, doi:10.1145/3386723.3387829.

[4] R.G. Engoulou, M. Bellaïche, S. Pierre, A. Quintero, "VANET security surveys," Computer Communications, **44**, 1–13, 2014, doi:10.1016/j.comcom.2014.02.020.

[5] D. Dinesh, M. Deshmukh, "Challenges in Vehicle Ad Hoc Network (VANET)," International Journal of Engineering Technology, **2**(7), 14, 2014.

[6] M.N. Mejri, J. Ben-Othman, M. Hamdi, "Survey on VANET security challenges and possible cryptographic solutions," Vehicular Communications, **1**(2), 53–66, 2014, doi:10.1016/j.vehcom.2014.05.001.

[7] S. Al-Sultan, M.M. Al-Doori, A.H. Al-Bayatti, H. Zedan, "A comprehensive survey on vehicular Ad Hoc network," Journal of Network and Computer Applications, 37, 380–392, 2014, doi:10.1016/j.jnca.2013.02.036.

[8] IEEE P1609.0/D5, September 2012: IEEE Draft Guide for Wireless Access in Vehicular Environments (WAVE) - Architecture., IEEE, Place of publication not identified, 2012.

[9] M.S. Sheikh, J. Liang, "A Comprehensive Survey on VANET Security Services in Traffic Management System," Wireless Communications and Mobile Computing, 2019, 1–23, 2019, doi:10.1155/2019/2423915.

[10] M.S. Sheikh, J. Liang, W. Wang, "Security and Privacy in Vehicular Ad Hoc Network and Vehicle Cloud Computing: A Survey," Wireless Communications and Mobile Computing, 2020, 1–25, 2020, doi:10.1155/2020/5129620.

[11] L. Zhu, F.R. Yu, Y. Wang, B. Ning, T. Tang, "Big Data Analytics in Intelligent Transportation Systems: A Survey," IEEE Transactions on Intelligent Transportation Systems, **20**(1), 383–398, 2019, doi:10.1109/TITS.2018.2815678.

[12] S. Kaffash, A.T. Nguyen, J. Zhu, "Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis," International Journal of Production Economics, **231**, 107868, 2021, doi:10.1016/j.ijpe.2020.107868.

[13] X. Cheng, L. Fang, L. Yang, S. Cui, Mobile Big Data, Springer International Publishing, Cham, 2018, doi:10.1007/978-3-319-96116-3.

[14] M. Chen, S. Mao, Y. Liu, "Big Data: A Survey," Mobile Networks and Applications, **19**(2), 171–209, 2014, doi:10.1007/s11036-013-0489-0.

[15] W.-H. Lee, C.-Y. Chiu, "Design and Implementation of a Smart Traffic Signal Control System for Smart City Applications," Sensors, **20**(2), 508, 2020, doi:10.3390/s20020508.

[16] S. Al-Sultan, M.M. Al-Doori, A.H. Al-Bayatti, H. Zedan, "A comprehensive survey on vehicular Ad Hoc network," Journal of Network and Computer Applications, 37, 380–392, 2014, doi:10.1016/j.jnca.2013.02.036.

[17] P. Walsh, "Support Vector Machine Learning for ECG Classification," 10.

[18] M.A. Alsheikh, S. Lin, D. Niyato, H.-P. Tan, "Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications," IEEE Communications Surveys & Tutorials, **16**(4), 1996–2018, 2014, doi:10.1109/COMST.2014.2320099.

[19] M. Ozay, I. Esnaola, F.T. Yarman Vural, S.R. Kulkarni, H.V. Poor, "Machine Learning Methods for Attack Detection in the Smart Grid," IEEE Transactions on Neural Networks and Learning Systems, **27**(8), 1773–1786, 2016, doi:10.1109/TNNLS.2015.2404803.

[20] S. So, P. Sharma, J. Petit, "Integrating Plausibility Checks and Machine Learning for Misbehavior Detection in VANET," in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, Orlando, FL: 564–571, 2018, doi:10.1109/ICMLA.2018.00091.

[21] K.M. Ali Alheeti, A. Gruebler, K.D. McDonald-Maier, "On the detection of grey hole and rushing attacks in self-driving vehicular networks," in 2015 7th Computer Science and Electronic Engineering Conference (CEEC), IEEE, Colchester: 231–236, 2015, doi:10.1109/CEEC.2015.7332730.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1428-1435 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Fabrication and Optimization of High Frequency ZnO Transducers for Both Longitudinal and Shear Emission: Application of Viscosity Measurement using Ultrasound

Hatem Dahmani[*,1], Ibrahim Zaaroura[2], Abbas Salhab[1,3], Pierre Campistron[1], Julien Carlier[1], Malika Toubal[1], Souad Harmand[2], Vincent Thomy[4], Marc Neyens[3], Bertrand Nongaillard[1]

[1]*Univ. Polytechnique Hauts-de-France, CNRS, Univ. Lille, ISEN, Centrale Lille, UMR 8520 – IEMN – Institut d'Électronique de Microélectronique et de Nanotechnologie, DOAE – Département d'Opto-Acousto-Électronique, F-59313 Valenciennes, France*

[2]*Univ. Polytechnique Hauts-de-France, CNRS, UMR 2201- LAMIH – Laboratoire d'Automatique de Mécanique et d'Informatique Industrielles et Humaines, F-59313 Valenciennes, France*

[3]*STMicroelectronics Crolles2, 850 rue Jean Monnet 38921 Crolles, France*

[4]*Univ. Lille, Centrale Lille, UVHC, ISEN, LIA LICS/LEMAC—IEMN UMR CNRS 8520, F-59000 Lille, France*

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 01 December, 2020
Online: 21 December, 2020

Keywords:
ZnO Transducer
High frequency
Shear and Longitudinal waves
Matching network
C-axis inclined ZnO
Volume and Shear viscosities

ABSTRACT

This paper covers the study of high-frequency (~ 1 GHz) ZnO piezoelectric transducer integrated on a silicon substrate able to generate both compressional and shear acoustic waves. First, to promote the longitudinal mode, an electrical matching of the transducer in this high-frequency range is effectuated. Second, to promote shear waves, new deposition conditions were applied, giving thin zinc oxide films of inclined c-axis. The RF microprobe was used to validate the transducer design and to conduct the viscosity measurements. Thus, the shear and the volume viscosity of a water droplet were measured.

1. Introduction

This paper is an extension of work originally presented in the International Conference on Design & Test of integrated micro & nano-System (DTS 2019) [1]. Piezoelectric sensors have many applications in different areas, like the medical, manufacturing, and research fields [2-5]. One of these applications is the piezoelectric ZnO transducer, which has been well known for about 50 years, mainly for its high-frequency applications. [6]. Compared to monocrystalline lithium niobate (LiNbO3) [7], ZnO has lower piezoelectric performances (coupling coefficients), but it is easier to be deposited with the classical sputtering process. The thickness of the deposited ZnO can be from 10s nm to several microns. Additionally, it is possible to lead to the right (C-axis) crystallite orientation for longitudinal acoustic wave emission with a relatively short process duration.

Prior publications have demonstrated that depending on the distance between the sample's axis and the sputtering target axis, the orientation of crystallites cannot be homogeneous over a large substrate surface [8]. The inclination of the C-axis can be modified by controlling this distance, which slightly increases the proportion of shear acoustic waves at the expense of compressional waves.

Only a few applications report using a transducer capable of generating longitudinal and shear acoustic waves simultaneously [9, 10]. The first step in this article is to present the manufacturing, the characterization and the electrical modeling of ZnO piezoelectric transducers that generate both shear and longitudinal waves in the solid. Then, an electrical matching was used in order to increase the longitudinal mode.

The deposition of layers with appropriate orientation for shear excitation has been mentioned in several articles using different methods and substrates [11-13]. Thus, to increase the shear mode, one of these methods, mainly based on the addition of a blind

[*]Corresponding Author: Hatem Dahmani, Email: hatem.issat342@gmail.com

positioned between the sputtering target and substrate, was used. The growth of ZnO films with inclined orientation makes it possible to generate shear waves.

Finally, to demonstrate the usability of this transducer**,** both longitudinal acoustic reflection and shear coefficients at the solid/liquid interface were used to characterize the viscous fluid (a sessile water droplet) properties from its shear and volume viscosity.

2. Manufacturing, characterization and modeling of ZnO transducers

2.1. The manufacturing steps of the ZnO transducers

The piezoelectric material used is ZnO. It is deposited between two metal electrodes. The deposition configuration is shown in Figure 1, and it is fabricated by a standard thin-film technology: deposition and lift-off.

The choice of ZnO as a piezoelectric material can be related to its ease of preparation and its increased scope in high-frequency applications. One can site, for example, the acoustic microscopy applications in 100s MHz to 1s GHz frequency range [14,15].

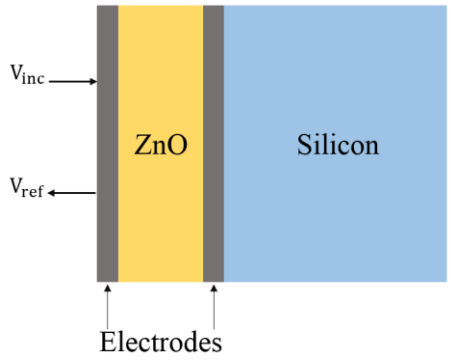


Figure 1: Multilayer Diagram of a Piezoelectric Transducer.

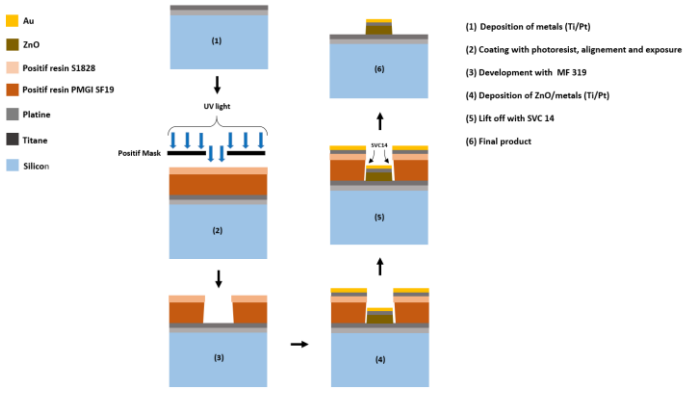
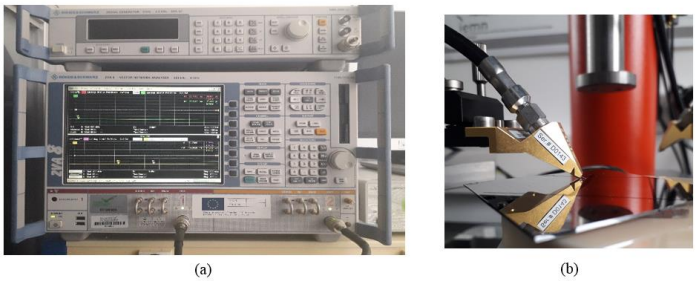


Figure 2: Piezoelectric transducer fabrication steps.

The ZnO piezoelectric transducer's fabrication starts with the deposition of 10 nm of a titanium layer by cathode sputtering on the surface of a 400 µm thick silicon wafer, followed by deposition of 80 nm of platinum. These two first layers will act as the lower electrode. A photolithography process is then performed to make the transducers pattern, where a thick 8 µm layer of PMGI SF19 (Microchem Corporation) is dispensed by spin coating on the lower electrode. Further, a layer of 1.5 µm of S1828 (Shipley Corporation) is appended. After ultraviolet exposure using a mask, a commercial developer MF 319 (Microposit) was used to obtain a pattern of different diameters in the PMGI layer. After that, a 2 µm thin ZnO layer was deposed. Finally, a 10 nm of titanium adhesion layer followed by 400 nm of gold were deposited. The gold layer is used to enable soldering of the contact electrode to a gold wire to connect the transducer to an external circuit. The last step is the lift-off process, where the UV exposed resin is removed to get the transducers pattern. The transducers have a diameter of 150 um. The microfabrication process is shown in Figure 2.

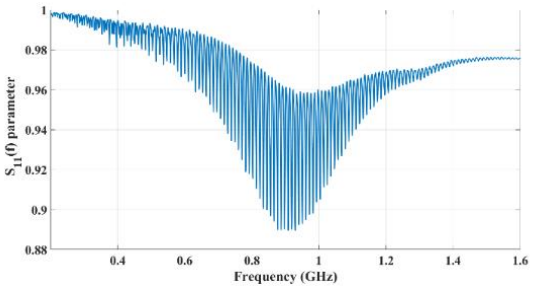
2.2. Characterization of the transducers

The ZnO transducers' characterization is carried out by a Rohde & Schwarz ZVA8 Vector Network Analyzer (VNA) as shown in Figure 3.a. The electrical contact is made using a radio-frequency measuring tip Z-probe, GS type (pitch = 1.25 mm) connected by a coaxial cable to the VNA (Figure 3.b). Since we are working within a high-frequency range in a Gigahertz order, a calibration must be done before each measurement to ensure reliable results. The silicon sample is fixed by suction (using a vacuum pump) on a circular plate.

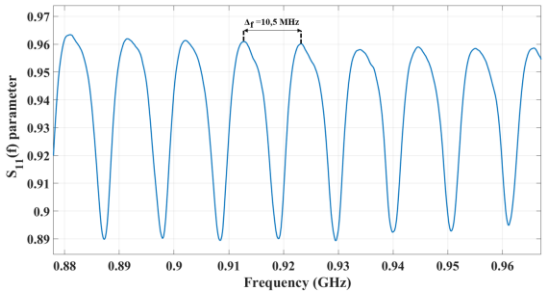


(a) (b)

Figure 3: Image of (a) Rohde & Schwarz ZVA8 Vector Network Analyzer (VNA), (b) tip Z-probe.



(a)



(b)

Figure 4: (a) Parameter S_{11} measured with the network analyzer, (b) Zoom around the longitudinal acoustic wave resonance.

The S_{11} parameter as a function of frequency, shown in Figure 4.a, is the output measured on a transducer located on the edge of the substrate. At this position, the crystal orientation of the ZnO is not perfectly vertical, thus, the generation of acoustic shear waves is clearly observed. This parameter given by Equation 1 is composed of two contributions [16,17]. The first one comes from the direct reflection of the electric wave on the transducer (S_{11}^{el}. The second is a result of all the acoustic wave reflections on the opposite side of the substrate. These acoustic reflections are then converted into electrical signal by the transducer.

$$S_{11}(f) = S_{11}^{el} + K_{11}S_{11}^{ac} \quad (1)$$

The coefficient K_{11} is the electro-acoustic-electrical coupling coefficient of the piezoelectric material used to convert energy from electric into acoustic for the emission then from acoustic into electrical for the reception.

The reflection transfer function of our system is represented by S_{11}. To obtain the temporal response, an inverse Fourier Transform is performed (Figure 5). In the time domain, the electrical term response appears shortly after the transducer's excitation while the acoustic term arises later.

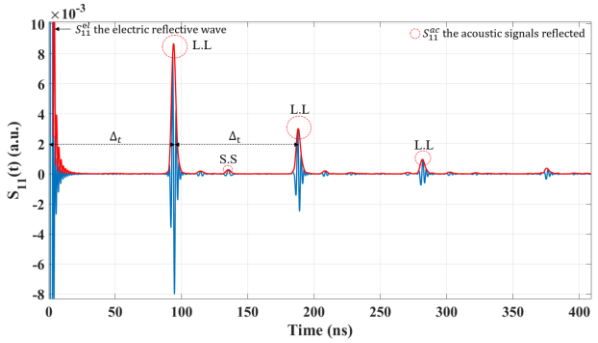


Figure 5: Impulse response $S_{11}(t)$ for f = [0.2 - 1.6 GHz].

The response shown in Figure 4 exhibits a rapid oscillation of $S_{11}(f)$ having a peak oscillation around the central frequency (f = 0.9 GHz). This is due to the fact that for each wavelength emitted in the substrate of a given thickness, the superposition of the emitted and received waves occurs with a variable phase shift. Thus, the total vibration amplitude acquired by the transducer (amplitude of the sum of the emitted and received waves) is greater for the case when the two waves (incident and reflected) are in phase and lower when they are in opposite phases (phase cancellation effect). This phenomenon is portrayed by the SWR standing wave ratio (between the highest and the lowest vibration levels). The frequency of these periodic interference ripples in $S_{11}(f)$ is $\Delta_f = c_{Si}/2e_{Si}$, where c_{Si} and e_{Si} are respectively the celerity of wave and the thickness of the silicon layer (400 µm) (Figure 4.b). The time-domain response corresponding to Figure 4.a is shown below in Figure 5. The multiple echoes are seen, the first one corresponding to the S_{11}^{el}. This electrical echo is directly due to the electrical mismatch between the transducer and the external circuit (50-Ohm load impedance of the VNA). The minor S_{11}^{el} delay corresponds to the electromagnetic wave propagation time in the cable portion between the network analyzer and the RF microprobe. The other echoes correspond to the different acoustic signals (S_{11}^{ac}) traveling back and forth in the silicon layer. They are separated by periods $\Delta_t = 1/\Delta_f = 94$ ns defined by the double Si wafer thickness's propagation time.

It is observed that the transducer emits both longitudinal and shear waves, which propagate in the silicon substrate at different speeds. The waves are denoted LL and SS for longitudinal/longitudinal or shear/shear propagation, respectively. The coupling of these 2 waves is very weak as the reflecting surface of the silicon is smooth and parallel to the transducer.

2.3. Transducer design using Auld's model

For checking the manufacturing quality of the transducers, we can model their impedance Zel and compare it to experiments. The piezoelectric layers being very thin compared to their lateral dimension (2 µm thick vs. 150 µm diameter), we can model them by a 1D model, like the Auld model [18].

As longitudinal and shear waves are emitted, we will use a modified Auld model that includes the two modes of emission. For that, it is necessary to consider the different mechanical impedances Z of the transducer's constituting layers. The analytical formulation for Zel of a transducer rigidly bounded to a semi-infinite solid is given by Equation 2; the mechanical impedances $Z^S_{(Air/Au/Ti)}$ and $Z^L_{(Air/Au/Ti)}$ in Equations 3 and 4 are hardly significant or negligible relatively to the other terms [16].

$$Z_{el} = \frac{1}{jC_0w} + \frac{K_S^2}{C_0w}A + \frac{K_L^2}{C_0w}B \quad (2)$$

where,

$$A = \frac{Z^S_{ZnO}}{\theta_S}\frac{2Z^S_{ZnO}[\cos(\theta_S)-1]+j\left[Z^S_{(Air/Au/Ti)}+Z^S_{(Si/Ti/Pt)}\right]\sin(\theta_S)}{Z^S_{ZnO}\left[Z^S_{(Air/Au/Ti)}+Z^S_{(Si/Ti/Pt)}\right]\cos(\theta_S)+j\left[(Z^S_{ZnO})^2+Z^S_{Air/Au/Ti)}.Z^S_{(Si/Ti/Pt)}\right]\sin(\theta_S)}$$

and

$$B = \frac{Z^L_{ZnO}}{\theta_L}\frac{2Z^L_{ZnO}[\cos(\theta_L)-1]+j\left[Z^L_{(Air/Au/Ti)}+Z^L_{(Si/Ti/Pt)}\right]\sin(\theta_L)}{Z^L_{ZnO}\left[Z^L_{(Air/Au/Ti)}+Z^L_{(Si/Ti/Pt)}\right]\cos(\theta_L)+j\left[(Z^L_{ZnO})^2+Z^L_{(Air/Au/Ti)}.Z^L_{(Si/Ti/Pt)}\right]\sin(\theta_L)}$$

L and S indices stand to the longitudinal and shear waves. $C_0 = \varepsilon_{ZnO}.S_{ZnO}/e_{ZnO}$ is the static capacitance of the ZnO transducer, ε_{ZnO} the dielectric permittivity of ZnO, e_{ZnO} and SZnO are the thickness and the active surface of the ZnO layer. K^2 is the square of the electromechanical coupling coefficient of the ZnO. $\theta = k_{ZnO}.e_{ZnO} = 2\pi f\,(e_{ZnO}/c_{ZnO})$ with k_{ZnO} the wave number in ZnO and c_{ZnO} the longitudinal or transverse wave's velocity the ZnO.

From an experimental point of view, the S-parameter is linked to the transducer impedance as follows:

$$S_{11} = \frac{Z_{el}-Z_0}{Z_{el}+Z_0} \quad (3)$$

Z_0 is the characteristic impedance of the network analyzer equal to 50 Ω.

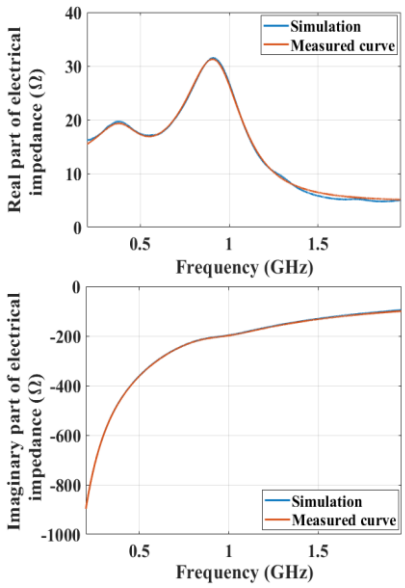


Figure 7: Simulated (blue curve) and measured (red curve) electrical impedance Z_{el} of the 150 µm diameter ZnO transducer.

The 150 µm diameter ZnO transducer's simulation results are shown in Figure 7 by the solid blue curves. The real part of Z_{el} (Figure 7.a) exhibits two peaks occurring at the resonance frequencies of the transverse (f_S= 0.38 GHz) and longitudinal (f_L= 0.9 GHz) waves.

A good fit between the simulation and the experimentation can be seen in Figure 7. So, the Auld model can be used to design a transducer for which the real part is close to 50 Ω by adjusting the thicknesses of the thin layers and the surface.

The manufacturing process is well calibrated which allows us to know the thicknesses of the different materials quite precisely. We were able, during various works, to measure the acoustic properties of ours deposited layers [17]. We therefore know the parameters on which the model depends, except the electro-mechanical coupling coefficients of the transducers which depend on the unknow orientation of the ZnO crystallites. We get them by an optimization routine using Nelder-Mead's algorithm. In the simulation, K_s and K_L were set to 0.10 and 0.23, respectively. The result of optimization for K_L is close to the theoretical one obtained for a perfect crystallite c-axis alignment (K_L = 0.27, according to [7]).

As shown in Figure 8, the pure electrical model we use represents the 2 modes of vibration of the transducer by the association of 2 RLC circuits shunted by the static capacitance C_0 . From a physical point of view, the capacity C_0 reflects the resonators electromagnetic behavior, while the two parallel resonant circuits exhibit the resonators acoustic behavior.

The electrical parameters of the circuit are determined by using the electrical impedance thanks to the measurement of the S_{11} parameter. As for the evaluation of the coupling coefficients, we obtained the electrical parameters by an optimization procedure.

The electrical parameters of the circuit are determined by using the S_{11} parameter:

$C_0 = 0.82$ pF, $L_L = 641$ nH, $C_L = 0.047$ pF, $R_L = 1645$ Ω, $L_S = 11.87$ µH, $C_S = 0.012$ pF, $R_S = 24$ kΩ.

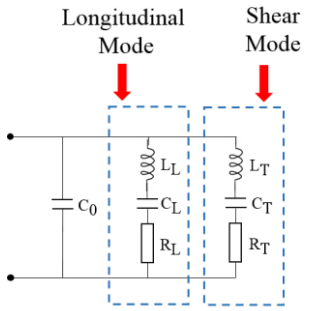


Figure 8: The equivalent electric circuit of the ZnO transducer.

3. Transducer optimization

3.1. Increasing the longitudinal acoustic waves emission by impedance matching

In any kind of application, we need to favor either one or both types of the generated waves. In the present study, regarding the matching network, promoting the longitudinal waves is sought [19].

The concept of electrical matching is to transfer the maximum power from the source to the transducer or, in other words, to eliminate electrical reflections on the transducer in a chosen frequency range. As piezoelectricity is a reciprocal phenomenon, the electrical matching network also acts as an acoustical matching layer. So, the electrical matching reduces the reflection of the acoustical waves returning on the piezoelectric layer, which increases the acoustical transmission coefficient at the piezoelectric device. Simultaneously, levels of both acoustic waves propagating in the device and electrical measurement signal increase.

There are two steps for the electrical impedance matching. The first one consists of connecting the transducer with a gold wire (25 µm of diameter). The second step uses an L-section impedance matching [20], consisting mainly of two reactive elements (capacitor and inductor). The capacity is used to adjust the real part of the impedance to 50 Ω, while the coil reduces the imaginary part. A capacitance has a significant loss effect at high-frequency, causing critical insertion losses [21]. For this reason, a chip inductor with a value of 8.6 nH was put in series with the load (transducer connected by the gold wire) at $f_{L,match}$ = 1.12 GHz. The real part of the load at this frequency is equal to 50 Ω.

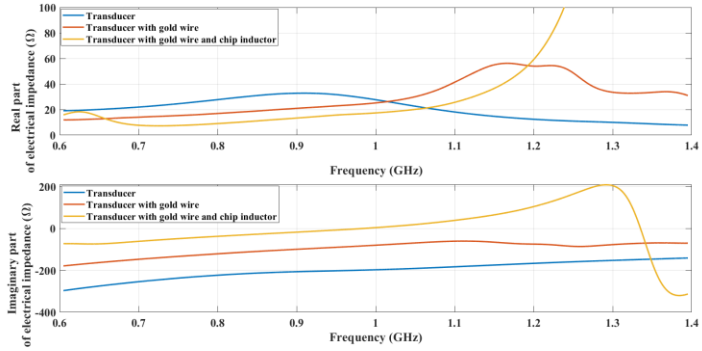


Figure 9: The real and imaginary parts of the electrical impedance of the ZnO transducer (blue curve), the transducer with gold wire (red curve) and transducer with gold wire and chip inductor (yellow curve).

It is clearly seen that the electrical response has been modified (Figure 9). This modification can be related to the losses and the parasite impedances of bonding wire and packaging. So, it has become complicated to choose a filter based on the equivalent electric circuit of the transducer, such as the Butterworth or Chebychev filters [22, 23, 24]. Therefore, the simplest solution is to use the transducer impedance's real and imaginary parts, including the wire.

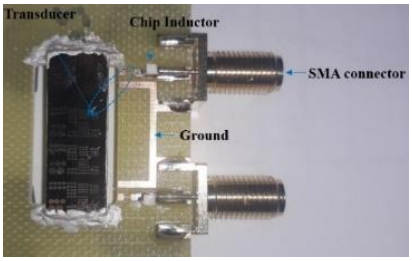


Figure 10: Image of the different elements of the circuit.

The expected result is the gain in longitudinal wave amplitude, which will make it easy to connect the transducer to the VNA or a generator via an SMA connector, as shown in Figure 10.

To verify that the electrical adaptation works as expected, we compare the insertion loss of the transducer before and after the adaptation, through an SMA connector. Afterwards, the responses in terms of received amplitude level will be compared. The electrical insertion losses L_e was calculated by the following equation:

$$L_e = -10 \log \left(1-\left|S_{11}^{el}\right|^2\right) \quad (4)$$

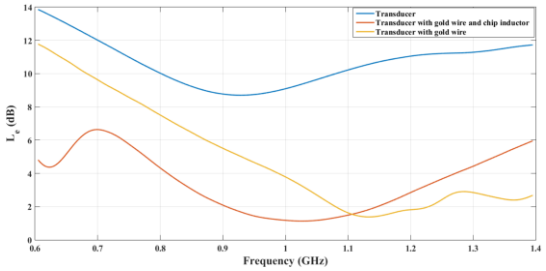


Figure 11: Measured electrical insertion losses.

As shown in Figure 11, the electrical adaptation has reduced the insertion losses. The matched transducer exhibits an amplitude gain of 8 dB and a bandwidth enlarged to 250 MHz.

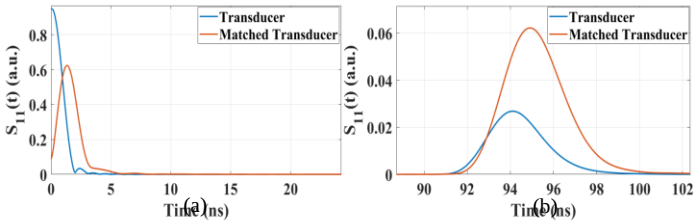


Figure 12: Impulse response $S_{11}(t)$, (a) the electrical reflection S_{11}^{el}, (b) the acoustic response S_{11}^{ac}.

The matching network's direct effect can be represented in Figure 12, showing the echoes before (blue curve) and after (red curve) the electrical matching. The attenuation of the reflected electrical signal is notably due to the conversion from electrical into acoustic energy. This conversion can be seen by the simultaneous amplitude decrease of the electrical echo S_{11}^{el} (Figure 12.a) and the increase of the first acoustical echo S_{11}^{ac} (Figure 12.b). However, this result only applies to $f_{L,match}$.

Although the electrical adaptation improves the transducer's acoustic sensitivity, the electrical reflection does not completely disappear because the use of an inductor is only at a single frequency $f_{L,match}$ and not over the bandwidth of the transducer.

3.2. Increasing the shear acoustic waves emission by deposition of ZnO inclined c-axis by R.F. magnetron sputtering

It appears difficult to use a matching network for increasing the level of shear waves as they are too weak. Previous measurements have been made with a transducer from a misaligned area (with respect to the ZnO target of the silicon wafer). By increasing that effect, it is possible to increase the shear level.

The ZnO thin film fabrication process was modified to balance the amplitude of generated longitudinal and shear waves by changing the material crystalline orientation. The film was deposited on three wafers with the following stack: Pt (80 nm) / Ti (10 nm) / Si (400 µm) by R.F. magnetron sputtering technique. The substrate was fixed in the chamber, whose total pressure was 0.02 mBar. A rectangular cover (the blind) is positioned between the target and the substrate, making it possible to obtain oblique incidence of the particles in order to locally influence the growth of the ZnO thin film (Figure 13). The blind was made of aluminum, had a height of 30 mm and a thickness of 1 mm. It was placed in the center of the substrate. The deposition rate of the ZnO films varied from 15 to 25 nm/min.

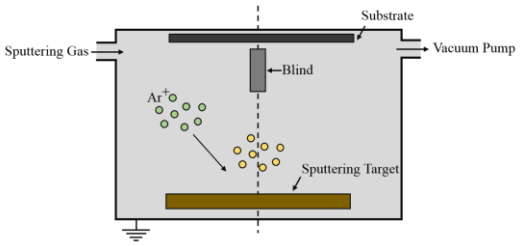


Figure 13: R.f. magnetron sputtering system with additional blind between substrate and target.

A cross-section SEM image of the ZnO deposited on the silicon substrate is shown in Figure 14. The inclination of the crystallite orientation is approximately 14.5° to the surface normal can be observed. So, shear and longitudinal waves can be excited. However, the inclination of the crystal orientation is strongly dependent on the distance between the blind and the transducer (in the substrate plane); therefore, the c-axis inclination will be higher in an area close to the blind or close to the edge of the substrate.

Figure 15.a shows the resonance diagram of the piezoelectric transducer; two distinguished downward peaks can be seen, the first one in the frequency range [0.35 - 1 GHz] which corresponds to critical transversal waves response compared to Figure 4 and the other peak in the range [1 - 2 GHz] corresponding to longitudinal

waves response. Two oscillations with different frequencies Δf = {S: 7.4 MHz, L: 10.5 MHz}, are presented in Figure 15.b. Each of them presents a different wave mode. This result directly reflects the transducers' ability sputtered with the inclined ZnO c-axis to generate both shear and longitudinal waves.

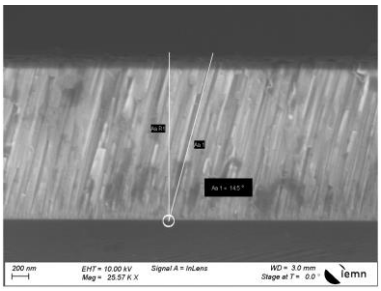
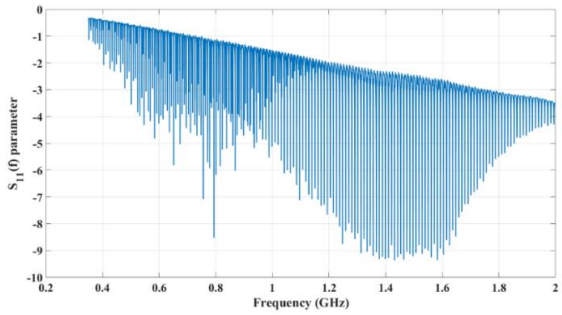
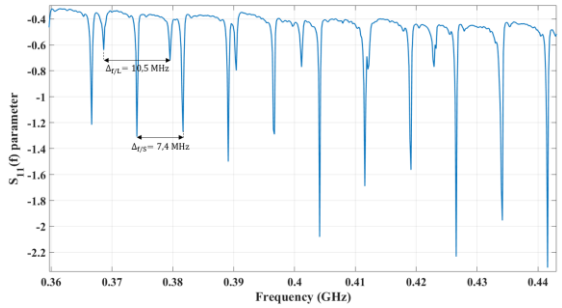


Figure 14: Cross-sectional SEM image of ZnO film with c-axis inclined of 14.5°.



(a)



(b)

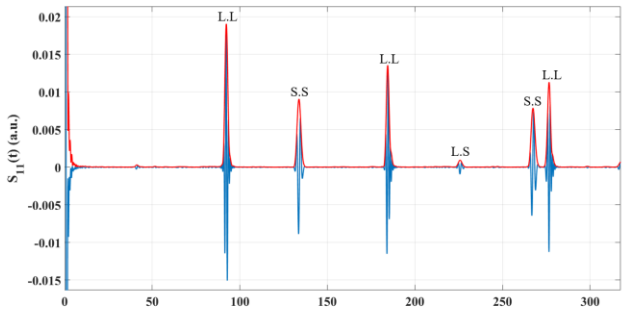
Figure 15: Impulse response $S_{11}(f)$, (a) for f=[0.35 - 2 GHz], (b) Zoom around the longitudinal and the shear acoustic waves resonance.

Figure 15 (b), as we expect each of them are close to the theoretical ones {Cs/2e = 5843/800e-6 = 7,3 MHz and Cl/2e = 8432/800e-6 = 10,54 MHz}.

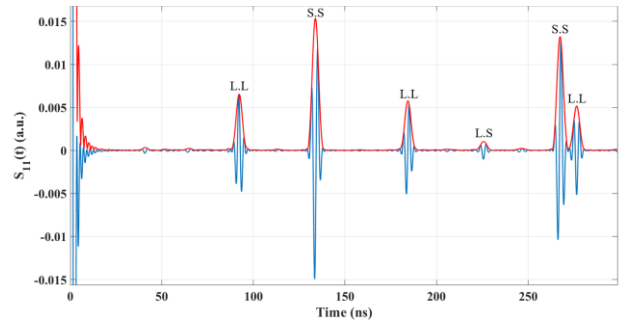
The time-domain representation is used below to check the level of both signals. Figure 16 shows the inverse Fourier transform of the S11 parameter. For [0.35 - 2 GHz], two different acoustic reflections LL, and SS, are the responses from the generated acoustic waves, with LL being a purely longitudinal response with longitudinal wave emitted and the longitudinal wave reflected. The longitudinal speed of sound in silicon (8432 m/s) is greater than the transversal one (5832 m/s); thus, the time of reception of the signal for the longitudinal waves should be the fastest (the first reflection on Figure. 16.a corresponds to LL). The second impulse SS is the purely transversal reflected signal with the time delay being significantly greater than LL. For a bandwidth of 1.65 GHz, the ratio of the shear wave to the longitudinal wave is 0.5.

We decreased the frequency range to [0.35 - 1 GHz] (Figure 16.b), the shear to shear (SS) reflection amplitude increases, and the longitudinal to longitudinal (LL) amplitude decreases, which what was expected to be seen since the lower frequency range favors the transversal waves (Figure 16.a). The ratio of shear to longitudinal amplitudes increases to 2.6.

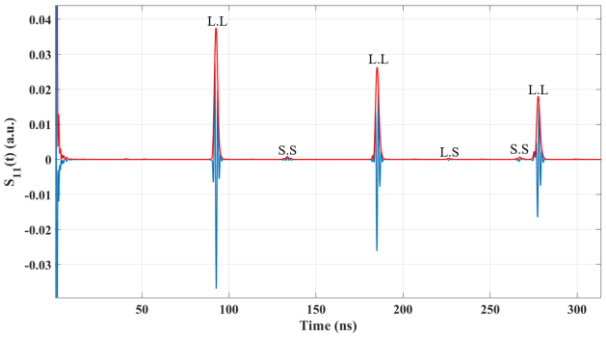
For f = [1 - 2 GHz] (Figure 16.c), the result is the profound decrease in the SS reflection amplitudes and the profound increase in the LL amplitudes, which is what to expect since the high frequencies favor the longitudinal waves. We now have a ratio of shear to longitudinal amplitudes decreased to 0.01.



(a)



(b)



(c)

Figure 16: Impulse response $S_{11}(t)$, the time delays $\Delta_{t/L} = 1/\Delta_{f/L} = 95$ ns and $\Delta_{t/S} = 1/\Delta_{f/S} = 135$ ns between them corresponds to the propagation delays of the longitudinal and transverse wave travelling twice the thickness of the wafer. (a) f = [0.35-2 GHz], (b) f = [0.35-1 GHz] and (c) f = [1-2 GHz].

Levels are weak, but all signals are out of the noise and can be used for measurements. If necessary, they can be increased using a convenient matching network, as demonstrated in part 1. Thus, we propose a method allowing the emission of both longitudinal and shear waves using only one fixed fabrication process.

However, the c-axis tilt is not the same on the wafer surface, so the transducers cannot generate the same wave level for each frequency range, so the ratio of shear to longitudinal amplitudes changes.

4. Application: Measurement of the viscosity of a liquid using both longitudinal and shear waves generated by the ZnO transducer

There are a vast number of applications that utilize physical parameters in relation to acoustic waves-material interactions. An application to measuring the fluid viscosity is presented below to demonstrate the usability of the proposed transducer design. Water is used as the reference fluid whose acoustic properties are given in [25, 26].

The total viscosity of water can be extracted from the shear and longitudinal waves generated at the liquid/ solid interface [27,28]. Our transducers can generate these two types of waves. We used the evaporation of 1 µL droplet of water to determine the two kinds of viscosities (namely shear and volume viscosities). The dynamic shear viscosity was extracted using the shear wave, while the volume viscosity from both waves.

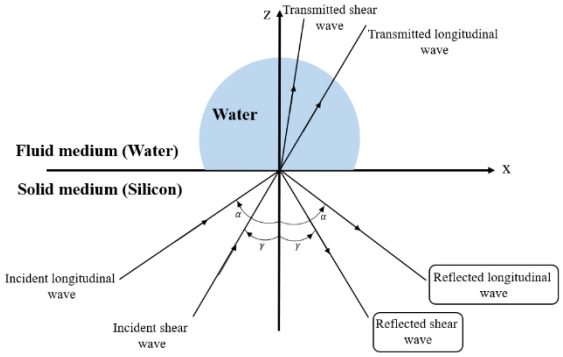


Figure 17: Incident acoustic waves interaction at droplet/silicon interface.

Figure 17 presents the direct contact of the acoustic waves to the water droplet/silicon interfaces. The acoustic waves reflect at the interface and generate two components, i.e., longitudinal and shear reflected waves.

Several mechanical properties at the silicon/water interface influence the measurement, and these are also strongly dependent on the temperature. The reflection coefficient at the solid/fluid interface is given by equation 5, Z_0 is the mechanical impedance of the silicon, considered purely elastic, so Z_0 is purely real [29], while water presents viscoelastic properties. As a consequence, its acoustic impedance Z_1^* is in a complex form.

$$r^* = \frac{Z_1^* - Z_0}{Z_1^* + Z_0} \tag{5}$$

where $Z = \rho C$, ρ is the density of the medium, and C is the acoustic velocity of the wave propagating inside of the medium. The reflection coefficients are determined for both longitudinal and shear reflected waves. So, the complex form of the shear reflection coefficient ($r_{SS}{}^*$) and the longitudinal reflection coefficient ($r_{LL}{}^*$) are shown in the Equations 6 and 7. These coefficients give access to the modulus and the phase angle of the two reflections acoustic wave at the interface.

$$r_{SS}^* = \frac{Z_w^{S*} - Z_{si}^S}{Z_w^{S*} + Z_{si}^S} = \frac{\rho_w C_S^* - Z_{si}^S}{\rho_w C_S^* + Z_{si}^S} \tag{6}$$

where,

$$C_S^* = \omega/k_S^* \quad \text{and} \quad k_S^* = \omega/C_S - i\alpha_S = \omega/C_S - i\sqrt{\frac{\rho_w \omega}{2\eta}}$$

$$r_{LL}^* = \frac{Z_w^{L*} - Z_{si}^L}{Z_w^{L*} + Z_{si}^L} = \frac{\rho_w C_L^* - Z_{si}^L}{\rho_w C_L^* + Z_{si}^L} \tag{7}$$

where,

$$C_L^* = \omega/k_L^* \text{ and } k_L^* = \omega/C_L - i\frac{\omega^2}{2\rho_w C_L^3}\left[\mu + \frac{4}{3}\eta\right]$$

where C_L^*, C_S^* are the complex longitudinal and shear velocity of water, ρ_w is the density, k_L^*, k_S^* the complex longitudinal and shear wave number, ω (rad/s) the angular velocity. η and μ are the shear and volume viscosity, respectively.

C_L and C_S are the effective longitudinal and shear phase velocities in water respectively, α_S is the shear attenuation. The kinetics of the shear and volume viscosity is shown in Figure 18. A sessile droplet of water evaporating at room temperature of 25°C is used. It is clearly seen that the consistency of the theoretical and experimental results.

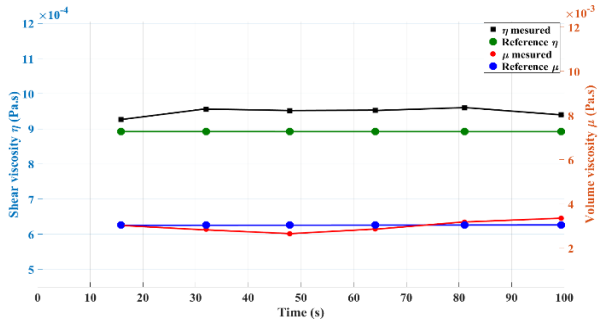


Figure 18: Kinetics of the shear and volume viscosity of a water droplet at T=25°C compared to the reference values.

It was essential to improve the shear waves emission to accurately measure the shear viscosity values for different complex fluids such as nanofluids. These nanofluids contain nanoparticles in suspension in a matrix fluid; that is why it is important to have a high sensitivity of these particles that settled above the surface.

5. Conclusion

This article is the extended version of a paper published in IEEE Xplore. We have studied the fabrication and the optimization of a ZnO high-frequency transducer able to generate both longitudinal and shear waves. To model the transducer's electrical response, we used the Auld model; we noticed a good agreement between electrical impedance measurements and simulations, allowing its use in the design of transducers. To

improve the sensitivity, we have used an electrical circuit including a chip inductor and an SMA connector so that it could be used without the requirements of an external RF microprobe. This matching circuit had been experimented with for the longitudinal mode and allowed us to gain ~ 8 dB on the electrical insertion losses.

In order to increase the shear wave level, the growth of ZnO films with tilting of the crystallite axis has been achieved using a blind of 10 mm that was positioned between the target and the substrate during the R.F. magnetron sputtering. The inclination of the crystal orientation was approximately 14.5°.

Both shear and longitudinal waves can be excited with this multi-mode transducer; the wave type can be selected by adjusting the emission frequency.

As to proof of the usability of this kind of measurement, an application has been proposed. We have exploited the simultaneous measurements of the two types of waves to extract both shear and volume viscosities of the water. Comparisons between the experimental and theoretical results show a good agreement.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was partially supported by the French Renatech network. A part of the cost has also been supported by STMicroelectronics and Hauts-de-France region. The authors wish to thank all the members of the cleanroom at IEMN- UMR CNRS 8520. They are also grateful to the professor Bertrand Nongaillard for his encouragement and support in this work. Finally, the authors also thank the laboratory engineers, Dr. Malika Toubal, and Dr. Nikolay Smagin, for their help in preparing the article.

References

[1] H. Dahmani, I. Zaaroura, P. Campistron, J. Carlier, M. Toubal, B. Nongaillard, "Study of a dual longitudinal and shear acoustic wave ZnO transducer in the gigahertz frequency range: Application to viscosity measurement and first step for device integration," in IEEE International Conference on Design and Test of Integrated Micro and Nano-Systems, DTS 2019, 2019, doi:10.1109/DTSS.2019.8914906.

[2] F.L. Di Scalea, H. Matt, I. Bartoli, A. Srivastava, G. Park, C. Farrar, "The fundamental response of piezoelectric guided-wave sensors and applications to damage and impact location," in Proceedings of the 3rd European Workshop - Structural Health Monitoring 2006, 1271-1280, 2006.

[3] C. Caliendo, "Longitudinal modes along thin piezoelectric waveguides for liquid sensing applications," Sensors (Switzerland), **15**(6), 12841-12856, 2015, doi:10.3390/s150612841.

[4] L.P. Suwal, R. Kuwano, "Disk shaped piezo-ceramic transducer for P and S wave measurement in a laboratory soil specimen," Soils and Foundations, **53**(4), 510-524, 2013, doi:10.1016/j.sandf.2013.06.004.

[5] F.A. Alexandre, P.R. Aguiar, R. Götz, M.A.A. Viera, T.G. Lopes, E.C. Bianchi, "A novel ultrasound technique based on piezoelectric diaphragms applied to material removal monitoring in the grinding process," Sensors (Switzerland), **19**(18), 2019, doi:10.3390/s19183932.

[6] J. de Klerk, "Thin film zinc oxide transducers for use in microwave devices," Ultrasonics, **8**(3), 159-164, 1970, doi:10.1016/0041-624X(70)90614-1.

[7] D. Royer, E. Dieulesaint, Elastic Waves in Solids I: Free and Guided Propagation, translated by D. P. Morgan, 2000.

[8] M. Link, Study and realization of shear wave mode solidly mounted film bulk acoustic resonators (FBAR) made of c-axis inclined zinc oxide (ZnO) thin films: application as gravimetric sensors in liquid environments, Ph. D Thesis, Henri Poincaré University, 2006.

[9] I. Zaaroura, M. Toubal, H. Reda, J. Carlier, S. Harmand, R. Boukherroub, A. Fasquelle, B. Nongaillard, "Evaporation of nanofluid sessile drops: Infrared and acoustic methods to track the dynamic deposition of copper oxide nanoparticles," International Journal of Heat and Mass Transfer, **127**, 1168-1177, 2018, doi:10.1016/j.ijheatmasstransfer.2018.07.102.

[10] C. Virgilio, Caractérisation du mouillage de surfaces micro / nanostructurées par méthode acoustique haute fréquence : application aux traitements humides dans l'industrie de la microélectronique Ph. D Thesis, Valenciennes and Hainaut-Cambresis University, 2017.

[11] M. Link, M. Schreiter, J. Weber, R. Primig, D. Pitzer, R. Gabl, "Solidly mounted ZnO shear mode film bulk acoustic resonators for sensing applications in liquids," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, **53**(2), 492-496, 2006, doi:10.1109/TUFFC.2006.1593389.

[12] S. Bensmaine, L. Le Brizoual, O. Elmazria, J.J. Fundenberger, B. Benyoucef, "Deposition of ZnO inclined c-axis on silicon and diamond by r.f. magnetron sputtering," in Physica Status Solidi (A) Applications and Materials Science, **204**(9), 3091-3095, 2007, doi:10.1002/pssa.200776322.

[13] M. Link, M. Schreiter, J. Weber, D. Pitzer, R. Primig, M.B. Assouar, O. Elmazria, "C-axis inclined ZnO films deposited by reactive sputtering using an additional blind for shear BAW devices," in Proceedings - IEEE Ultrasonics Symposium, **1**, 202-205, 2005, doi:10.1109/ULTSYM.2005.1602831.

[14] K. Yamada, H. Shimizu, M. Minakata, "PLANAR-STRUCTURE FOCUSING LENS FOR OPERATION AT 200 MHZ AND ITS APPLICATION TO THE REFLECTION-MODE ACOUSTIC MICROSCOPE.," in Ultrasonics Symposium Proceedings, 745-748, 1986, doi:10.1109/ultsym.1986.198834.

[15] B.T. Khuri-Yakub, "Scanning acoustic microscopy," Ultrasonics, **31**(5), 361-372, 1993, doi:10.1016/0041-624X(93)90070-G.

[16] Y. Deblock, P. Campistron, M. Lippert, C. Bruneel, "Electrical characterization of plate piezoelectric transducers bonded to a finite substrate," The Journal of the Acoustical Society of America, **111**(6), 2681-2685, 2002, doi:10.1121/1.1416904.

[17] P. Campistron, J. Carlier, N. Saad, J. Gao, M. Toubal, L. Dupont, G. Nassar, B. Nongaillard, "High frequency ultrasound, a tool for elastic properties measurement of thin films fabricated on silicon," in Advanced Materials Research, **324**, 277-281, 2011, doi:10.4028/www.scientific.net/AMR.324.277.

[18] B.A. Auld, R.E. Green, Acoustic Fields and Waves in Solids: Two Volumes, Physics Today, 1974.

[19] J. Gao, J. Carlier, S. Wang, P. Campistron, D. Callens, S. Guo, X. Zhao, B. Nongaillard, "Lab-on-a-chip for high frequency acoustic characterization," Sensors and Actuators, B: Chemical, **177**, 753-760, 2013, doi:10.1016/j.snb.2012.11.037.

[20] D.M. Pozar, Microwave Engineering, 4th Edition, 2012.

[21] R. Fiore, "Capacitors in broadband applications," Applied Microwave and Wireless, **13**(5), 40, 2001.

[22] R.N. Thurston, "Effect of Electrical and Mechanical Terminating Resistances on Loss and Bandwidth According to the Conventional Equivalent Circuit of a Piezoelectric Transducer," IRE Transactions on Ultrasonic Engineering, **7**(1), 16-25, 1960, doi:10.1109/T-PGUE.1960.29256.

[23] D.C. Youla, A new theory of broadband matching, IEEE Transactions on Circuit Theory, 1964.

[24] W.-K. Chen, The Circuits and Filters Handbook, 2003.

[25] L. Korson, W. Drost-Hansen, F.J. Millero, "Viscosity of water at various temperatures," Journal of Physical Chemistry, **73**(1), 34-39, 1969, doi:10.1021/j100721a006.

[26] R.A. Horne, D.S. Johnson, "The viscosity of water under pressure," Journal of Physical Chemistry, **70**(7), 2182-2190, 1966, doi:10.1021/j100879a018.

[27] J. Holmes, G. Parker, W. Povey, "Temperature dependence of bulk viscosity in water using acoustic spectroscopy," in Journal of Physics: Conference Series, **269**(1), 2011, doi:10.1088/1742-6596/269/1/012011.

[28] A.J. Zuckerwar, R.L. Ash, "Variational approach to the volume viscosity of fluids," Physics of Fluids, **18**(4), 2006, doi:10.1063/1.2180780.

[29] M. Trempa, C. Reimann, J. Friedrich, G. Müller, D. Oriwol, "Mono-crystalline growth in directional solidification of silicon with different orientation and splitting of seed crystals," Journal of Crystal Growth, **351**(1), 131-140, 2012, doi:10.1016/j.jcrysgro.2012.04.035.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1436-1457 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Parameter Estimation for Industrial Robot Manipulators Using an Improved Particle Swarm Optimization Algorithm with Gaussian Mutation and Archived Elite Learning

Abubakar Umar[1], Zhanqun Shi[1, *], Lin Zheng[2], Alhadi Khlil[1], Zulfiqar Ibrahim Bibi Farouk[3]

[1]*School of Mechanical Engineering, Hebei University of Technology, Tianjin, 300401, China*

[2]*School of Mechanical and Electrical Engineering, Dalian Minzu University, Dalian, 116600, China*

[3]*School of Mechanical Engineering, Tianjin University, Tianjin, 300350, China*

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 01 December, 2020
Online: 21 December, 2020

Keywords:
Robotics
Dynamic model
Swarm intelligence

ABSTRACT

This work presents the analysis and formulation for optimizing the dynamic model and parameter estimation of all the six joints of a 6DOF industrial robot manipulator by utilizing swarm intelligence to optimize two excitation trajectories for the first three links at the arm and the last three links at the wrist of the robot manipulator. Numerical techniques were used to reduce the observation matrix to a minimum linear combination of parameters, thereby maximizing the identifiable parameters, and the Linear Least Square method was used for parameter identification. An improved particle swarm optimization algorithm with mutation and archived elite learning was proposed for solving the dynamic optimization problem of the industrial robotic manipulator. The basic parameters of the algorithm have been optimized for robotic manipulator analysis. The proposed algorithm is computationally economical while completely dominating other Evolutionary algorithms in solving robot optimization problems. The algorithm was further used to analyze 36 benchmark functions and produced competitive results.

1. Introduction

This paper is an extension of the work originally presented in the 2019 IEEE International Conference on Artificial Intelligence and Computer Applications [1], where a novel mutating Particle swarm optimization (MuPSO) based solution was proposed for analyzing the inverse kinematic problem of multi-degree-of-freedom robot manipulator. It was observed that the basic parameters of the Particle swarm optimization (PSO) algorithm are incapable of analyzing robot optimization problems, and it was established that the PSO parameters need to be modified for robot analysis. Further investigations were published in [2] where various parameters were used to analyse the inverse kinematic problem of four popular robot configurations and a new range of parameters for solving robot optimization problems was established, thereby an optimized mutating PSO algorithm for robot analysis was developed. If the algorithm runs into stagnation while solving a robot optimization problem, the mutation function is used to generate a completely new swarm and all the information regarding the previous swarm is lost. It is keen to note that if a particle from the previous swarm is retained, that particle becomes the global best particle for the new swarm, making the entire swarm converge again at the previous solution. It was observed that it would be beneficial to save previously non-dominated solutions in an archive for future reference. This work therefore seeks to further modify the mutating PSO algorithm by implementing it with an archived elite learning system, and using the new eMuPSO algorithm for dynamic parameter estimation of a 6 Degree-of-freedom (DOF) industrial manipulator and finally compare the improved algorithm with other state of the art algorithms in analyzing 36 benchmark functions. The rest of the paper is structured as follows; section 2 presents previous literature, section 3 introduces the new Mutating PSO algorithm, section 4 presents the results of parameter estimation of robot manipulators, and compares the new PSO algorithm with state of the art swarm-based algorithms on thirty-six benchmark functions and finally section 5 concludes the findings.

2. Literature

In industrial robotics, a dynamic model determines the actuator force required to achieve the desired joint configuration, coupling

*Corresponding Author: Zhanqun Shi, Hebei University of Technology, Tianjin-China, Tel.: +86-2260438217, z_shi@hebut.edu.cn

effects and non-linearity of the system introduces dynamic uncertainties. These errors degrade the robot's performance. Therefore, re-calibrating robot parameters are necessary for efficient robot analysis and control, that can only be achieved experimentally. In dynamic model estimation, obtaining experimental data requires that firstly, a linearized model of the robot manipulator is developed, then a trajectory that excites the robot's rigid body dynamics is developed, afterwards input torque and output joint configuration data is experimentally obtained and finally the dynamic parameters are estimated. The accuracy of the estimation method is largely dependent on the chosen trajectory and measuring accuracy. The actuator torque can be measured from input electric currents while optic encoders are incorporated for measuring joint output conditions. The recent improvement in technology makes available encoders and reduction gears that are capable of preventing cumulative errors and minimizing gearbox backlash. Therefore, carefully selecting an exciting trajectory is paramount in dynamic model estimation. Reference [3] observed that the excitation trajectory can be optimized by minimizing the condition number of the observation matrix. The observation matrix is a function of only the joint variables; it enables the solution of nonlinear systems to be presented as linear. The condition number of the observation matrix is a measure of noise immunity, a value closer to unity results in a better signal to noise ratio (SNR). The presence of unidentifiable parameters results in a high condition number of the observation matrix. References [4, 5] proposed techniques for categorizing dynamic parameters into identifiable parameters, unidentifiable parameters, and identifiable parameters in linear combinations. These works suggested that eliminating unidentifiable parameters from the observation matrix enhances estimation accuracy. In [6, 7] a set of Minimal Linear Combinations (MLC) of identifiable dynamic parameters from the total amount of parameters to be identified was generated. Reference [8] proposed formulating the periodic Finite Fourier series (FFS) as an excitation trajectory. FFS inhibits noise interference, if the identification experiment is repeated several times, averaging the measured data in the time domain improves SNR. Reference [9] used a modified Fourier series (MFS), where the FFS is implemented on a circular trajectory in the Cartesian space such that the initial and final conditions of the joints are the same, permitting continuity in the optimization experiment without necessarily stopping and restarting the manipulator. The particle swarm optimization algorithm (PSO) is a stochastic metaheuristic swarm-based evolutionary algorithm (EA), each particle in the swarm is attracted towards the global best particle's position and its personal best position denoted as (Gbest and Pbest). The velocity and position of canonical PSO are updated according to (1) and (2).

$$X_i^d(t+1) = X_i^d(t) + V_i^d(t+1)\,, \tag{1}$$

$$V_i^d(t+1) = wV_i^{\dim}(t) + c_1 r_1\left(Pbest_i^{\dim}(t) - X_i^{\dim}(t)\right)\cdots + c_2 r_2\left(Gbest^{\dim}(t) - X_i^{\dim}(t)\right)\,, \tag{2}$$

where i = 1, 2, 3…$nPop$ denotes for the index of each particle, $nPop$ is the swarm size, and dim = 1,2…dof denotes for the dimensionality of the solution space, dof is the degrees of freedom of the robot. V_i^{dim} and X_i^{dim} stand for the position and velocity vectors of the ith particle, respectively. $V_i^{dim} = [v_i^1, v_i^2, \ldots, v_i^{dof}]$, $X_i^{dim} = [x_i^1, x_i^2, \ldots, x_i^{dof}]$. w is the inertia weight. c_1 and c_2 are cognitive and social learning coefficients. r_1 and r_2 are two uniformly distributed random numbers within the range of [0,1]. Over the years, PSO has undergone various modifications either to improve its internal dynamics (enhance convergence and exploitation characteristics) or to satisfy the requirement of a specific real-world optimization problem or both.

2.1. PSO with Improved Dynamics

An index based ring topology PSO was proposed by [10] for solving multimodal multi-objective optimization problems having more than one Pareto optimal solution corresponding to the same objective function. Reference [11] studied the impact of communication topology in PSO, the star topology (Gbest) where each particle communicates with every other particle, was found better in analyzing unimodal separable problems while the ring topology (LBest) where each particle communicates with only its two neighbor particles, is preferred for multimodal, non-separable and composite functions. The ring topology is more capable of jumping between optimum basins and it was also found to never converge at a local optimum, but with a considerably higher computational cost. Memetic algorithms (MA) are hybrid EAs that are a fusion of global and local search techniques. EAs are usually employed for global search while unconstrained optimizers are employed for local search. Adaptive MAs have been successfully combined with other EAs but not PSO, [12] proposed a co-evolutionary memetic PSO for solving multi-objective optimization problems. Feature selection (FS) is a combinatorial problem for large dimension data processing which requires large memory and high computational cost, [13] proposed a variable-length PSO for FS. The Proposed algorithm is flexible and simplifies large data analysis, and can jump out of local optima while narrowing the search space. The cooperative search strategy which prevents particle from being trapped in local optima was merged with PSO by [14] for unconstrained optimization, where cooperative multiple swarms were used to improve the convergence and efficiency of the canonical PSO.

2.2. Problem Oriented PSO

Medicine and medical healthcare are amongst the top beneficiaries of artificial intelligent swarm algorithms. Health care services are shifting from inpatients to outpatients. Reference [15] showed that accident and emergency centers in hospitals are currently being optimized to ensure 98% of patients get the required attention within 24 hours. The increasing demand for primary health care outlets has led to a requirement to optimize the operational efficiency of primary healthcare centers. Grouping the patients into an appropriate category optimizes time and manpower whereas patient no shows after grouping reduces operational efficiency of clinics. In [16], using PSO to improve the accuracy of traditional clustering techniques was proposed for analyzing real-time patient attendance data and group them into clusters, while [15] compared PSO with opposition based learning and self-adaptive cohort intelligence for identifying significant features capable of predicting no shows. Gene selection from microarray data for cancer diagnosis and testing also involve large data sets and are computationally expensive, support vector machines algorithm is a fast and efficient classification model

which requires its parameters tuned to satisfy specific problems. Reference [17] compared the performance of PSO-SVM and Memetic SVM for classification and feature selection of cancer cells. Alzheimer's disease is a primary stage of dementia which is a type of memory disorder that affects the brain. The structural changes of the brain's internal regions are the most commonly measured in diagnosing AD. Reference [18] compared the accuracy of PSO and other metaheuristic algorithms in segmenting the brain sub-regions. Air pollution is a major contributor to human mortality and a potential danger to the environment and ecological system [19]. Thermal power plants such as coal, petroleum, and natural gas are the main sources of air pollution, mercury contamination has been identified to be the most acute air pollutant produced by power plants [20]. In [21], a novel combination of modified genetic algorithm and the improved particle swarm optimization was proposed for minimizing the fuel cost of generating plants and emission simultaneously, by managing and controlling the integration of renewable energy and thermal power production, where an optimal generation plan for maximizing system efficiency can be achieved. Reference [22] proposed using an adaptive neuro-fuzzy inference system with PSO (ANFIS-PSO) for predicting mercury emission in power plants. Reference [23] sought to minimize the construction cost of reinforced concrete retaining walls (RCRW). These structures include bridges, railways, dams, etc., capable of withstanding the pressure resulting from the difference in level by an embankment, excavation, or natural processes. PSO was used to determine the optimum solution between popular techniques. In rural areas where the utility grid is unavailable, renewable energy is an attractive alternative for water pumping applications. The power-voltage curve of a photovoltaic cell was found to have multiple maximum power points under partial shading conditions, making the solar tracking mechanism unstable. In [24], the performance of PSO was compared with salp swarm algorithm for maximum power point tracking of solar panels. Electric motors are most popular for electrical-mechanical power conversion and employed in several industrial applications including robotics, because of their simplicity, durability, and low maintenance cost. They usually require a controller for high performance and efficiency. The electrical parameters of the electrical motor are very essential to design, performance assessment, and feasibility of the control technique, any difference in the actual motor parameters adversely affects the system performance. Reference [25] proposed using PSO for estimating the electrical parameters of induction motors.

3. Mutating PSO with Elite Archive Learning

The proposed PSO is equipped with a modified set of parameters and governing equations. The mutation function is used to generate a new swarm around the vicinity of the most promising particles when the algorithm runs into stagnation. The operation of the proposed mutating PSO with elite learning (eMuPSO) is divided into two, during the early stages, the algorithm searches the solution space for the minimum solution, then when it stagnates, elite solutions are being recalled from the archive and merged with the best solutions of the current run to generate a new population. The new population would be created either through Random or Gaussian mutation depending on the application. It was found that a mix of both was suitable for kinematic analysis while the Gaussian mutation was suitable for dynamic analysis.

3.1. Modified Parameters

It was established in [1] and [2] that the basic parameters of canonical PSO are not capable of ensuring converging solutions for robot optimization problems especially when the DOF is greater than three. The basic parameters of the mutating PSO were modified to satisfy the requirement for solving robot optimization problems. This was achieved by testing the performance of various PSO parameters on four popular robot configurations, and a relationship between the inertia weight and the social learning coefficient was derived. More details on modifying PSO parameters can be found in [2]. A non-linearly decreasing inertia weight is implemented with values between (2.1 – 0.6), the cognitive learning coefficient is constantly at 2.24, while the social learning coefficient is non-linearly increasing between (1.8 – 3.9). The equations for updating w and c2 are given in (3) and (4).

$$w_{iter} = w_{initial} * n^{iter} , \quad (3)$$

$$c_{2iter} = c_{2initial} / m^{iter} , \quad (4)$$

where *iter* is the iteration number, and *n* and *m* are coefficients. The coefficients *n* and *m* can be determined by setting *iter* to the maximum value. Robot configurations have complex dynamics described by highly coupled and non-linear, second-order differential equations, there are possibilities of numerous strong local minimizers in the solution space.

3.2. Mutation

In [26] it was shown that structural bias is a characteristic of swarm algorithms that tends to confine the algorithm to a region of the total search area. Mutation is a valid technique for overcoming such constraints. It was established in [2] that even the most intelligent swarm-based solutions are capable of running into stagnation, random mutation was used to push the algorithm out of stagnation and enhance convergence. The Gaussian mutation was used in [27] to improve swarm performance. Reference [28] also used a Gaussian mutation throw point strategy to redistribute the swarm at sparse positions. In the proposed algorithm, when the swarm stagnates at a local minimum solution, a Gaussian mutation is used to push the algorithm out of stagnation and redistribute the swarm to enhance unbiased coverage of the search space. The radius of the swarm is monitored according to (5). When the swarm radius becomes too small, it signifies that the swarm has lost diversity and might be stagnating, then the mutation function is initiated. The condition in (6) states that if there is no change in the swarm radius or when the change in the swarm radius is negligible, then the counter *g* increases sequentially otherwise the counter is reset to zero. A second condition in (7) terminates the algorithm when the minimum solution is achieved or the swarm is mutated when *g* is greater than a threshold value *G*.

Assuming there are *k* elite particles, the entire swarm population would be divided into *k+1* parts and each *k* part would be replaced by off-springs of the kth elite particle, while the *k+1* part consists of particles generated from a random distribution, thereby maintaining diversity in the swarm. such that if x^{dim} is the

dimth dimension of the swarm, then all the elements in the *dim*th dimension shall mutate according to (8).

$$rad_{iter} = GlobalBestCost_{iter} - mean(BestCost)_{iter} , \quad (5)$$

$$g = \begin{cases} g+1, & if, (rad_{iter-1} - rad_{iter}) = 0 \\ 0 , & Else \end{cases} , \quad (6)$$

$$f(MuPSO) = \begin{cases} end & if \quad 1e^{-8} \\ mutate & if \quad g \geq G \end{cases} , \quad (7)$$

$$^{\dim} x_{mutate} = {}^{\dim} x_{elite} + gaussian(\sigma) , \quad (8)$$

where *rad* is the radius of the swarm, *iter* refers to the current iteration, $^{dim}x_{mutate}$ and x_{elite} are the mutated off-spring and the elite solution in the *dim*th dimension of the swarm, and σ is a variable between [0-1] multiplied by size of the search space.

3.3. Elite Archive

When the algorithm is searching for the best solution, the experiences of the best particle is transferred to the next iteration through the global best information, but at the beginning of subsequent generation or after mutation, all the information regarding the best particle from the previous generation is lost. This inspired the introduction of an archive where the experiences of the elite non-dominated solutions are saved and can be called in the future to enhance the search. Reference [29] applied a dynamic archive maintenance strategy to improve the diversity of solutions in multi-objective particle swarm optimization. In [30] an external archive was employed to preserve the non-dominated solutions visited by the particles to enable evolutionary search strategies to exchange useful information among them, and [31] used a grid-based approach for the archiving process and ε-dominance method to update the archive, which helps the algorithm to increase the diversity of solutions. Reference [32] used the cooperative archive to exploit the valuable information of the current swarm and archive. Information about the elite particles from dynamic sub swarms was used in [33] to improve the following sub-swarm, while [32] introduced a new velocity updating technique that explores the external archive of non-dominated solutions in the current swarm. In this article, an elite archive learning is used to refine the solution in the final stages of the algorithm. Elites from previous searches are saved as global elites, while elites of the current search are saved as local elites. When the algorithm stagnates, a combination of both elite vectors is used to generate a new population, and the global elites are updated with new solutions from the current local solutions

3.4. Robot Dynamics

Industrial robot manipulator

A 6 DOF industrial robot manipulator was analyzed, the structure of the robot manipulator is shown in Figure 1 and its D-H parameters are given in Table 1. A homogeneous transformation matrix describes the state (orientation and position coordinates) of a robot link with respect to the previous link in the Cartesian coordinate space and can be used to describe the state of the end-effector (tool) of the manipulator with respect to the base (global coordinate) frame. If all the joint parameters are known (D-H Parameters), the homogeneous matrix of each successive pair of frames for the forward and inverse directions can be obtained from (9). The robot manipulator's end-effector coordinates would be a product of post multiplication as shown in (10).

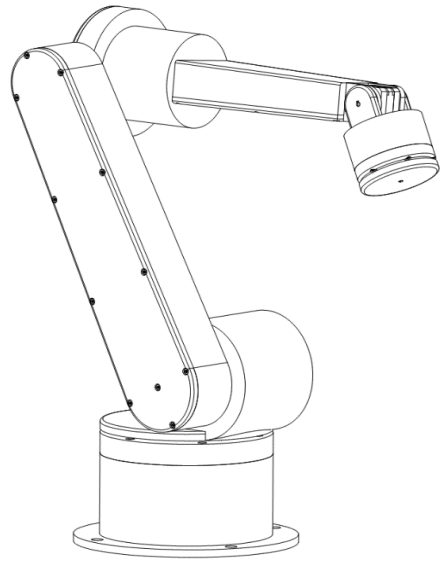


Figure 1: A 6 degree of freedom industrial robot manipulator

Table 1: D-H parameters of 6 degree of freedom industrial robot manipulator

Joint	Link	Off-set	Joint	Off-set
1	300	320	(-165 : +165)	-pi/2
2	700	0	(-110 : +110)	0
3	0	0	(-110 : +70)	-pi/2
4	0	200+497.5	(-160 : +160)	pi/2
5	0	0	(-120 : +120)	-pi/2
6	0	97.5+30	(-242 : +242)	0

$$T_{\dim}^{\dim-1} = \begin{bmatrix} c\theta_{\dim} & s\theta_{\dim}c\alpha_{\dim-1} & s\theta_{\dim}s\alpha_{\dim-1} & a_{\dim}c\theta_{\dim} \\ -s\theta_{\dim} & c\theta_{\dim}c\alpha_{\dim-1} & c\theta_{\dim}s\alpha_{\dim-1} & a_{\dim}s\theta_{\dim} \\ 0 & -s\alpha_{\dim-1} & c\alpha_{\dim-1} & d_{\dim} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (9)$$

$$T_4^0 = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} = T_1^0 T_2^1 T_3^2 T_4^3 \quad (10)$$

where $c\theta$ and $s\theta$ refers to $cos\theta$ and $sin\theta$ respectively.

Dynamic parameter estimation

The Linear-least-square (LLS) method together with an optimized excitation trajectory was used for dynamic parameter estimation of the robot manipulator. LLS requires that the model equations are linear with respect to dynamic parameters and it is

sensitive to noise in the measured data. [34] observed that the actuator forces of an industrial manipulator are linear functions of dynamic parameters, [4] inferred that reformulating the Newton-Euler dynamic model such that the link inertia tensors are expressed about the link coordinate frames instead of the center of the links' mass, would result in a linearized Newton-Euler formulation. Equations (11-14) presents a summarized Newton-Euler formulation in the form of observation matrix. For detailed information on linearizing the Newton-Euler equation, see [4] and [34].

$$\Phi = \left(\Omega^T \Omega\right)^{-1} \Omega^T * \Gamma \ , \tag{11}$$

$$\Gamma = \left(\left(\Omega^T \Omega\right)^{-1} \Omega^T\right)^{inv} \Phi \ , \tag{12}$$

$$\Gamma = \begin{vmatrix} \tau_{\dim} \\ \vdots \\ \tau_{dof} \end{vmatrix} = \begin{vmatrix} \omega_{\dim}\left(q,\dot{q},\ddot{q}\right) \\ \vdots \\ \omega_{dof}\left(q,\dot{q},\ddot{q}\right) \end{vmatrix} * \begin{vmatrix} \phi_{\dim} \\ \vdots \\ \phi_{dof} \end{vmatrix} , \tag{13}$$

$$\phi_{\dim} = \begin{bmatrix} mm & mc_x & mc_y & mc_z & I_{XX} & I_{YY} & I_{ZZ} & I_{XY} & I_{XZ} & I_{YZ} \end{bmatrix}^T , \tag{14}$$

$$fitness = cond(\Omega) + \frac{1}{\varphi_{\min}(y)} + \varphi_{\max} y \ , \tag{15}$$

where Γ is the compound torque for the manipulator comprising of individual joint torques τ, the compound observation matrix Ω comprises of individual joint vectors $\omega(q, \dot{q}, \ddot{q})$, and is a function of the joint position, velocity, and acceleration only. ϕ is a vector of unknown robot parameters for each link that require to be identified, it includes link mass *mm*, first moments $[mc_x\ mc_y\ mc_z]^T$ along the x-y-z axis, and six inertia tensor I_{ij}, these compose the compound vector of unknown parameters Φ. The intelligent swarm-based algorithm would be required to minimize the condition number of the observation matrix while maximizing the smallest singular value and minimizing the largest singular value. Therefore, the fitness function would be as described in (15).

Exciting trajectory

A modified Fourier series (MFS) formulation was used to define the excitation trajectory with a cubic polynomial. The equation of motion for the trajectory of the first three joints are given in (16-18), a circular trajectory path was implemented so as to allow continuous trajectory tracking of the robot during experimentation. The trajectory is to be optimized by applying the eMuPSO algorithm to minimizing the condition number of the observation matrix. The trajectory is implemented in Cartesian space, and joint positions are determined from inverse kinematic solutions. The wrist of the manipulator has a z-x-z orientation, therefore the excitation trajectory along the *x* and *y* axis can be derived independently from (16) and (17) while that of the *z* axis is estimated from (19), where *d6* is the offset length of the sixth link of manipulator (see DH parameters) and t_{14} is the position coordinate of transformation matrix T_4^0 from the *Base* to *Link-04*.

$$Q_x(it) = d0_j^{it} + \sum_{j=1}^{N} \begin{pmatrix} \cos(P_{it}) * a0_j^{it} + \dots \\ \frac{b0_j^{it}}{freq * j} \sin(freq * j * tym) - \dots \\ \frac{c0_j^{it}}{freq * j} \cos(freq * j * tym) \end{pmatrix} , \tag{16}$$

$$Q_y(it) = e0_j^{it} + \sum_{j=1}^{N} \begin{pmatrix} \sin(P_{it}) * a0_j^{it} + \dots \\ \frac{b0_j^{it}}{freq * j} \sin(freq * j * tym) - \dots \\ \frac{c0_j^{it}}{freq * j} \cos(freq * j * tym) \end{pmatrix} , \tag{17}$$

$$Q_{z,arm}(it) = f0_j^{it} + \sum_{j=1}^{N} \begin{pmatrix} \frac{b0_j^{it}}{freq * j} \sin(freq * j * tym) - \dots \\ \frac{c0_j^{it}}{freq * j} \cos(freq * j * tym) \end{pmatrix} , \tag{18}$$

$$Q_{z,wrist}(it) = f0_j^{it} + \sqrt{d6^2 - \left(t_{14} - Q_x(it)\right)^2} \ , \tag{19}$$

$$Vector(it) = \begin{bmatrix} vec_{11} \\ vec_{21} \\ vec_{31} \end{bmatrix} = \begin{bmatrix} t_{14} \\ t_{24} \\ t_{34} \end{bmatrix} - \begin{bmatrix} Q_x(it) \\ Q_y(it) \\ Q_z(it) \end{bmatrix} , \tag{20}$$

$$\theta_4 = a\tan\left(vec_{31}, vec_{21}\right) , \tag{21}$$

$$\theta_5 = a\tan\left(norm\left(\left[0, vec_{21}, vec_{31}\right]^T\right), vec_{21}\right) , \tag{22}$$

$$\theta_6 = a\tan\left(\left(vec_{31}(it) - vec_{31}(it-1)\right),\ \left(vec_{21}(it) - vec_{21}(it-1)\right)\right) , \tag{23}$$

$$\left.\begin{array}{l} \min_{a_j, b_j, c_j} \ cond(\Omega) \quad such \quad that; \\ \quad \{s(Q(it))\} \subset S \\ if \quad Q_x = [q_{x1}, q_{x2}, \cdots q_{xk}] \\ \quad 0 \le Q_x(it) \le 2\pi \\ \quad 0 \le \dot{Q}_x(it) \le 0 \\ \quad 0 \le \ddot{Q}_x(it) \le 0 \end{array}\right\} , \tag{24}$$

where *Q* describes the trajectory in Cartesian space such that $[Q_x, Q_y, Q_z]^T$ are the *x-y-z* coordinates of *Q*, *P* is the polynomial, the variables *a0* is the radius, *b0* and *c0* are the coefficients of the MFS while the variables *d0*, *e0*, and *f0* are the x-y-z position coordinates of the excited trajectory. $j = 1,2 \dots N$ denotes the harmonics of the MFS, while the index of time segment/period is denoted as *it*; *freq* is the fundamental frequency and *tym* is the period. The joints 4, 5, and 6 coincident at the joint 5, described by the transformation matrix T_4^0 in (10). The inverse kinematic solution of these joints can be estimated given the orientation matrix, but in this case only the position vector is known (i.e. $[Qx, Qy, Qz]^T$). Therefore, the rotation angles of the joints at the wrist would be estimated from

(20-23), and the orientation matrix is evaluated by substituting θ in (9) then inverse kinematics can be subsequently evaluated. To achieve continuity along the tool-path, the initial and final conditions of the trajectory must be the same as described in the conditions in (24). The initial and final values for the circular trajectory *P* are set as *0* and *2π* radians, 3 harmonics for the coefficients of MFS were used (N=3), and the fundamental frequency is *0.6283* with a period of *10* seconds.

4. Results

The eMuPSO algorithm was first implemented together with the LLS for optimizing the observation matrix and dynamic parameter estimation of all the six links of the 6DOF robot manipulator. The ideal parameters of the manipulator are given in (25-32). Then the proposed algorithm was finally implemented in evaluating thirty-six benchmark functions including twenty-four variable-dimension benchmark functions and twelve constant-dimension benchmark functions. The variable dimension benchmark functions consist twelve functions each, of unimodal and multi-modal roles (a full description of all benchmark function used are given in the appendix).

The results were compared with the Standard PSO (PSO) [35] and four other swarm-based EAs including the Whale Optimization Algorithm (WOA) [36], Grey Wolf Optimization (GWO) [37], Grass-hopper Optimization Algorithm (GOA) [38], and the Differential Evolution (DE) algorithms [39].

4.1. Dynamic Parameter Estimation

Generating initial elite particles

Parameter estimation was in two stages; initial elite particles were first generated which were then used to generate the swarm for parameter estimation. This reduces the probability for the algorithm to run into stagnation. A swarm of ten particles was first generated for each variable *(a0, b0, ..., f0)*. From inverse kinematics, the joint conditions (position, velocity and acceleration) corresponding to the generated trajectory of each set of particle is estimated and compared with the stated constraints in (24). Trajectories that do not meet the constraints are flagged and a penalty was introduced.

The fitness function for this swarm is a measure of the degree of compliance of the generated trajectory with the stated robot constraints. If a trajectory is found to completely satisfy these constraints, the variables corresponding to such trajectory is elevated to elite status and a new particle is generated to replaces that particle, else, the entire swarm's velocity and position is updated in accordance with PSO. Elite particles should be generated within 5-10 iteration if the minimum and maximum limits of the variables are carefully selected. This algorithm is terminated when 5 elite particles have been generated. The variable limits are set at ±10. While the variables *a* and *b* span between negative and positive limits, others must always be positive.

$$\mathrm{Mm} = \begin{bmatrix} \mathrm{mm1} \\ \mathrm{mm2} \\ \mathrm{mm3} \\ \mathrm{mm4} \\ \mathrm{mm5} \\ \mathrm{mm6} \end{bmatrix} = \begin{bmatrix} 5.029912009 \\ 4.530794148 \\ 3.056783430 \\ 1.815211338 \\ 0.990530741 \\ 0.186899621 \end{bmatrix}, \tag{25}$$

$$\mathrm{Mc} = \begin{bmatrix} \mathrm{mcx1} & \mathrm{mcy1} & \mathrm{mcz1} \\ \mathrm{mcx2} & \mathrm{mcy2} & \mathrm{mcz2} \\ \mathrm{mcx3} & \mathrm{mcy3} & \mathrm{mcz3} \\ \mathrm{mcx4} & \mathrm{mcy4} & \mathrm{mcz4} \\ \mathrm{mcx5} & \mathrm{mcy5} & \mathrm{mcz5} \\ \mathrm{mcx6} & \mathrm{mcy6} & \mathrm{mcz6} \end{bmatrix} \cdots = \begin{bmatrix} 4.3956 & -15.8576 & 10.1804 \\ -44.1094 & 93.9465 & 183.7506 \\ 0.37299 & 0.001086 & 100.6463 \\ -0.01376 & 12.03674 & 18.1007 \\ -0.00300 & -0.00352 & -21.2584 \\ 0.004296 & 0.000028941 & 4.9163 \end{bmatrix}, \tag{26}$$

$$\mathrm{I}_1 = \begin{bmatrix} \mathrm{Ixx1} & \mathit{Ixy1} & \mathit{Ixz1} \\ \mathit{Ixy1} & \mathit{Iyy1} & \mathit{Iyz1} \\ \mathit{Ixz1} & \mathit{Iyz1} & \mathit{Izz1} \end{bmatrix} \cdots = \begin{bmatrix} 20975.4835 & 0 & 0 \\ 0 & 13651.6206 & 0 \\ 0 & 0 & 19154.2091 \end{bmatrix}, \tag{27}$$

$$\mathrm{I}_2 = \begin{bmatrix} \mathrm{Ixx2} & \mathit{Ixy2} & \mathit{Ixz2} \\ \mathit{Ixy2} & \mathit{Iyy2} & \mathit{Iyz2} \\ \mathit{Ixz2} & \mathit{Iyz2} & \mathit{Izz2} \end{bmatrix} \cdots = \begin{bmatrix} 103901.5055 & 0 & 0 \\ 0 & 79173.3684 & 0 \\ 0 & 0 & 30809.7396 \end{bmatrix}, \tag{28}$$

$$\mathrm{I}_3 = \begin{bmatrix} \mathrm{Ixx3} & \mathit{Ixy3} & \mathit{Ixz3} \\ \mathit{Ixy3} & \mathit{Iyy3} & \mathit{Iyz3} \\ \mathit{Ixz3} & \mathit{Iyz3} & \mathit{Izz3} \end{bmatrix} \cdots = \begin{bmatrix} 9855.9631 & 0 & 0 \\ 0 & 9759.3916 & 0 \\ 0 & 0 & 7360.3570 \end{bmatrix}, \tag{29}$$

$$\mathrm{I}_4 = \begin{bmatrix} \mathrm{Ixx4} & \mathit{Ixy4} & \mathit{Ixz4} \\ \mathit{Ixy4} & \mathit{Iyy4} & \mathit{Iyz4} \\ \mathit{Ixz4} & \mathit{Iyz4} & \mathit{Izz4} \end{bmatrix} \cdots = \begin{bmatrix} 16365.8046 & 0 & 0 \\ 0 & 1616.0497 & 0 \\ 0 & 0 & 17081.9923 \end{bmatrix}, \tag{30}$$

$$I_5 = \begin{bmatrix} Ixx5 & Ixy5 & Ixz5 \\ Ixy5 & Iyy5 & Iyz5 \\ Ixz5 & Iyz5 & Izz5 \end{bmatrix} \cdots = \begin{bmatrix} 1058.2078 & 0 & 0 \\ 0 & 996.4668 & 0 \\ 0 & 0 & 1033.5495 \end{bmatrix}, \quad (31)$$

$$I_6 = \begin{bmatrix} Ixx6 & Ixy6 & Ixz6 \\ Ixy6 & Iyy6 & Iyz6 \\ Ixz6 & Iyz6 & Izz6 \end{bmatrix} \cdots = \begin{bmatrix} 103.8927 & 0 & 0 \\ 0 & 103.8903 & 0 \\ 0 & 0 & 204.5344 \end{bmatrix}, \quad (32)$$

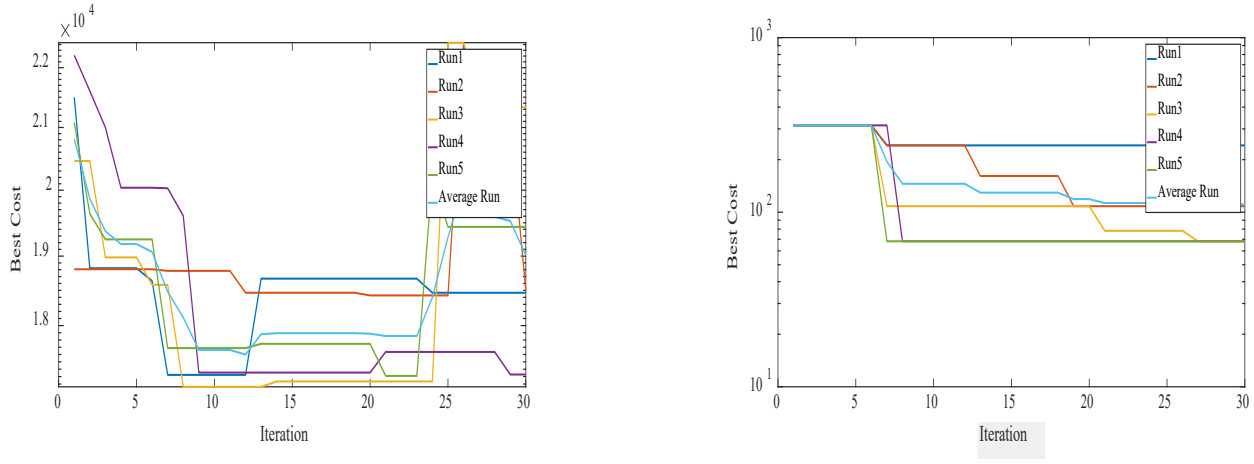
Optimizing the excitation trajectory

A swarm of 20 particles is generated from the previously generated *initial elite particles* through Gaussian mutation. Similar to previous procedures, the boundary conditions are set, joint conditions corresponding to each generated trajectory is estimated, again penalties are introduced to make trajectories that do not satisfy the constraints in (24) less likely to be selected as exemplars. The fitness function for this swarm is evaluated according to (15). The swarm of the algorithm is mutated at 10 iterations after stagnation (i.e. G=10).

Parameter estimation

The dynamic parameters of a 6DOF manipulator was estimated in two stages, the last three links corresponding to the wrist were first estimated followed by the first three links corresponding to the arm of the manipulator. A numeric technique for reducing the observation matrix to a minimum linear combination of identifiable parameters was implemented as described in [40] where three empty matrices were created corresponding to *identifiable* parameters, *unidentifiable* parameters, and *identifiable parameters in linear combinations*. Understanding that the rank of the observation matrix is equal to the number of identifiable parameters, the norm of each column of the observation matrix is determined independently. The columns with zero norms represent coefficients of *unidentifiable* parameters and are moved to the matrix of *unidentifiable* parameters, else it is added to the matrix of *identifiable* parameters. Note that the initial rank of the empty matrix of *identifiable* parameters is zero, if the newly added column does not increase the rank of the matrix of *identifiable* parameters, it is moved to the matrix of *identifiable parameters in linear combination*. The eMuPSO was used to select the trajectory that best optimizes the observation matrix. Removing *unidentifiable* parameters from the observation matrix increases estimation accuracy. In the first estimation stage (wrist), out of the 30 unknown parameters at the wrist, 15 were independently identifiable, 7 were identifiable in linear combination, while 8 were unidentifiable. Out of the 7 parameters that were identifiable in linear combination, it was observed that for most elite particles, 5 were identifiable with lower errors. The 15 independently identifiable parameters include *Iyy4*, *Ixz4*, *mx5*, *Ixx5*, *Iyy5*, *Izz5*, *Ixy5*, *Ixz5*, *Iyz5*, *mcx6*, *mcy6*, *Iyy6*, *Izz6*, *Ixz6*, and *Iyz6*. The 7 parameters identifiable in linear combination include *mcy4*, *mcy5*, *mcz5*, *mm6*, *mcz6*, *Ixx6*, and *Iyz6*. It was found that *Ixx6* and *Iyz6* were identified with very large errors for most elite particles. In the second estimation stage (arm), 20 parameters were independently identifiable, 2 were identifiable in linear combinations while 8 were unidentifiable. All parameters were identified with little errors. The 22 identified parameters include *mcy1*, *Iyy1*, *mm2*, *mcx2*, *mcy2*, *mcz2*, *Iyy2*, *Izz2*, *Ixz2*, *Iyz2*, *mm3*, *mcx3*, *mcy3*, *mcz3*, *Ixx3*, *Iyy3*, *Izz3*, *Ixy3*, *Ixz3*, *Iyz3*, *Ixz1*, and *Ixx2*.

Figure 2.0 (a) and (b) show the convergence plot at the wrist and at the arm of the manipulator respectively. The eMuPSO was populated with twenty particles in thirty iterations and five runs. The Figure 2.0 shows all five runs, the sixth being the average of all five runs. It would be observed that in the third run of Figure 2.0 (a), the algorithm converges to the global minimum solution in eight iterations, then it is mutated in the 14th iteration where it stagnates until it is mutated again in the 24th iteration. The best solutions are stored in the archive before mutation. The Figure 2.0 (b) show less agitation of the algorithm signifying less mutation and simpler analysis.



(a) (b)

Figure 2: Convergence plot of the observation matrices (a) at the wrist, and (b) at the arm.

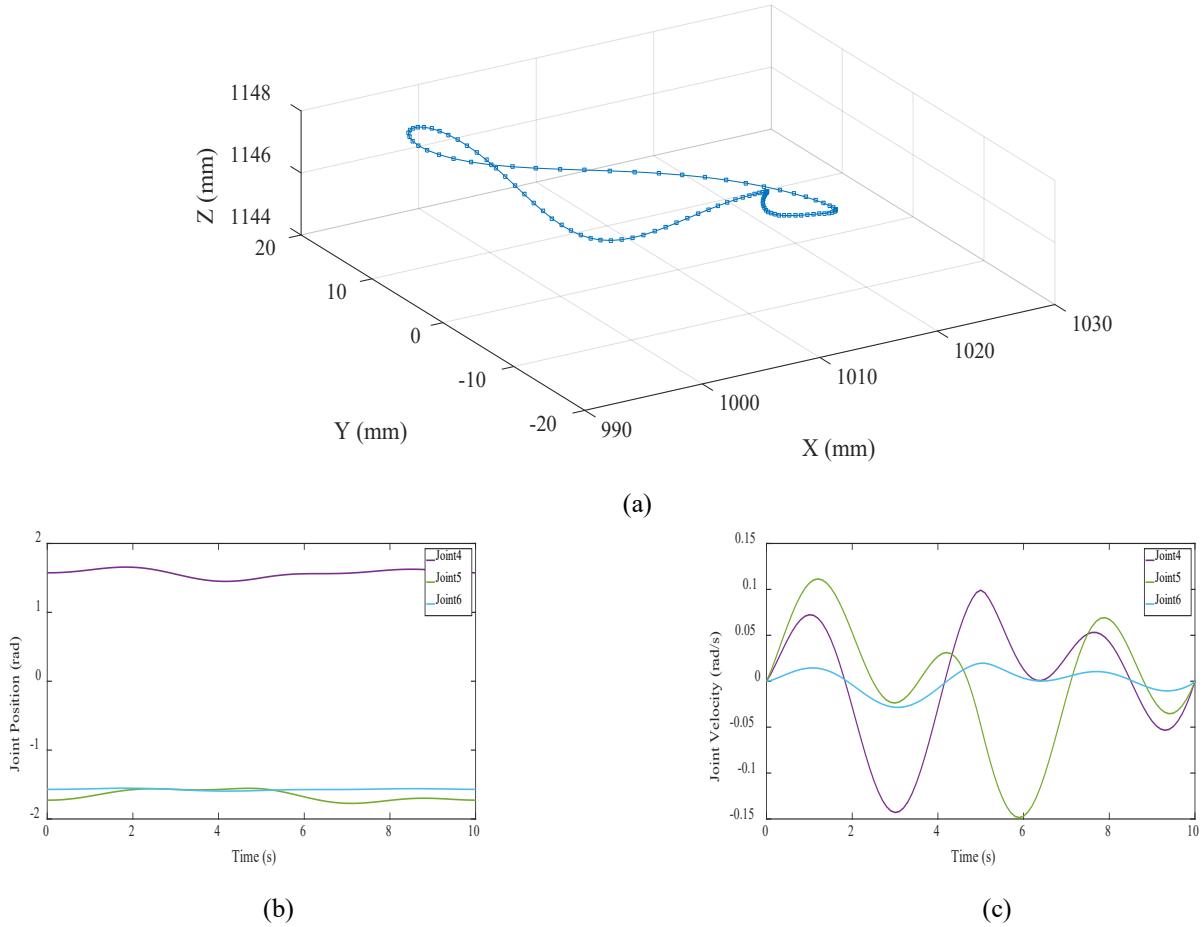
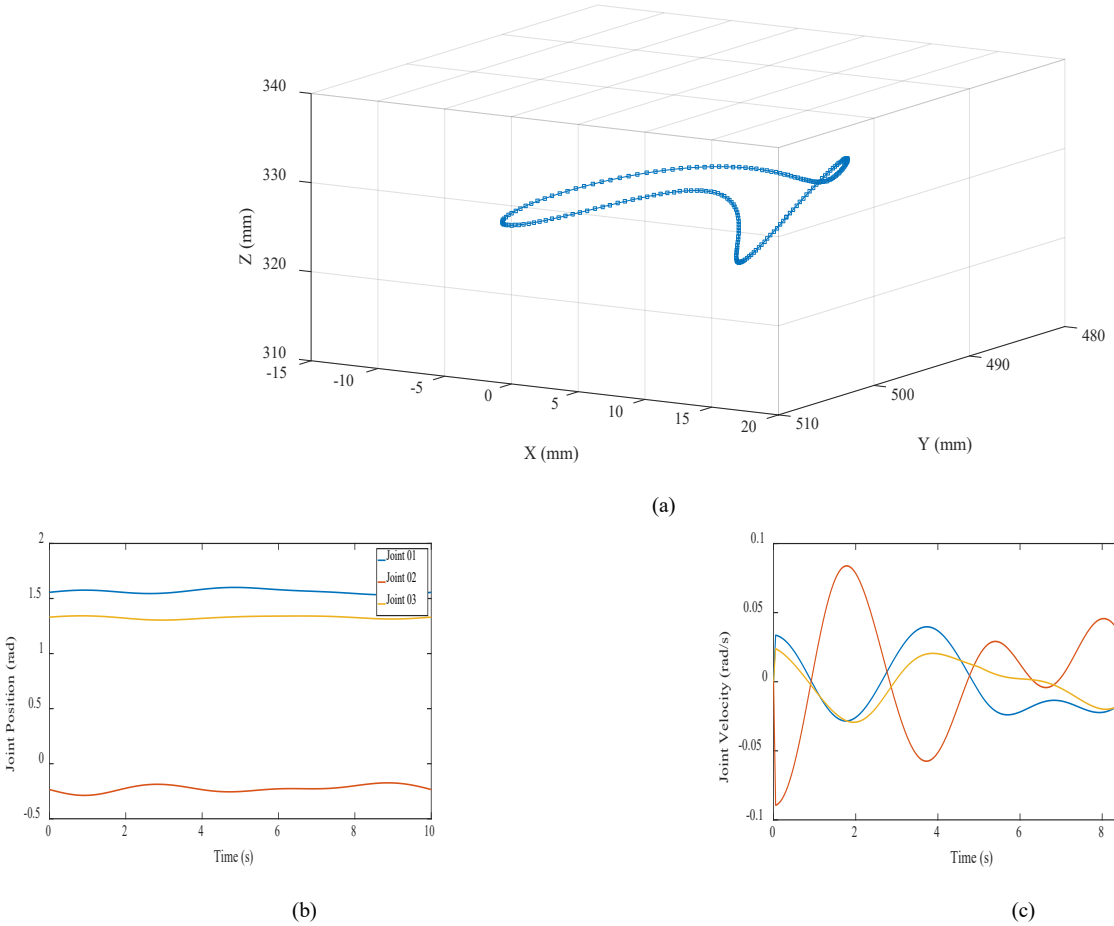


Figure 3: Excitation trajectory for the links at the wrist of manipulator. (a) Trajectory in Cartesian space (b) Joint displacement (c) Joint velocities

Table 2: Dynamic parameter estimation for the last three joints of a 6 DOF industrial robot

Parameter	Elite Particles (wrist)					
	1	2	3	4	5	6
a0(1)	-3.8711	-8.7566	-6.1829	-4.824	-6.7566	-4.4276
a0(2)	9.8782	1.1156	5.7223	-2.4838	3.1156	4.0066
a0(3)	0.42207	4.5867	-4.0733	-3.0035	3.748	-5.7368
b0(1)	2.1907	2.8328	4.9482	2.6421	4.7319	6.1336
b0(2)	-3.5264	-7.3354	-2.4158	-1.0079	-5.3354	5.1722
b0(3)	-2.293	-7.4121	0.31497	-3.5628	-5.4121	0.91808
c0	6.5802	5.4755	6.423	6.8105	5.0595	2.9898
d0	6.5802	3.1097	6.423	0.85837	5.0595	2.9898
e0	0	0	0	0	0	0
f0	t_{34}	t_{34}	t_{34}	t_{34}	t_{34}	t_{34}
fitness	418.354	401.380	357.993	351.894	349.889	297.128
condition number	154.591	152.878	156.902	157.041	155.063	154.087
Error(20)	10.8120	188.679	14.5859	99.1558	86.8486	25.3959
Error(22)	2220.44	16789.5	529.495	4568.45	8777.48	44.3582



(a) (b) (c)

Figure 4: Excitation trajectory for the links at the arm of manipulator. (a) Trajectory in Cartesian space (b) Joint displacement (c) Joint velocities

Table 3: Dynamic parameter estimation for the first three joints of a 6 DOF industrial robot

Parameter	Elite Particles (Arm)					
	1	2	3	4	5	6
a0(1)	-1.2742	5.2376	-3.2742	-2.8265	0.72577	-3.2742
a0(2)	-6.5754	6.2129	-4.5754	-6.5754	-8.5754	-8.5754
a0(3)	-9.0638	8.2042	-7.0638	-9.0638	-6.8925	-6.7108
b0(1)	-0.4967	6.2173	1.5033	-0.4967	1.5033	-2.4967
b0(2)	4.5411	-2.6712	2.6293	2.9774	2.5411	6.5411
b0(3)	-1.6233	-7.73	-0.74646	1.2535	-3.6233	0.37674
c0	3.0295	1.4518	1.0295	1.7199	1.0295	1.0295
d0	0	0	0	0	0	0
e0	173.11	101.03	184.34	187.21	188.04	166.14
f0	5.7001	6.0889	1.0295	1.7199	1.0295	1.0295
fitness	293.392	199.588	145.3132	103.235	70.9066	64.7651
condition number	53.5757	44.7731	26.41273	17.8764	13.4862	11.8133
Error(20)	0.00270	0.00286	0.035218	0.01955	0.03211	0.00822
Error(22)	0.00460	0.00766	0.074774	0.03507	0.08359	0.01544

Tables 2 and 3 presents the parameters of the non-dominated elite solutions for the joints at the arm and wrist of the manipulator, including the three harmonics each for *a0* and *b0*, with corresponding fitness values, condition number of the observation matrix and cumulative errors for the best 20 identified parameters and for all 22 identified parameters. Figure 3.0 (a) presents the resultant trajectory in Cartesian space for the 5th elite particle at the robot's wrist, showing a somewhat circular continuous trajectory with fluctuations that are typical of FFS. Figures 3.0 (b) and (c) show the position and velocity of the three joints at the robot's wrist in joint space, while Figure 4.0 shows the equivalent for the 1st elite particle at the robot's arm. From the results presented in Table 2 (at the wrist), it would be observed that while the *6th elite particle* has the lowest fitness value and the lowest cumulative estimation error for all the 22 identified parameters, the *2nd elite particle* presents the trajectory with the lowest condition number, and the *1st elite particle* presents the lowest cumulative estimation error for the 20 best identified parameters. Likewise, in Table 3 (at the arm), although all the 22 identifiable parameters were identified with higher accuracy, yet it would also be observed that the *6th elite particle* has the lowest fitness and presents the trajectory with the lowest condition number, but the *1st elite particle* presents the lowest cumulative estimation errors for both the 20 best identified parameter and the 22 identifiable parameters. This supports the observation of [41] that the best solution does not guarantee the global minimum solution because the dynamic optimization problem is not convex.

4.2. Benchmark Function Analysis

A total of thirty-six benchmark functions (BF) were evaluated, and the results presented in Tables 4-6. Reference [2] showed that accuracy and convergence time are paramount in robot analysis, the average and standard deviations alone do not completely describe the results, and some solutions with good averages are incapable of finding the global minimum solution (inaccurate). Therefore, the minimum values are also presented in the tables. The solution that best minimizes the function is presented in bolded font. From the Table 4, in the unimodal BF analysis, all the algorithms satisfactorily found the minimized solution for the function f6; eMuPSO, PSO, and DE equally minimized the function f7 while GWO and WOA equally minimized the function f11. WOA and GWO had the best performance, producing the best minimizing solutions for five out of twelve unimodal BF each. WOA dominated other algorithms in f1, f3, and f12 functions, tied with GWO in f11 and tied with all other metaheuristics in f6. While GWO dominated other metaheuristics in f8, f9, and f10. Therefore, WOA and GWO dominated other algorithms in three unimodal BF, tied in two BF and were dominated in seven BF each as elaborated in Table 7. The PSO dominated other metaheuristics in f4 and f5, while eMuPSO was dominant in f2. Both tied with other metaheuristics in f6 and f7 and were dominated in eight and nine unimodal BF respectively. From the Table 5, in the multimodal BF analysis, it can be observed that the eMuPSO and PSO tied in f1, f4, and f12; PSO, WOA, and GWO tied in f2; WOA and GWO tied in f3; eMuPSO, PSO and GOA tied in f5; all the metaheuristics except GWO and DE tied in f6, eMuPSO and DE tied in f9, and all the metaheuristics except GOA tied in f11. eMuPSO had the best performance producing the best minimum results on eight out of twelve multimodal BF, followed by PSO, WOA, and GWO producing the best minimum solutions for seven, five, and four multimodal BF respectively. eMuPSO dominated other metaheuristics on f10 and tied with other metaheuristics on seven multimodal BF. PSO did not dominate other metaheuristics on any of the multimodal BF, but tied on seven and was dominated in five BF, WOA was dominant on f7, tied with other metaheuristics on 4 Multimodal BF and was dominated in 7 BF. GWO was dominant on f8, tied with other metaheuristics on 3 BF, and was dominated on 8 BF. From the Table 6, in the constant dimension BF analysis, eMuPSO and PSO tied in producing the best minimizing solution for f1 and f6, all metaheuristics except GOA and GWO tied on f2; eMuPSO, PSO and DE tied on f3, f5, f7, f8, f9, f11. All the metaheuristics except WOA and GWO tied on f4 and f12. eMuPSO once again had the best performance producing the best minimum solutions for all twelve BF, dominated other metaheuristics in f10, and tied in eleven BF, followed by PSO and DE producing the best-minimized solution for eleven and nine BF respectively. Compared to other algorithms, WOA and GWO performed well in the unimodal BF analysis but performed poorly in the constant-dimension BF analysis. After numerous experiments, at various dimension sizes, it can be suggested that the WOA algorithm is convenient for analyzing optimization problems with high dimensionality (especially unimodal roles), but may not necessarily maintain its lead when the dimensionality is low, while the proposed eMuPSO is convenient for analyzing optimization problems with lower dimensionality. Compared to other EAs, the eMuPSO and PSO algorithms performed averagely in the unimodal BF evaluations but were excellent in analyzing multimodal and constant-dimension BF. DE performed poorly in the unimodal and multimodal BF, performed well in the constant-dimension BF analysis.

The Wilcoxon signed-rank test was performed at a 5% significance level to compare the results of each run and decide on the significance of the results, and the p-values are reported in Table 8. In the statistical analysis, the best algorithm in each test function is the algorithm with the best average value, the best algorithm is compared with other algorithms independently. For instance, if eMuPSO is the best algorithm, pairwise comparison is done between eMuPSO/PSO, eMuPSO/GOA, and so on. Observe that since the best algorithm cannot be compared with itself, N/A with bold font size has been assigned to it. From Table 8, in the unimodal statistical analysis (F1-F12) WOA algorithm was most significant in six out of twelve BF as elaborate in Table 9. SPSO and GWO were significant in four functions each, while DE and eMuPSO were significant in three and two of the functions respectively. In the multimodal statistical analysis (F13-F24), the eMuPSO and DE were most significant in six out of the twelve BF each, followed by WOA with five BF, while SPSO and GWO were most significant in 3 of the BF each. In the constant-dimension statistical analysis (F25-F36), the eMuPSO and DE remained most significant for ten out of the twelve BF each, followed by the SPSO with 8 BF, GOA was most significant in 2 of the BF, while others were significant in only one each of the benchmarks functions.

Table 4: Unimodal benchmark function analysis.

Unimodal Benchmark Functions		**Evolutionary Algorithms**					
		eMuPSO	PSO	WOA	GWO	GOA	DE
F1: Sphere	*Min*	2.00E-12	4.90E-44	**0**	2.85E-240	2.25E-06	0.001903
	Ave	9.46E-08	5.44E-42	0	1.91E-232	8.66E-05	0.031038
	StDev	3.13E-07	2.23E-41	0	0	9.17E-05	0.039287
F2: Beale	*Min*	**1.67E-08**	7.70E-08	0.000176	6.42E-05	0.000149	3.147E-08
	Ave	0.0178315	0.015488	0.026001	0.005977	0.035998	1.43E-07
	StDev	0.0308013	0.029336	0.044659	0.0154745	0.038128	2.13E-07
F3: Matyas	*Min*	1.69E-12	2.12E-24	**0**	1.33E-143	1.42E-06	3.50E-07
	Ave	3.42E-09	1.07E-18	0	1.57E-120	1.55E-05	6.23E-06
	StDev	7.91E-09	5.33E-18	0	4.10E-120	1.51E-05	5.08E-06
F4: Rosenbrook	*Min*	10.167783	**0.087952**	14.80708	15.116454	13.29022	6.665086
	Ave	28.023866	6.257364	15.61825	16.217369	258.643	28.08251
	StDev	32.409539	3.514322	0.417166	0.6497502	480.0779	31.98284
F5: Step 2	*Min*	3.91E-10	**0**	1.38E-05	1.91E-07	7.31E-07	0.004025
	Ave	8.03E-08	0	4.99E-05	0.1749605	0.000109	0.047807
	StDev	1.77E-07	0	3.02E-05	0.1875981	0.000145	0.055443
F6: Scahffer 2	*Min*	**0**	0	0	0	0	0
	Ave	0	0	0	0	0	0
	StDev	0	0	0	0	0	0
F7: Scahffer 3	*Min*	**-3.37E-4**	-3.37E-4	-3.37E-4	-3.37E-4	-3.37E-4	-3.37E-4
	Ave	-3.37E-4	-3.37E-4	-3.36E-4	-3.37E-4	-3.37E-4	-3.37E-4
	StDev	1.65E-19	1.65E-19	3.00E-06	3.45E-08	1.08E-09	1.65E-19
F8: Scwefel 1.2	*Min*	0.0003765	1.92E-11	0.317571	**7.14E-99**	1.759081	0.003903
	Ave	0.0071479	4.45E-09	207.2391	4.23E-82	33.02237	0.079942
	StDev	0.0069211	6.37E-09	408.9857	2.15E-81	28.64997	0.090924
F9: Scwefel 2.21	*Min*	0.370036	9.76E-10	1.62E-09	**8.56E-66**	0.018632	0.179791
	Ave	1.0249904	1.39E-08	4.32931	1.20E-62	1.084041	2.315132
	StDev	0.5007591	1.17E-08	9.450776	4.98E-62	1.537078	2.441406
F10: Scwefel 2.22	*Min*	0.0027291	5.00E-26	**2.0800E-322**	4.67E-136	0.018623	0.077971
	Ave	0.0326953	2.66E-24	2.7900E-321	1.61E-133	0.192329	1.517522
	StDev	0.0266489	5.88E-24	0	5.42E-133	0.578804	2.606056
F11: De Jong 4	*Min*	2.35E-22	1.53E-79	**0**	0	1.40E-13	5.83E-10
	Ave	3.14E-14	4.80E-73	0	0	2.75E-09	0.117002
	StDev	9.95E-14	1.73E-72	0	0	3.50E-09	0.640849
F12: Axis Parallel	*Min*	4.79E-12	2.48E-46	**0**	2.61E-241	3.93E-05	6.28E-05
	Ave	2.12E-08	2.71E-43	0	2.61E-235	0.001067	0.000993
	StDev	6.13E-08	9.27E-43	0	0	0.001414	0.001253

Table 5: Multimodal benchmark function analysis.

Multimodal Benchmark Functions		**Evolutionary Algorithms**					
		eMuPSO	PSO	WOA	GWO	GOA	DE
F13: Easom	*Min*	**0**	0	3.26E-15	7.77E-12	2.11E-16	NaN
	Ave	0	0	0.023333	1.00E-09	0.086667	NaN
	StDev	0	0	0.043018	9.16E-10	0.034575	NaN
F14: Griewank	*Min*	2.00E-07	**0**	0	0	0.000577	0.011976
	Ave	0.0343134	0.015987	0.001742	0.0001222	0.04842	0.147971
	StDev	0.0330935	0.017756	0.009541	0.0006694	0.06238	0.143891
F15: Rastrigin	*Min*	8.9546265	7.959672	**0**	0	31.83894	9.732789
	Ave	17.378784	16.01883	0	0	67.56469	36.96119
	StDev	5.8526738	5.267084	0	0	29.12814	22.58322
F16: Pen Holder	*Min*	**-9.04E-08**	-9.04E-08	2.87E-07	0.0393752	0.026031	NaN
	Ave	0.0168779	0.001703	0.029721	0.0652148	0.082208	NaN
	StDev	0.0251307	0.004706	0.02437	0.0179396	0.034506	NaN
F17: Test-tube Holder	*Min*	**-2.04E-08**	-2.04E-08	-2.04E-08	-2.04E-08	-2.04E-08	-2.040E-08
	Ave	-2.04E-08	-2.04E-08	0.000198	-1.63E-08	0.002207	-1.50E-05
	StDev	0	0	0.000604	4.80E-09	0.00908	3.01E-05
F18: Egg Holder	*Min*	**0**	0	0	1.03E-10	0	NaN
	Ave	69.652218	52.92821	1.14E-13	25.398919	149.1868	NaN
	StDev	77.530967	74.41267	1.16E-13	50.110379	130.2925	NaN
F19: Damavandi	*Min*	2	2	**0.009163**	2.0000051	2	2.000002
	Ave	2	2	0.278923	2.2578664	2.000002	2.000018
	StDev	3.27E-10	0	1.197583	1.412306	3.83E-06	1.32E-05
F20: Cross-Leg Table	*Min*	0.0037107	0.049876	0.099255	**0**	0.09978	0.0150
	Ave	0.0603736	0.084285	0.099835	0.0832343	0.099862	0.011415
	StDev	0.0271336	0.019337	0.000111	0.0378599	2.45E-05	0.006958
F21: Bucking 4	*Min*	**0**	1.47E-90	2.20E-11	1.58E-08	3.80E-08	0
	Ave	0	5.03E-79	1.17E-06	1.15E-06	0.00028	0
	StDev	0	2.35E-78	1.28E-06	1.26E-06	0.000495	0
F22: Bucking 6	*Min*	**0.0023858**	0.003192	0.003289	0.092071	0.0065	0.013404
	Ave	0.1010663	0.119555	0.124073	0.152622	0.098243	0.097272
	StDev	0.0585834	0.074511	0.08727	0.0926189	0.066986	0.055002
F23: Cosine	*Min*	**0**	0	0	0	1.94E-15	0
	Ave	0.0084647	0	0	0	0.08306	0
	StDev	0.0080534	0	0	0	0.05875	0
F24: Modified Rosenbrook	*Min*	**3.4040243**	3.404024	3.404024	3.4040243	3.404024	3.404024
	Ave	4.8692154	5.80161	3.404024	6.8672033	7.000974	3.404024
	StDev	1.9585576	1.991086	4.00E-08	1.3815921	1.219484	1.88E-15

Table 6: Constant-dimension benchmark function analysis.

Constant-Dimension Benchmark Functions		Evolutionary Algorithms					
		eMuPSO	SPSO	WOA	GWO	GOA	DE
F25: Paviani	*Min*	**-2.34E-06**	-2.34E-06	4.52E-06	1.91E-05	-2.34E-06	NaN
	Ave	-2.34E-06	-2.34E-06	0.000137	4.28E-05	-2.34E-06	NaN
	StDev	5.70E-15	3.54E-15	0.00013	2.09E-05	2.03E-11	NaN
F26: Modified Ackley	*Min*	**0**	0	0	1.53E-10	1.78E-15	0
	Ave	0	0	2.01E-10	4.45E-09	3.00E-14	0
	StDev	0	0	3.77E-10	3.79E-09	2.01E-14	0
F27: Camel 6 Hump	*Min*	**2.22E-16**	2.22E-16	1.11E-15	1.11E-13	4.44E-16	2.22E-16
	Ave	2.22E-16	2.22E-16	5.03E-13	3.93E-10	8.44E-16	2.22E-16
	StDev	0	0	1.25E-12	4.48E-10	3.37E-16	0
F28: Branin Rcos	*Min*	**-2.26E-11**	-2.26E-11	-2.26E-11	1.84E-10	-2.26E-11	-2.26E-11
	Ave	-2.26E-11	-2.26E-11	6.93E-09	1.94E-08	-2.26E-11	-2.26E-11
	StDev	0	0	1.32E-08	2.29E-08	4.85E-15	0
F29: Hartmann 3	*Min*	**-1.78E-05**	-1.78E-05	-1.78E-05	-1.78E-05	-1.78E-05	-1.78E-05
	Ave	-1.78E-05	-1.78E-05	0.000528	0.0005009	0.051517	-1.78E-05
	StDev	0	0	0.001523	0.0017303	0.196121	0
F30: Hartmann 6	*Min*	**0**	0	9.46E-07	4.68E-08	6.22E-15	4.44E-16
	Ave	0.0435941	0.047557	0.041273	0.056637	0.043647	0.11493
	StDev	0.0582734	0.059241	0.059382	0.0745266	0.058344	0.021707
F31: Shekel 5	*Min*	**0**	0	2.36E-06	2.51E-06	6.04E-14	0
	Ave	1.5943441	2.514812	0.81061	0.6767119	4.505148	0
	StDev	2.7615053	3.018129	2.487387	1.7547512	3.369024	0
F32: Shekel 7	*Min*	**-20.80588**	-20.80588	-20.80588	-20.80588	-20.80588	-20.80588
	Ave	-19.2602	-19.6892	-20.62807	-20.80587	-17.28696	-20.80588
	StDev	2.9070217	2.583984	0.970316	1.00E-05	3.854556	3.61E-15
F33: Shekel 10	*Min*	**-2.20E-07**	-2.20E-07	2.50E-06	1.27E-06	-2.20E-07	-2.20E-07
	Ave	0.4489944	1.336099	0.455673	1.37E-05	4.623238	-2.20E-07
	StDev	1.7463952	2.782463	1.750715	6.81E-06	3.885799	5.42E-16
F34: Perm	*Min*	**1.00E-30**	1.81E-26	0.067631	0.0001552	0.00023	0.0049
	Ave	0.042594	0.03135	2.678428	1.3509178	0.046792	0.006286
	StDev	0.095434	0.089466	3.288481	2.9794594	0.093625	0.007091
F35: Goldstein Price	*Min*	**-8.08E-14**	-8.08E-14	5.46E-11	5.70E-10	-7.02E-14	-8.08E-14
	Ave	-7.87E-14	-7.86E-14	5.70E-07	1.04E-06	2.7	-7.93E-14
	StDev	9.85E-16	8.92E-16	1.13E-06	9.80E-07	14.78851	8.83E-16
F36: Langerman 5	*Min*	**0.3033287**	0.303329	0.694475	0.3033287	0.303329	0.303329
	Ave	0.5761576	0.61325	1.021407	0.5757514	0.742518	0.329675
	StDev	0.3298947	0.33584	0.186392	0.3026761	0.321734	0.10027

Table 7: Summary of benchmark function evaluations.

Summary of Benchmark Function Analysis		Evolutionary Algorithms					
		eMuPSO	PSO	WOA	GWO	GOA	DE
Minimum function Evaluations	*Unimodal*	1/2/9	2/2/8	3/2/7	3/2/7	0/1/11	0/2/10
	Multimodal	1/7/4	0/7/5	1/4/7	1/3/8	0/3/10	0/3/9
	Constant Dim.	1/11/0	0/11/1	0/1/11	0/0/12	0/2/10	0/9/3
	Total	3/20/13	2/20/14	4/7/25	4/5/27	0/6/30	0/14/22

Table 8: p-Values obtained from Wilcoxon signed rank pairwise test.

Wilcoxon signed rank pairwise test (Benchmark Functions)	**Evolutionary Algorithms (P_val)**					
	eMuPSO	SPSO	WOA	GWO	GOA	DE
F1: Sphere	1.73E-06	1.73E-06	NA	1.73E-06	1.73E-06	1.73E-06
F2: Beale	NA	1.73E-06	1.73E-06	1.73E-06	1.73E-06	1
F3: Matyas	1.73E-06	1.73E-06	NA	1.73E-06	1.73E-06	1.73E-06
F4: Rosenbrook	1.92E-06	NA	1.73E-06	1.73E-06	1.73E-06	1.92E-06
F5: Step 2	1.73E-06	NA	1.73E-06	1.73E-06	1.73E-06	1.73E-06
F6: Scahffer 2	NA	1	1	1	1	1
F7: Scahffer 3	NA	1	1.73E-06	1.73E-06	1.73E-06	1
F8: Scwefel 1.2	1.73E-06	1.73E-06	1.73E-06	NA	1.73E-06	1.73E-06
F9: Scwefel 2.21	1.73E-06	1.73E-06	1.73E-06	NA	1.73E-06	1.73E-06
F10: Scwefel 2.22	1.73E-06	1.73E-06	NA	1.73E-06	1.73E-06	1.73E-06
F11: De Jong 4	1.73E-06	1.73E-06	NA	1	1.73E-06	1.73E-06
F12: Axis Parallel	1.73E-06	1.73E-06	NA	1.73E-06	1.73E-06	1.73E-06
F13: Easom	NA	1	1.67E-06	1.73E-06	1.98E-07	NaN
F14: Grievant	1.73E-06	5.94E-05	1	1	1.73E-06	1.73E-06
F15: Rastrigin	1.73E-06	1.73E-06	NA	1	1.73E-06	1.73E-06
F16: Pen Holder	NA	1.01E-06	1.73E-06	1.73E-06	1.73E-06	NaN
F17: Test-tube Holder	0.015625	0.015625	1.73E-06	1.73E-06	5.06E-06	1
F18: Egg Holder	NA	0.000375	0.014795	0.002585	4.73E-06	NaN
F19: Damavandi	3.11E-05	3.11E-05	NA	2.84E-05	3.11E-05	3.11E-05
F20: Cross-Leg Table	1.92E-06	1.73E-06	1.73E-06	7.69E-06	1.73E-06	1
F21: Bucking 4	1.73E-06	1.73E-06	1.73E-06	1.73E-06	1.73E-06	1
F22: Bucking 6	NA	0.19861	0.14704	0.010444	0.53044	0.734325
F23: Cosine	NA	1	1	1	1.73E-06	1
F24: Modify Rosenbrook	NA	3.38E-05	1.73E-06	1.73E-06	1.72E-06	1
F25: Paviani	1.73E-06	1	1.73E-06	1.73E-06	1.73E-06	NaN
F26: Modified Ackley	NA	1	2.56E-06	1.73E-06	1.73E-06	1
F27: Camel 6 Hump	NA	1	1.73E-06	1.73E-06	1.50E-06	1
F28: Branin Rcos	NA	1	1.73E-06	1.73E-06	0.007813	1
F29: Hartmann 3	NA	1	1.73E-06	1.73E-06	1.73E-06	1
F30: Hartmann 6	0.877403	0.106394	1	0.544006	0.082206	0.000616
F31: Shekel 5	0.25	0.000244	1.73E-06	1.73E-06	1.73E-06	1
F32: Shekel 7	NA	0.0625	1.73E-06	1.73E-06	1.71E-06	1
F33: Shekel 10	0.375	0.007813	1.73E-06	1.73E-06	1.73E-06	1
F34: Perm	0.24519	0.236936	1.73E-06	0.000388	0.093676	1

F35: Goldstein Price	0.000509	0.004651	1.73E-06	1.73E-06	1.73E-06	1
F36: Langerman 5	0.165027	0.025637	1.73E-06	0.001382	0.000148	1

Table 9: Summary of Wilcoxon signed rank pairwise test evaluations.

Summary of Wilcoxon signed rank pairwise test		**Evolutionary Algorithms**					
		eMuPSO	PSO	WOA	GWO	GOA	DE
Wilcoxon Signed Rank test	*Unimodal*	0/2/10	2/2/8	4/2/6	2/2/8	0/1/11	1/2/9
	Multimodal	2/4/6	0/3/9	1/4/7	0/3/9	0/1/11	2/4/6
	Constant Dim.	0/10/2	1/7/4	0/1/11	0/1/11	0/2/10	1/9/2
	Total	2/16/18	3/12/21	5/7/24	2/6/28	0/4/32	4/15/17

4.3. Convergence Analysis

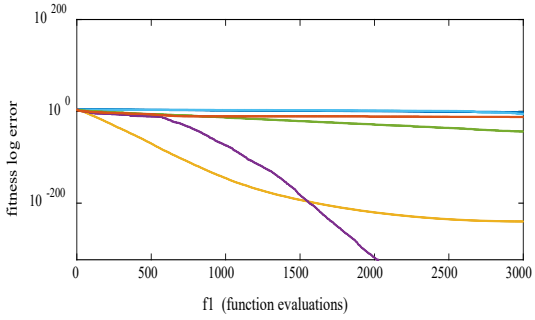
In Figure 5, the unimodal convergence plot analysis, it can be observed that eMuPSO was dominant in f2, tied with other metaheuristics in f6 and f7, produced the second-best minimum solution for f4 and f5, and the third-best minimum solution for f8. A large overshoot can also be observed in the results of f7. In Figure 6, the multimodal convergence plots, eMuPSO dominated others in f10 and tied in f1, f4, f5, f6, f9, f10, f11. eMuPSO produced the second-best minimized solution in f8. All the algorithms except WOA stagnated at the local minimum in f7. In Figure 7, the Constant-dimension benchmark function convergence plots, eMuPSO best minimized all twelve BMF dominating others in f10. Overshoot can be observed in the results of eMuPSO in f25, f28, f29, f33, and f35. Similarly, an overshoot can also be observed in the results of WOA in f28 and f29. In the convergence plots presented, it can be seen that the GOA and GWO produced the slowest converging results, this is most evident in f13, f16, f18, f22, and most of the constant-dimension benchmark functions. This suggests good exploration capabilities of the swarms. Whereas eMuPSO, PSO, and DE produced the fastest converging results. In robot analysis speed and accuracy are required, therefore the best algorithm has to make a balance between fast convergence speed and good exploration tendencies.

5. Conclusion

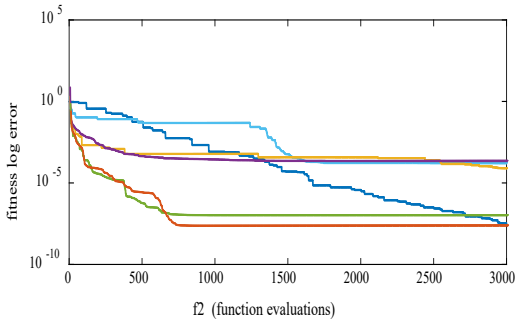
An enhanced PSO algorithm was proposed with a Gaussian mutation and an archive elite learning (eMuPSO), the parameters and dynamic update technique were optimized for robot analysis. The parameter identification problem of robot manipulators is generally computationally demanding, therefore finding a computationally efficient solution is important for which swarm-based algorithms are suitable. In Dynamic parameter estimation, excitation trajectory was optimized using the proposed eMuPSO, it allowed the swarm to be populated with fewer particles thereby requiring less computational time. The mutation function allowed a wholesome search of the solution space. The LLS method was subsequently used to estimate the dynamic parameters of a 6DOF industrial robot manipulator where it was seen that estimation errors are larger at the wrist of the manipulator. Parameter estimation accuracy is better at the arm with less dexterity, while the total estimation accuracy for the 6DOF manipulator is deteriorated by the cumulative estimation errors at the wrist with increased dexterity. This suggests that larger DOF manipulators would be suitable for manufacturing, agriculture and other industrial applications, but may not be suitable for high end application that are keen on accuracy like medical operations, airplane manufacture, etc. The eMuPSO particularly gives a scientist the advantage of choosing from the set of elite particles, the solution that best solves the desired problem. The eMuPSO is therefore capable of analyzing both kinematic and dynamic robot problems of robot manipulators. Further experimentation showed that the landscape of the solution space for robot optimization problems are non-linear, non-separable and multi-modal with multiple possibilities of sub-optimal solutions (local minimizers) some of which are very steep. When a solution approaches the vicinity of these steep local minimizers, it easily falls into its trap and it is usually very difficult to break out of such stagnation without artificial perturbation, resulting to the failure of most evolutionary algorithms in analyzing robot optimization problems. The performance of the eMuPSO can be related to its origins. The parameters of the eMuPSO was derived experimentally by comparing the performance of various robot manipulator configurations as described in [2]. The robot optimization problem is a non-linear and multi-modal problem with dimension size usually less than twenty. Robot manipulator problems are generally computational demanding, therefore a fast convergence time and high accuracy are important. The eMuPSO best showed its veracity in analyzing multimodal and constant dimension BF. PSO and DE performed well in constant dimension BF. It would also be observed that the proposed eMuPSO cannot drill a solution below 1E-30, this is also suitable for robot analysis because the solution to robot optimization problems is usually in the range of 1E-10.

For a successful dynamic parameter estimation analysis, the following should be carefully observed.

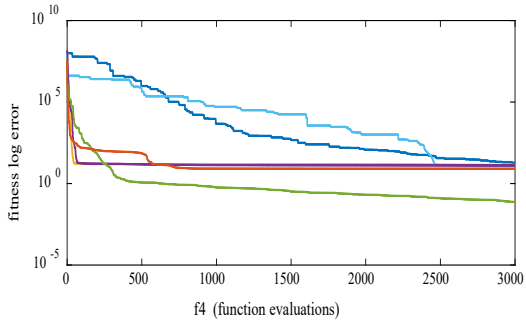
- It is recommended to keep either *d0* or *e0* equal to zero to avoid singularity so that the excitation trajectory is limited two quadrants on the x-y plane.
- It would be keen to differentiate between dynamic and kinematic DH-parameters. When determining the position coordinated of the robot manipulator, the sum of lengths of the 3rd and 4th links make up *d4*, while during dynamic analysis this results in actuator torque errors. Likewise, the length *d6*.



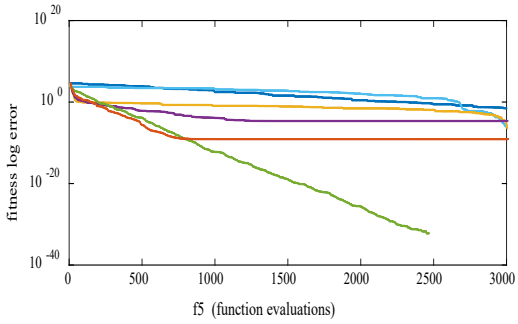
(a) Sphere



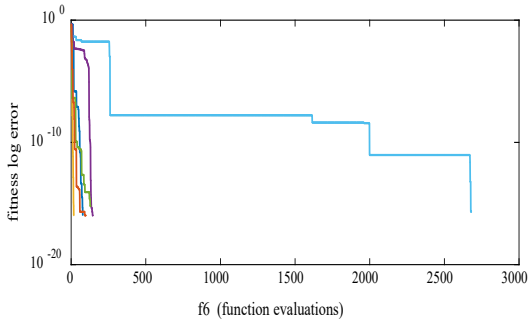
(b) Beale



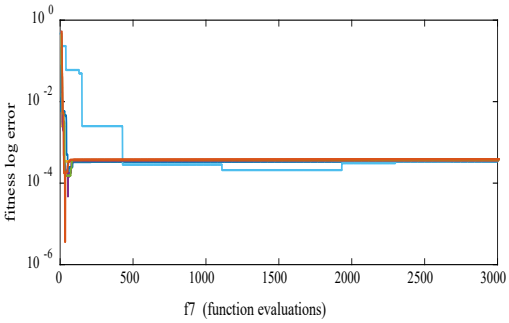
(c) Rosenbrook



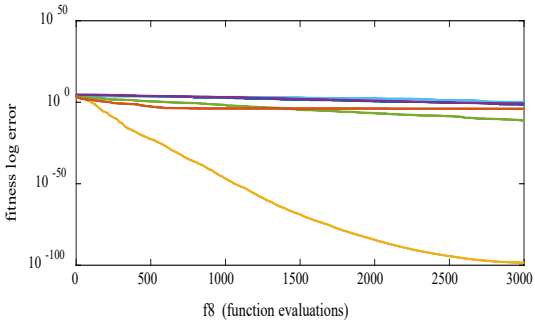
(d) Step 2



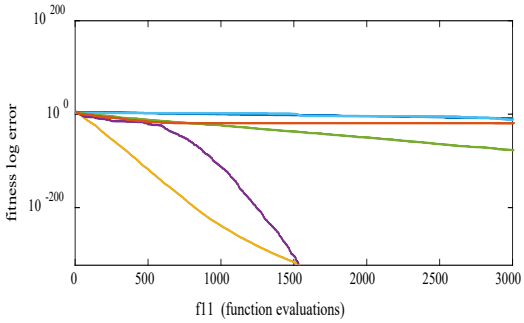
(e) Schaffer 2



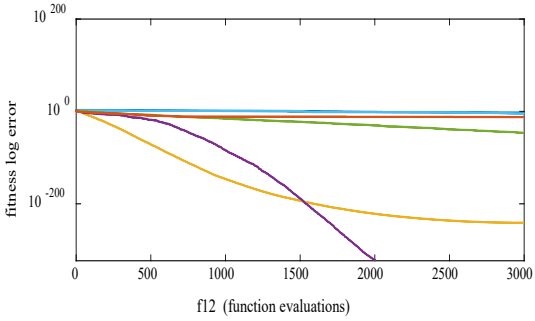
(f) Schaffer 3



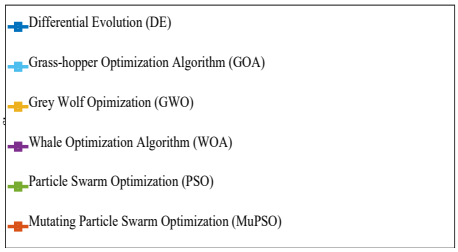
(g) Scwefel 1.2



(h) De Jong 4



(i) Axis Parallel



(j) Legend

Figure 5: Convergence analysis for unimodal benchmark functions.

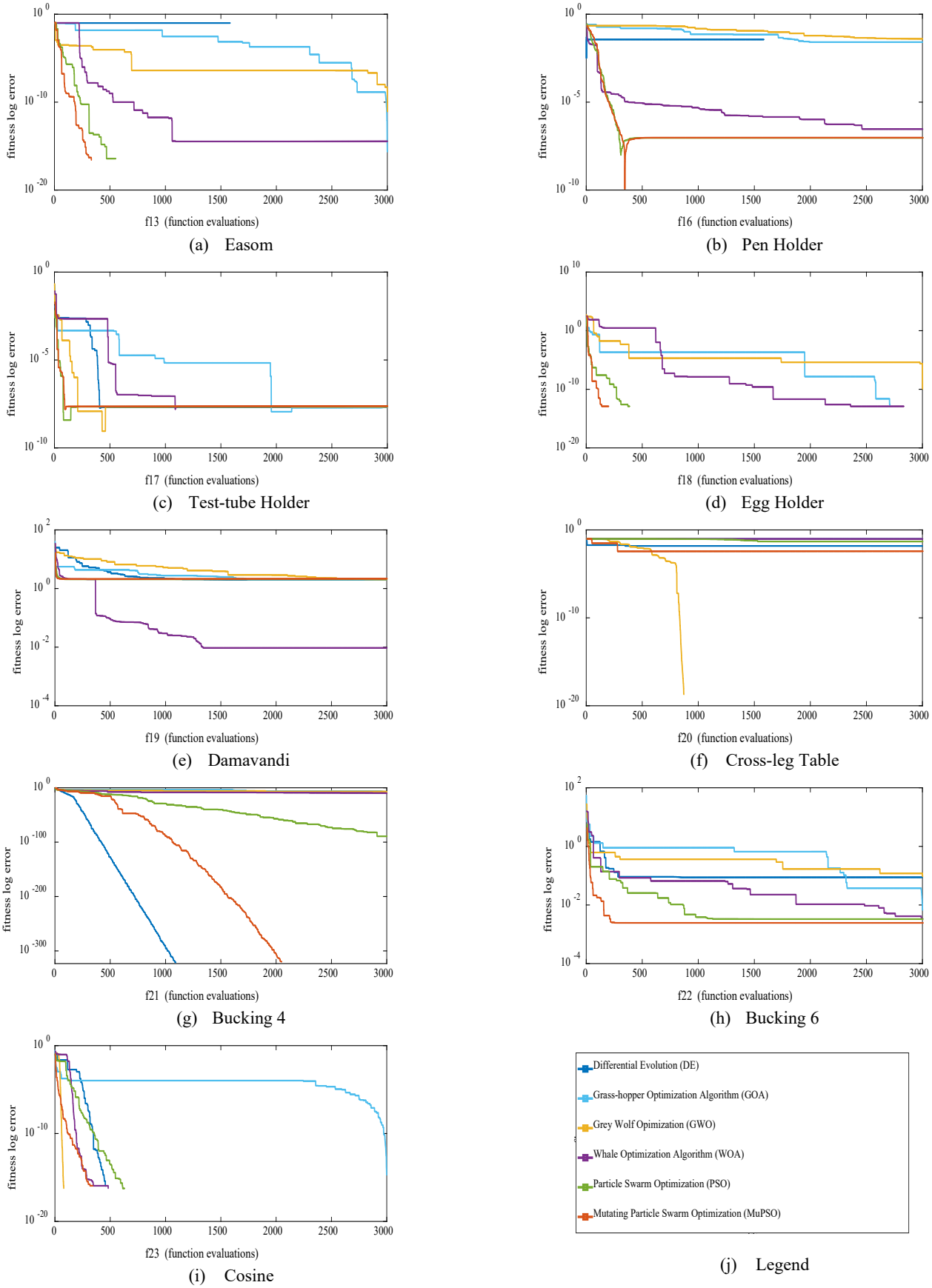
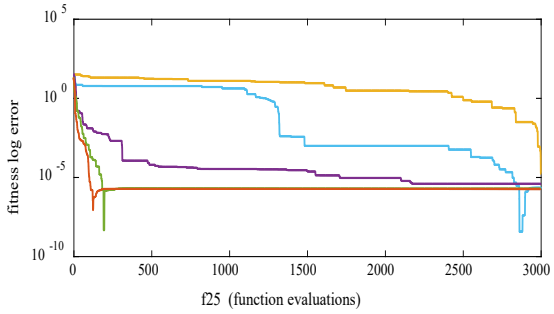
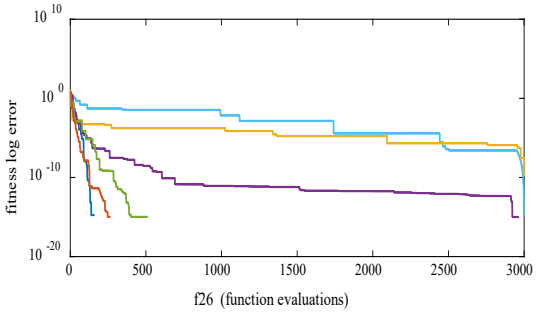


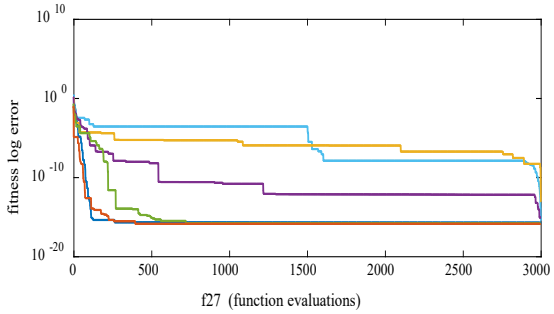
Figure 6: Convergence analysis for multimodal benchmark functions.



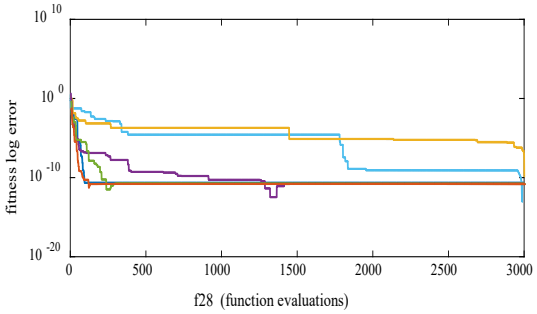
(a) Paviani



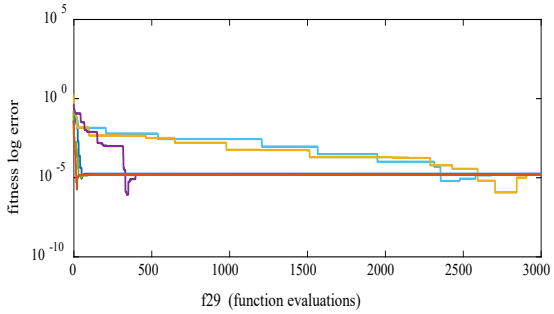
(b) Modified Ackley



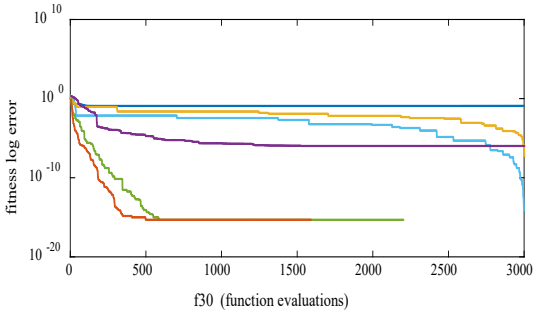
(c) Camel 6 Hump



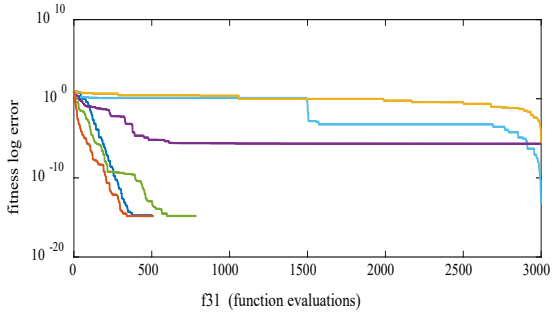
(d) Branin R cos



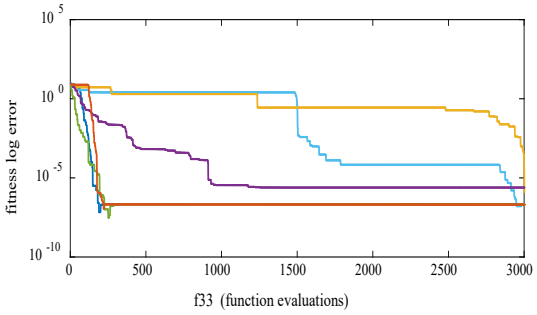
(e) Hartmann 3



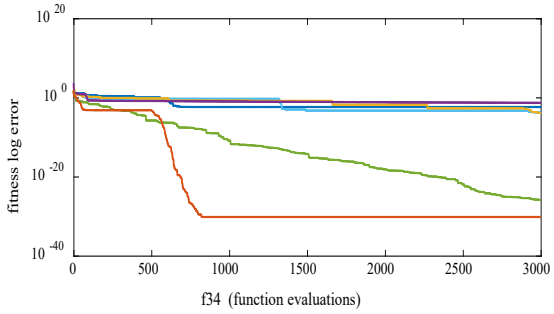
(f) Hartmann 6



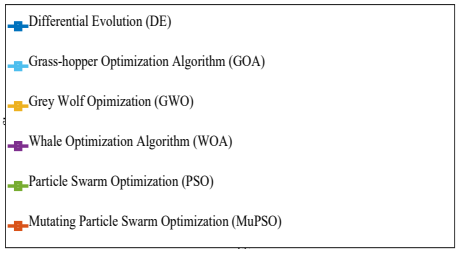
(g) Shekel 5



(h) Shekel 10



(i) Perm



(j) Legend

Figure 7: Convergence analysis for constant-dimension benchmark functions.

- When determining the 6th joint angle from the excitation trajectory vector with (27), the initial joint position is unknown, so it is recommended to assume the initial joint position at the first segment (*@ it = 1*) is equal to the joint position at the last time segment to reduce the overshoot in velocity.

Funding

This research was funded by the National Natural Science Foundation of China, grant number 51875166.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

I would like to first acknowledge my late mum whose unflinching support kept me going against all odds and who unfortunately is no more with us to witness my achievements, my lovely wife and children who were patient with me through these trying times and finally to Prof. Shi for giving me the opportunity and finding me worthy.

References

[1] A. Umar, Z. Shi, W. Wang, Z.I.B. Farouk, "A Novel Mutating PSO Based Solution for Inverse Kinematic Analysis of Multi Degree-Of-Freedom Robot Manipulators," Proceedings of 2019 IEEE International Conference on Artificial Intelligence and Computer Applications, ICAICA 2019, 459–463, 2019, doi:10.1109/ICAICA.2019.8873449.

[2] A. Umar, Z. Shi, A. Khlil, Z.I.B. Farouk, "Developing a New Robust Swarm-Based Algorithm for Robot Analysis," Mathematics, **8**(2), 158, 2020, doi:10.3390/math8020158.

[3] B. Armstrong, "On Finding Exciting Trajectories for Identification Experiments Involving Systems with Nonlinear Dynamics," The International Journal of Robotics Research, **8**(6), 28–48, 1989, doi:10.1177/027836498900800603.

[4] P.K. Khosla, "Categorization of Parameters in the Dynamic Robot Model," IEEE Transactions on Robotics and Automation, **5**(3), 261–268, 1989, doi:10.1109/70.34762.

[5] K. Bouabaz, Q. Zhu, "Improved numerical technique for industrial robots model reduction and identification," in 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), IEEE, Hefei: 1032–1038, 2016, doi:10.1109/ICIEA.2016.7603734.

[6] S.K. Lin, C.J. Fang, "Efficient formulations for the manipulator inertia matrix in terms of minimal linear combinations of inertia parameters," Journal of Robotic Systems, **16**(12), 679–695, 1999, doi:10.1002/(SICI)1097-4563(199912)16:12<679::AID-ROB2>3.0.CO;2-Y.

[7] S.K. Lin, "Minimal Linear Combinations of the Inertia Parameters of a Manipulator," IEEE Transactions on Robotics and Automation, **11**(3), 360–373, 1995, doi:10.1109/70.388778.

[8] J. Swevers, C. Ganseman, J. De Schutter, H. Van Brussel, "Experimental robot identification using optimised periodic trajectories," Mechanical Systems and Signal Processing, **10**(5), 561–577, 1996, doi:10.1006/mssp.1996.0039.

[9] W. Wu, S. Zhu, X. Wang, H. Liu, "Closed-loop dynamic parameter identification of robot manipulators using modified fourier series," International Journal of Advanced Robotic Systems, **9**, 2012, doi:10.5772/45818.

[10] C. Yue, B. Qu, J. Liang, "A Multiobjective Particle Swarm Optimizer Using Ring Topology for Solving Multimodal Multiobjective Problems," IEEE Transactions on Evolutionary Computation, **22**(5), 805–817, 2018, doi:10.1109/TEVC.2017.2754271.

[11] T. Blackwell, J. Kennedy, "Impact of Communication Topology in Particle Swarm Optimization," IEEE Transactions on Evolutionary Computation, **23**(4), 689–702, 2019, doi:10.1109/TEVC.2018.2880894.

[12] U. Bartoccini, A. Carpi, V. Poggioni, V. Santucci, "Memes evolution in a memetic variant of particle swarm optimization," Mathematics, **7**(5), 1–13, 2019, doi:10.3390/math7050423.

[13] B. Tran, B. Xue, M. Zhang, "Variable-Length Particle Swarm Optimization for Feature Selection on High-Dimensional Classification," IEEE Transactions on Evolutionary Computation, **23**(3), 473–487, 2019, doi:10.1109/TEVC.2018.2869405.

[14] S.K.S. Fan, C.H. Jen, "An enhanced partial search to particle swarm optimization for unconstrained optimization," Mathematics, **7**(4), 2019, doi:10.3390/math7040357.

[15] M. Aladeemy, L. Adwan, A. Booth, M.T. Khasawneh, S. Poranki, "New feature selection methods based on opposition-based learning and self-adaptive cohort intelligence for predicting patient no-shows," Applied Soft Computing Journal, **86**, 105866, 2020, doi:10.1016/j.asoc.2019.105866.

[16] W. Liu, Z. Wang, X. Liu, N. Zeng, D. Bell, "A Novel Particle Swarm Optimization Approach for Patient Clustering from Emergency Departments," IEEE Transactions on Evolutionary Computation, **23**(4), 632–644, 2019, doi:10.1109/TEVC.2018.2878536.

[17] S.K. Baliarsingh, W. Ding, S. Vipsita, S. Bakshi, "A memetic algorithm using emperor penguin and social engineering optimization for medical data classification," Applied Soft Computing Journal, **85**, 105773, 2019, doi:10.1016/j.asoc.2019.105773.

[18] D. Chitradevi, S. Prabha, "Analysis of brain sub regions using optimization techniques and deep learning method in Alzheimer disease," Applied Soft Computing Journal, **86**, 105857, 2020, doi:10.1016/j.asoc.2019.105857.

[19] H. Paramesh, "Air Pollution and Allergic Airway Diseases: Social Determinantsand Sustainability in the Control and Prevention," Indian Journal of Pediatrics, **85**(4), 284–294, 2018, doi:10.1007/s12098-017-2538-3.

[20] S. Zhao, D. Pudasainee, Y. Duan, R. Gupta, M. Liu, J. Lu, "A review on mercury in coal combustion process: Content and occurrence forms in coal, transformation, sampling methods, emission and control technologies," Progress in Energy and Combustion Science, **73**, 26–64, 2019, doi:10.1016/j.pecs.2019.02.001.

[21] A. Goudarzi, Y. Li, J. Xiang, "A hybrid non-linear time-varying double-weighted particle swarm optimization for solving non-convex combined environmental economic dispatch problem," Applied Soft Computing Journal, **86**, 105894, 2020, doi:10.1016/j.asoc.2019.105894.

[22] S. Shamshirband, M. Hadipoor, A. Baghban, A. Mosavi, J. Bukor, A.R. Várkonyi-Kóczy, "Developing an ANFIS-PSO model to predict mercury emissions in combustion flue gases," Mathematics, **7**(10), 1–16, 2019, doi:10.3390/math7100965.

[23] N. Moayyeri, S. Gharehbaghi, V. Plevris, "Cost-based optimum design of reinforced concrete retaining walls considering different methods of bearing capacity computation," Mathematics, **7**(12), 2019, doi:10.3390/MATH7121232.

[24] J. Arfaoui, H. Rezk, M. Al-Dhaifallah, F. Elyes, M. Abdelkader, "Numerical performance evaluation of solar photovoltaic water pumping system under partial shading condition using modern optimization," Mathematics, **7**(11), 1–19, 2019, doi:10.3390/math7111123.

[25] H. Rezk, A.A. Elghany, M. Al-Dhaifallah, A.H.M. El Sayed, M.N. Ibrahim, "Numerical estimation and experimental verification of optimal parameter identification based on modern optimization of a three phase induction motor," Mathematics, **7**(12), 2019, doi:10.3390/MATH7121135.

[26] A. V. Kononova, D.W. Corne, P. De Wilde, V. Shneer, F. Caraffini, "Structural bias in population-based algorithms," Information Sciences, **298**, 468–490, 2015, doi:10.1016/j.ins.2014.11.035.

[27] H. Liang, K. Zhang, J. You, H. Yu, "Multi-objective Gaussian particle swarm algorithm optimization based on niche sorting for actuator design," Advances in Mechanical Engineering, **7**(12), 1–7, 2015, doi:10.1177/1687814015620326.

[28] Y. Sun, Y. Gao, "A multi-objective particle swarm optimization algorithm based on Gaussian mutation and an improved learning strategy," Mathematics, **7**(2), 2019, doi:10.3390/math7020148.

[29] S. Cheng, H. Zhan, Z. Shu, "An innovative hybrid multi-objective particle swarm optimization with or without constraints handling," Applied Soft Computing Journal, **47**, 370–388, 2016, doi:10.1016/j.asoc.2016.06.012.

[30] Q. Lin, J. Li, Z. Du, J. Chen, Z. Ming, "A novel multi-objective particle swarm optimization with multiple search strategies," European Journal of Operational Research, **247**(3), 732–744, 2015, doi:10.1016/j.ejor.2015.06.071.

[31] V.K. Patel, V.J. Savsani, "A multi-objective improved teaching-learning based optimization algorithm (MO-ITLBO)," Information Sciences, **357**(June), 182–200, 2016, doi:10.1016/j.ins.2014.05.049.

[32] A. Lin, W. Sun, H. Yu, G. Wu, H. Tang, "Adaptive comprehensive learning particle swarm optimization with cooperative archive," Applied Soft Computing Journal, **77**, 533–546, 2019, doi:10.1016/j.asoc.2019.01.047.

[33] X. Xia, Y. Tang, B. Wei, L. Gui, "Dynamic multi-swarm particle swarm

optimization based on elite learning," IEEE Access, **7**, 184849–184865, 2019, doi:10.1109/ACCESS.2019.2960890.
[34] C.G. Atkeson, C.H. An, J.M. Hollerbach, "Estimation of Inertial Parameters of Manipulator Loads and Links," The International Journal of Robotics Research, **5**(3), 101–119, 1986, doi:doi.org/10.1177/027836498600500306.
[35] J. Kennedy, R. Eberhart, "Particle Swarm Optimization," in Proceedings of the 1995 IEEE International Conference on Neural Networks. Part 1 (of 6), IEEE: 1942–1948, 1995.
[36] S. Mirjalili, A. Lewis, "The Whale Optimization Algorithm," Advances in Engineering Software, **95**, 51–67, 2016, doi:10.1016/j.advengsoft.2016.01.008.
[37] S. Mirjalili, S.M. Mirjalili, A. Lewis, "Grey Wolf Optimizer," Advances in Engineering Software, **69**, 46–61, 2014, doi:10.1016/j.advengsoft.2013.12.007.
[38] S. Saremi, S. Mirjalili, A. Lewis, "Grasshopper Optimisation Algorithm: Theory and application," Advances in Engineering Software, **105**, 30–47, 2017, doi:10.1016/j.advengsoft.2017.01.004.
[39] R. Storn, "Differential evolution design of an IIR-filter," Proceedings of the IEEE Conference on Evolutionary Computation, 268–273, 1996, doi:10.1109/icec.1996.542373.
[40] K. Radkhah, D. Kulic, E. Croft, "Dynamic Parameter Identification for the CRS A460 robot," in IEEE International Conference on Intelligent Robots and Systems, IEEE, San Diego, CA, USA: 3842–3847, 2007, doi:10.1109/IROS.2007.4399314.
[41] D. Kostić, B. de Jager, M. Steinbuch, R. Hensen, "Modeling and identification for high-performance robot control: An RRR-robotic arm case study," IEEE Transactions on Control Systems Technology, **12**(6), 904–919, 2004, doi:10.1109/TCST.2004.833641.

Appendix A1: Unimodal benchmark functions.

Unimodal functions	**Test function**	**Dim**	**Range**	**f_{min}**
F1: Sphere	$f_1(x)=\sum_{i=1}^{n} x_i^2$	30	-100:100	0
F2: Beale	$f_2(x)=\left(1.5-x_i+x_i x_{i+1}\right)^2+\left(2.25-x_i+x_i x_{i+1}^2\right)^2+\left(2.625-x_i+x_i x_{i+1}^3\right)^2$	30	-4.5:4.5	0
F3: Matyas	$f_3(x)=0.26\left(x_i^2+x_{i+1}^2\right)0.48 x_i x_{i+1}$	30	-10:10	0
F4: Rosenbrook	$f_4(x)=\sum_{i=1}^{n}\left[100\left(x_{i+1}-x_i^2\right)^2+\left(x_i-1\right)^2\right]$	30	-30:30	0
F5: Step2	$f_5(x)=\sum_{i=1}^{n}\left(\left\lvert x_i+0.5\right\rvert\right)^2$	30	-100:100	0
F6: Scahffer 2	$f_6(x)=0.5+\frac{\sin^2\left(x_1^2-x_2^2\right)^2-0.5}{1+0.001\left(x_1^2-x_2^2\right)^2}$	30	-100:100	0
F7: Scahffer 3	$f_7(x)=0.5+\frac{\sin^2\left(\cos\left\lvert x_1^2-x_2^2\right\rvert\right)^2-0.5}{1+0.001\left(x_1^2-x_2^2\right)^2}$	30	-100:100	0.0015
F8: Schwefel 1.2	$f_8(x)=\sum_{i=1}^{D}\left(\sum_{j=1}^{D} x_j\right)^2$	30	-100:100	0
F9: Schwefel 2.21	$f_9(x)=\max_{1\le i\le D}\left\lvert x_i\right\rvert$	30	-100:100	0
F10: Schwefel 2.22	$f_{10}(x)=\sum_{i=1}^{D}\left\lvert x_i\right\rvert+\prod_{i=1}^{n}\left\lvert x_i\right\rvert$	30	-10:10	0
F11: De Jong	$f_{11}(x)=\sum_{i=1}^{D} i x_i^4$	30	-5.12:5.12	0
F12: Axis Parallel	$f_{12}(x)=\sum_{i=1}^{D} i x_i^2$	30	-5.12:5.12	0

Appendix A2: Multimodal benchmark functions.

Multimodal functions	Test function	Dim	Range	f_{min}
F13: Easom	$f_{13}(x)=-\cos(x_i)\cos(x_{i+1})\exp\left(-(x_i-\pi)^2-(x_2-\pi)^2\right)$	30	-100 : 100	-1
F14: Griewank	$f_{14}(x)=\sum_{i=1}^{n}\frac{x_i^2}{4000}-\prod_{i=1}^{n}\frac{\cos(x_i)}{\sqrt{i}}+1$	30	-600:600	0
F15: Rastrigin	$f_{15}(x)=\sum_{i=1}^{n}\left[x_i^2-10\cos(2\pi x_i)+10\right]$	30	-5.12:5.12	0
F16: Test-tube Holder	$f_{17}(x)=-4\left\lvert\exp\left(\left\lvert\cos\left(\frac{x_i^2+x_{i+1}^2}{200}\right)\right\rvert\sin(x_i)\cos(x_{i+1})\right)\right\rvert$	30	-10:10	-10.87
F17: Egg Holder	$f_{18}(x)=\sum_{i=1}^{n-1}x_{i+1}+47\sin\left(\sqrt{\left\lvert\frac{x_i}{2}+x_{i+1}\right\rvert}\right)-x_i\sin\left(\sqrt{\lvert x_i-x_{i+1}-47\rvert}\right)$	30	-512:512	959.64
F18: Damavandi	$f_{19}(x)=\left(\sum_{i=1}^{n}\left(1-\left\lvert\frac{[\sin(\pi x_i-2)\sin(\pi x_{i+1}-2)]}{\pi^2(x_i-2)(x_{i+1}+1)-2}\right\rvert\right)^5 * \left(2+(x_i-7)^2+2(x_{i+1}-7)^2\right)\right)/n$	30	0:14	0
F19: Cross-leg Table	$f_{20}(x)=\sum_{i=1}^{n}\frac{-1}{\left(\left\lvert\exp\left(\left\lvert 100-\frac{\sqrt{x_i^2+x_{i+1}^2}}{\pi}\right\rvert\right)\sin(x_i)\sin(x_{i+1})\right\rvert+1\right)^{0.1}}$	30	-10:10	0
F20: Bucking 4	$f_{21}(x)=100x_{i+1}^2+0.01\lVert x_i+10\rVert$	30	-15:-5 -3:3	0
F21: Bucking 6	$f_{22}(x)=100\sqrt{\lVert x_{i+1}-0.01x_i^2\rVert}+0.01\lVert x_i+10\rVert$	30	-15:-5 -3:3	0
F22: Cosine	$f_{23}(x)=-0.1\sum_{i=1}^{n}\cos(5\pi x_i)-\sum_{i=1}^{n}x_i^2$	30	-1:1	0.4
F23: Modified Rosenbrook	$f_{25}(x)=74+100(x_{i+1}-x_i^2)+(1-x)^2-400\exp\frac{-(x_i-1)^2+(x_{i+1}+1)^2}{0.1}$	30	-2:2	0
F24: Schwefel 2.26	$f_{26}(x)=-\frac{1}{D}\sum_{i=1}^{D}x_i\sin\sqrt{\lvert x_i\rvert}$	30	-500:500	-418.98

Appendix A3: Constant-dimension benchmark functions.

Constant-Dimension functions	Test function	Dim	Range	f_{min}
F25: Paviani	$f_{27}(x)=\sum_{i=1}^{10}\left[(\ln(x_i-2))^2+(\ln(10-x_i))^2\right]-\left(\prod_{i=1}^{10}x_i\right)^{0.2}$	10	2.00:10	-45.778

F26: Modified Ackley	$f_{28}(x)=\sum_{i}^{D}\left(\exp^{-0.2}\sqrt{x_i^2+x_{i+1}^2}+3\left(\cos(2x_i)+\sin(2x_{i+1})\right)\right)$	2	-35:35	-3.9172
F27: Camel 6 Hump	$f_{29}(x)=\left(4-2.1x_i^2+\frac{x_i^4}{3}\right)x_i^2+x_ix_{i+1}+\left(4x_{i+1}^2-4\right)x_{i+1}^2$	2	-5:5	-1.0316
F28: Branin Rcos	$f_{30}(x)=\left(x_{i+1}-\frac{5.1x_i^2}{4\pi^2}+\frac{5x_i}{\pi}-6\right)^2+10\left(1-\frac{1}{8\pi}\right)$ $*\cos(x_i)\cos(x_{i+1})\ln\left(x_i^2+x_{i+1}^2+1\right)+10$	2	-5:10 0:15	0.3978
F29: Hartman 3	$f_{31}(x)=-\sum_{i=1}^{4}c_i\exp\left[-\sum_{j=1}^{3}a_{ij}\left(x_j-p_{ij}\right)^2\right]$	3	1:3	-3.8627
F30: Hartman 6	$f_{32}(x)=-\sum_{i=1}^{4}c_i\exp\left[-\sum_{j=1}^{6}a_{ij}\left(x_j-p_{ij}\right)^2\right]$	6	0:1	-3.3223
F31: Shekel 5	$f_{33}(x)=-\sum_{i=1}^{5}\frac{1}{\sum_{j=1}^{4}\left(x_j-a_{ij}\right)^2+c_i}$	4	0:10	-10.149
F32: Shekel 7	$f_{34}(x)=-\sum_{i=1}^{7}\frac{1}{\sum_{j=1}^{4}\left(x_j-a_{ij}\right)^2+c_i}$	4	0:10	-10.399
F33: Shekel 10	$f_{35}(x)=-\sum_{i=1}^{10}\frac{1}{\sum_{j=1}^{4}\left(x_j-a_{ij}\right)^2+c_i}$	4	0:10	-10.531
F34: Perm	$f_{36}(x)=\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left(j^i+0.5\right)\left(\frac{x_j}{j}\right)^{i-1}\right]^2$	4	-4:4	0
F35: Goldstein Price	$f_{37}(x)=$ $\left[1+\left(x_i+x_{i+1}+1\right)^2\left(19-14x_i+3x_i^2-14x_{i+1}+6x_ix_{i+1}+3x_{i+1}^2\right)\right]*$ $\left[30+\left(2x_i-3x_{i+1}\right)^2\left(18-32x_i+12x_i^2+48x_{i+1}+36x_ix_{i+1}+27x_2^2\right.\right.$	2	-2:2	3
F36: Langerman 5	$f_{38}(x)=-\sum_{i=1}^{m}c_i\exp\left(-\frac{1}{\pi}\right)\sum_{j=1}^{D}\left(x_j-a_{ij}\right)^2\cos\left(\pi\sum_{j=1}^{D}\left(x_j-a_{ij}\right)^2\right)$	10	0:10	-1.4

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1458-1465 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Comparative Study Between Three Methods for Optimizing the Power Produced from Photovoltaic Generator

El hadji Mbaye Ndiaye[1,*], Mactar Faye[1,2], Alphousseyni Ndiaye[1,2]

[1]*Research team Efficiency and Energy System, Alioune Diop University of Bambey, BP 30, Senegal*

[2]*Laboratory of Water, Energy, Environment and Industrial Processes, ESP, Dakar-Fann, S-10700, Senegal*

A R T I C L E I N F O

Article history:
Received: 17 October, 2020
Accepted: 26 November, 2020
Online: 21 December, 2020

Keywords:
MPPT
PSO
ANFIS
Photovoltaic Generator
Artificial Intelligence

A B S T R A C T

The technics of maximum power point tracking is widely used in solar photovoltaic energy and electric power system applications. Traditionally, these technics are based on conventional methods like perturb and observe and incremental conductance. In this work, three methods based on particle swarms optimization, incremental conductance and adaptive neuro fuzzy inference system are presented. A comparative study is carried out. The study of this paper shows that there is a limitation in the incremental conductance method. To overcome the shortage of this last method, particle swarms and adaptive neuro fuzzy optimization methods are used. The behaviors of the three methods are compared and evaluated in simulation under matlab/simulink. Results demonstrate that the adaptive neuro fuzzy inference system is effective for photovoltaic power optimization even for non-uniform climatic conditions. It has the best performances followed by the particle swarms method.

NOMENCLATURE

ANFIS	Adaptive Neuro Fuzzy Inference System	Ipv [A]	PV current
PVS	Photovoltaic Solar	RMSE [-]	Root Mean Square Error
MPP	Maximum Power Point	MSE [-]	Mean Square Error
InC	Incrementale Conductance	MAPE [%]	*Mean Absolute Percentage Error*
PI	Proportional Integral	AIC [-]	Akaike Information Criterion
PID	Proportional Integral Derivated	C1 [µF]	Boost input capacity
MsF	Membership	C2 [µF]	Boost output capacity
MPPT	MPP Tracking	f [kHz]	Frequency of switching the Mosfet
P&O	Perturb and Observ	α [-]	Duty cycle
PVM	Photovoltaic Module	L [mH]	Inductance
GA	Genetic Algorithm	Vpv [V]	PV voltage
FIS	Fuzzy Inference System	Pref [W]	Mesured Power
FL	Fuzzy Logic	Imax [A]	Current at Maximum Power
STC	Standard Test Condition	Vmax [V]	Voltage at Maximum Power
PWM	Pulse Width Modulation	Popt [W]	Optimized Power with ANFIS
PSO	Particle Swarm Optimization	m [-]	coefficient of inertia
ANN	Artificial Neural Network	Q_i [-]	acceleration coefficient
ACO	Ant Colony Optimization	r_i [-]	random number
AI	Artificial Intelligence	Ir [W/m²]	Irradiation
DC	Direct Current	T [°C]	Temperature
Ppv [W]	Photovoltaic power	F [W]	Fitness function
		R_{MPPT} [%]	MPPT efficiency

*Corresponding Author: El hadji Mbaye Ndiaye, Université Alioune DIOP de Bambey, elhadjimbaye.ndiaye@uadb.edu.sn

1. Introduction

Today's world sees rapid industry development and this makes us more energy dependent. Fossil sources occupy the largest share of electricity production [1], [2]. The use of conventional energies (around 87% of global energy consumption) has an undesirable impact on the environment in terms of greenhouse gas (GHG) emissions and safety (nuclear accidents). 36.5 billion tons of CO_2 is emitted in 2020 [3]. To overcome these problems, it is necessary to resort to alternative energies [4–9]. Among them is Photovoltaic Modules (PVM) solar energy. Its annual growth rate over the last ten years is estimated at more than 40% [10]. Despite their low efficiency, PVM still have a high price. To get around this efficiency problem, techniques for optimizing the power generated by the PVM are proposed [11], [12–16]. This technique, makes it possible to run after the maximum power that the module is capable of supplying [17]. The generated power depends on weather conditions irradiation and temperature [18–20]. Under these latter conditions, the electrical characteristics of the PVM have only one optimal. So not too much trouble for the technique to converge the system to the MPP. Under non-uniform conditions of irradiation and temperature, the electrical characteristics of the PVM present several optimal points. The algorithm used must therefore be able to distinguish the global optimum from the local optima [21]. It therefore requires a sophisticated algorithm which will be able to make a global exploration of the search space in order to make the system converge towards the global MPP [22].

The most used of the classical methods are the P&O method and the InC method [5], [23–26]. Several researchers have used the P&O technique in their work. This method, based on the voltage disturbance and the observation of the variation of the power, is very widespread in the literature. Originally, it was designed to exceed the limits of other types of deterministic controls such as Hill Climbing, Open Circuit Voltage Fraction, etc. It also presents limits linked mainly to the response time and the numerous oscillations around the MPP. The InC method is proposed [3, 18, 25]. In [27], according to the results, the incremental conductance algorithm performs better than the perturb and observe algorithm. To overcome the problems linked to the limits of the methods mentioned above under variable of climatic conditions, Soft-Computing methods based on Meta-heuristic algorithms and Artificial Intelligence (AI) algorithms are proposed. MPPT techniques are based almost exclusively on these techniques. They are very numerous and diverse. They range from Evolutionary Algorithms (Genetic Algorithm), Meta-heuristic algorithms (Optimization by Particle Swarms) and AI algorithms (Artificial Neural Networks, Fuzzy Logic, ANFIS) [21], [10, 11, 28–31].

Artificial Neural Networks (ANN) have been used in several works to optimize the power delivered by a PVM. Its operating principle is inspired by that of the human brain. With their great generalization capacity, they are used for solving complex optimization problems [32], [33]. This is the case in [34] where MPPT method based on ANN is compared to MPPT method based on P&O. The results show that ANN method is more robust than P&O method, regardless of the operating conditions of the PVM. Its limits lie in its lack of interpretation and the difficulty of determining the appropriate number of Layer/Neurons. In addition, these latter limits represent strong points for the Fuzzy Logic (FL) algorithm. The interpretive inability encountered with ANN is resolved by the use of LF. It uses its linguistic variables to overcome this problem [35], [36]. Using the classical logic process, it has facility for extension and interaction [37], [38].

FL also has limits which can be circumvented. Which makes them two complementary techniques and the combination of which gives the Neuro-Fuzzy technique including ANFIS.

In reference [39], two ANFIS models are proposed for grid-connected PV system current injection and battery control. The results show that the proposed the models offered allow the battery charging/discharging process to be supervised and at the same time injecting good quality energy into the grid.

In [40] two MPPT techniques based on ANFIS and P&O are compared. Comparison results show that the technique based on the ANFIS algorithm is more robust.

In [5] Five MPPT techniques are proposed. A comparative study is carried out between them. The simulation results show that the best performances are obtain with ANFIS with an efficiency of 99.4%, against 98.1% and 97.5% respectively for FL and P&O.

The power optimization technique using the ANFIS algorithm is more robust than other techniques such as FL, ANN, InC and P&O. It overcomes the problems encountered with ANN and FL as it is a complementary technique linking the two [28, 38, 41].

Other types of techniques based on algorithms whose principle is inspired by the evolution of nature are presented in the literature [11, 42]. Among them there are methods based on the Particle Swarms Optimization (PSO) algorithm [43, 44].

In [45], a comparative study between techniques using PSO, InC and P&O is carried out. Results of simulation show that the PSO technique is the more robust with the most low response time compared to the two others techniques (InC and P&O). In [46], authors proposed a comparative study between PSO, P&O and FL techniques to optimize the MPP of the PVM. Results show that the MPPT technique based on PSO outperformed FL and P&O techniques.

The work in this paper is consisting to do a comparative study between MPPT techniques based on ANFIS, PSO and InC algorithms put under the same conditions in order to indicate the most efficient and the most robust for power optimization problems. A validation of these three commands is done using an experimental database. The scientific contribution of this paper is to design a MPPT technique based on ANFIS algorithm using database collected to solar power plant in tropical zone.

The rest of the article is organized as follows. Section II presents the MPPT proposed approaches. Section III the proposed approach for the MPPT techniques. Section IV presents the results of simulations and discussions. Finally, a conclusion is made in section V.

2. Proposed approach

In solar PV system, power delivered by the PVM is not always the maximum. This is due to the phenomenon of intermittence. As a result, the operating point of the PVM is not even the MPP. It so requires a technique which is able to extract the MPP of the PVM. This technique, called MPPT optimizes the power through the generation of a duty cycle for controlling the static converter.

The block diagram of the studied system is given in Figure 1.

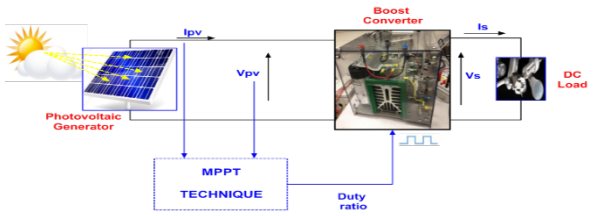


Figure 1: Schematic diagram of the MPPT technique

2.1. Incremental Conductance method based MPPT

The InC algorithm is among the most efficient of the classical algorithms. It is based on the cycle of the change in pressure to the change in voltage of the PVM (equation 1). At MPP, the slope is zero. The tracking is done according to the position of the operational point (or slope) (dPpv / dVpv) relative to the MPP (equation (3)). The latter depends on the value of the conductance (Ipv / Vpv). The sign of the latter indicates whether the MPP is reached or not (equation (2)). It is compared to its increment. This amounts to saying that the MPP depends on the voltage variation and that of the current [18,26],[47,48]. Thus, the algorithm increments or decrements the duty cycle of the static converter to continue the MPP. The flowchart is shown in Figure 2.

$$\frac{dP_{pv}}{dV_{pv}} = I_{pv} + V_{pv}\frac{dI_{pv}}{dV_{pv}} \quad (1)$$

The MPP is found when:

$$\frac{dP_{pv}}{dV_{pv}} = 0 \rightarrow \frac{dI_{pv}}{dV_{pv}} = -\frac{I_{pv}}{V_{pv}} \quad (2)$$

The sign of the slope indicates the direction of evolution of the MPP according to equation (3).

$$\begin{aligned} &\frac{dP_{pv}}{dV_{pv}} > 0, left\ side\ of\ MPP \\ &\frac{dP_{pv}}{dV_{pv}} < 0, for\ V_{pv} > V_{MPP}, right\ side\ of\ MPP \\ &\frac{dP_{pv}}{dV_{pv}} = 0, on\ the\ MPP \end{aligned} \quad (3)$$

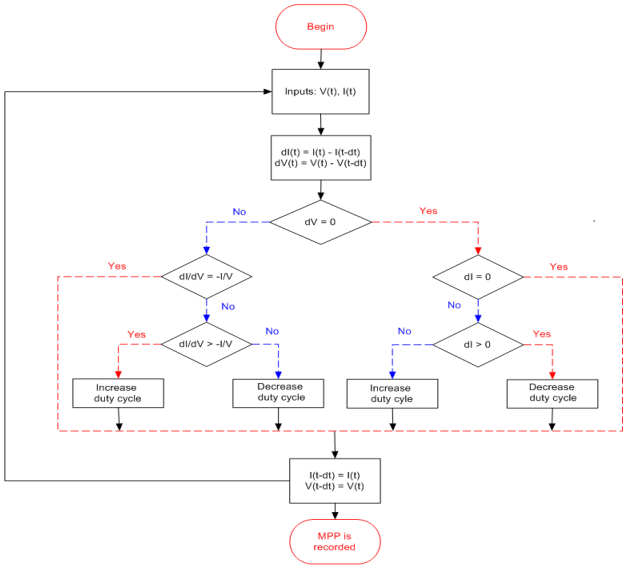


Figure 2: Incremental Conductance flowchart based MPPT

2.2. PSO method based MPPT

Particle Swarm Optimization is an evolutionary meta-heuristic approach. It is used for solving optimization problems. Its principle is based on the behavior of particles (individuals) [49]. In this paper we use a swarm of birds (Figure 3). In this type of swarm, collective intelligence is involved. Particles converge towards those with the best performance [50].

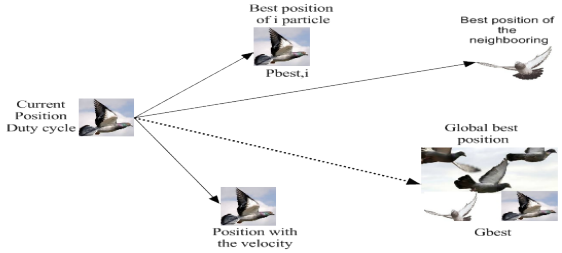


Figure 3: Movement of a particle in the swarm of birds

The duty cycle α of the boost converter represents the particles. Velocity and position are initialized to begin and the movement of the particles are described by equations (4) and (5). For evaluating the best position of the particles, fitness function is calculated according to the equation (6). Vpv and Ipv are calculated for each particle i with a fixed position in the search space [αmin, αmax]. The algorithm converges the system towards the global optimum. For this the duty cycle is initiated. This ratio, depending on P_{besti} and G_{best}, is corrected if it deviates from the best overall duty cycle as in the principle of Figure 3. The particles are then evaluated in terms of position and velocity according to equations (7) and (8). Updates are made to re-evaluate the optimum duty cycle for controlling the boost converter. Table 1 gives the parameters of the PSO. These parameters are defined for the PSO simulations. The flowchart is shown in Figure 4.

Table 1: PSO implementation parameters

Parameters	Values
Q_1	1,2000
Q_2	2
m	0,4000
Iterations	25

$$\alpha_i^{t+1} = \alpha_i^t + \left(\frac{d\alpha}{dt}\right)_i^{t+1} \quad (4)$$

$$\left(\frac{d\alpha}{dt}\right)_i^{t+1} = w\left(\frac{d\alpha}{dt}\right)_i^t + Q_1 r_1 (P_{besti} - \alpha_i^t) + Q_2 r_2 (G_{best} - \alpha_i^t) \quad (5)$$

$$F_i(\alpha^t) = P_{pv,i}^t \quad (6)$$

$$P_{besti} = \alpha_i^t \quad if \quad F_i(\alpha^t) \geq F_i(P_{besti}) \quad (7)$$

$$G_{best} = \max(P_{besti}) \quad (8)$$

F is the fitness function,

α is the duty cycle of the boost converter,

t is the time of simulation

r_1 and r_2: uniformly pull in [0,1].

w: coefficient of inertia.

Qi: acceleration coefficient

P_{besti} is the personal best position of particle *i*

G_{best} is the best position of the particles in the entire population.The MPPT technique based on PSO algorithm is described by equations (4) to (8).

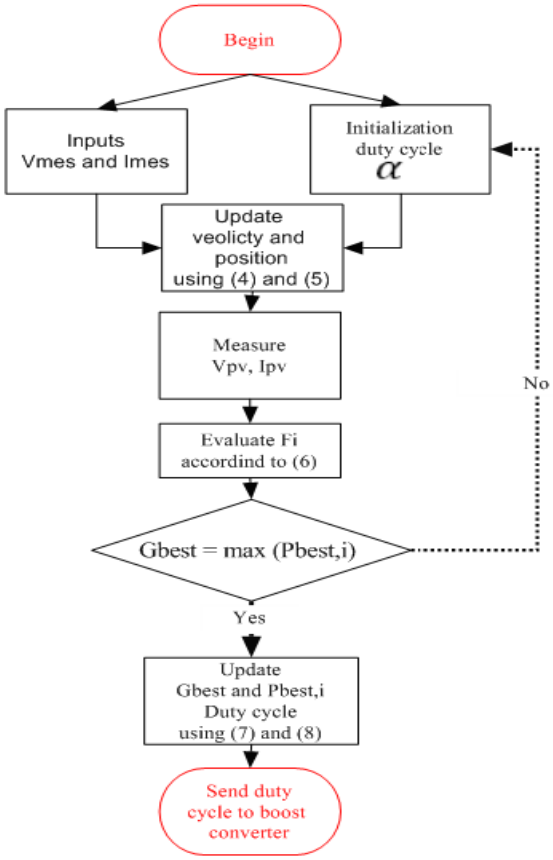


Figure 4: PSO flowchart based MPPT

2.3. ANFIS method based MPPT

ANFIS is an adaptive neuro-fuzzy inference system that has five layers in its structure. These layers refine the fuzzy rules already established by human experts and readjust the overlap between the different fuzzy subsets [51–54].

The neural structure replaces the hidden layers with fuzzy rules. This further simplifies learning and interpreting the results obtained. The structure proposed in this work receives the voltage and current from the PVM as inputs and supplies the output with an optimal power comparable to that of the PVM.

The architecture of the MPPT ANFIS technique is given in Figure 5.

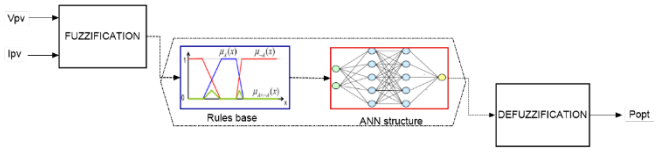


Figure 5. Architecture of the MPPT ANFIS technique

Layer 1: Each Neuron calculates the degree of truth of a fuzzy subset by its transfer function. It is called fuzzification layer. Numeric input values are converted to linguistic variables (equations 9 and 10) for high interpretability.

$$O_i^1 = \begin{cases} \mu_{Ai}(V_{pv}) & for \quad i = 1,2 \\ \mu_{B(i-2)}(I_{pv}) & for\ i = 3,4 \end{cases} \tag{9}$$

With:

$$\mu_{Ai}(V_{pv}) = exp\left[-\left(\frac{Vpv-a_i}{b_i}\right)^2\right] \tag{10}$$

Layer 2: It calculates the degree of activation of antecedents (premises). This is the layer of fuzzy rules (equation 11).

$$w_i = \mu_{Ai}(V_{pv}) * \mu_{Bi}(I_{pv}) \tag{11}$$

Layer 3: It normalizes the degree of activation of the rules: it is the normalization layer (equation 12).

$$G_i = \frac{w_i}{w1+w2} \tag{12}$$

Layer 4: It determines the parameters of the consequence of the fuzzy rules. Previous linguistic variables will be translated again into numeric values before being sent to the last layer. It is defuzzification (equation13).

$$O_i^4 = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{13}$$

Layer 5: It calculates the overall output of the system: it is the output layer. Equation (14) gives the expression of the optimal power generated by the PVM with the ANFIS technique for three fuzzy subsets.

$$P_{opt} = \sum_{i=1}^{nit}\left\{P_{opt,i}.\frac{e^{-\left[\left(\frac{Vpv-a_i}{b_i}\right)^2+\left(\frac{Ipv-a_{i-2}}{b_{i-2}}\right)^2\right]}}{\sum_{i=1}^{3}\left(e^{-\left[\left(\frac{Vpv-a_i}{b_i}\right)^2+\left(\frac{Ipv-a_{i-2}}{b_{i-2}}\right)^2\right]}\right)}\right\} \tag{14}$$

where ai and bi are the parameters of the premise of the MsF, and Popt is the optimal power delivered by the ANFIS controller.

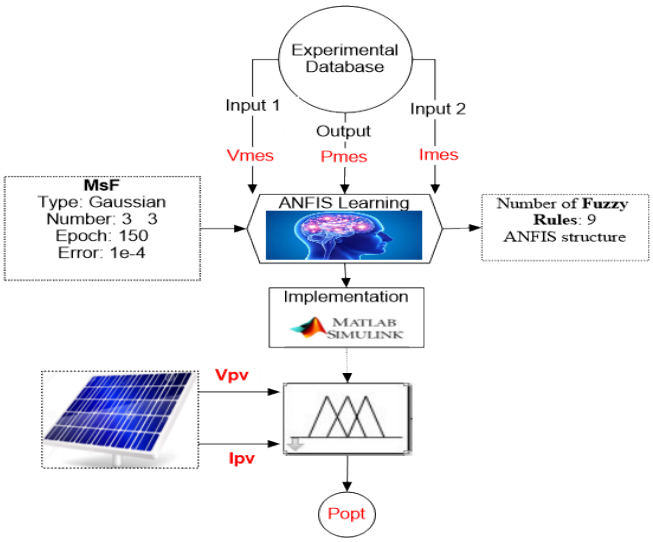
With nit the number of iterations when learning the ANFIS algorithm. Figure 6 gives the flowchart of the technique implemented in simulink and table 2 the learning parameters. The matlab "anfisedit" interface is used for learning the MPPT ANFIS command with a real database made up of two inputs (Vpv and Ipv) and one output. The number and type of MsF are fixed as well as the number of iterations. The hybrid learning algorithm is used and an error tolerance of 1e-4 is arbitrarily set.

Table 2 : ANFIS learning parameters

Parameters	Values
FIS	Takagui-Sugeno
MsF (Type)	Gaussian
MsF (Number)	3 3
Fuzzy rules	9
Epochs	10
RMSE (Training)	0.000123
RMSE (Checking)	0.002012

3. Simulation Results and Discussions

In this section, the results of simulations under matlab/simulink are presented. These simulations are performed with a real database. The latter first allowed to digitally characterize the PVM before being used for the validation of the three techniques presented in previous sections. The PVM consists of two PVs in series. The photovoltaic platform shown in Figure 7 is used in this study. The characteristics of the photovoltaic system being identified (Sharp Module) are given in Table 3. This platform is located at the Polytechnic high school of Cheikh Anta Diop University, Dakar, Senegal. This country is a tropical zone with an adequate rate of sunshine (5.7 kWh/m²/day) for the installation of solar PV plant.



ANFIS flowchart based MPPT

Table 3: Characteristics of the PV

Parameters	Values
V_{CO}	20 V
V_{MPP}	16 V
I_{SC}	2.5600 A
I_{MPP}	2.4300 A
P_{MPP}	38.3800 W

Experimental bench

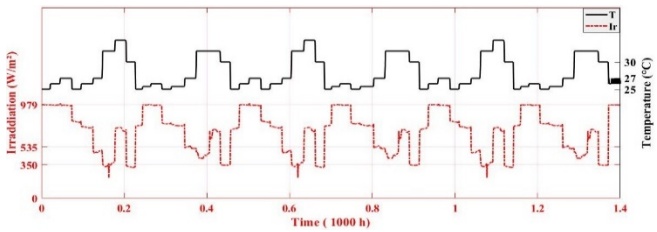


Figure 8: Experimental database of irradiation and temperature

In this subsection, the results obtained by the three MPPT methods are visualized. A comparative study is then carried out in order to detect the best order. Table 4 shows the simulation parameters and the Figure 9 gives the Simulink model.

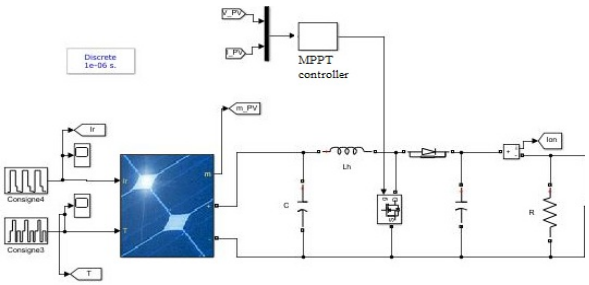


Figure 9: System simulated under matlab/simulink

Table 4: System electrical parameters

Parameters	Values
Input capacity	200.6230 µF
Inductance	1 mH
Output capacity	480 µF
Load	34.8000 Ω
Switching frequence	15 kHz

The RMSE and MAPE criteria are evaluated at the level of equations (17) and (18). They characterize the difference between the power generated by the control and the real power (Figure 14). The efficiency of the MPPT technique is given by equation (19).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_{ref} - P_{pv_i}\right)^2} \quad (17)$$

$$MAPE = 100 * \frac{1}{n}\sum_{i=1}^{n}\left|\frac{P_{ref} - P_{pv_i}}{P_{ref}}\right| \quad (18)$$

$$E_{MPPT}(\%) = \frac{P_{PV,MPPT}}{P_{Mesure}} \quad (19)$$

Figure 12 and Figure 13 respectively show the currents and voltages of the PVM obtained with the different MPPT techniques. They follow changes in weather conditions (Figure 8) which have a considerable influence on the point of operation of the PVM. In Figure 12, we see that the currents obtained with the PSO and ANFIS techniques are almost identical while that obtained with the InC technique is much lower. The opposite phenomenon is observed on the voltage curves in Figure 13. Indeed, the MPPT InC technique pursues the MPP first by comparing the voltage and the current of the PVM. Then, as the principle is based on conductance, the control performs compensation to achieve the desired MPP. Only, as it evaluates each time the variation of the power compared to the voltage, it often diverges with an enormous overshoot of the voltage. In addition, being also a static control, it takes a relatively long time to adapt to a climatic disturbance. This induces a small and slow variation in its duty cycle (Figure 10). As a result, it has difficulty extracting maximum power for these non-uniform irradiation and temperature conditions. Figure 11 represents the power curves of the three controls with respect to the measured power (reference). It reveals overruns for the PSO

and InC techniques while the ANFIS technique follows the set-point with an accuracy of 93.46%.

Furthermore, the performance criteria presented in Figure 14 show that the ANFIS technique is better with an extremely low RMSE (0.0194), no overshoot (D=0) and a higher efficiency of around 99.9984%. It is followed by the PSO technique with a RMSE of 1.7235 and an efficiency of 94.8748%.

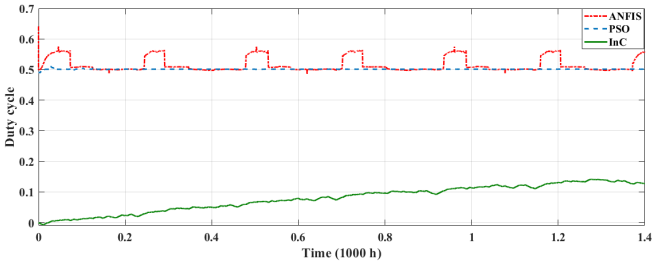


Figure 10: Variation in the duty cycle

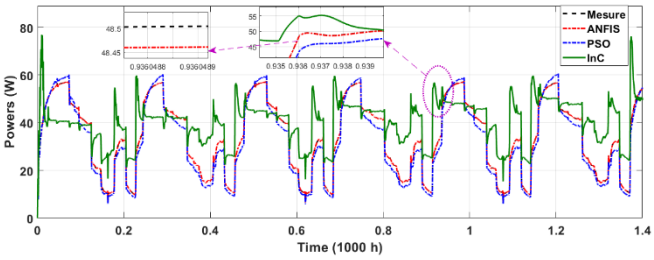


Figure 11: PVM powers

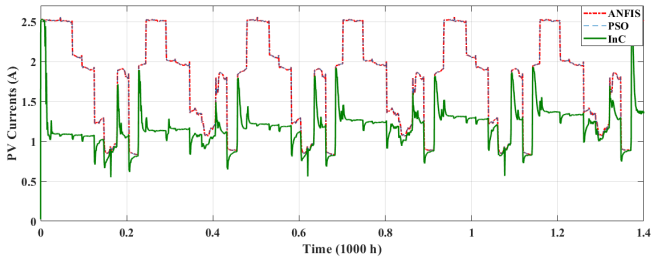


Figure 12: Currents of the PVM

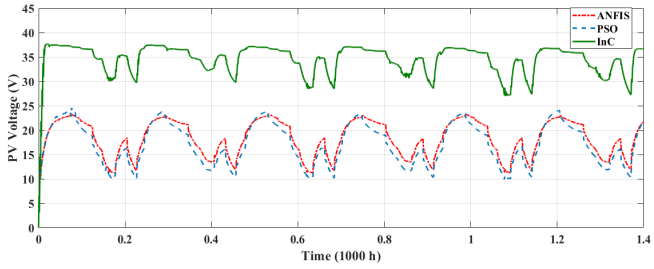


Figure 13: PVM voltages

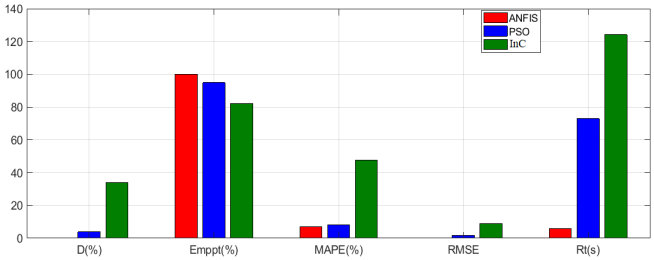


Figure 14: MPPT controller's performance parameters

These results corroborate those found in the literature which highlight the oscillating nature of the power obtained with the InC technique due to their inability to detect with precision the overall maximum in a situation of non-uniform climatic conditions [3]. They also confirm the results presented in [38] where the authors made a comparison between ANFIS and InC.

4. Conclusion

MPPT techniques are used to optimize power generated by PVM. In this work, a comparative studied between PSO, ANFIS and InC is done. An experimental validation of these MPPT techniques is also done using real database. The simulation results show that the MPPT technique based on ANFIS algorithm has the best performances. However, if we have to choose between these methods, the two parameters to consider are the complexity of the implementation and the performance of the method. For the first criterion, the choice will be relatively focused on the InC because of its great ease of implementation. For the second criterion, ANFIS will be chosen because of its best performance in terms of response time and accuracy. But we can conclude that the ANFIS and PSO methods give good results compared to the InC method.

Conflict of Interest

The authors declared that there is no conflict of interest.

Acknowledgments

The authors thank the efficiency and energy system research team of Alioune Diop University of Bambey (Senegal) and the Laboratory of Water, Energy, Environment and Industrial Processes, ESP (Dakar-Senegal).

References

[1] W.H. Ali, P. Cofie, J.H. Fuller, S. Lokesh, E.S. Kolawole, "Performance and Efficiency Simulation Study of a Smart-Grid Connected Photovoltaic System," Energy and Power Engineering, 71–85, 2017, doi:10.4236/epe.2017.92006.

[2] E. hadji M. Ndiaye, A. Ndiaye, M. Faye, S. Gueye, "Intelligent Control of a Photovoltaic Generator for Charging and Discharging Battery Using Adaptive Neuro-Fuzzy Inference System," International Journal of Photoenergy, **2020**, 144–156, 2020.

[3] S. Shabaan, M.I.A. El-sebah, P. Bekhit, "Maximum power point tracking for photovoltaic solar pump based on ANFIS tuning system," Journal of Electrical Systems and Information Technology, **5**(1), 11–22, 2018.

[4] N. Vela, F. Chenlo, "Early degradation of silicon PV modules and guaranty conditions," **85**, 2264–2274, 2011, doi:10.1016/j.solener.2011.06.011.

[5] M.A. Enany, M.A. Farahat, A. Nasr, "Modeling and evaluation of main maximum power point tracking algorithms for photovoltaics systems," Renewable and Sustainable Energy Reviews, **58**, 1578–1586, 2016.

[6] F. Chekired, A. Mahrane, M. Chikh, Z. Smara, "Optimization of energy management of a photovoltaic system by the fuzzy logic technique," Energy Procedia, **6**, 513–521, 2011, doi:10.1016/j.egypro.2011.05.059.

[7] V. Indragandhi, R. Logesh, V. Subramaniyaswamy, V. Vijayakumar, P. Siarry, L. Uden, "Multi-objective optimization and energy management in renewable based AC/DC microgrid," Computers and Electrical Engineering, **70**, 179–198, 2018, doi:10.1016/j.compeleceng.2018.01.023.

[8] A. Ndiaye, "Etude et conception de commandes intelligentes dédiées aux systèmes d'injection d'énergie photovoltaïque dans le réseau électrique de distribution," Thèse de Doctorat de Ecole Supérieur Polytechnique de Dakar, 2014.

[9] A.A. Fröhlich, E.A. Bezerra, L.K. Slongo, "Experimental analysis of solar energy harvesting circuits efficiency for low power applications," Computers and Electrical Engineering, **45**, 143–154, 2015, doi:10.1016/j.compeleceng.2014.09.004.

[10] F.K. Slimane Hadji, Jean-Paul Gaubert, "Theoretical and experimental analysis of genetic algorithms based MPPT for PV systems," Energy Procedia, **74**, 772–787, 2015, doi:10.1016/j.egypro.2015.07.813.

[11] M.K. M. Zagrouba, M. Bouaïcha, A. Sellami, "Optimisation par les algorithmes génétiques et modélisation par la méthode LPV d'un système

photovoltaïque," Vème Congrès International Sur Les Energies Renouvelables et l'Environnement, 1–6, 2010.
[12] S. Titri, C. Larbes, K. Youcef, K. Benatchba, "A new MPPT controller based on the Ant colony optimization algorithm for Photovoltaic systems under partial shading conditions," Applied Soft Computing Journal, **58**, 465–479, 2017.
[13] E. hadji M. Ndiaye, A. Ndiaye, M. Faye, "Experimental Validation of PSO and Neuro-Fuzzy Soft-Computing Methods for Power Optimization of PV installations," in 8th IEEE International Conference on Smart Grid, IEEE Xplore, Elhadjimbaye2020: 189–197, 2020, doi:10.1109/icsmartgrid49881.2020.9144790.
[14] B. Long, L. Huang, H. Sun, Y. Chen, F. Victor, K. To, "An intelligent dc current minimization method for transformerless grid-connected photovoltaic inverters," ISA Transactions, (xxxx), 2018, doi:10.1016/j.isatra.2018.12.005.
[15] M. Farhat, "Advanced ANFIS-MPPT Control Algorithm for Sunshine Photovoltaic Pumping Advanced ANFIS-MPPT Control Algorithm for Sunshine Photovoltaic Pumping Systems," 2012 First International Conference on Renewable Energies and Vehicular Technology Advanced, (April 2012), 2014.
[16] H. Bounechba, A. Bouzid, H. Snani, A. Lashab, "Electrical Power and Energy Systems Real time simulation of MPPT algorithms for PV energy system," International Journal of Electrical Power and Energy Systems, **83**, 67–78, 2016, doi:10.1016/j.ijepes.2016.03.041.
[17] S. Hadji, J. Gaubert, F. Krim, "Real-time Genetic Algorithms-based MPPT: Study and comparison (Theoretical an Experimental) with conventional methods," Energies, (Ic), 2018, doi:10.3390/en11020459.
[18] R.O. Nabil Karami, Nazih Moubayed, "General reviews and classification of different MPPT techniques," Renewable and Sustainable Energy Reviews, **68**(September 2016), 1–18, 2017.
[19] P. Hunoor, S.R. Savanur, "Design and Analysis of ANFIS Controller to Control Modulation Index of VSI Connected to PV Array," European Journal of Advances in Engineering and Technology, **2**(5), 12–17, 2015.
[20] P.F. Torres, A.F.P. Costa, V.L. Chaar Junior, W.L. Monteiro, M.A.B. Galhardo, J.T. Pinho, W.N. Macêdo, "A mobile educational tool designed for teaching and dissemination of grid connected photovoltaic systems," Computers and Electrical Engineering, **76**, 168–182, 2019, doi:10.1016/j.compeleceng.2019.03.017.
[21] F. Belhachat, C. Larbes, "Global maximum power point tracking based on ANFIS approach for PV array con fi gurations under partial shading conditions," Renewable and Sustainable Energy Reviews, **77**(February), 875–889, 2017, doi:10.1016/j.rser.2017.02.056.
[22] W. Issaadi, S. Issaadi, A. Khireddine, "Comparative study of photovoltaic system optimization techniques : Contribution to the improvement and development of new approaches," Renewable and Sustainable Energy Reviews, **82**(August 2017), 2112–2127, 2018, doi:10.1016/j.rser.2017.08.041.
[23] E. Fadil, H. El, A. Yahya, "Maximum Power Point Tracking Algorithm for Photovoltaic Systems under Partial Shaded Conditions," IFAC-PapersOnLine, **49**(13), 217–222, 2016.
[24] T.H. Kwan, X. Wu, "High performance P & O based lock-on mechanism MPPT algorithm with smooth tracking," Solar Energy, **155**, 816–828, 2017, doi:10.1016/j.solener.2017.07.026.
[25] H. Bounechba, A. Bouzid, K. Nabti, H. Benalla, "Comparison of perturb & observe and fuzzy logic in maximum power point tracker for PV systems," Energy Procedia, **50**, 677–684, 2014.
[26] D. Gueye, A. Ndiaye, F. Mactar, "Design Methodology of Novel PID for Efficient Integration of PV Power to Electrical Distributed Network," International Journal of Smart Grid, **2**(1), 2018.
[27] I.V. Banu, R.Ä. Beniug, M. Istrate, "Comparative Analysis of the Perturb-and-Observe and Incremental Conductance MPPT Methods," THE 8th INTERNATIONAL SYMPOSIUM ON ADVANCED TOPICS IN ELECTRICAL ENGINEERING May 23-25, 2013 Bucharest, Romania, (July 2014), 2013, doi:10.1109/ATEE.2013.6563483.
[28] Y.K. Semero, J. Zhang, D. Zheng, S. Member, "PV Power Forecasting Using an Integrated GA-PSO-ANFIS Approach and Gaussian Process Regression Based Feature Selection Strategy," CSEE JOURNAL OF POWER AND ENERGY SYSTEMS, **4**(2), 210–218, 2018, doi:10.17775/CSEEJPES.2016.01920.
[29] S. Paul, J. Thomas, "Comparison of MPPT using GA optimized ANN employing PI controller for solar PV system with MPPT using Incremental Conductance," International Conference on Power, Signals, Controls and Computation (EPSCICON), (January), 8–10, 2014.
[30] S. Hadji, F. Krim, J. Gaubert, U. De Poitiers, "Development of an algorithm of maximum power point tracking for photovoltaic systems using Genetic Algorithms," 7th International Workshop on Systems, Signal Processing and Their Applications (WOSSPA), 43–46, 2011.
[31] P. Jood, S.K. Aggarwal, V. Chopra, "Performance assessment of a neuro-fuzzy load frequency controller in the presence of system non-linearities and renewable penetration," Computers and Electrical Engineering, **74**, 362–378, 2019, doi:10.1016/j.compeleceng.2019.02.009.
[32] A. Harrag, H. Bahri, "Novel neural network IC-based variable step size fuel cell MPPT controller Performance , efficiency and lifetime improvement," International Journal of Hydrogen Energy, **42**(5), 3549–3563, 2016, doi:10.1016/j.ijhydene.2016.12.079.
[33] S. Saravanan, R.B. N, "RBFN based MPPT algorithm for PV system with high step up converter," Energy Conversion and Management, **122**, 239–251, 2016.
[34] S. Messalti, A. Harrag, A. Loukriz, "A new variable step size neural networks MPPT controler: Review, simulation and hardware implementation.," Renewable and Sustainable Energy Reviews, **68**(August 2015), 221–233, 2017.
[35] F. Chekired, Z. Smara, A. Mahrane, M. Chikh, S. Berkane, "An energy flow management algorithm for a photovoltaic solar home," Energy Procedia, **111**(September 2016), 934–943, 2017, doi:10.1016/j.egypro.2017.03.256.
[36] M. Nabipour, M. Razaz, S.G.H. Seifossadat, S.S. Mortazavi, "A new MPPT scheme based on a novel fuzzy approach," Renewable and Sustainable Energy Reviews, **74**(February), 1147–1169, 2017, doi:10.1016/j.rser.2017.02.054.
[37] O. Kraa, M. Becherif, M.Y. Ayad, R. Saadi, M. Bahri, A. Aboubou, I. Tegani, "A Novel Adaptive Operation Mode based on Fuzzy Logic Control of Electrical Vehicle," Energy Procedia, **50**(0), 194–201, 2014, doi:10.1016/j.egypro.2014.06.024.
[38] E.M. Ndiaye, A. Ndiaye, M.A. Tankari, G. Lefebvre, "Adaptive Neuro-Fuzzy Inference System Application for The Identification of a Photovoltaic System and The Forecasting of Its Maximum Power Point," 7th International IEEE Conference on Renewable Energy Research and Applications, ICRERA 2018, **5**, 1–7, 2018.
[39] M.M. Ismail, A.F. Bendary, "Smart Battery Controller using ANFIS for Three Phase Grid Connected PV Array System," Mathematics and Computers in Simulation, 2018.
[40] K. Amara, A. Fekik, D. Hocine, M. Lamine, "Improved performance of a PV solar panel with Adaptive Neuro Fuzzy Inference System ANFIS based MPPT," 7th International IEEE Conference on Renewable Energy Research and Applications, ICRERA 2018, **5**, 1098–1101, 2018.
[41] B. Kebe, O. Ba, B. Niang, L. Thiaw, "Etude , synthèse et implémentation d ' un contrôleur neuro- floue pour la commande de groupe turbo-alternateur," **2**(January), 46–53, 2018.
[42] M. Elloumi, R. Kallel, G. Boukettaya, "A comparative study of GA and APSO algorithm for an optimal design of a standalone PV/Battery system," 15th International Multi-Conference on Systems, Signals & Devices (SSD), 1104–1109, 2018.
[43] A. El Dor, "Perfectionnement des algorithmes d ' optimisation par essaim particulaire : applications en segmentation d ' images et en électronique," Thèse de Doctorat de l'Unicersité Paris-Est, 2013.
[44] K. Khezzane, F. Khoucha, "Application de la Technique PSO pour la Poursuite du PPM d ' un Système Photovoltaïque," The 3nd International Seminar on New and Renewable Energies, 1–6, 2014.
[45] E.A. Gouda, M.F. Kotb, D.A. Elalfy, "Modelling and Performance Analysis for a PV System Based MPPT Using Advanced Techniques," EJECE, European Journal of Electrical and Computer Engineering, **3**(1), 1–7, 2019.
[46] O. Ben Belghith, L. Sbita, F. Bettaher, "MPPT Design Using PSO Technique for Photovoltaic System Control Comparing to Fuzzy Logic and P & O Controllers," Energy and Power Engineering, **8**, 349–366, 2016, doi:10.4236/epe.2016.811031.
[47] T.T. Guingane, Z. Koalaga, E. Simonguy, F. Zougmore, D. Bonkoungou, T.T. Guingane, Z. Koalaga, E. Simonguy, F. Zougmore, D.B. Modélisation, "Modélisation et simulation d ' un champ photovoltaïque utilisant un convertisseur élévateur de tension (boost) avec le logiciel MATLAB / SIMULINK," JOURNAL INTERNATIONAL DE TECHNOLOGIE, DE L'INNOVATION, DE LA PHYSIQUE, DE L'ENERGIE ET DE L'ENVIRONNEMENT, **2**(1), 2016.
[48] I.V. Banu, M. Istrate, "Modeling of Maximum Power Point Tracking Algorithm for Photovoltaic Systems," ICEPE, (July 2014), 2012, doi:10.1109/ICEPE.2012.6463577.
[49] S. Motahhir, A. El Hammoumi, A. El Ghzizal, "The Most Used MPPT Algorithms : Review and the Suitable Low-cost Embedded Board," Journal of Cleaner Production, 2019, doi:10.1016/j.jclepro.2019.118983.
[50] S. Krishnamurthy, J.P. Ram, "EL-PSO based MPPT for Solar PV under Partial Shaded Condition," Energy Procedia, **117**, 1047–1053, 2017, doi:10.1016/j.egypro.2017.05.227.
[51] J.R. Jang, "ANFIS : Adaptive-Network-Based Fuzzy Inference System," IEEE, **23**(3), 1993.
[52] A. Fallah, F. Zarei, H. Zarrabi, M.J. Lariche, "ANFIS-GA modeling of dynamic viscosity of N- Alkane in different operational conditions," Petroleum Science and Technology, **0**, 1–7, 2018, doi:10.1080/10916466.2018.1458117.
[53] R.K. Kharb, F. Ansari, S.L. Shimi, "Design and Implementation of ANFIS based MPPT Scheme with Open Loop Boost Converter for Solar PV

Module," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, **3**(1), 6517–6524, 2014.

[54] A. Sharifian, M.J. Ghadi, S. Ghavidel, L. Li, J. Zhang, "A new method based on Type-2 fuzzy neural network for accurate wind power forecasting under uncertain data," Renewable Energy, **120**, 220–230, 2018, doi:10.1016/j.renene.2017.12.023.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1466-1473 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

An Investigation of the Effect of Optimal Plane Spacing Between Electrode Planes for the EIT Industrial Applications

Yew Lek Chong, Renee Ka Yin Chin*

Faculty of Engineering, University Malaysia Sabah, Kota Kinabalu, Sabah, 88400, Malaysia

ARTICLE INFO

Article history:
Received: 10 November, 2020
Accepted: 10 December, 2020
Online: 21 December, 2020

Keywords:
Electrical Impedance Tomography
Plane Spacing
Image quality

ABSTRACT

In this paper, the effect of plane spacing between electrode planes on Electrical Impedance Tomography (EIT) reconstructed images is investigated. Image properties of models for various plane spacings between electrode planes on EIT imaging were investigated by applying conventional measurement strategies. Sensitivity analysis and spatial resolution analysis were used to study the influence of the different plane spacing between electrode planes on imaging properties. In the sensitivity analyses, the results indicate that there are insignificant differences in sensitivity level for the models with different plane spacings, regardless of measurement strategies applied. From the spatial resolution analyses, the findings are conclusive as there are visible differences in the spatial resolution across the off-electrode plane. A comparative study using reconstructed images was also done. The true distributions with the different number of objects are used as references to assess resulting reconstructed images obtained from models with different plane spacing between electrode planes. Results indicate the model with plane spacing between electrodes planes, which is one quarter to the height of the model, provides the better quality of reconstructed images, in terms of estimations of dimension, and colour contrast of the imaged object.

1. Introduction

Electrical Impedance Tomography (EIT) is useful for visualising the conductivity distributions of an imaged space. Due to its non-intrusiveness, low cost and versatility, EIT has been widely used in medical [1] and industrial [2] applications. Recently, EIT is gaining more attention in industrial applications as it provides intuitions and insights to the reaction and dynamic of the process, as well as useful information of flow characteristic in a pipeline or process vessel. However, EIT imaging suffers from relatively low spatial resolution, especially towards the centre of the imaged space, which is ultimately impacting the quality of the reconstructed images. This is due to the ill-posed nature of the EIT problem [3]. As a result, numerous efforts have been made to further understand this problem and explore ways to improve spatial resolution in EIT imaging.

In an EIT system, the planar arrangement is a commonly employed electrode configuration, where electrodes are mounted in an equally spaced manner, on the perimeter of the process tank or vessel, and the centre of electrodes are aligned on a defined height. In industrial applications, an effective electrode configuration system greatly depends on the vessel or tank size. The criteria of the electrode configuration include the plane spacing between electrode planes and the number and size of electrodes. So far, there is no past work that studied plane spacing of the multiple electrode planes in EIT applications systematically. Instead, the past publications are more focus on the electrode parameters, such as size and number of electrodes, in medical and industrial applications.

In [4], the authors investigated the optimal distance of the multiple sensor plane using a 3D EIT imaging technique based on simulation and experimental study. The investigation showed that the quality of the reconstructed images was accessed based on the correlation coefficient and the relative error. The results showed that 3.5 cm plane spacing between electrode planes, for a model with a diameter of 16 cm and a height of 21.5 cm, produced the highest value in correlation and achieved the lowest values in the relative errors as well as provided optimal performance in the distribution tests. In [5], the authors investigated the performance of drive and measurement electrode patterns using various plane spacing between electrode planes. The results highlighted that the Signal to Error (SNR) decreases as the spacing between two electrode planes decreases, regardless of the transverse and

*Corresponding Author: Renee Ka Yin Chin, Email: reneekychin@ums.edu.my

longitudinal current patterns applied. The transverse current patterns provided a better resolution than longitudinal current patterns but were more susceptible to noise and error, leading to poorer sensitivity in detecting impedance changes in the centre of imaged space. However, in industrial applications, the plane spacing between electrode planes may become a minor concern for the researchers.

Electrode parameters, such as size and number of electrodes, have been investigated in the past few decades. In [5], an experiment was carried out to examine the effect of the height of electrodes on the quality of reconstructed images. The findings indicated that the sensitivity or detectability of the EIT system reduced or remained the same as the height of electrodes was increased to twice the height of the imaged object or target. In [6], the authors investigated the optimum sized electrodes for Electrical Resistance Tomography (ERT). The findings highlighted that the optimum size of electrode coverage is 60% and 80% of the wall surface area of the model. In [7], [8], the authors investigated the electrode properties of ERT. They found that dot or point electrodes could eliminate the effect of contact impedance. Dot or point electrodes are more commonly applied in medical applications. However, it is impractical when implementing dot or point electrodes in industrial applications, as it is more costly to fabricate [9].

Electrodes used in applications are often in rectangular shape. This is not ideal, as narrow angles on the electrode corners result higher current densities in these areas. Circular electrodes are ideal for this purpose due to the uniformity of its shape [10]. However, forward solving requires discretisation of the model into tetrahedral elements in EIT. An accurate modelling of circular electrodes requires a large number of small elements to best represent the shape of electrodes, which results in an increased number of computational elements, and more importantly, these elements do not contribute to the representation of the imaged space. When the coarser mesh structure is used, the inaccurate representation of the shape of the electrode will occur and it leads to errors in forward and inverse solving. As a compromise, rectangular electrodes are often used.

In this paper, the influence of different plane spacing between electrode planes on EIT applications in industrial processes (e.g. pipeline flow and suspension monitoring, and mixing processes) on the quality of reconstructed images is studied. The main purpose of this investigation is to investigate a minimum requirement of space between electrode planes, such that information is not lost. This investigation was carried out through investigating the effect of different electrode configurations on imaging properties and reconstructed images. The rest of the paper consists of five (5) sections. The model setup for various plane spacing between electrode planes is described in Section II. Section III discusses the findings obtained from the comparative studies on image properties. Section IV describes the experimental setup for image reconstruction, as well as its results. Finally, Section V provides the conclusion of this study.

2. Model Setup

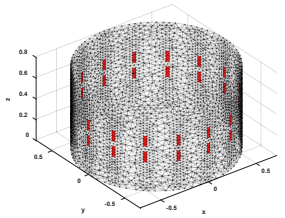
In this work, a cylindrical model with a radius of 71.4 cm and a height of 80 cm is used. The dimensions of the model are chosen to be sufficiently big, such that it can accommodate a variety of different electrode configurations, including spacing between electrode planes, which is the main focus of the work presented in this paper.

The electrode configuration used for this work consists of 32 electrodes arranged in two planes. Each plane consists of 16 electrodes arranged equi-spaced on the periphery of the model. The electrodes are rectangular in shape, with the dimensions of 3 cm in width and 10 cm in height. In this study, the spacing between planes is set as 5 cm (the smallest gap), 10 cm, 15 cm, 20 cm, 25 cm and 30 cm (largest gap). A summary of the plane spacing between electrode planes and the height of the centre of the upper and lower electrode planes is provided in Table I.

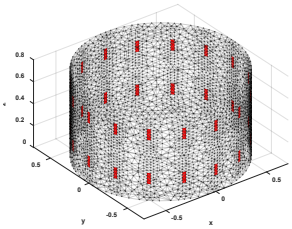
Table 1: Plane Spacing Between Two Electrode Planes and its Center of Upper and Lower Electrode Plane

Number of electrodes per plane	Plane spacing between two electrode planes	Center of upper electrode plane (cm)	Center of lower electrode plane (cm)
16	5 cm	47.5	32.5
	10 cm	50.0	30.0
	15 cm	52.5	27.5
	20 cm	55.0	25.0
	25 cm	57.5	22.5
	30 cm	60.0	20.0

The work presented in this paper is simulated using Electrical Impedance Tomography and Diffuse Optical Tomography Reconstruction Software (EIDORS), which is an open-source toolkit available for MATLAB [11]. The discretisation of the models was done by applying the Finite Element Method (FEM) through Netgen [12]. The FEM models with various spacing between electrodes planes are as depicted in Figure 1. It is should be noted that the number of elements for each of the models is the same as they are based on the same cylindrical model. The work presented in this paper is through simulation as it is a more feasible option due to the requirement of using electrodes with different plane spacing.



(a) 5 cm plane spacing between electrode planes model



(b) 30 cm plane spacing between electrode planes model

Figure 1: Examples of FEM models with varying plane spacing between electrode planes.

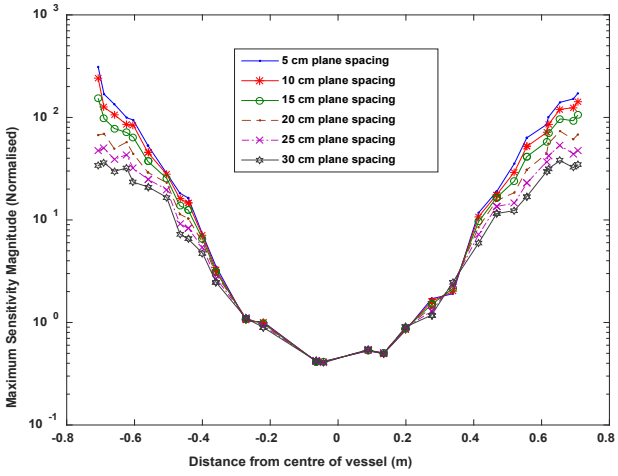
All measurements were simulated using the adjacent current injection protocol, which produces *N*(N-3)* number of

measurements, where *N* is the number of electrodes. In the case of two planes of 16-electrodes per plane model, 416 full-frame measurements were generated. Another set of measurement data were repeated to simulate by using another opposite current injection protocol, which produces *N*(N-4)* number of measurements. There are 384 full-frame measurements generated for a model with two planes of 16-electrodes per plane.

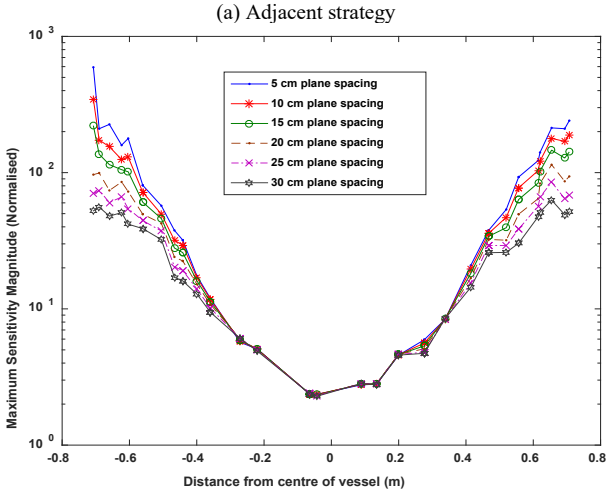
3. Experimental Results

3.1. Sensitivity Analysis

Sensitivity analysis provides information on the potential success of detecting any changes in conductivity in the imaged space. In the discretised model, each element has a sensitivity value for any measurement acquired, which makes up the sensitivity (Jacobian) matrix. For this comparative study, the maximum sensitivity for each discretised element for a full set of measurement, normalised to the volume of the element, is computed and compared, as shown in Figure 4 and Figure 5.



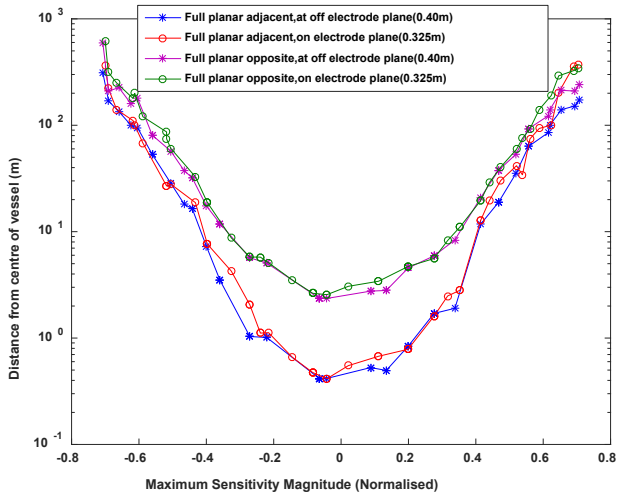
(a) Adjacent strategy



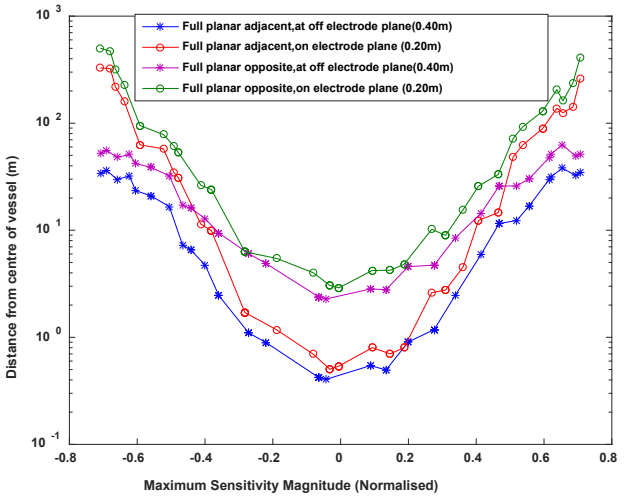
(b) Opposite strategy

Figure 4: Sensitivity comparison for models with different plane spacing electrode planes at the off-electrode plane (the height is taken at 0.40 m from the bottom of the model) by applying the (a) adjacent strategy and (b) opposite strategy.

Figure 4(a) shows the comparison of maximum sensitivity magnitude for the models with different spacing between electrode planes across an off-electrode plane when applying adjacent current injection protocol. From Figure 4(a), it can be seen that there are insignificant differences, as the spacing between electrode planes increases, especially towards the centre of the imaged space. Similar results are observed using the opposite current injection strategy, as reflected in Figure 4(b). This may due to the planar measurement strategy used, whereby measurements are only taken on the plane where the current source and sink pairs are located.



(a) The model with 5 cm plane spacing between electrode planes



(b) The model with 30 cm plane spacing between electrode planes

Figure 5: Sensitivity comparison for the models both (a) 5 cm and (b) 30 cm plane spacing between electrode planes, respectively, at the off-electrode plane and on the electrode plane.

From Figure 5, sensitivity is higher for opposite strategy due to the current penetrating through the imaged space, as opposed to the adjacent strategy, where current flowing typically tends to be more confined nearer to the periphery of the model [13]. The other obvious observation is that the sensitivity is higher on the electrode plane in comparison to those obtained off-electrode plane. This is expected, as the current signal is stronger on the electrode plane compared to the off-electrode plane [14]. This observation is similar to a previous study [15], where it indicates that the possibility of detecting a change in conductivity reduces when further away from the electrode plane.

Comparing Figure 5(a) and Figure 5(b), it can be observed that when the space between electrode planes are smaller (5 cm), there are insignificant differences between sensitivity on- and off-electrode planes. This is likely to be due to the closeness of the locations of the two electrode planes. In comparisons, when the space between electrode planes is bigger (30 cm), there are noticeable differences between sensitivity on- and off-electrode planes. Although the EIT image reconstruction problem is often treated as a 2D problem, where images and data are captured on a specific plane, it is essentially a 3D problem.

3.2. Spatial Resolution Analysis

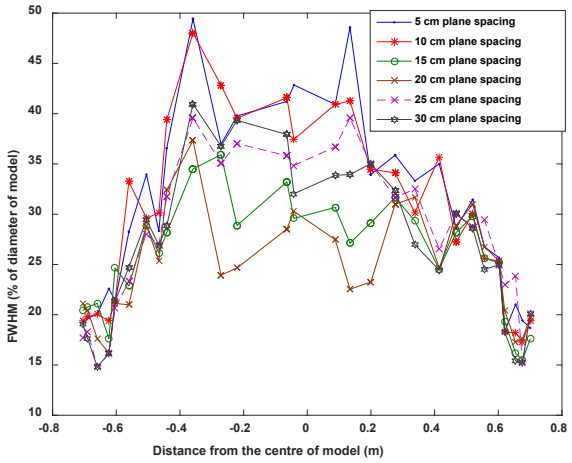
Spatial resolution analysis utilises information given in the resolution matrix, which contains the correlation between elements in a discretised model. In this analysis, the Full-Width Half-Maximum (FWHM) technique [16] is applied to investigate the spatial resolution of models with different spacing between electrode planes. The FWHM method applied in this paper is based on the technique, which was first proposed by [17] and later adopted by [18]. In spatial resolution analysis, the minimum distance between two point sources, which shares the same solution value in an image is measured. The same solution value shared is known as half the peak value. Spatial resolution for each point is computed and compared. The results are as shown in Figure 6 and Figure 7.

Figure 6 shows the spatial resolution comparing models with different spacing between electrode planes using both adjacent and opposite strategies. It is shown that the regions near the wall of the model, in general, have better resolution (low FWHM) in comparison with the area towards the centre of the model, which resulted in higher FWHM. This general trend obtained is consistent with the sensitivity analysis obtained in Figure 4(a) and Figure 4(b), where they indicate that sensitivity is lower in the centre of the model but is higher near the wall region of the model. With a higher possibility of detecting a change in conductivity, and thus it increases the chance of change detected to be reconstructed.

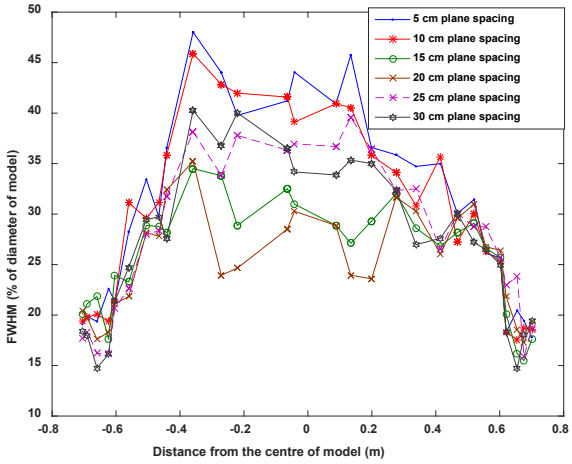
It is worth noting that unlike the sensitivity analysis, spatial resolutions do not show symmetry across the plane, and irregularities (i.e. peaks or valleys) are observed. One of the factors that affect spatial resolution is the size of the discretised element, and unlike sensitivity analysis, the computed spatial resolution for each discretised element is not normalized to the volume of the element, as it will provide misrepresentation on the visibility contrast.

From Figure 6(a), it can be observed that models with 5 and 10 cm plane spacing between electrode planes produced the worst resolution overall, especially in the centre region of the model. However, it is worth noticing that model with 20 cm plane spacing between electrode planes resulted in better spatial resolution in general. This finding is not anticipated as the sensitivity analyses are shown in Figure 4, where it indicates that there are insignificant differences in maximum sensitivity magnitude across the off-electrode plane, for both current injection strategies.

Referring to Figure 7, comparing spatial resolution for the on- and off-electrode planes, it is observed that there are insignificant differences in the resolution for an on-electrode plane, especially with the gap between the two electrode planes is small (5 cm). This result is consistent with the results observed in its corresponding comparative analysis, as shown in Figure 5. The spatial resolution analysis in Figure 7 also shows that the overall resolution is evenly matched for usable information which is available for image reconstruction.



(a) Adjacent strategy



(b) Opposite strategy

Figure 6: Spatial resolution comparison for models with different plane spacing by applying adjacent strategy in (a) and opposite strategy in (b) at 0.40 m height from the base of the model.

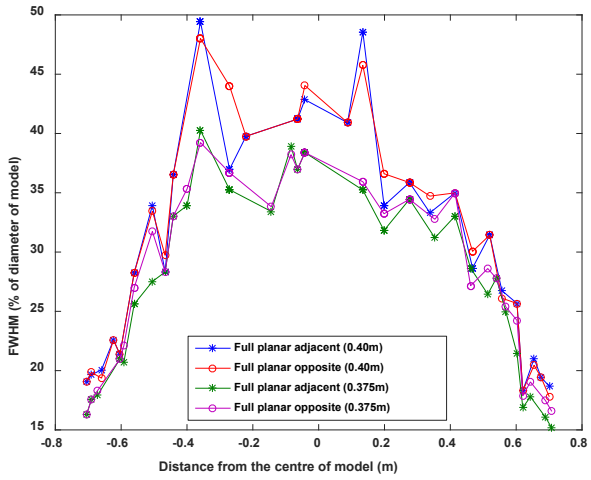
It is interesting to see in Figure 7(a) that there is very little difference in spatial resolution when comparing spatial resolutions for the adjacent and opposite strategies. This is unexpected, as shown in Figure 5, where there is a distinguishable difference in sensitivity between the two different measurement strategies. One of the reasons is likely to be due to both strategies producing a similar amount of usable (stable) information for reconstruction, which resulted in the spatial resolution for these two strategies to be evenly matched.

4. Experimental Set-up

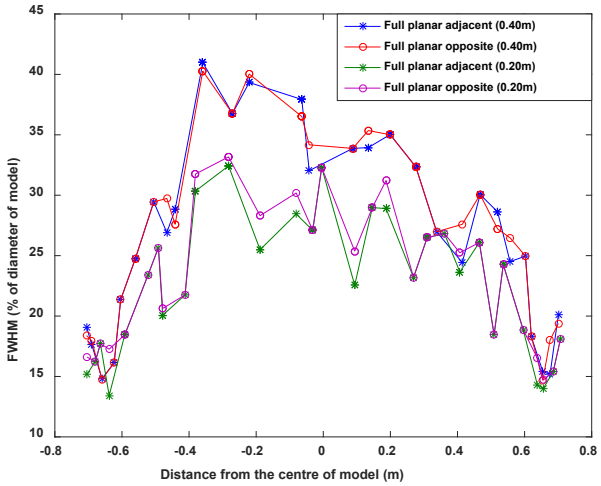
4.1. Experimental Set-Up

Measurements for image reconstruction are simulated using the EIDORS package [11], which is MATLAB-based. As

mentioned earlier, this study is more feasible to be done in simulation, as laboratory validation requires a process tank that allows multiple settings for electrodes configurations. For validation, two test distributions are used for experiment and comparison.



(a) The model with 5 cm plane spacing between electrode planes

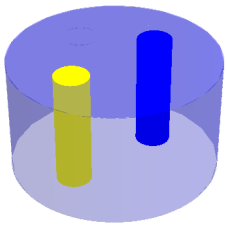


(b) The model with 30 cm plane spacing between electrode planes

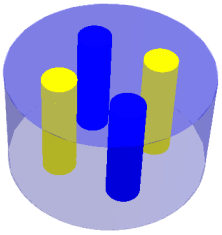
Figure 7: Spatial resolution comparison for the models both (a) 5 cm and (b) 30 cm plane spacing between electrode planes, respectively, at the off-electrode plane and on the electrode plane.

Measurements are simulated using a finely discretised FEM model. The level of discretization is chosen to be significantly higher than the discretized model used for forward and inverse solving, as it is intended to mimic a physical model. Two measurements are taken for each test distribution, for each electrode configuration, and each measurement strategy.

In this work, two test distributions are used for image reconstruction, as shown in Figure 8. The first test distribution consists of two cylinders, each with a radius of 12 cm. The conductivity of the background is 0.012 S/m (approximately the conductivity value of tap water). One of the cylinders has a higher conductivity than background conductivity, and the other is non-conducting. The arrangement of test objects is as shown in Figure 8(a). The second test distribution is similar to the first test distribution, with the addition of one conducting cylinder. The arrangement of test objects is as shown in Figure 8(b). All reconstructed images were obtained by applying the Non-Linear Gauss-Newton (NLGN) algorithm in inverse solving.



(a) True distribution with two cylindrical objects



(b) True distribution with four cylindrical objects

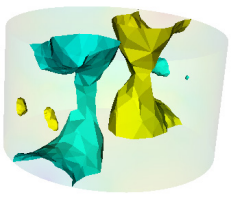
Figure 8: True distribution models with (a) two cylindrical objects, (b) four cylindrical objects and (c) four floating objects. The colour scale is applied, where yellow colour indicates the conducting object and blue colour indicates the non-conducting object.

4.2. Results and Discussions

The image reconstructions obtained from models with different spacing between electrode planes are shown in Figure 9 to Figure 12. The colour scale is maximised to show the best contrast. The reconstructed images are assessed and discussed based on the dimensions of the reconstructed objects and colour contrast of the reconstructed conductivity distribution.

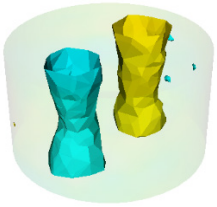
Although sensitivity analyses indicated there are insignificant differences in the maximum sensitivity magnitude as the gap between electrode planes increases, this is not reflected in spatial resolution analysis and reconstructed images. When comparing the reconstructed images from Figure 9-12, it can be seen that there are visible differences in terms of shape and estimation of dimensions of the objects.

In Figures 9-12, it can be seen that the models with 5cm and 10 cm plane spacing between electrode planes generally produced the worst reconstruction images. The shape and the dimension of the cylindrical test objects are not reconstructed properly, especially towards the top and the bottom of the imaged space. There are also noticeable artefacts visible in the imaged space. The results are similar for experiments done using both the adjacent and opposite strategies.

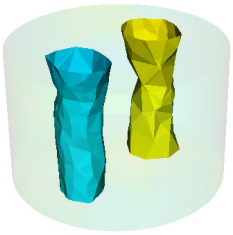


(a) 5 cm plane spacing between electrode planes

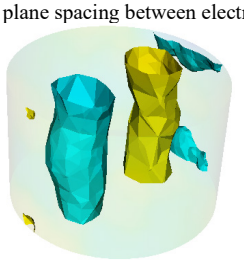
(b) 10 cm plane spacing between electrode planes



(c) 15 cm plane spacing between electrode planes



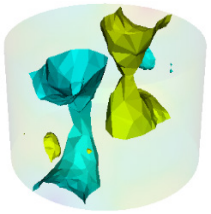
(d) 20 cm plane spacing between electrode planes



(e) 25 cm plane spacing between electrode planes

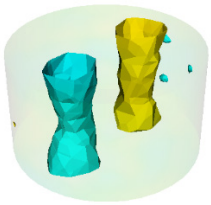
(f) 30 cm plane spacing between electrode planes

Figure 9: Reconstructed images comparison for models with various plane spacing (a)-(f) by applying the adjacent current protocol.

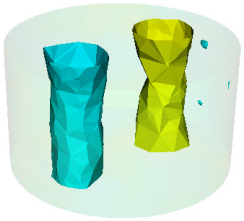


(a) 5 cm plane spacing between electrode planes

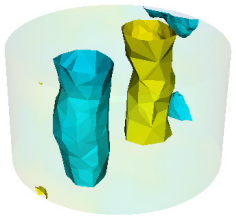
(b) 10 cm plane spacing between electrode planes



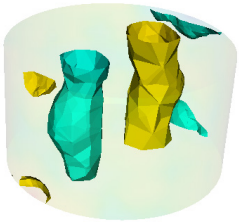
(c) 15 cm plane spacing between electrode planes



(d) 20 cm plane spacing between electrode planes

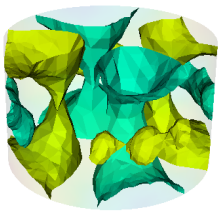


(e) 25 cm plane spacing between electrode planes

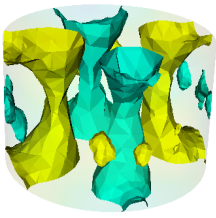


(f) 30 cm plane spacing between electrode planes

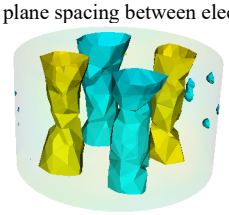
Figure 10: Reconstructed images comparison for models with various plane spacing (a)-(f) by applying the opposite current protocol.



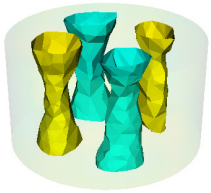
(a) 5 cm plane spacing between electrode planes



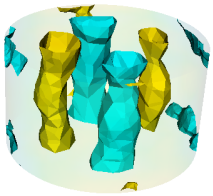
(b) 10 cm plane spacing between electrode planes



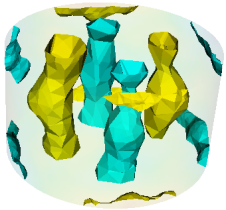
(c) 15 cm plane spacing between electrode planes



(d) 20 cm plane spacing between electrode planes

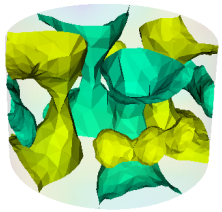


(e) 25 cm plane spacing between electrode planes

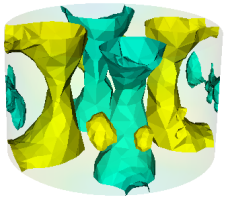


(f) 30 cm plane spacing between electrode planes

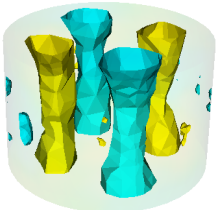
Figure 11: Reconstructed images comparison for models with various plane spacing (a)-(f) by applying the adjacent current protocol.



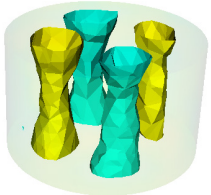
(a) 5 cm plane spacing between electrode planes



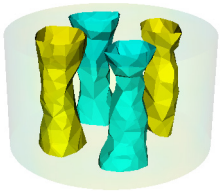
(b) 10 cm plane spacing between electrode planes



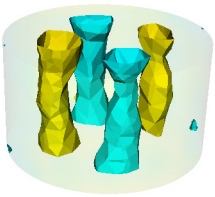
(c) 15 cm plane spacing between electrode planes



(d) 20 cm plane spacing between electrode planes



(e) 25 cm plane spacing between electrode planes



(f) 30 cm plane spacing between electrode planes

Figure 12: Reconstructed images comparison for models with various plane spacing (a)-(f) by applying the opposite current protocol.

Models with 15 cm and 20 cm gaps generally produced better reconstructed images. With the exception of the reconstructed image using a test distribution with four cylindrical objects, the application of the adjacent strategy for the model with 20 cm gap, it can be seen in Figures 9-12 that the cylindrical test objects are reconstructed in the right configuration, with similar dimensions to the true distribution. There are also little to no artefacts visible in the imaged space.

For the widest gaps, namely the 25 cm and 30 cm electrode spacing models, the images reconstructed using the opposite strategy are better compared to those obtained using the adjacent strategy. The images obtained using the adjacent strategy has a significant amount of artefacts, and it is noticeable that the shape of the test objects are not reconstructed correctly where the gap of the electrodes are (i.e. space between electrode planes). It can be seen in Figures 9(e-f) and 11 (e-f) that the dimensions are overestimated in that space. In contrast, reconstructed images obtained using the opposite strategy turn out better, with the test objects shown to be cylindrical objects, and with less artefacts. This result is consistent with the results observed in the sensitivity analyses.

5. Conclusion

A summary of the results is tabulated in Table 2. This study is conducted through two methods: statistical method, through comparing sensitivity and spatial resolution of different models used; and through reconstructed images.

Table 2: Summarization of the results on image properties

Analysis	Best model	Worst model
Sensitivity Analysis	Results are inconclusive; Opposite strategy is better than adjacent strategy	
Spatial Resolution Analysis	15 and 20 cm gap (both strategies)	5 and 10 cm gap (both strategies)
Reconstructed Images	15 and 20 cm gap (opposite strategy)	5 and 10 cm gap (both strategies)

From Table 2, it can be seen that the results obtained through spatial resolution analysis are reflected in the reconstructed images, where the best quality and worst quality images are

reflective of the results obtained through spatial analysis. In this case, however, the sensitivity analysis results were inconclusive, as the sensitivity level in the centre region of the imaged space is indistinguishable.

It is worth noting that the results given in Table 2 are relative to the dimension of the model used for this study. When the model is scaled up or scaled down, an investigation similar to that carried out in this study should be conducted to ensure that the scaling effect does not deteriorate the quality of the reconstructed images.

Through this work, it is shown that it is important to investigate the spacing between electrode plane, to ensure that the information is not lost due to the inability to capture any changes that occur. It is impractical to have electrode planes close to each other, as it will result in a waste of resources. Therefore, a minimal distance ought to be set in order to ensure that any changes in conductivity can be measured and the information can be used to reconstruct and visualize the changes.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to express their gratitude to the financial support provided by the KPT-RACE (RACE0010-TK-2013) grant, KPT-FRGS (FRGS0425-TK-1/2015) grant and UMS-SGA (SGA0007-2019) grant for this work.

References

[1] H. X. Wang, C. Wang, W. L. Yin, "Optimum design of the structure of the electrode for a medical EIT system," Measurement Science and Technology, **12**(8), 1020–1023, 2001. https://doi.org/10.1088/0957-0233/12/8/305

[2] J. Nowakowski, E. Hammer, D. Sankowski, D. Styra, R. Wajman, R. Banasiak and A.Romanowski, "New concept of ECT/ERT/GRT tomography for multiphase flow measurements," Automatyka (pó³- rocznik AGH), **14**, 741-748, 2010.

[3] W. Lionheart, N. Polydorides and A. Borsic, Part 1: The Reconstruction Problem, In Holder D. (ed) Electrical Impedance Tomography: Methods, History and Applications, Bristol, 3-64, 2005.

[4] Z. Hao, S. Yue, B. Sun, and H. Wang, "Optimal distance of multi-plane sensor in three-dimensional electrical impedance tomography," Computer Assisted Surgery, **22**, 326–338, 2017.

[5] J. Hope, F. Vanholsbeeck, and A. McDaid, "Drive and measurement electrode patterns for electrode impedance tomography (EIT) imaging of neural patterns activity in peripheral nerve," Biomedical Physics and Engineering Express, **4**(6), 2018. https://doi.org/10.1088/2057-1976/aadff3

[6] J. C. Newell, Y. Peng, P. M. Edic, R. S. Blue, H. Jain, and R. T. Newell, "Effect of electrode size on impedance images of two-and three-dimensional objects," in 1998 IEEE Transactions on Biomedical Engineering, **45**(4), 531–534, 1998. https://doi.org/10.1109/10.664209

[7] P. A. T. Pinheiro, W. W. Loh, F. J. Dickin, "Optimal sized electrodes for electrical resistance tomography," Electronics Letters, **34**(1), 69-70, 1998. https://doi.org/10.1049/el:19980092

[8] F. Dickin, M. Wang, "Electrical resistance tomography for process applications," Measurement Science and Technology, **7**(3), 247–260, 1996. https://doi.org/ 10.1109/10.664209

[9] L. D. Suits, T. Sheahan, J. Lee, J. Santamarina, "Electrical Resistivity Tomography in cylindrical Cells—Guidelines for hardware pre-design," Geotechnical Testing Journal, **33**, 2010. https://doi.org/10.1520/GTJ102366.

[10] M. Hanke, B. Harrach and N. Hyvonen," Justification of Point Electrode Models in Electrical Impedance Tomography," Mathematical Models and Methods in Applied Sciences, **21**(06), 1395–1413, 2011. https://doi.org/10.1142/S0218202511005362

[11] K. S. Cheng, D. Isaacson, J.C. Newell, D.G. Gisser," Electrode models for electric current computed tomography," in 1989 IEEE Trans. Biomed Engr., **36**, 918–924, 1989.

[12] N. Polydorides, W.R.B. Lionheart, "A Matlab toolkit for three-dimensional electrical impedance tomography: a contribution to the Electrical Impedance and Diffuse Optical Reconstruction Software project," Meas. Sci. Technol., **13**, 1871–1883, 2002.

[13] J. Schoberl, Netgen: Automatic Mesh Generator, 2008.

[14] P. O. Gaggero, "Miniaturization and distinguishability limits of electrical impedance tomography for biomedical application", Ph. D Thesis, University of Neuchatel, 2011.

[15] P. Metheral, D. C. Barber, R. H. Smallwood, "Three-dimensional electrical impedance tomography," **380**, 509-5012, 1996. https://doi.org/https://doi.org/10.1038/380509a0

[16] J.L. Wheeler, W. Wang, M. Tang, "Á comparison of methods for measurement of spatial resolution in two-dimensional circular EIT images," Physiol. Meas., **23**, 169-175, 2002. https://doi.org/10.1088/0967-3334/23/1/316

[17] S.C. Murphy, "Ad hoc electrode arrangement for electrical tomography," Ph. D Thesis, University of Manchester, United Kingdom, 2008.

[18] W. Menke, Geophysical Data Analysis: Discrete Inverse Theory (Rev. Ed), Academic Press: San Diego, 1989.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1474-1479 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Social Engineering to Establish Digital Culture in Higher Education

Tedi Priatna[1,*], Dian Sa'adillah Maylawati[2,3], Hamdan Sugilar[4], Muhammad Ali Ramdhani[2]

[1]*Department of Islamic Education, UIN Sunan Gunung Djati Bandung, Indonesia*

[2]*Department of Informatics, UIN Sunan Gunung Djati Bandung, Indonesia*

[3]*Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia*

[4]*Department of Mathematics Education, UIN Sunan Gunung Djati Bandung, Indonesia*

ARTICLE INFO

Article history:
Received: 18 September, 2020
Accepted: 12 December, 2020
Online: 21 December, 2020

Keywords:
Digital Culture
Digital Behavior
Higher Education
Information Technology
Organizational Culture
Social Engineering

ABSTRACT

This study aims to discuss the establishment of digital culture in higher education as an effort of higher education to align themselves with the development of science and technology so that that higher education is always at the forefront in the enforcement and provide high-quality education. The method used in this study is a qualitative method, which analyzes the causal relationship between the dimensions studied. Observation and literature review used as the main source of data/ information analyzed in this study. The analysis results showed that social engineering is an effective medium of modifying values, subjective norms, and rules for the creation of digital culture. The implication of this research shows that social engineering is an important activity that must be scheduled periodically by higher education as a medium for strengthening organizational culture to increase the capacity of human resources.

1. Introduction

World civilization moves together with the rapid development of information technology. The utilization of digital technology in all aspects of social life enters almost all lines of activities that regulate community life, including in the education field [1, 2]. In Indonesia, higher education as an educational institution has the task of carrying out the *Tridharma Perguruan Tinggi*/ Higher Education *Tridharma*, namely education, research, and community service [3].

For carrying out functions optimally, higher education required to improve their business processes by adapting and adopting technology continually. The mastery of science and technology is substantive and elementary in the development of superior and modern universities. The diffusion of technology in business processes of higher education believed can create better performance in carrying out higher education *Tridharma*. Technology diffusion is an adaptation of higher education to its environment, by changing values, and making the work patterns of the academic community more systematic and efficient. One form of change from the diffusion of technology is the change of work culture into digital work patterns after this referred to as digital culture.

The digital culture was born from the phenomenon of digitalization in human civilization. Digital culture occurs because of the increasing uniqueness and importance of the technology application that encourages people to think about the importance of transforming significantly and dramatically [4]. Where finally, technology has an important role and becomes a necessity in the development of social culture [5,6]. Digitalization is a simplification of information material in the form of binary code or the integration of communication technology with computer logic. Data conversion methods have advantages in increasing the flexibility of handling, storing, processing, and sending data. The flexibility offered by the digitization process facilitates the process of transmitting and manipulating information materials that have an economic impact on a network, where information materials can be disseminated more efficiently among network users [7].

*Corresponding Author: Tedi Priatna, Email: tedi.priatna@uinsgd.ac.id

Several previous research that relevant to social engineering related to digital culture in higher education includes: (1) preparation of higher education in facing disruption of the digital era is a challenge in university branding, including adapting to the concept of virtual universities, virtual classes, artificial intelligent, and techno-entrepreneurship [8]; (2) guaranteeing the quality of education in the era of digital disruption in higher education is considered necessary, higher education which is one of the places to improve the quality of human resources needs to implement an internal guarantee system that is in line with the rapid development of digital technology [9]; (3) digital culture in higher education that the most frequently and widely used, especially for millennial students is e-learning/ digital learning/ virtual learning/ digital literacy [10–12], although on the other hand, learning in the classroom directly is still needed [13, 14]; (4) several models are created related to digital culture that will transform higher education into digital universities that are ready to compete in the era of technological disruption [15,16], even online education policies and practices have supported the concept of digital university past, present and in the future [17]; (5) even in some studies discussing that digital culture in higher education can support the development of the digital economy [18], because the educational revolution towards technological progress is a revolution for the long term [19, 20]. The novelty/ originality of this research lies in the focus of the study, which examines digital formation models for the development of the implementation of *Tridharma* in higher education.

2. Research Method

This research is policy research, where the product of this research expected to use as a foundation and consideration in the decision making of top university leaders. The methodology used in this study is qualitative research [21, 22] by collecting the literature review, observation, and Focus Group Discussion (FGD) [23]. The primary data source of this study refers to the results of observations, and literature reviews that have relevance to the objective conditions on the research object. Researchers use causal logic as a rational argument that connects the concepts which establish the organizational culture. Researchers in this study act as participant observers. This research conducted at UIN Sunan Gunung Djati Bandung (after this referred to as a university). This research puts digital culture as part of organizational culture. The main foundation used in this research is the need to shift the work culture that is manual towards digital culture as a consequence of technological developments and the demands of educational service users for technology-based prime services.

3. Result and Discussion

At the university, which is the object of research, 58 application programs have been designed and built as a modernization/ digitalization of the general business processes that run by the university. The design of the application program is aiming to optimize business processes carried out by the university. Figure 1 provides a detailed relationship between the main business process, supporting activities, and actors that run in the university.

Based on the results of discussions and interviews found the fact is not all application programs are running optimally. This case is allegedly due to low desire and the user's ability to run application programs. The integration of digital culture in the university's organizational culture is one solution to this problem. Organizational culture, according to many studies, can improve organizational performance [24–27].

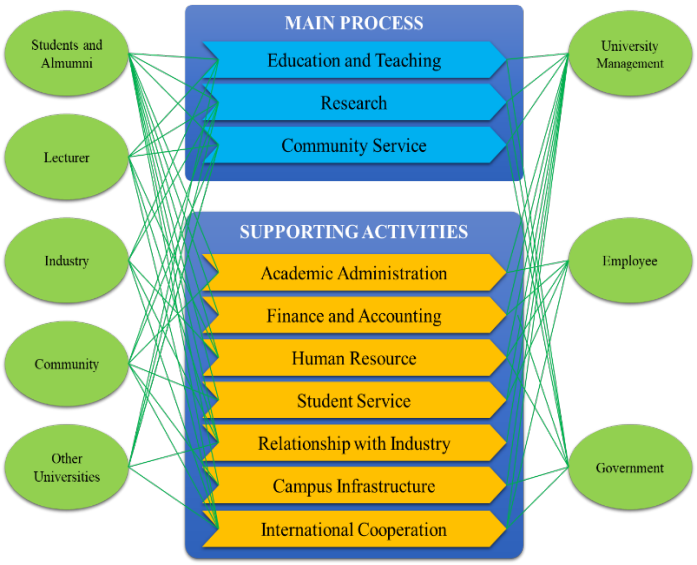


Figure 1: Process and Stakeholder Relations [28]

3.1. Organizational Culture

Organizational culture is a shared perception or a system of shared meanings shared by its members. The function of organizational culture is to determine the roles that distinguish organizations from others, define shared goals that are greater than individual interests, maintain social stability, enhance identity for members of the organization, and provide control mechanisms that provide guidelines for attitudes and behavior [29].

Organizational culture at the university will foster an individual's identity as a process of personal assimilation as part of the academic community. The similarity of embedded values will make it easier for every academic community to understand and appreciate every activity carried out by the university. An understanding of the organizational culture at the university will facilitate the resolution of internal problems such as giving awards, work ethics, and career development. This understanding will also help universities in overcoming external issues so that universities can do the best service.

Factors that lead to the development of organizational culture consists of internal and external factors, modifying/ changing the culture of the organization will create existing human resources will be of higher quality [30]. The organizational culture that is the color of the research object is the work culture of the Ministry of Religion (MORA) of Republic Indonesia. Structurally, the university which used as a research locus is a work unit that is under the authority of MORA. The work culture adopted at MORA includes integrity, professionalism, innovation, responsibility, and exemplary [31]. The alignment of work culture to a digital culture presented in Figure 2.

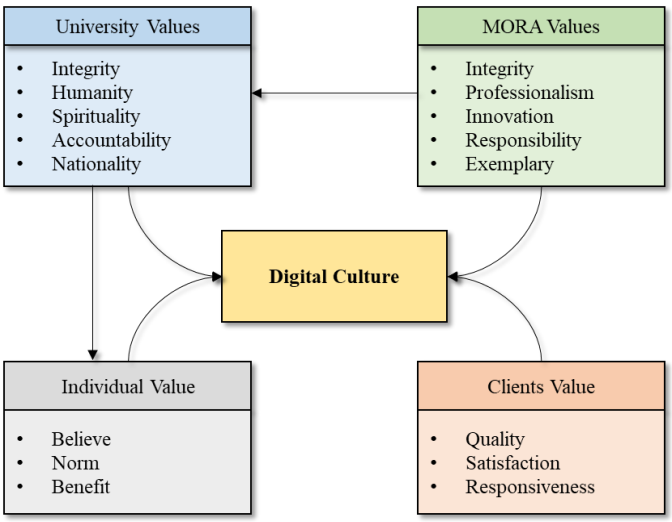


Figure 2: Alignment of Work Culture in Digital Culture

3.2. Digital Culture

Digital culture is a behavioral habit that utilizes digital technology as part of its work. The formation of digital culture is essential in sustaining the main tasks of the university. In general, individual activities motivated by the culture that influences their behavior. Culture requires individuals to behave and give instructions to them about what must be followed and learned. An influential culture can create a bond between the university and the academic community and inspire to create university excellence. The digital culture built by values, norms/ ethics, and regulations, which expressed in the form of behavior. Figure 3 describes the dimensions of forming behavior in digital culture.

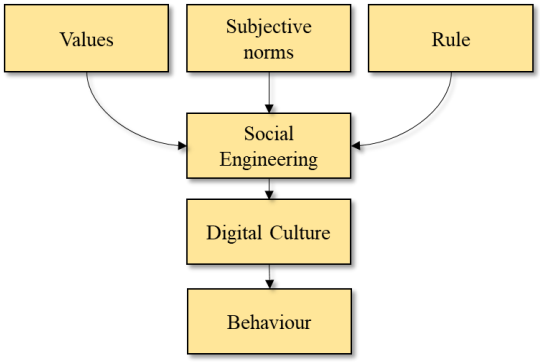


Figure 3: Dimensions of Forming Behavior in Digital Culture

3.2.1. Value

Values are all things related to human behavior regarding good or bad as measured by religion, tradition, ethics, morals, and culture that apply in society [32]. The definition of value used in this article is an individual's understanding of the elements of goodness that give meaning to his status.

Digital culture in the academic community requires value as an individual motivation to conduct digital-based behaviour. The value dedicated to the academic community includes strengthening personal beliefs about the value of goodness, which realized by working professionally and innovatively as part of worship. Digital culture will provide position and imaging for the academic community as individuals who are good at carrying out the activities in the social community.

3.2.2. Subjective Norm

Subjective norms are one's perceptions of the thoughts of those who considered to have a role and have hope for him to do something and the extent of the desire to fulfill these expectations. The concept of subjective norms is a representation of the demands or pressures of the environment lived by the individual and showed the individual's belief in the existence of approval or not from social figure [33], [34]. Norms generally formed on the convergence of the values of goodness to individuals in the group. The norm that developed in digital culture is the desire always to improve things. Norms implemented lead to forms of appreciation or social sanctions for the implementation of digital culture.

3.2.3. Regulation

Regulations are written decisions regarding binding behavior, which contain provisions regarding rights, obligations, functions, status, or order. Rules established and issued by the authorized institution [35]. Because it is binding, this regulation can shape digital culture as part of personal obligations in carrying out its activities on campus. The effectiveness of university regulations can do by applying formal awards and sanctions for the implementation or violation of its commitments as an academic community.

3.2.4. Behaviour

Behaviour in general terminology can translate as all actions or actions carried out by humans, which can be observed directly or indirectly [36]. Behaviour is a function of the relationship between the individual and his environment. Different environments can determine differences in human behaviour between one another [37]. Based on a biological perspective, the behaviour is an activity or activity of the organism in question. Behavioural symptoms that can observed directly from organisms influenced by environmental and genetic factors (heredity). The mechanism of fusion and the meeting between the two elements is called the learning process [38].

Behaviour is a mental condition (thinking, opinion, attitude, and etcetera) that reacts to the status that exists outside the object. The reaction can be in the form of an active response accompanied by action. Simply stated, behaviour can refer to as an action taken by a person or group of people based on particular interests or needs based on the values, knowledge, and norms of the group concerned.

Educational psychologist Benjamin Bloom divides human behavior into three elements [39], among others:

1) Cognitive. The cognitive item includes an individual's beliefs that are related to the way individuals perceive the object of attitude to the responses they see and know (knowledge) based on their views, thoughts, beliefs, emotional needs, personal experiences, and information from others.

2) Affective. The affective element refers to the subjective psychological aspects of individuals towards the object of attitude, both positive and negative. Emotional reactions are primarily determined by what believed to be good or harmful to the object of the attitude.

3) Psychomotor. The psychomotor element is also called the behavioural component, which is the attitude component related to predisposition or the tendency to act on the object of the attitude it faces.

In general, human behaviour is an instrument of response (operant response) as a response that is born and develops on specific stimuli. Therefore, certain conditions (operant conditioning) need to be created to form the type of response/ behaviour [39]. The stages of developing behaviour in operant conditioning carried out as part of digital culture are:

1) The introduction of reinforcement elements in the form of rewards/ prizes for the behaviour to be established;

2) Analysis of the components forming the desired behaviour, which arranged in sequence for the formation of the desired behaviour;

3) For the implementation phase the sequence of the components is accompanied by the determination of tentative (temporary) objectives and identification of reinforcements for each component;

4) Carry out a sequence of components that have arranged as a behaviour formation activity.

In the implementation of digital culture, it is necessary to emphasize the importance of collaboration between individuals, so the planned goals will run according to what desired.

3.3. Social Engineering

Humans created as individual creatures and social beings; socializing is a characteristic of humans and is a natural tendency in living the activities of human life. In the context of individual beings, human beings given physical characteristics and characters that make the difference between humans and other individuals; if they are part of a unified social group, they will experience the process of socialization [40]. In the process of socialization with groups/ communities, individuals often have different perspectives that can trigger conflicts of interest, which referred to as social problems. The social problem is the gap between how an individual behaves who seen as opposed to the norm agreed upon by the community members [41]. Social issues are closely related to norms, moral values, and social institutions [42].

The results of research observations can see the gap in the use of information systems applications that have been provided by the university. It is found groups who optimally use information systems, on the other hand, found the academic community who rarely use the information system as a work support tool. In addition to the problem of professionalism, in the context of organizational culture, researchers see this gap as a social problem. To overcome social issues, it is necessary to do the social transformation that emphasizes individual adaptation to social norms and institutions (digitizing policies) that apply. The next stage is changing one's behaviour as aligning itself with these values.

A more systematic and planned social transformation generally carried out through a process of social engineering. In social engineering, the object is the community to form a better system and system as desired by the engineer (the social engineer). Social engineering begins with the emergence of social problems, namely the imbalance between *das-sein* and *das-sollen*, or what happens is not following the expectations of the organization (university). Social engineering is a social planning that leads to the realization of social transformation, which is supported by the internalization of humanization values [40].

Social engineering is a process of social change that is systematic and planned to deal with social problems that implemented using various strategies and stages to make social change happen as desired. For social engineering to run effectively and efficiently, the planning, implementation, and evaluation activities of social engineering activities must involve all stakeholders [43]. Social engineering is not only used for inculcation and strengthening patterns of habits and behavior, but also used to direct the desired goals, eliminate habits that deemed unnecessary/ unsuitable, create new patterns of behavior, and etcetera [44]. Culture can form through formal assignments and habituation processes [45].

Education/ training, from a particular perspective, can be understood as social engineering, where education/ training will provide a fundamental influence in the process of social change. In education, there is a planting of universal values that will change the human paradigm so that humans can have the ability to think critically objectively [40, 46]. Education/ training in this context can be understood as a process of increasing understanding of professionalism and innovation that realized in digital culture.

The education/ training process can be useful if a learning module is available [47], open space for discussion/ debate [48], good implementation process, and involve all parties. The process of education/ training to create a work culture is to form workers' attitudes towards work (work attitudes). This attitude is determined and influenced by three factors, namely: involvement in work, job satisfaction, and work commitment. These three factors are interrelated with one another [49].

In higher education, to establish the digital culture of academic community (especially lecturer, educational personnel, and student as main stakeholders), social engineering is become one of important strategy. Research found that social responsibility attitude and behavior can influence the user satisfaction in higher education, especially for student [50]. More than just being "literate" in technology, social engineering including habits, attitudes and behavior of the academic community in higher education is a trigger to improve the digital culture. Starting from the policies and support of top management, to fostering habits in the use of technology and the creation of a digital culture atmosphere.

4. Lack of the Study

This study specific to the perspective of digitalization in Higher Education. While, one perspective could be the digitization of education by itself. Within this topic not only social engineering is important but also technical engineering of the used technology. Another perspective could be worldwide digital technology transformation which then becomes a central perspective for studying by the teachers and students. With this central perspective teachers and lecturers could be acquainted with the development of new digital technologies which are of will be used in the coming years.

5. Conclusion

A digital culture is a form of higher education adaptation in facing various dynamics of technological development. Digital culture is a modification of the organizational culture that has existed at higher education, with reinforcement in technology/ digital-based work behavior. This research creates a formulation model to enhance the digital culture for the development of main activities in higher education. This model of digital culture implements the university value, MORA value, individual value, and client (technology user) value. The process of cultural formation can carry out with social engineering by modifying values, subjective norms, and work rules. Then, social engineering can change the digital culture to be behavior. For the implementation of digital culture, social institutions take precedence over the formal administrative approach. Therefore, for the further works, the social engineering can be implemented and the behavior of digital culture in higher education can be evaluated and enhanced.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to appreciate and thank UIN Sunan Gunung Djati Bandung academic community for supporting this research.

References

[1] H. Harun, "Kesiapan Pendidikan Tinggi dalam Menghadapi Revolusi Industri 4.0 dalam Perspektif Kelembagaan," in Prosiding Seminar Nasional, Universitas Cokroaminito Palopo, Palopo: 187–194, 2018.

[2] V.G. Gerasimova, M.R. Melamud, D.R. Tutaeva, Y.D. Romanova, N.A. Zhenova, "The adoption of e-learning technology at the faculty of distance learning of plekhanov russian university of economics," Journal of Social Studies Education Research, **9**(2), 172–188, 2018. https://jsser.org/index.php/jsser/article/view/254

[3] B. Lian, "Tanggung Jawab Tridharma Perguruan Tinggi Menjawab Kebutuhan Masyarakat," in Seminar Nasional Program Pascasarjana Universitas PGRI Palembang, 100–106, 2019.

[4] C. Gere, Digital culture, Reaktion Books, 2009.

[5] V. Miller, Understanding digital culture, SAGE Publications Limited, 2020.

[6] L. Muliawaty, K. Alamsyah, U. Salamah, D.S. Maylawati, "The concept of big data in bureaucratic service using sentiment analysis," International Journal of Sociotechnology and Knowledge Development, 2019, doi:10.4018/IJSKD.2019070101.

[7] M. Ezmieralda, "Budaya Digital dan Perubahan Konsumsi Media Masyarakat," in Seminar Nasional FISIP-UT 2010, Universitas Terbuka, Tangerang Selatan, 2010.

[8] M.N. Yunus, U., Heriyati, P., Lindawati, A., & Willyarto, Branding Perguruan Tinggi di Era Digital, Penerbit Qiara Media, 2019.

[9] B.S. Putra, A. Rohman, K. Diantoro, A. Soderi, W.R. Puteri, "IMPLEMENTASI SPMI SEBAGAI UPAYA PERGURUAN TINGGI MENGHADAPI ERA 5.0," Quality Assurance Practice, **1**(1), 2019.

[10] L.R. Salazar, E. Diego-Medrano, "The Role of Millennial Students in Shaping the Digital Learning Culture in Higher Education," Examining Millennials Reshaping Organizational Cultures: From Theory to Practice, 123, 2018.

[11] P.S. Kiser, M. Abdoney, E.A. Teaff, "Black Mirrors and Digital Culture: Digital Citizenship through Pop Culture," 2019.

[12] J. Potter, J. McDougall, Digital media, culture and education: Theorising third space literacies, Springer, 2017.

[13] C. Costa, M. Murphy, A.L. Pereira, Y. Taylor, "Higher education students' experiences of digital learning and (dis) empowerment," Australasian Journal of Educational Technology, **34**(3), 2018. https://doi.org/10.14742/ajet.3979

[14] S. Lange, C. Robertson, "How is digital culture influencing your practice?," Spark: UAL Creative Teaching and Learning Journal, **2**(2), 139–143, 2017.

[15] J. Khalid, B.R. Ram, M. Soliman, A.J. Ali, M. Khaleel, M.S. Islam, "Promising digital university: a pivotal need for higher education transformation," International Journal of Management in Education, **12**(3), 264–275, 2018. https://doi.org/10.1504/IJMIE.2018.092868

[16] M.A. Peters, P. Jandri, "Peer production and collective intelligence as the basis for the public digital university," Educational Philosophy and Theory, **50**(13), 1271–1284, 2018.

[17] A.G. Picciano, Online education policy and practice: The past, present, and future of the digital university, Taylor & Francis, 2016.

[18] E.A. Khitskov, S. V Veretekhina, A. V Medvedeva, O.L. Mnatsakanyan, E.G. Shmakova, A. Kotenev, "Digital transformation of society: problems entering in the digital economy," Eurasian Journal of Analytical Chemistry, **12**(5b), 855–873, 2017.

[19] A. Kalache, "An education revolution in response to the longevity revolution," Revista Brasileira de Geriatria E Gerontologia, **22**(4), 2019. https://doi.org/10.1590/1981-22562019022.190213

[20] J. Gagliardi, A. Parnell, J. Carpenter-Hubin, "The Analytics Revolution in Higher Education," Change: The Magazine of Higher Learning, **50**(2), 22–29, 2018, doi:10.1080/00091383.2018.1483174.

[21] Sharan B. Merriam, Qualitative Research: A Guide to Design and Implementation, 2009, doi:10.1017/CBO9781107415324.004.

[22] J.A. Maxwell, L.E. Reybold, Qualitative Research, 2015, doi:10.1016/B978-0-08-097086-8.10558-6.

[23] M. a Masadeh, "Focus Group : Reviews and Practices," International Journal of Applied Science and Technology, 2012.

[24] M.A. Mahfouz, D.A. Muhumed, "Linking organizational culture with financial performance," Bussecon Review of Social Sciences (2687-2285), **2**(1), 38–43, 2020, doi:10.36096/brss.v2i1.171.

[25] M.R. Saif, N.K. Bin Yeop, "The Effect of Organizational Culture on Knowledge Management and Managerial Performance of Government Department in Dubai," European Journal of Multidisciplinary Studies, **5**(1), 47–53, 2020, doi:10.1002/kpm.344.

[26] E. Cera, A. Kusaku, "Factors Influencing Organizational Performance: Work Environment, Training-Development, Management and Organizational Culture," European Journal of Economics and Business Studies, **6**(1), 16–27, 2020. https://dx.doi.org/10.26417/ejes.v6i1.p16-27

[27] D. Lee, "Impact of organizational culture and capabilities on employee commitment to ethical behavior in the healthcare sector," Service Business, **14**(1), 47–72, 2020, doi:10.1007/s11628-019-00410-8.

[28] W. Darmalaksana, M.A. Ramdhani, R. Cahyana, A.S. Amin, "Strategic Design of Information System Implementation at University," International Journal of Engineering & Technology, 2018, doi:10.14419/ijet.v7i2.29.14257.

[29] E.A. Sudanang, S.E. Priyanto, "Pengaruh Kepuasan Kerja dan Budaya Organisasi terhadap Produktivitas Kerja Karyawan di Horison Apartemen dan Kondotel Yogyakarta," Kepariwisataan: Jurnal Ilmiah, **14**(1), 31–36, 2020.

[30] M.H. Yusuf, "Pengembangan Budaya Organisasi dalam Lembaga Pendidikan," Tarbawi: Jurnal Pendidikan Islam, **14**(1), 81–96, 2017. https://doi.org/10.34001/tarbawi.v14i1.613

[31] Z.S. Thontowi, M. Qowim, A. Dardiri, "Implementasi Lima Nilai Budaya Kerja di Kantor Kementerian Agama Kabupaten Banyumas," FIKROTUNA, **9**(1), 1160–1178, 2019.

[32] Q.Y. Zaqiah, A. Rusdiana, Pendidikan Nilai: Kajian Teori dan Praktik di Sekolah, 2014.

[33] D. Kurniawati, S. Sumarji, "Pengaruh Pengetahuan, Persepsi, Sikap, dan Norma Subyektif terhadap Perilaku Konsumen Muslim dalam Konsumsi

Produk Pangan Halal di Kabupaten Nganjuk," Manajemen Agribisnis: Jurnal Agribisnis, **18**(1), 49–52, 2020.
[34] R.R. Agustin, N. Diana, A. Afifudin, "Pengaruh Ekspektasi Pendapatan, Motivasi, Pendidikan Kewirausahaan, dan Norma Subyektif terhadap Minat untuk Berwirausaha," Jurnal Ilmiah Riset Akuntansi, **9**(04), 2020.
[35] N. Qamar, F.S. Rezah, Ilmu dan Teknik Pembentukan Peraturan Perundang-Undangan, Social Politic Genius (SIGn), Makasar, 2020.
[36] S. Notoatmodjo, Pendidikan dan Perilaku Kesehatan, Rineka Cipta, Jakarta, 2003.
[37] S.P. Robbins, T.A. Judge, Perilaku Organisasi, Gramedia, Jakarta, 2010.
[38] T.H. Basri, Hubungan Tingkat Pengetahuan tentang Makanan Berserat Dengan Pola Konsumsi Makanan Berserat pada Mahasiswa Angkatan 2008-2011 Fakultas Kedokteran Universitas Lampung, Universitas Lampung, 2012.
[39] S. Notoatmodjo, Promosi Kesehatan: Teori dan Aplikasinya, Rineka Cipta, Jakarta, 2007.
[40] T. Syarofi, Studi Analisis Pemikiran Jalaludin Rahmat tentang Social Engineering dan Relevansinya dengan Tujuan Pendidikan Islam, Institut Agama Islam Negeri Walisongo, 2010.
[41] M. Soelaiman, Teori Sosial Dasar: Teori dan Konsep Ilmu Sosial, Rafika Aditama, Bandung, 2001.
[42] P. Jones, L. Bradbury, S. Le Bouttilier, Pengantar Teori-teori Sosial: Dari Teori Fungsionalisme Hingga Postmodernisme, Yayasan Obor Indonesia, Jakarta, 2016.
[43] E. Susilowati, Windiani, "Pendekatan Rekayasa Sosial dalam Pengelolaan Sumberdaya Air: Studi Kasus Pengadaan Tanah untuk Pembangunan Bendungan Semantok di Kabupaten Nganjuk-Jawa Timur," in Seminar Nasional Strategi Pembangunan Nasional Menghadapi Revolusi Industri 4.0, 1–12, 2018.
[44] A.L. Diab, "Peranan Hukum sebagai Social Control, Social Engineering, dan Social Welfare," Jurnal Al-'Adl, **7**(2), 53–66, 2014.
[45] M.S. Mazaya, H. Ainissyifa, "Character education through the obligation to wear santri's dress," in Proceedings of the International Conference on Islamic Education (ICIE 2018), Atlantis Press, Paris, France: 309–311, 2018, doi:10.2991/icie-18.2018.56.
[46] T. Borodina, A. Sibgatullina, A. Gizatullina, "Developing Creative Thinking in Future Teachers as a Topical Issue of Higher Education," Journal of Social Studies Education Research, **10**(4), 226–245, 2019. https://doi.org/10.1590/1981-22562019022.190213
[47] M.S. Mazaya, "Effective practical learning model for the subject of basic information technology," Journal of Physics: Conference Series, **1157**, 42003, 2019, doi:10.1088/1742-6596/1157/4/042003.
[48] M.S. Mazaya, "Logical framework for smart discussion in learning process," Journal of Physics: Conference Series, **1157**, 42002, 2019, doi:10.1088/1742-6596/1157/4/042002.
[49] M.R.K. Umam, "Peran Islamic Management Principles pada Pembentukan Corporate Culture untuk Meningkatkan Kinerja Sumber Daya Manusia pada Organisasi Bisnis," Eka Cida, **1**(1), 110–124, 2016.
[50] D. Gallardo-Vazquez, J.A. Folgado-Fernandez, F. Hipolito-Ojalvo, L.E. Valdez-Juarez, "Social Responsibility Attitudes and Behaviors' Influence on University Students' Satisfaction," Social Sciences, **9**(2), 8, 2020. https://doi.org/10.3390/socsci9020008

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1480-1487 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Simulation Based Energy Consumption Optimization for Buildings by Using Various Energy Saving Methods

Omar Khaled Barakat[1], Ahmed El-Biomey Mansour [1], Mahmoud Mohamed Abd Elrazik[2], Ashraf Aboshosha[3], Amir Yassin Hassan[4,*]

[1]*Department of Electrical Engineering, Faculty of Engineering, Al-Azhar University, Cairo, 11651, Egypt*

[2]*Building physics Institute, Housing and Building National Research Center, Giza, 12511, Egypt*

[3]*Department of Radiation Engineering, NCRRT, Atomic Energy Authority, Nasr City, Cairo, 11651, Egypt*

[4]*Department of Power Electronics and Energy Conversion, Electronics Research Institute, Giza, 12511, Egypt*

ARTICLE INFO

Article history:
Received: 23 September, 2020
Accepted: 04 December, 2020
Online: 21 December, 2020

Keywords:
Energy efficiency
Power saving
Smart grid
Design builder software

ABSTRACT

The world has turned in the modern era to reduce peak times for electricity consumption, where the peak times cause excessive load on the electricity grid and power stations. In Egypt air conditioning loads are considered the main contributing reasons for the peak time's problem due to presence of large unshaded glazed surfaces, which cause high thermal loads inside the buildings. In this study, various methods of energy rationalization in buildings are evaluated by using many scenarios, from the obtained results of this research study. It is found that the use of shading on windows is an effective strategy where the percentage of rationalization using scenarios integration reached 40%, and the financial returns have also been calculated based on this work.

Nomenclature

AHU	*Air Handling Unit*	LED	*Light Emitting Diodes*
BMS	*Building Management System*	PF	*Power Factor Correction*
CVRMSE	*Coefficient of Variation of Root Mean Squared Error*	SHGC	*Solar Heat Gain Coefficient*
DB	*Design Builder*	THD	*Total Harmonic Distortion*
GUI	*Graphical User Interface*	VFD	*Variable Frequency Drive*
HVAC	*Heating, Ventilation, and Air Conditioning Systems*	VSD	*Variable Speed Drive*

1. Introduction

The rise of 4G industry era in our nowadays life led to a great demand of efficient energy systems specially the requirements of the large-scale industrial systems. This led to a significant rise in electricity prices; therefore energy conservation became a crucial policy [1]. It is an influential reason for preserving the environment and climate [2]. Therefore the world has turned to energy efficiency [3]. With the continuous development of technology, the smart home applications appeared in its contribution to improve energy efficiency effectively. As the smart grid deals with the flow of energy in distribution networks whose objectives are to balance supply and demand [4]. The more energy transfers from one form to another, the more losses occur. With the increasing of energy transformations, it is found a reduced gab between production and the consumer also an increase in the efficiency of devices and equipment, the loss decreases, the consumption be more efficient, rationalization increases and helps in reducing the demand for increased power plants [5].

Energy management is the assessment and measurement of electrical energy consumption in buildings [4]. As the global economy develops, it is necessary to strengthen and improve research and development in building energy management. The association of energy classification methods in new buildings is also necessary to develop building policies that are of great importance. Therefore, the main functions of energy management systems are to monitor, control and improve energy flow in general [6]. Energy management systems start from the use of energy-

*Corresponding Author: Amir Yassin Hassan, 33 El-Tahrir st., Dokki, Giza, Egypt., Email: amir@eri.sci.eg

saving devices, through smart buildings to supervisory control applications [7].

Energy auditing is also a review and evaluation of building energy use and identification of energy saving measures to reduce consumption [6]. This is through the analysis of public service bills such as electricity bills [4]. Throughout this framework a professional power saving protocol is applied to reduce the electrical energy consumption in public buildings. This protocol includes - but not limited to - power factor correction, variable speed control, earth leakage current treatment, lighting management, heating, ventilation, and air conditioning (HVAC) system, load management and intelligent energy system. Correction of power factor has many economic benefits, one of those, is saving the financial fine that the electricity company imposes when the power factor drops below 0.90 [8], [9]. Technically, it reduces the overload on transformers and generators and reduces the overload on cables and switches [10]. The variable speed drive (VSD) and frequency speed drive (VFD) when used in the Air Handling Unit (AHU), it can save energy up to 18% and save energy at least 7.5% when used in Chiller pumps [10], [11]. Earthing leakage current treatment is obtained by isolating cables and conductors according to the Egyptian code for electrical installations and connections in buildings. Assuming that the ground leakage current in a residential or administrative unit is 30 mA, this means that the monthly energy loss is about 4.75 kilowatt hours [8].

The use of light emitting diodes (LED) bulbs technology, which is considered an improvement in lamp efficiency, leads to limited energy savings, but when maximizing daylight, it is one of the measures to achieve a low energy design [12]. Also, a study was conducted to study the effect of changing the power factor in the LED bulbs on the total harmonic distortion (THD), the study was conducted on a house and a hotel, and the result was that it contributes to increasing harmonics [13]. As well as in another study on the effect of changing the feeding voltage of the LED bulbs (160-240) volts on the quality of the bulb and the color characteristics of the bulb, it was found that the bulb of good and approved industry is not affected by this change [14]. A comparison of the LED bulbs feeding was done with a constant voltage (DC) and a variable voltage (AC) and the result was that the lamp efficiency increased from 5% to 10% in the case of feeding using DC more than AC [15]. Also, two of the typical methods for achieving effective energy management and lighting control are working to integrate natural and industrial lighting or using highly efficient bulbs [16].

HVAC systems are responsible for the occupants 'comfort in the building by saving internal temperature, humidity, and the total air quality inside the building. Therefore, they are responsible for a large part of the building's energy consumption [17]. It is noted that the use of technological developments increased the efficiency of HVAC systems and thus reduced energy consumption, for example, the use of automatic natural ventilation by using suction fans and automatic windows. When a strong wind or storm occurs, it closes automatically [17]. Various concepts and elements have been developed to reduce energy use for cooling and heating in buildings [16] Such as:

- Mixed method: Use natural ventilation when the external temperature is moderate and use mechanical if necessary.
- Air distribution system: there are fans that work when there are people to reduce the thermal loads inside the building. Therefore, it is noted that the application of this system in the German commercial bank building, Commerzbank, where the architect Norman Foster has used [16]:

Natural lighting from the surrounding windows with hinges and can be opened and shaded by a Building Management System (BMS) system.

Natural ventilation for approximately 60% of the total hours of use, and mechanical ventilation is used in extreme conditions. There is also a system for night ventilation where the temperature is moderate in summer for pre-cooling of the thermal mass and the possibility of using the heat switch for the ground and groundwater.

2. Methodology

The research aim is the possibility of using simulation programs to determine the energy consumption per square meter annually, whether the building is finished or in the design stage and Economic study of the methods of rationalization in the simulation program and the application of the most feasible method in practice. The workflow of this study will be structured as follows:

- Determining the study case to implement the energy rationalization mechanism.
- Preliminary evaluation of the case study using smart electricity meters.
- Evaluating using the DB simulation program with studying the following:
  - The effect of rationalization on change the Set Point of the Cooler device on consumption rationalization.
  - Effect of rationalization in lighting.
  - Rationalization by fixed shading.
  - The effect of rationalization in glass types on the rationalization of energy.
  - The effect of rationalization in the use of thermal insulation.
- presenting of simulation results for rationalization scenarios.
- Making a Comparison between actual consumption and results of simulation program consumption (DB) and then calculating the error.
- Discussing the rationalization scenarios used in the study.
- Evaluation of the economic return.

3. Model Description

Air conditioning engineers ASHRAE (the Energy Accounts Committee) completed a new project in the late 1990's to create mathematical algorithms that can be used for various building models (including the contents of a building's cover, appliances, and equipment in the building).It is also available to the simulation community [18]. Therefore, the Design Builder (DB) program is an easy-to-use design environment that helps in designing the construction model. Therefore, it loads heating, cooling and

lighting calculations using heat. Likewise, energy conservation indicators, annual consumption, thermal comfort indicators, temperature and humidity, heat transfer from walls and windows, air leakage and architectural shading calculations can be analyzed every hour and a study of rationalization methods more efficiently and with high accuracy [19]. DB is a GUI for Energy plus [20].

The case study was chosen on the western front of building of the National Center for Housing and Building Research, which is a government administrative building located in Giza Governorate, as shown in Table 1 and Figures 1 and 2. That by selecting the case that contains lighting and air conditioning loads only as shown in Table 1. As well as working hours start from 8 Am in the morning until 5 Pm in the afternoon.

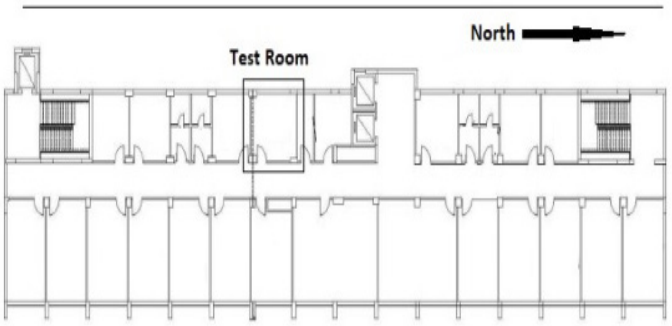


Figure 1: Case Study Location

Table 1: General Description of the Case Study

Item	Description
The building's name	The National Building and Housing Research Center (HBRC) is a government administrative building of the Ministry of Housing.
Building location	87 Tahrir Street - Dokki - Giza Governorate
The shape of the building design	As Figure 4
Case study location	As Figure 3 shows the location of the room on the fifth floor, where it overlooks the western façade, and there is another floor above this floor.
Room Area	14.5 m^2
Window	Open Area (Window): **2.95 m^2** Wall Area: **12.5 m^2** Window to wall ratio: 23% Sill Height for window (m): 1.3 Window Height (m): 1 Shading: no
Type window	TV: 0.88; SHGC: 0.82; U-Factor: 6.17; type: clear Single (6mm)
Construction	Physical properties of thermal Resistivity $m^2.C°/w$ / Specific heat (J/Kg.C°) / Conductivity (W/M.C°) / Density ρ ($k\ g/m^3$) / From the outside to the inside 0.37 / 840 / 1.6 / 1800 / Pink brick 0.02 / 730 / 1.25 / 1200 / Mortar paint 0.38 / Total resistivity
Activity	Density (people / m^2): 0.10 -0.11 Lighting energy (w/ m^2): 21.9 Computer Gain (w/ m^2): zero Office Equipment Gain (w/ m^2): zero
Climate region	Delta and Cairo region .As shown in the Figure No (3)
Room Details	cooling loads Equipment description / Quantity of equipment (unit) / Total power Unit (Watt) Air Condition window / 1 / 2000 lighting loads Quantity of fixture (units) / Fixture Type / Total power for Room (Watt) 2×4 lamp / Fluorescent lamp / 310 Watt ×9 Hours/day =2790 Watt

Figure 2: The Building Design

4. Simulation Results

4.1. Modeling and Validation Check

Case study simulation was done on DB program as shown in Figure 3. The validity of the simulation program was also confirmed by comparison between the simulation program results and the laboratory measurement of the energy consumption of the air conditioner by a smart open source electricity meter that measures every half a second and maintains these measurements in a built-in memory as shown in Figure 4. Also, the smart meter measurement was also compared with a certified clamp meter measurement and the error rate was 0.02%. As for lighting consumption, it was constant throughout working hours. Laboratory measurements of air conditioning loads were done on August 25, 2019 (8Am – 5Pm) and the measurement was done without using shading. Also on September 4, using 55% fixed shading. The beginning of laboratory registration was from 10 in the morning until 5 in the afternoon. Therefore, Table 2 Shows the result of comparison and the error rate between the simulation program and the laboratory measurement. From equation 1 the error associated with the energy models may be declared valid the CVRMSE within ±30% when using hourly data or 5% to 15% when using monthly data [20].

$$\text{CVRMSE} = 100 \times \frac{\sqrt{\overline{(Y_{measured} - Y_{modelled})^2}}}{\overline{Y}_{measured}} \quad (1)$$

Where, $Y_{measured}$ is the measure data point, $Y_{modelled}$ is the modeled data point and $\overline{Y}_{measured}$ is the average of all measured data points.

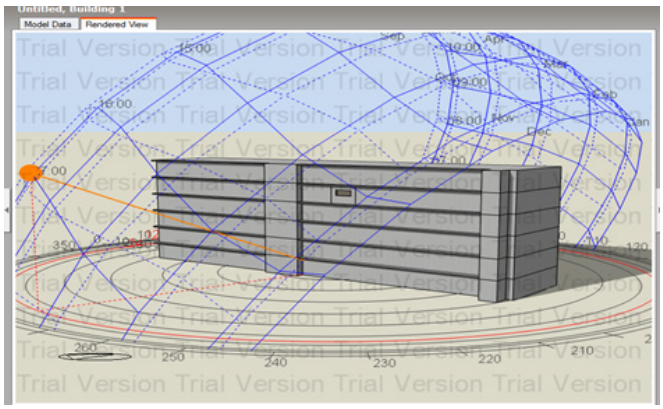


Figure 3: The Building Modeling in the Simulation Program

Figure 4: The Smart Electricity Meter for Measuring and Recording Energy

Table 2: The Error Percentage between Simulation and Laboratory Measurement

Time	Error (%)	Simulation Air condition consumption (KW)	actual Air condition consumption (KW)
Total Daily (8Am – 5Pm)	1.46%	22.74572	22.4177

4.2. Simulation result with the basic case

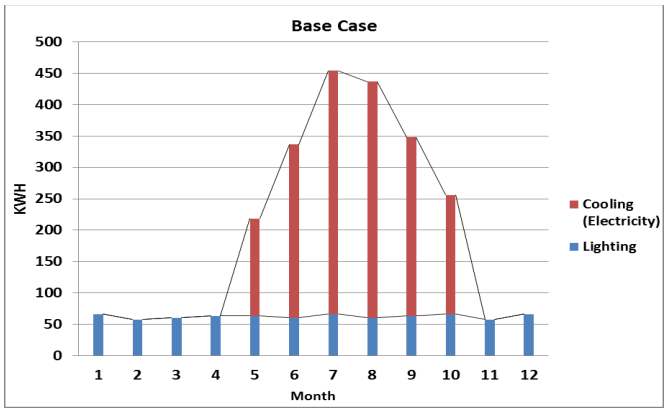


Figure 5: The Monthly Consumption of the Basic Case

Table 3: The Annual Consumption with the Basic Study Case

Month	Lighting (KWh)	Cooling (KWh)
1	66.22	0
2	57.58	0
3	60.46	0
4	63.34	0
5	63.34	155.30
6	60.46	275.82
7	66.22	387.90
8	60.46	376.13
9	63.34	284.59
10	66.22	190
11	57.58	0
12	66.22	0
Summation	751.50	1669.78
Total	2421.29	

4.3. Simulation Result of the Consumption Rationalization by Changing the Set Point of the Air Conditioning Device:

Table 4 and Figure 6 illustrate the electricity consumption while simulating with changing the Set Point of the air conditioning device.

Table 4: The Electrical Energy Consumption for Each Set Point

NO.	Set point (°c)	Total (KWH)	Consumption increase %	CO2 increase	kwh/m2
1	21°C	3357.38	38.66%	38.66%	230
2	22°C	3078.35	27.14%	27.14%	211
3	23°C	2808.18	15.98%	15.98%	193
4	24°C	2548.40	5.25%	5.25%	175
5	24.5° C	2421.29	0.00%	0.00%	166

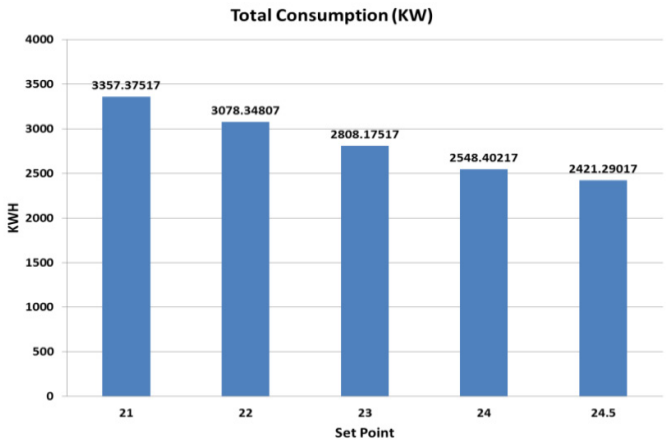


Figure 6: The Electrical Energy Consumption at Set Point

4.4. Simulation Result of the Consumption Rationalization by Lighting Control

Table 5 and Figure 7 show the electricity consumption while simulating with rationalization by lighting control.

Table 5: The Ratios of Rationalization Using Lighting Control Systems

NO.	Type of Saving use	KWh/m^2	Co_2 saving	Rationalization rate
1	STEPED Controller	98.15	40.9%	40.9%
2	Linear Controller	115.93	30.2%	30.2%
3	LED	113.81	31.5%	31.5%
4	LED+SENSOR	100.36	39.6%	39.6%

4.5. Simulation Result of the Consumption Rationalization by Changing the Glass Type

Table 6 and Figure 8 indicate the simulation results while using rationalization by changing the glass type.

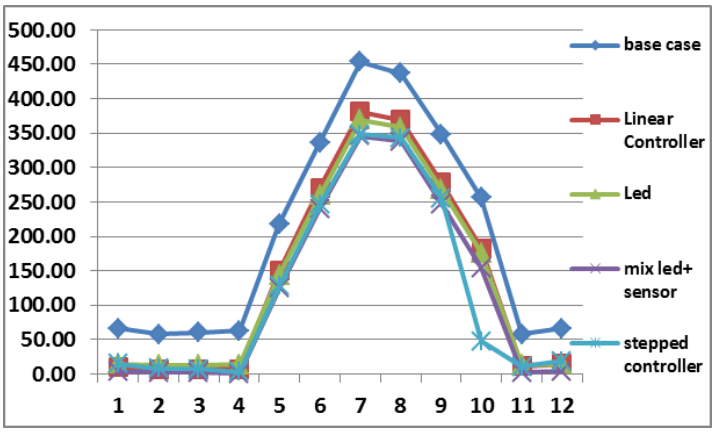


Figure 7: Simulation Results in Electrical Energy Consumption for Each Type of Lighting Control

Table 6: The Simulation Results in Electrical Energy Consumption for Each Type of Glass Used

Item	Single(base case)	Dbl Blue 6mm/13mm Air	Thin film car core
Annual (KWh)	2421.29	2271.1611	2144.1506
SHGC	0.83	0.497	0.28
Rationalization rate %	-	6.3%	11.6%
Co2 saving %	-	6.3%	11.6%

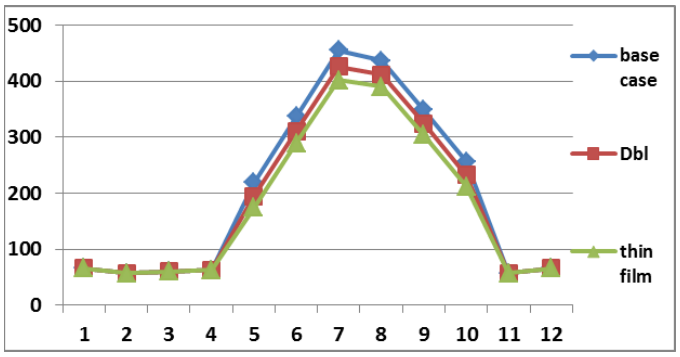


Figure 8: The Simulation Results in Electrical Energy Consumption for Each Type of Glass Used

4.6. Simulation Result of the Consumption Rationalization Using Thermal Insulation of Facade Walls

Table 7 and Figure 9 show the simulation results of rationalization with thermal insulation of facade walls.

Table 7: The Cooling Consumption Based on Wall Resistance

R	Cooling (KWh)
0.26	1815
0.292	1556
0.31	1501
0.32	1485
0.339	1485
0.35	1497
0.39	1561
1	1724
2.853	1907.3216

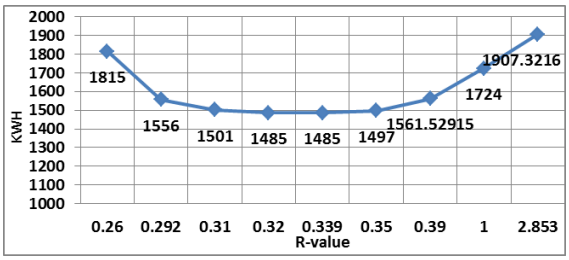


Figure 9: The Energy Consumption Using Thermal Insulation

4.7. Simulation Result of the Consumption Rationalization by Using Shading Control

Table 8 and Figure 10 illustrate the simulation result swith shading control.

Table 8: The Rationalization Scenarios Using Shading Methods

No.	The type of shading used	KWh/m^2	CO2 kg/kWh equiv.	rationalization percentage
1	base case vertical shading	166	-	-
2	When the window shading ratio 55%	143	14.1%	14.1%
3	High reflectance - low transmittance shade	156	%6.0	%6.0
4	Drapes - closed weave light	162	%2.3	%2.3
5	Blind with high reflectivity slats	161	3.2%	3.2%
6	Transparent insulation	164	1.3%	1.3%
7	Electrochromic reflective 6mm	158	4.9%	4.9%
8	horizontal fixed shading 1.0 m	149	10.5%	10.5%
9	horizontal fixed shading 1.0 m and From both sides 1.0 m	147	11.7%	11.7%

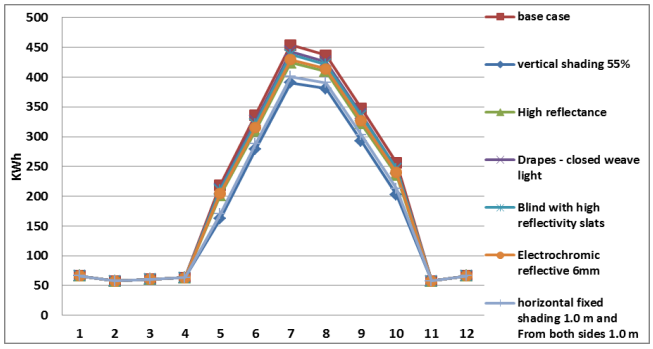


Figure 10: The Rationalization Scenarios Using Shading Methods

4.8. Simulation Result with Integration of Rationalization Methods

Table 9 and Figure 11 show the simulation result with integration of rationalization methods (mix 1 – mix2 – mix3)

Table 9: Integrations Between Scenarios

Scenarios Integration	saving %	KWh/m^2	Payback period (Year)
Led	31.5%	114	1.1
mix1 (Led +Daylight sensor)	39.6%	100	3.6
Mix2 (Led +shading 55%)	43.0%	95	3.4
mix3 (Led + Day light sensor+ shading 50%)	50.5%	82	5.1

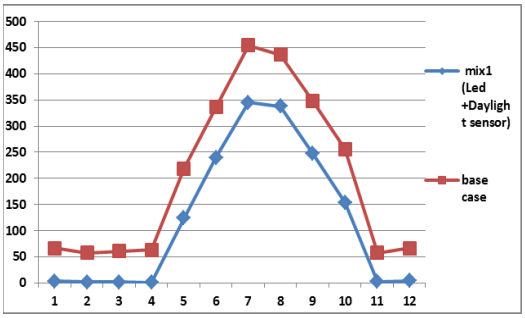


Figure 11: (a) Simulation Result for Comparison Between Total Consumption for Base Case and When Using Integration mix 1 (Led +Daylight Sensor)

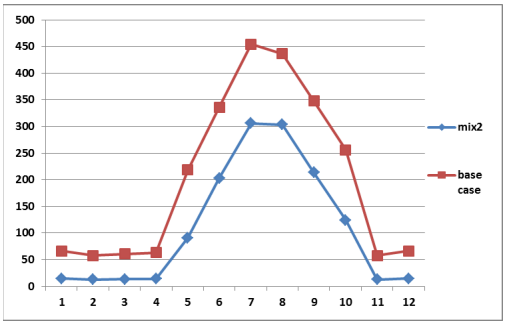


Figure11: (b) Simulation Result for Comparison Between Total Consumption for Base Case and When Using Integration mix 2 (Led + Vertical Shading at 55%).

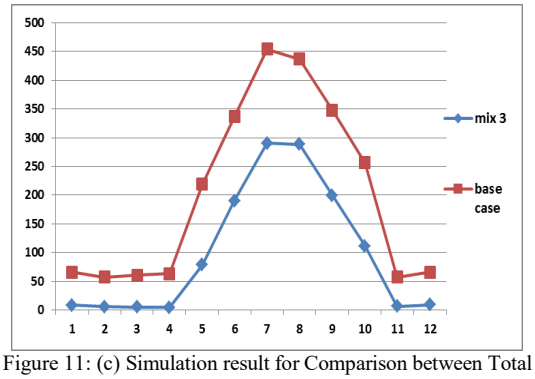


Figure 11: (c) Simulation result for Comparison between Total Consumption for Base Case and When Using Integration mix 3 (Led + Day Light Sensor + Vertical Shading at 55%)

5. Discussions

5.1. Electricity Consumption for the Basic Case

After making the comparison and proving the error rate within the mathematically allowed, an annual consumption of the research study, shown in Figure 5 and Table 3, has been made. Also, it was taken into consideration that this case uses air conditioning in the period of 6 summer months, starting from May to October only. As for the lighting consumption, it is used throughout the year, and it turns out that the average consumption of electrical energy is 166 KWh/m^2. Where also by calculates the emission of carbon dioxide due to consumption 1383.8Kg CO2-eq / kWh considering that each kilowatts consumption is 0.57151 kgCo2 [21]. Also for calculate the emissions per square meter is 94.9kg-CO2-eq / m2.

5.2. The Effect of Set Point Change of Air Conditioning Device on the Rationalization of Energy Consumption

The effect of electrical consumption when using Set point for the air conditioning device on the following temperatures has been studied: (21°c - 22°c - 23°c - 24°c – 24.5°c). Table 4 shows the electrical consumption for each Set Point. Therefore, it was found that setting the temperature to 24.5 ° C saves energy consumption 38% compared to setting the temperature at 21 ° C as in Figure 6. Therefore, it is considered one of the best rationalization and not expensive. The annual consumption of the base case was also calculated by setting the temperature to 24.5° .

5.3. The Effect of Lighting Control on the Rationalization of Energy Consumption

In this case, the daylight factor was suitable. Therefore, the focus was on using rationalization by replacing fluorescent bulbs into low-watt LED bulbs [22]. Also, using a daylight sensor using the Step 1 method, as there is only one office in the case study, and from Table 5 it was found that rationalization of changing old lamps to LED technology saves 31% of the total consumption. Also, rationalization using a daylight sensor saves 36%, as in Figure 7.

5.4. The effect of glass types on the rationalization of energy consumption

Table 6 shows a study of rationalization in consumption by reducing solar heat gain coefficient (SHGC). Also in Figure 8, it is noted that the lower the SHGC, the lower consumption.

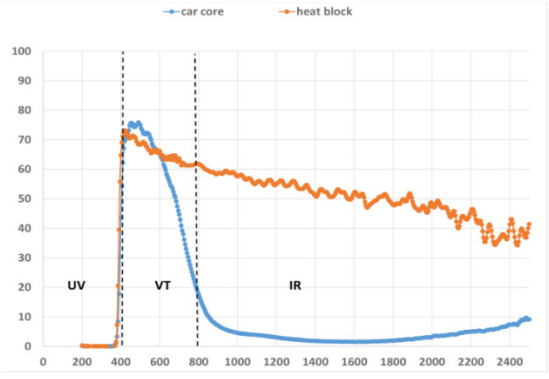


Figure 12: A Thin Film Car Core Characteristic

Also, with advanced technology, day by day, the focus will be on making materials that reduce SHGC and increase the visible transmission light as shown in Figure 12 for a thermoforming glass sample that shows its ability to reduce SHGC and increase the visible transmission light.

5.5. The Effect of Thermal Insulation on the Rationalization of Energy Consumption

From Table 7 it was found that using insulation in the western facade increases the energy consumption in summer due to the heat retention within the research study. Also Figure 9 shows the best resistance of the external wall

5.6. The Effect of Shading Control on the Rationalization of Energy Consumption

Figure 10 and Table 8 show the scenarios for rationalization by shading. Therefore, it was found that vertical shading by 55% as in Figure 13 of the window area is better in rationalization than horizontal and vertical shading on both sides of the window length of 1 m . Vertical shading is characterized by having a motor that can control the rate of openings for windows. Therefore, it is possible to maximize the benefit of the sun in winter and increase shading in summer. Also, observed decrease of the value of the flowing current in the wire was 14%.Also, the reason for choosing 55% shading is because at this point the lighting intensity was measured at the desktop level by lux meter in 4-sept -2019 during working hours (8am - 5pm) and the minimum was 300 lux. Accordingly, this shading was bought as in the Figure 13 and it is controlled according to occupant's desire. As lighting consumption constant, sureness of simulation program validity is by comparison of electric energy consumption that resulted from cooling as in the table 10

Figure 13: 55% Shading During the Summer Period

Table 10: The Comparison Between The Readings Recorded in the Simulation Program and Measured by Smart Meter

Time	Simulation (KWH)	Measure (KWH)	Error %
10 a.m.	2.312434	2.21625	5.16%
11 a.m.	1.775199	1.5	14.76%
12 p.m.	1.708757	1.575	7.17%
1 p.m.	1.681947	1.8375	8.34%
2 p.m.	1.725026	1.8	4.02%
3 p.m.	1.72124	2.0625	18.30%
4 p.m.	1.704831	2.0625	19.18%
Total	12.629434	13.05375	3.3%

5.7. The Effect of Integration on the Rationalization of Energy Consumption and Return on Investment (ROI)

Table 9 shows the methods of integration, the percentage of rationalization in each scenario, and the period for recovering capital from the energy saving (Assuming electricity tariff in Egypt is 1.5 EGP/kWh), the more integration scenarios of rationalization, the greater the savings rate as in Figure 11. It can be noted that the non-use of photovoltaic energy due to our approaching to the lowest station that can be connected to the distribution network in Egypt is 5 kilowatts. Therefore, it is not economically feasible in this case. The economic feasibility to rationalization can be measured by the following equation [18]:

$$\text{Expected payback period} = \frac{\text{initial cost+operation and maintenance costs}}{\text{cost operation saving}} \quad (2)$$

6. Conclusion

Optimizing the utilization of energy in buildings tends to save more power. The "Design Builder" simulation program is used for simulating the study building. In this paper, various energy saving methods are used to optimize the energy consumption in buildings. It can be found that the more integration between saving scenarios, the higher economic return. The behaviour of occupants poses a great challenge in implementing the rationalization methods. It can be noted that the percentage of rationalization that reached is large, that is due to lighting and air-condition loads only. Controlling the temperature of the refrigeration device represents the more simple and effective rationalization method with low cost and significant rationalization ratio. Also it is recommended to replace old bulbs with high efficiency bulbs because they have an economic return and reduce the consumption of cooling loads indirectly. Also rationalization can be achieved by the use of shading and lighting control using the daylight sensor. Shading helps in converging the average electrical energy consumption in the summer and winter seasons, as well as helps in reducing the consumed power by the air conditioners. Rationalization in air conditioning loads helps in achieving a balance in the average monthly consumption during the year where an annual energy saving about 39.6% is achieved. Comparison between the simulation results and the measured results shows that the simulation program can be reliable in studying the economic feasibility of rationalization scenarios before implementation.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Special thanks to Electronics Research Institute and the National Center for Housing and Building Research.

References

[1] M.H.M. Alham, Optimal operation of electric power systems considering wind power uncertainty incorporating energy storage system and demand side management, Ph.D. thesis, Cairo Universty, 2016.

[2] P.A. Owusu, S. Asumadu-Sarkodie, "A review of renewable energy sources, sustainability issues and climate change mitigation," Cogent Engineering, **3**(1), 1–14, 2016, doi:10.1080/23311916.2016.1167990.

[3] M.A. Tovar Reaños, N.M. Wölfing, "Household energy prices and inequality: Evidence from German microdata based on the EASI demand system," Energy Economics, **70**, 84–97, 2018, doi:10.1016/j.eneco.2017.12.002.

[4] N. Bin Ishak, Energy management practices and implementation in campus building, Master Thesis, University Malaysia Pahang, 2016.

[5] T. Davis, J.Hanania, K. Stenhouse, L.Vargas Suarez, J. Donev, Energy loss, Energyeducation, 2015.

[6] S.K. Singh, B.K. Tiwari, A.K. Pandey, "Energy conservation: Analysis improvement through energy audit," 3rd International Conference on Innovative Applications of Computational Intelligence on Power, Energy and Controls with Their Impact on Humanity, CIPECH 2018, 203–209, 2018, doi:10.1109/CIPECH.2018.8724220.

[7] B. Zhou, W. Li, K.W. Chan, Y. Cao, Y. Kuang, X. Liu, X. Wang, "Smart home energy management systems: Concept, configurations, and scheduling strategies," Renewable and Sustainable Energy Reviews, **61**, 30–40, 2016, doi:10.1016/j.rser.2016.03.047.

[8] HBRC, Egyptian Code for Improving Energy Efficiency in Commercial Buildings, Egypt, 2017.

[9] EEHC, Annual Report of Egypt Electricity 2017/2018, The Egyptian Electricity Holding Company, Ministry of Electricity and Renewable Energy, Egypt, 2018.

[10] K. Braun, E. Eaves, C. Giambri, D. Chapman, H. Heavner, J. Woodward, J. Nagel, K. Gipson, "Reducing electrical energy consumption of AHU fans through the integration of variable frequency drives," 2016 IEEE Systems and Information Engineering Design Symposium, SIEDS 2016, 61–65, 2016, doi:10.1109/SIEDS.2016.7489328.

[11] A. Kasman, W. Ruslan, W. Basuki, "Performance Analysis of Chiller System," International Journal of Applied Engineering Research ISSN 0973-4562 , **14**(13), 3017–3021, 2019.

[12] M. Ljubenovic, P. Mitkovic, B. Stojanovic, M. Ignjatovic, J. Janevski, P. Zivkovic, "Intelligent Skin and Occupancy in the Context of Increasing Energy Efficiency in Buildings," ANNALS of Faculty Engineering Hunedoara – International Journal of Engineering, 201–208, 2018.

[13] M.S. Islam, N.A. Chowdhury, A.K. Sakil, A. Khandakar, A. Iqbal, H. Abu-Rub, Power quality effect of using incandescent, fluorescent, CFL and LED lamps on utility grid, 2015 1st Workshop on Smart Grid and Renewable Energy, SGRE , 2015, doi:10.1109/SGRE.2015.7208731.

[14] I.A.V. Damanik, N.I. Sinisuka, "The effect of voltage variation (160-240volt) on lighting quality and color properties of LED lamps in Indonesia," 2016 3rd IEEE Conference on Power Engineering and Renewable Energy, ICPERE 2016, 179–183, 2016, doi:10.1109/ICPERE.2016.7904865.

[15] A. Jhunjhunwala, K. Vasudevan, P. Kaur, B. Ramamurthi, Kumaravel, S. Bitra, K. Uppal, "Energy efficiency in lighting: AC vs DC LED lights," 2016 1st IEEE International Conference on Sustainable Green Buildings and Communities, SGBC 2016, 7–10, 2016, doi:10.1109/SGBC.2016.7936068.

[16] M. sanusi, A. abd alghani , "The use of modern technologies Technologies for Developing Banks Buildings in Egypt", International Journal in Architecture, Engineering and Technology, "(article in Arabic)", Egypt, 2018, doi:10.21625/baheth.v1i1.192.

[17] B. Tetik, Energy performance of smart buildings: simulating the impact of active systems and passive strategies, Master's thesis, Pontificia Universidad Catolica del Peru, 2014.

[18] A.M.A. Abbady, Improvement of thermal energy efficiency of office buildings by integration of passive and active, Master's thesis, Cairo University, Egyp, 2019.

[19] Team DesignBuilder, DesignBuilder v6 Simulation Documentation, DesignBuilder program, 2019.

[20] J. Anderson, Modelling and performance evaluation of net zero energy buildings, Master's thesis, University of Wollongong, 2016.

[21] Egyptian Electric Utility and Consumer Protection Regulatory Agency (EgyptERA), Electricity report monthly, EgyptERA, EGYPT, 2020.

[22] The Ministry of Electricity and Renewable Energy in Egypt, Energy Efficiency for Lighting & Appliances Project in egypt, MOEE, Egypt, 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1488-1498 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Impact of Changing Microstructural Compositions of Lime Based Mortar on Flexibility: Case Study of Sustainable Lime-Cement Composites

Sule Adeniyi Olaniyan*

Department of Architecture, Ladoke Akintola University of Technology, Ogbomoso, 210211, Nigeria

ARTICLE INFO

Article history:
Received: 02 April, 2020
Accepted: 03 December, 2020
Online: 21 December, 2020

Keywords:
Composite
Flexibility
Modulus of Elasticity
Porosity
Sustainability

ABSTRACT

Lime mortar, an age-long building material is primarily popular for its flexibility, the basis of its ability to accommodate masonry deformation, hence durability. However, lime's characteristic delayed setting/hardening time, low mechanical strength and poor internal cohesion often characterised by volumetric changes have put its use into decline. These shortcomings have therefore relegated relevance of this fundamental flexibility feature which underscores limes' excellent performance and durability characteristics. The research therefore attempted to leverage this feature through evaluation of synergised lime composites, using cement as a partial replacement for lime. This is with a view to integrating advantageous features of 'lime and cement' as a composite, at the expense of their known individual drawbacks. The methodology involved mortars with the same Binder/Aggregate (B/A) mix ratio (1:3) using five different compositions of 'cement-lime' binders (i.e. 1:1, 1:2, 1:3, 2:1 and 3:1). The research focused on comparative evaluations of each composition in both fresh and hardened states, with the latter covering twelve-month curing period. While Water/Binder ratio constituted the fresh state assessment parameters, mechanical characteristics and microstructural features were evaluated in the hardened state. Results of the investigation show that progressive addition of cement significantly changes pore size distribution (PSD) of lime mortar from predominant pore sizes between (0.5 – 5 µm) and (5 – 20 µm) into (10 nm – 2 µm) range. This alteration is associated with porosity reduction by up to 11%. Significant improvements in the mechanical strengths of the composite is recorded as both the compressive and flexural strengths of the composite with 75% of cement is 18 and 6 times higher respectively, compared with the reference mortar. However, progressive addition of cement is proportional to the E-value of the composite (with a clear linear relationship), leaving a negative impact on the flexibility. Nonetheless, all the composites investigated exhibit elastic behaviours relative to the basic lime mortar. In particular, composite with cement addition up to 33% of the binder compositions exhibits deformation tendencies under compression. However, mortars with higher cement compositions (i.e. above 33%) would strain linearly until failure occurs suddenly with minimal deformation. Substitution of lime with cement therefore has a significant impact on the microstructural compositions of lime mortar, and subsequent improvement on the performance of the composite. Despite the improved mechanical strengths, inherent flexibility of lime is maintained though negatively impacted, subject to the amount of lime substituted. Hence, relative to specific purposes, lime revival can be promoted in form of sustainable lime-cement composites.

1. Introduction

Lime mortar, an age-long building material [1], is primarily popular for its flexibility, the basis of its ability to accommodate masonry deformation. It has the flexibility to cushion masonry joints to absorb strains, prevent cracking and result in medium to high flexural bond strengths [2]. By virtue of its reasonably high flexibility, lime mortar exhibits low elastic modulus, thereby displaying capability to deform more on load application relative

*Corresponding Author: Sule Adeniyi Olaniyan, Email: saolaniyan@lautech.edu.ng

to Portland cement [3]. The modulus of rupture and the bond strength of an appropriate, well cured mature lime mortar are such that movement joints are not normally required in new (traditional) construction and any movement experienced (i.e. structural, seasonal and thermal) is taken up by minute adjustment over many joints due to their 'plastic' and 'self-healing' properties [4]-[6]. In addition, lime mortar possesses excellent permeability feature via its relatively large interconnected pore structures. These pore structures allow ice crystal growth in frost periods, thereby accommodating the crystals within the pore structures without causing deterioration of the matrix [7]-[9]. This feature enhances durability of lime mortars in a building fabric against environmental conditions. Lime mortar also exhibits phenomenal 'breathability' through which moisture and vapour transfer from the external environment are freely dissipated via its permeable material, in view of its capillary porosity. This enhances the performance of the materials and structure holistically [2, 7, 10, 11]. Lime mortars also have superior water retention as values from 94.2 to 99.5% have been consistently measured against those of 60-80% of Portland cement equivalents [2]. This property enhances workability which improves contact between mortar and substrate, thereby increasing bond.

With growing emphasis on the need for reduced energy consumption and minimised atmospheric CO2 concentration [12], continued use of lime mortar in building has significant environmental benefits: manufacture of limes consumes less energy and produces less greenhouse gases (compared with Portland cement); its exposure to the atmosphere as lime based mortars absorb most or all of the carbon dioxide that was driven off during its calcination, a phenomenon called re-carbonation; masonry laid using lime based mortar has lower bond strength (than cement) that the units can be prised off easily thereafter, thereby facilitating recycling of the materials, and; building structures finished with lime mortars are usually characterised with low thermal conductivity as this affects the interior surface temperatures of buildings, and may therefore perform better as an insulating material [8, 11, 13].

Despite the properties as stated above, lime mortar is connected with exaggeratedly long setting and hardening periods, low internal cohesion, volumetric changes (i.e. shrinkage, particularly, aerial lime), relatively low mechanical strengths and a high water absorption capacity through capillarity. These have substantially impacted negatively on project delivery periods and significantly resulted in its relegation and relative disuse [9], [14]-[17]. To leverage some of the remarkable features particularly, flexibility, this situation therefore leads to the study of blended binding materials, in form of lime-cement composites.

Cement is a hydraulic material in nature because of its ability to set and harden under water by virtue of a chemical reaction with it [18, 19]. It is characterised with high compressive strength and exhibits higher thermal expansion coefficient than most masonry. It hinders accommodation of movements resulting from creep or thermal effects, and exhibits negligible plastic deformation under load. Thus, it fails by brittle fracture [4], [20]-[26]. Effect of cement on porosity and pore size distribution of lime mortars is crucial since changes in microstructural features of lime can substantially modify its overall performances. In conservation works especially, cement is considered chemically incompatible with lime based mortars as it has low permeability, responsible for introduction of soluble salts and characterised with a high modulus of elasticity that is unfit for accommodation of masonry deformations [27]-[32]. It is also established that Portland cement reacts with carbonic acid to form alkali carbonate or bicarbonate salts which are undesirable. However if Portland cement is used in a small amount, this will remain a minor problem as previous works show that an increase in cement content does not proportionally increase the amount of soluble salts [30, 33].

The evolving lime-cement composite would therefore share in the advantages of lime and cement mortars, avoiding or reducing their individual disadvantages [34]. These blended materials will be characterized by faster setting than lime and better flexibility than cement mortars, which generally improves their application [34]-[40]. The underlying purpose of this effort is driven by the need to improve lime mortar performances and reduce carbon dioxide emissions related to the production of cement [41]. It would also facilitate protection of the environment and conservation of energy resources along with the advent of a sizable market, for renewed interest in the use of lime as a building material [14, 42, 43].

2. Experimental Procedure

2.1. Materials and mortar Preparation

Materials used for the present research have been products commercially available in the market. Natural Hydraulic Lime of the class NHL-5.0 (St Astiers, UK), characterised with the lime's highest compressive strength, based on the 28-day test [44] was adopted for this research. This lime type (NHL-5) was chosen based on its relatively short setting time for optimal strength yield, its availability, ease of handling, as well as the need to maximise lime performance behaviour [14, 45]. Table 1 shows Lime's average Particle Size Distribution obtained through laser diffraction method, using Laser Difractometry Xmastersize. This method is based on measurement of the laser beam scattered by the particle analysed. The distribution indicates particles with main equivalent diameter of 33.99 µm (by volume) and 10.19 µm (by Surface Area). 90% of these particles have sizes below 91.88 µm (by volume). Lime's chemical compositions (by elements) determined by Energy Dispersive Spectrometry are given in Table 2. Cement (CEM I 52.5) for this investigation was obtained from 'Hanson Cement', United Kingdom. Its average Particle Size Distribution shown in Table 1 indicates particles with main equivalent diameter of 21.08µm (by volume) and 10.64µm (by Surface Area). 90% of these particles have sizes below 41.55µm (by volume). Cement's chemical compositions (by elements) are given in Table 2. Siliceous fine kiln dried sand, obtained from Fife Silica Sands (a division of Patersons of Greenoackhill Ltd, United Kingdom) constituted the aggregates. The aggregate was passed through a sieve analysis in accordance with the requirements of [46], and the particle size distributions in compliance with ASTM C 136 [47] are shown in Figure 1. The sand had Particle Size Distribution of 0-2 mm (i.e. 0.05 mm < Ø < 2 mm) which is considered suitable, in accordance with ASTM C 33 [48].

Following the required standard (BSI, 2000) [49], the mortars were prepared as the binder-aggregate (B/A) ratio was maintained at 1:3 by volume. This was chosen from the commonest dosage

Table 1: Particle Size Diameters of the tested materials

Parameters		d(v,0.5)	d(v,0.1)	Mode	d(v,0.9)	D[4,3] (main equivalent diameter by volume)	D[3,2] (main equivalent diameter by Surface Area)
Specified Particle Diameter (µm)	**Lime**	10.78	5.28	6.50	91.88	33.99	10.19
	Cement	20.14	5.40	31.78	41.55	21.08	10.64

Table 2: Chemical Composition of the materials

Material	Elemental Chemical Composition of the materials (by % weight of the dry specimen)											
	Ca	O	Si	C	Sb	Al	Fe	Mg	S	K	Na	Ti
Lime	47.6	37.6	5.2	4.4	3.3	0.7	0.5	0.5	0.2	-	-	-
Cement	75.6	9.2	4.7	-	5.7	1.1	0.3	0.4	1.8	1.2	0.1	-
Sand	0.2	53.5	43.0	-	-	1.6	0.4	-	-	1.1		0.2

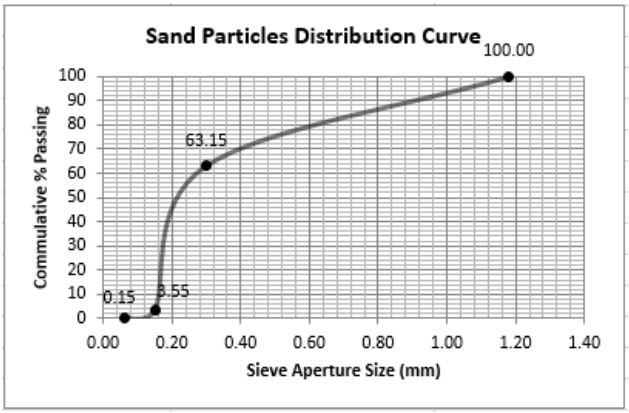


Figure 1: Grain size distribution of the aggregate

reported in the literature [15, 23, 33, 50]. To avoid measurement imprecision during batching processes, volume proportions of components were converted to weights [2]. Mortar mixtures were prepared using the correct amount of water required to obtain adopted workability of 145±5 mm (measured by the flow table test – BS EN 1015-3 (BSI, 2000) [49] as BS EN 1015-6 (BSI, 1999) [51] specifies a flow value of '140-200 mm' for 'plastic mortar'. Arising from visual and physical assessments of the mixes during the trial experimentation however, a flow value of 140 – 150mm (i.e. 145+5mm) was adopted. As observed, either higher or lower value tends towards stiffness or fluidity respectively. This was determined in accordance with (BSI, 2005) [52].

Using the stated B/A ratio (i.e. 1:3), each mortar formulation was prepared with progressively increasing/decreasing cement contents as indicated in Table 3. Mixing was done in the laboratory mixer of 30 litres maximum capacity. The mixing procedure was performed in a number of stages: Aggregates were placed first, followed by other dry materials (i.e. lime and cement, pre-mixed earlier, where applicable) and these were blended consistently for 60 seconds as best practice dictates that the NHL powder should be thoroughly mixed through the dry sand, ensuring batch colour consistency prior to gradually adding water. Water was added slowly during 30 seconds and mixing continued for another 30 seconds. Mixing was stopped for 90 seconds as mortar adhering to the wall and bottom of the mixer bowl was scraped off. Mixing then resumed to obtain consistent mixture. The entire mixing period lasted about 5minutes. For every mortar mix, minimum of three prismatic specimens of 40×40×160 mm were prepared, the average value of which represented the 'actual value' for consideration during the specimen evaluations afterwards (i.e. for microstructural analysis, and mechanical characteristics evaluations: Flexural and Compressive tests; Moduli of Elasticity determination). The specimens were compacted with a vibration table after mould filling in prismatic casts (BSI, 2010b) [53], removed from the moulds 2 days later and left to cure at the laboratory ambient conditions of 21+4ºC (temperature) and 40+5% (relative humidity), until the test dates of 1, 3, 6, 9 and 12 months.

2.2. Analytical methodology

This involved microstructural characteristics evaluation, mechanical properties assessment and some other relevant preliminary material testing. Microstructural characteristics of the mortar samples were evaluated in terms of the total porosity (in %), median pore diameter (by volume in nm), bulk density and pore size distribution, using Mercury Intrusion Porosimetry (MIP) technique. This was achieved with AutoPore IV 9500 by Micrometrics (with pressure range up to 60000 psi). The test was carried out with samples of approximately 1.5g that were extracted from the core of the crushed prisms. Under short term loading and at a relatively low rate of load application (approximately 2mm/min), the chance of micro crack propagation is minimum [33, 54]. These samples were obtained at the test ages of one and six months, and dried in an oven at temperature of 75 ± 1 ºC for 24 hours before the test, to ensure that the sample is devoid of moisture contents (which may otherwise affect its microstructural properties, thereby affecting the results). The mechanical properties were evaluated with regard to the three-point flexural tests, and compressive strength. While the flexural strength tests were performed on the ELE AutoTest 2000 apparatus with a load application pace of 50 N/s, compressive strength tests were conducted on the two fragments of each specimen (resulting from the preceding flexural test) using INSTRON 3367 with 30kN load capacity, moving at a loading rate of 2mm/min.

Table 3. Lime-Cement Mortars' Compositions by materials

Cement Contents	**0%**	**25%**	**33%**	**50%**	**66%**	**75%**
Mortar Reference I.D.	**L13**	**LC31**	**LC21**	**LC11**	**LC12**	**LC13**
Volumetric Ratio (L-C-Sd)	1-0-3	3-1-12	2-1-9	1-1-6	1-2-9	1-3-12
Lime (L): Volume (Volume in 'ml'/ mass in 'g')	1 (1700/ 1172)	3 (770/ 528)	2 (665/ 458)	1 (850/ 586)	1 (510/ 352)	1 (260/ 176)
Cement (C): Volume (Volume in 'ml'/ mass in 'g')	0 (0)	1 (415/ 442)	1 (540/ 574)	1 (830/ 883)	2 (995/ 1060)	3 (1250/ 1325)
Sand (Sd): Volume (Volume in 'ml'/ mass in 'g')	3 (4350/ 6444)	12 (4350/ 6444)	9 (4240/ 6283)	6 (4350/ 6444)	9 (3920/ 5800)	12 (4350/ 6444)

The results reported in this work were all taken as an average value of six similar specimen fragments. Additionally, INSTRON 3367 plots stress/strain graph on the screen, with the value for Modulus of Elasticity generated automatically. Other Preliminary Material Testing carried out included: application of Scanning Electron Microscopy/Energy Dispersive X-ray Spectrometry using 'Carl Zeiss EVO 50' Scanning Electron Microscope to examine and analyse the microstructure, morphology and chemical composition characterizations of the experimental materials (Lime and cement); Laser Difractometry Xmastersize with air dispersion was adopted for determination of the Pore Size Distribution of the tested materials (Lime and cement), and; Consistency of fresh mortars was examined using the flow table test in accordance with (BSI, 2000) [49].

3. Results and Discussion

3.1. Water/Binder Ratio

Table 4 shows the results of the W/B ratios obtained for lime, and lime-cement mortars batched using varying binders' contents. Water/Binder (W/B) ratio was adjusted appropriately to ensure the same mortar flow of 145±5 mm for all the mixes. For the reference mortar (lime mortar, L13), W/B ratio was recorded at 1.53.

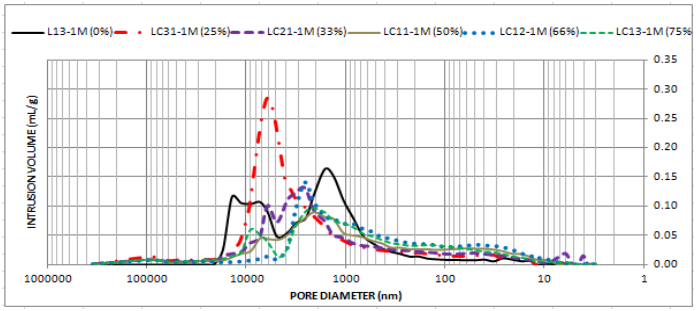
With progressive addition of cement content, a clear reduction of water demand was recorded, although non-linear. Thus, lime mortar exhibits its ability to retain water in the mix as it imparts plasticity and enhances workability [2, 4]. Besides, this reduction in water demand could be attributed to quicker formation of more hydration products of Calcium Silicate Hydrates (C–S–H) and resulting densification of the mortar matrix. It can therefore be inferred that increase in the percentage of lime with respect to cement mass requires a higher percentage of water in order to obtain a paste with the same consistency. This is because lime has smaller particle size with a higher specific surface. This leads to the mortar's finer pore system which results in higher suction and stronger water-retaining characteristics [55, 56].

Table 4: Lime-Cement Mortars' Compositions by materials and Water/Binder ratios.

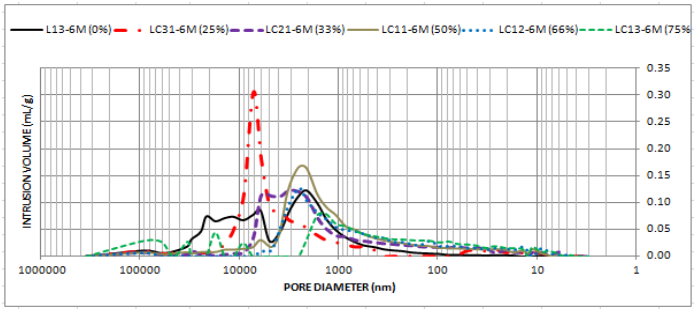
Cement Contents	**Mortar Reference I.D.**	**Water Content** (g)	**Water/ Binder Ratio**
0%	**L13**	1790	1.53
25%	**LC31**	1880	1.94
33%	**LC21**	1775	1.72
50%	**LC11**	1800	1.23
66%	**LC12**	1680	1.19
75%	**LC13**	1710	1.14

3.2. Microstructural features

Microstructural characteristics such as porosity and Pore Size Distribution (PSD) were evaluated for different mortars. Figures 2 (a) and (b), and Table 5 show the PSD curves and other microstructural data obtained by Mercury Intrusion Porosimetry (MIP) for the six different composites at 1 and 6 months of curing.



(a)



(b)

Figure 2: Pore size distribution for cured lime mortars with increasing cement contents: (a) At 1-month curing; (b) At 6-month curing

For lime only, at 1-month curing, the graph indicates a bimodal PSD with predominant sizes between (0.5-5 µm) and (5-20 µm) (Figure 2(a)). In this case, the mortar exhibits both gel pores (i.e. 1 nm – 3 µm) and substantial proportion of capillary pores (i.e. 3-30 µm). These large pores result from loss of excess unbound kneading water due to strong water-retaining characteristic of lime, and slow carbonation process [57]. The pores can also be related to fine cracks induced by associated drying shrinkage which can even expand and widen.

After 6 months, in spite of the continuous hydration and carbonation processes, the median pore diameter increased (from 2.16 to 3.96 µm) (Table 5) indicating more presence of larger pore sizes (i.e. between 0.5 µm and 40 µm) with the shift of the PSD curves to the left (Figure 2 (b)). However, a reduction is observed in the total pore volume (i.e. from 27.77 to 23.42%), which indicates filling of some of the capillary voids due to deposition of the hydration and carbonation products. The bulk density was also reduced (from 1.75 to 1.72 g/ml) due to evaporation of excess kneading water and subsequent drying shrinkage.

However, progressive addition of cement to the mortar resulted in unimodal PSD with lower median pore diameters and formation of finer pores at 6 months (Table 5). The graphs are shifted towards the right (relative to lime mortars), an indication of the dominance of smaller pore sizes.

The observed decrease in the median pore diameters is a consequence of lower W/B ratios used to achieve similar consistency. Also, it can be related to subsequent filling of larger pores by the hydration products particularly, C–S–H gel [58, 59, 60]. This can explain why the addition of cement led to predominance of smaller pore sizes in lime mortar. Usually, kneading water decreases with increasing cement content in the paste (as demonstrated in Section 3.2). This led to the densification of the mortar matrixes and consequent reduced porosities (21.45% for LC31) at 6 months (Table 5).

3.3. Mechanical Properties

The results of the mortars' mechanical assessments (flexural strength, compressive strength and modulus of elasticity) for a 12-month curing period are summarised in Appendix I. Each result was taken as an average value of three similar specimens for flexural strength and five similar specimens for both compressive strength and modulus of elasticity. The coefficients of variation (COV) fall substantially within the lower range (0 – 17%). This suggests consistent results. However, there were also few cases of higher COV, which indicate some degree of scatter in those cases. Appropriate error bars are also displayed on the respective graphs.

For lime only, Figure 3 and (Appendix I) show a flexural strength value of 0.70 MPa in the first month of curing which decreased to 0.61 MPa at the end of 12 months. This may be due to formation of different pore sizes resulting from loss of excess unbound kneading water as revealed in the microstructural analyses above.

Table 5: Extracted Mercury Intrusion Porosimetry data for composite mortars containing increasing cement contents

Specimen Reference I.D.	Curing Period	Median Pore Diameter (Volume) [nm]	Bulk Density at 0.52psia [g/mL]	Porosity [%]
L13 (0%)	1 Month	2163.6	1.75	27.77
	6 Months	3955.0	1.72	23.42
LC31 (25%)	1 Month	4276.0	1.78	29.17
	6 Months	6616.9	1.80	30.03
LC21 (33%)	1 Month	3036.1	1.78	27.54
	6 Months	2579.4	1.84	23.43
LC11 (50%)	1 Month	1487.3	1.83	24.35
	6 Months	1909.7	1.87	25.67
LC12 (66%)	1 Month	989.4	1.86	25.51
	6 Months	1340.8	1.92	20.81
LC13 (75%)	1 Month	1372.1	1.86	24.64
	6 Months	1033.7	1.90	21.45

As anticipated, progressive addition of cement to the mortar increased the flexural strength throughout the curing period, though not in a linear trend. Cement with faster hydration, quicker formation of C-S-H, reduced pore sizes and volume, attained higher flexural strength at a relatively shorter time. Notably, LC12 recorded the highest values of 4.12 MPa and 4.23 MPa at 1 month and 12 months respectively.

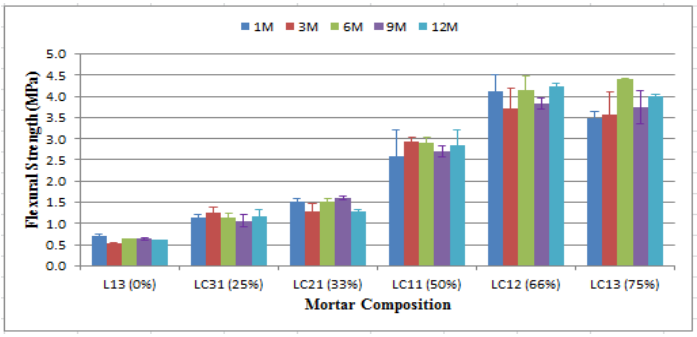


Figure 3: Flexural Strength developments for lime mortars with increasing cement contents

The compressive strength developments followed a similar trend to that of flexural strength. A marginal increase in the strength from 0.61 MPa at 1 month curing to 0.66 MPa at 12 months in the mix containing lime only was observed (Figure 4 and Appendix I).

Expectedly, these values significantly increased with progressive addition of cement. At 1 and 12 months of curing, highest values of 9.17 MPa and 12.34 MPa were recorded for the mix with triple cement content (LC13). This represents a significant increase (in each case) above the mix with no cement content (L13), over the same curing period.

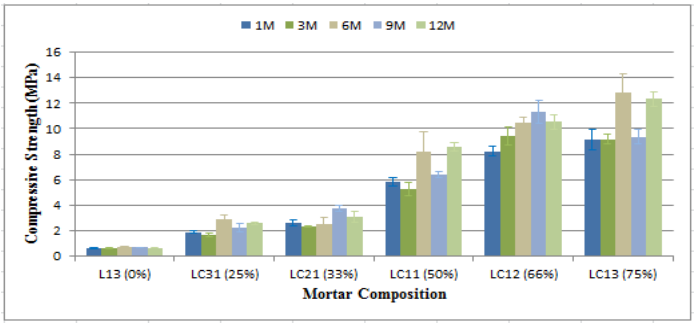


Figure 4: Compressive Strength developments for lime mortars with increasing cement contents

Despite increases recorded in both flexural and compressive strengths with increasing cement contents across the mixes, no significant increase was observed within each mix over the curing period. Being a hydraulic lime, the initial strength recorded may be related to the products of hydration reactions (i.e. formation of calcium silicate hydrates (C–S–H) and calcium aluminate hydrates (C–A–H)). Subsequently the remaining bulk calcium hydroxide would be expected to react with atmospheric carbon dioxide through carbonation over time, for further strength developments. However, rate of carbonation is found to be strongly dependent on relative humidity [61]-[63]. The laboratory curing condition of 45% relative humidity is considered possibly low and could have impacted negatively on the carbonation process thereby leading to the insignificant strength gains over time. Nonetheless, observed impact of the low humidity on the results may be applicable and relevant in some geographical regions, where the humidity is not always very high (e.g. tropical regions like Nigeria, among others).

The compressive strength recorded in each case is related to the mortars' practical performance. This is as reflected in their moduli of elasticity (E) values as shown in Appendix I and Figures 5 to 7.

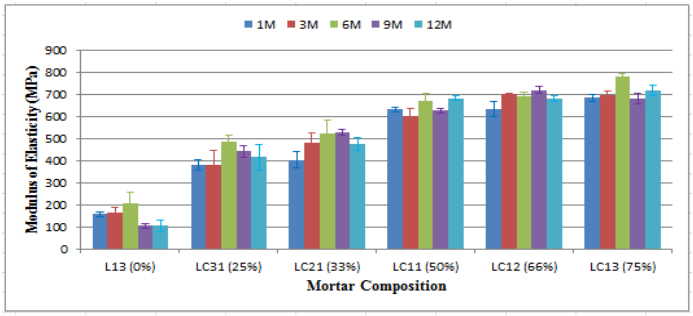


Figure 5: Moduli of elasticity (E) values for lime mortars with increasing cement contents

From Figure 5, L13 with the least compressive strength correspondingly recorded lowest 'E' value over the 12-month curing period. Also, LC13 with the highest compressive strength correspondingly recorded the highest 'E' value over the same curing period. Both LI3 and LC13 have the tendencies to display contrasting elastic behaviours (high and low) respectively. This is supported with the mortars' Compressive/Flexural strength ratios (f_c/f_f) and Moduli of Elasticity relationships with increasing cement contents as illustrated in Figure 6. It is well established that low compressive to flexural strength ratio (f_c/f_f) of similar materials is proportional to the Modulus of Elasticity (E-value) which is inversely proportional to its elastic behavior [64, 65]. Figure 6 presents the relationships between (f_c/f_f) and the E-values with cement contents for studied lime mortars. The dotted arrow on the graph indicates the general tendency of the E-values with respect to the increase in cement contents.

From Figure 6, it can be observed that both (f_c/f_f) and the E-values are proportional to cement content. However, these relationships are not linear over the 12-month curing period. Figures 7 (a) and (b) show further analyses to determine any potential pattern of relationships between the parameters

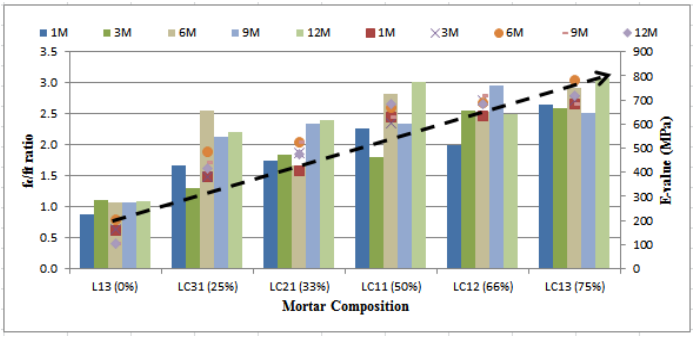
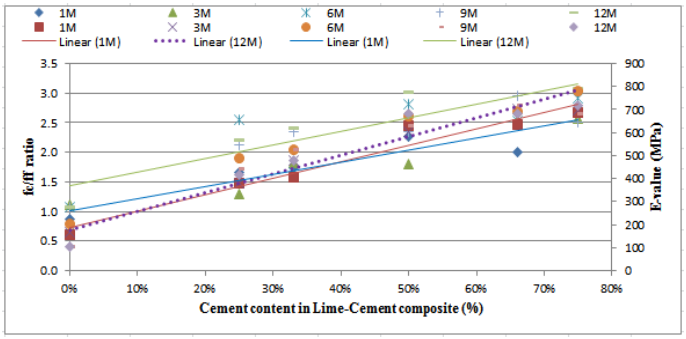


Figure 6: Compressive/Flexural strength ratios (fc/ff) and Moduli of Elasticity versus lime and cement mortars over 12-month curing (general relationships)

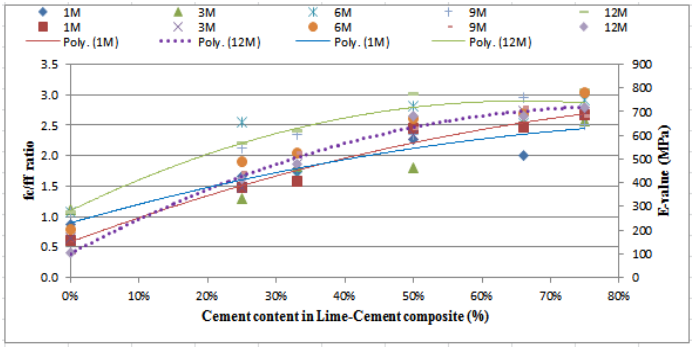
As can be observed from Figures 7 (a) and (b), good relationships could be established between (f_c/f_f) and cement

contents (with correlation coefficients above 0.80 obtained in each case). The relationship is described by the quadratic equation ($R^2 = 0.89$):

$$f_c/f_f = -4.14c^2 + 5.46c + 1.09 \quad (1)$$



f_c/f_f: Linear equation for the relationship at:
1-month ($f_c/f_f = 2.04c + 1.02$); $R^2 = 0.8715$;
12 months ($f_c/f_f = 2.28c + 1.44$); $R^2 = 0.765$
E-value: Linear equation for the relationship at:
1-month (Tc = 712.91E + 187);
12 months (Tc = 807.79E + 178.65); $R^2 = 0.9166$
(Note: Tc: Thermal Conductivity; c is cement content in %; E is E-value) (a)



f_c/f_f: Quadratic equation for the relationship at:
1-month ($f_c/f_f = -1.52c^2 + 3.21c + 0.89$); $R^2 = 0.8972$; 12 months ($f_c/f_f = -4.14c^2 + 5.46c + 1.09$); $R^2 = 0.8985$
E-value: Quadratic equation for the relationship at:
1-month ($E = -462.6c^2 + 1068.5c + 148.86$); $R^2 = 0.9674$; 12 months ($E = -964.31c^2 + 1549.1c + 99.155$); $R^2 = 0.9856$
(Note: Tc: Thermal Conductivity; c is cement content in %; E is E-value)
(b)

Figure 7: Compressive/Flexural strength ratios (f_c/f_f) and Moduli of Elasticity versus cement content for lime and cement mortars: (a) Through linear relationships; (b) Relationships using polynomial curves

However, a strong relationship could be established between the E-value and cement content. This relationship is described by the quadratic equation with a high correlation coefficient ($R^2 = 0.99$):

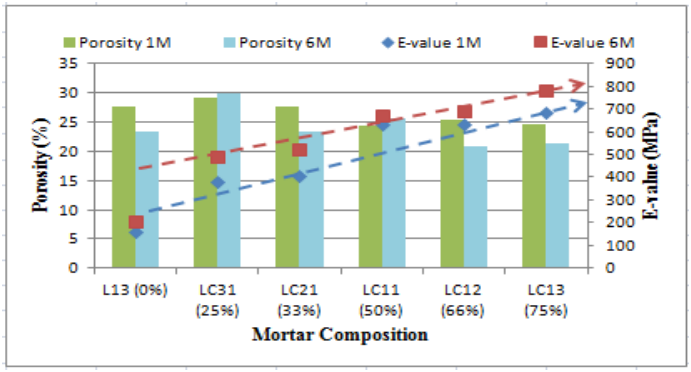
$$E = -964.31c^2 + 1549.1c + 99.155 \quad (2)$$

where E is E-value, and c is cement content in %.

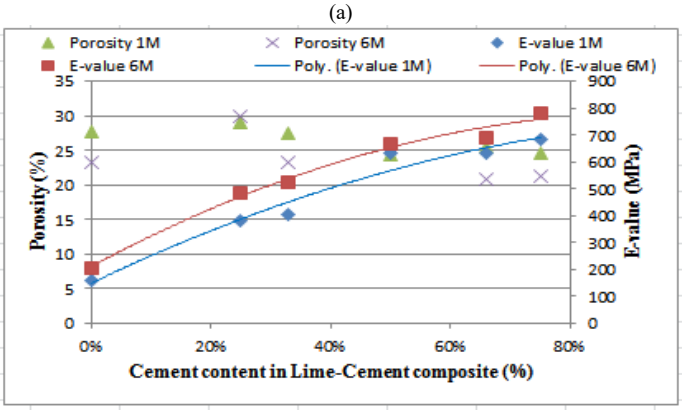
4. Further Results Synthesis: Changing Microstructural Compositions vis-à-vis Lime Composites' Flexibility

The changing microstructural characteristic features of the lime-based mortars are related to the mortar's modulus of elasticity (E-value).

Figure 8 (a) presents the effect of cement content, on both porosity and E-value. While porosity is inversely proportional to cement content, E-value is directly proportional. The dotted arrows on the graph indicate the increasing pattern of the E-values for 1 and 6 months curing.



(a)



E-value: Quadratic equation for the relationship at:
1-month (E = $-462.6c^2 + 1068.5c + 148.861$); $R^2 = 0.9674$;
6-months ($E = -626.12c^2 + 1199.6c + 210.65$); $R^2 = 0.9868$;
*(Note:*p *is porosity;c is cement content in %)*

(b)

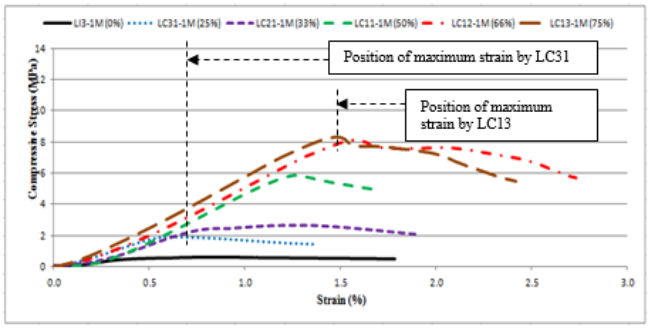
Figure 8: Relationships between increase in cement content, porosity and E-value of Lime mortar at 1 and 6-month curing: (a): General relationships; (b): Relationships using polynomial curves

An attempt was made to approximate these experimental results by establishing mathematical relationships between cement content and porosity as well as E-values, using quadratic equations (Figure 8 (b)). Considering composite performance at 6 months, a clear relationship can be established between cement content (c) and E-value (E) as expressed in the following equation:

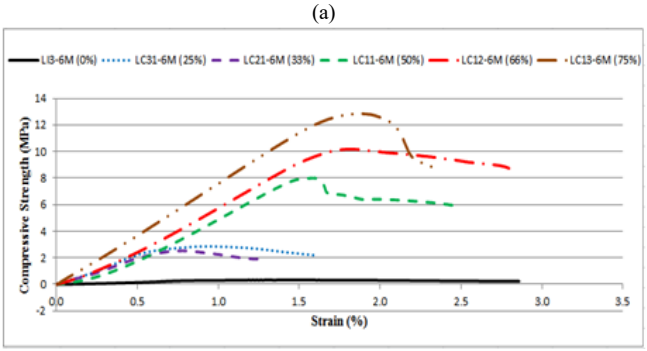
$$E = -626.12c^2 + 1199.6c + 210.65 \quad (3)$$

The coefficient of correlation in this case is very high (0.99), indicating a very strong relationship

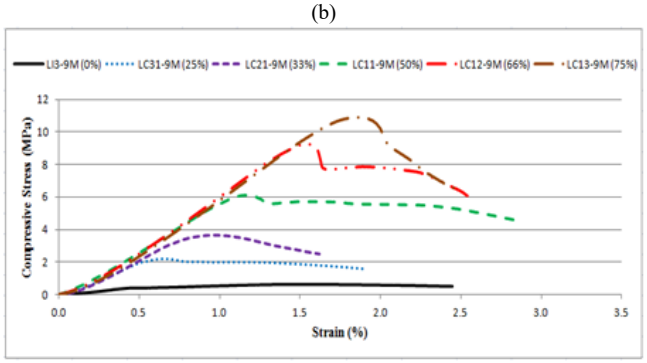
Lime mortar is primarily popular for its flexibility, the basis of its ability to accommodate masonry deformation. With the presence of lime (in varying compositions), both LI3 and LC13 (with the least lime content) have the tendencies to display contrasting elastic behaviours (high and low) respectively, as demonstrated in their stress-strain relationships at 1, 6, 9 and 12 months of curing respectively (Figures 9 (a) to (d)), with varying capacities to absorp deformation. The trend is also applicable to other mixes.



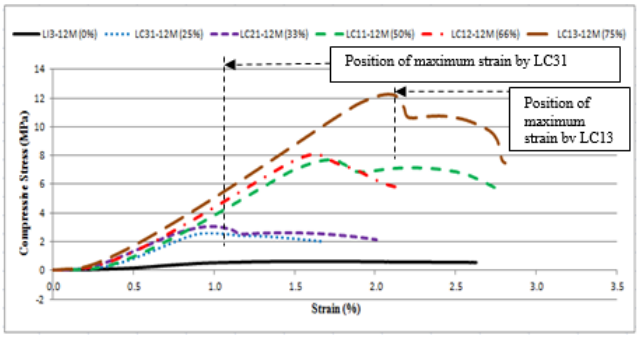
(a)



(b)



(c)



(d)

Figure 9: Compressive Strength versus Strain relationships for lime and lime-cement mortars at different curing ages: (a): At 1-month curing; (b): At 6-month curing; (c): At 9-month curing; (d): At 12-month curing

The smaller the slope of the curve, the better the elastic behaviour. All mortars exhibit comparably varying slopes relative to lime mortar's (L13). Compressive stress varies proportionally with increasing cement content as shown on the individual behavioural patterns for lime-cement composites (Figures 9 (a) to (d)). For example, the specimen with the highest cement content (75% cement content i.e. LC13) sustained about 8 MPa and 12 MPa maximum compressive stresses after 1 and 12 months respectively. This differs significantly from the maximum compressive stresses of 2 MPa and 3 MPa sustained by the composite with the least cement content (25% cement content i.e. LC31). However, increasing cement content is associated with higher slope of the curve, having attendant negative consequences on the strain, and the ability of the sample to sustain deformation. This results in decreasing flexibility.

Thus, LC13 curve with the highest gradient is characterised with the least flexibility, and thus, the least capability to sustain deformation before failure. Nevertheless, cement content at about 33% of the binders produces composite with comparable flexibility relative to lime mortar. This is evident in LC31 and LC21. In general, cement addition up to 33% of the binder compositions would enable the mortars under compression to exhibit some deformation tendencies. The smaller the slope of the curve, the better the elastic behaviour. All mortars exhibit comparably varying slopes relative to lime mortar's (L13). Compressive stress varies proportionally with increasing cement content as shown on the individual behavioural patterns for lime-cement composites (Figures 9 (a) to (d)). For example, the specimen with the highest cement content (75% cement content i.e. LC13) sustained about 8 MPa and 12 MPa maximum compressive stresses after 1 and 12 months respectively. This differs significantly from the maximum compressive stresses of 2 MPa and 3 MPa sustained by the composite with the least cement content (25% cement content i.e. LC31). However, increasing cement content is associated with higher slope of the curve, having attendant negative consequences on the strain, and the ability of the sample to sustain deformation. This results in decreasing flexibility. Thus LC13 curve with the highest gradient is characterised with the least flexibility, and thus, the least capability to sustain deformation before failure. Nevertheless, cement content at about 33% of the binders produces composite with comparable flexibility relative to lime mortar. This is evident in LC31 and LC21. In general, cement addition up to 33% of the binder compositions would enable the mortars under compression to exhibit some deformation tendencies. However, mortars with higher cement compositions (i.e. above 33%) would strain linearly until failure occurs, and failure appears suddenly with minimal deformation.

5. Conclusion

This study has presented empirical data to reveal synergies obtainable by taking the advantage of respective useful features of each of lime and cement, combined as mortar constituents. The study sought to evolve low carbon composite construction materials in forms of 'lime-cement' mortars with cement serving as a partial replacement for lime in each case. Through this process, potential benefits derivable from blending lime and cement are revealed, particularly, maximizing 'flexibility' feature of lime mortar. This is a primary lime property essentially responsible for its age-long durability. However, due to other prominent drawbacks associated with lime, its overall usage has gone to decline. Thus, it is clear from this study that the patterns exhibited by all the composites investigated suggest that they are relatively characterized by elastic behaviours comparable to the basic lime-based mortar. In particular, cement addition up to 33%

of the binder compositions would enable the mortars under compression exhibit some deformation tendencies. However, mortars with higher cement compositions (i.e. above 33%) would strain linearly until failure occurs suddenly, with minimal deformation. This study therefore is an attempt to revive sustainable lime mortar as a direct response to the climate change challenge in view of the material's sustainability characteristics. Conclusively, this effort would facilitate protection of the environment and conservation of energy resources along with the advent of a sizable market, for renewed interest in the use of lime as a building material.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgment

The authors gratefully acknowledge the financial supports from 'TETFUND AST&D' Scheme, an educational development organ of The Federal Government of Nigeria. Thanks, are also due to Hanson Cement, United Kingdom, for the provision of the cement used for this work.

References

[1] D. Zhang, J. Zhao, D. Wang, C. Xu, M. Zhai, X. Ma, "Comparative study on the properties of three hydraulic lime mortar systems: Natural hydraulic lime mortar, cement-aerial lime-based mortar and slag-aerial lime-based mortar" Constr. Build. Mater, **186**, 42–52, 2018 https://doi.org/10.1016/j.conbuildmat.2018.07.053

[2] R. Hanley, S. Pavía, "A study of the workability of natural hydraulic lime mortars and its influence on strength", Mater. Struct, **41**(2), 373-381, 2008.

[3] S. Pavia, O. Brennan, "Portland cement-lime mortars for conservation", 3rd Historic Mortars Conference (HMC13) held at the University of the West of Scotland Glasgow, Scotland, 2018.

[4] E.A.W. Hendry, "Masonry walls: material and construction", Constr. Build. Mater, **15**, 323–33, 2001. https://doi.org/10.1016/S0950-0618(01)00019-8

[5] R. J. Ball, A. El-Turki, W.J. Allen, J.A. Nicholson, G.C. Allen, "Deformation of NHL3.5 and CL90/PC hybrid mortars", Proceedings of the Institution of Civil Engineers – Constr. Mater., **162**(1), 29-35, 2009. Doi: 10.1680/coma.2009.162.1.29

[6] P. Faria, F. Henriques, V. Rato, "Comparative evaluation of lime mortars for architectural conservation", J Cult Herit, **9**(3), 338-346, 2008.

[7] A.M. Forster, "An assessment of the relationship between the water vapour permeability and hydraulicity of lime based mortars with particular reference to building conservation materials science", PhD Thesis, Heriot-Watt University, Edinburgh, 2002.

[8] G. Cultrone, E. Sebastián, M.O. Huertas, "Forced and natural carbonation of lime-based mortars with and without additives: Mineralogical and textural changes". Cement and Concrete Res., **35**(12), 2278-2289, 2005.

[9] A. Isebaert, W. De Boever, V. Cnudde, L. Van Parys, "An empirical method for the estimation of permeability in natural hydraulic lime mortars" Mater. Struct., **49**(11), 4853-4865, 2016. https://doi.org/10.1016/S0950-0618(01)00019-8

[10] A.J. Klemm, D.E. Wiggins "Lime mortar and sacrificial protection of heritage stonework" Proceedings of the Institution of Civil Engineers - Engineering History and Heritage, **168**(4), 167-176, 2015.

[11] A.J. Edwards "Properties of Hydraulic and Non-hydraulic Limes for Use in Construction" PhD Thesis, Napier University, Edinburgh, 2005.

[12] K. AL-Bashaireh "Lime Mortar and Plaster from The House XVII-XVIII, Umm El-Jimal, Jordan: Archaeometric Analysis. Shedet, Annual Peer-Reviewed Journal issued by The Faculty of Archaeology, Fayoum University, **6**, 212–224, 2019. https://doi.org/10.36816/shedet.006.12

[13] S.A. Olaniyan, A.J. Klemm, F.C. Almeida, "Evolving Low Carbon Sustainable Building Material: Making Case for Cement-Lime Composites" 9th International Concrete Conference on Environment, Efficiency and Economic Challenges for Concrete, Dundee, 2016.

[14] S.A. Olaniyan, "Sustainable Lime Based Mortars: Performance Assessment of Composites for Building Construction" PhD Thesis, Glasgow Caledonian University, Glasgow, Scotland (UK), 2017.

[15] L. Ventola, M. Vendrell, P. Giraldez, "Newly-designed traditional lime mortar with a phase change material as an additive" Constr. Build. Mater, **47**, 1210-1216, 2013.

[16] F. Pacheco-Torgal, J. Faria, S. Jalali, "Some considerations about the use of lime–cement mortars for building conservation purposes in portugal: A reprehensible option or a lesser evil?" Constr. Build. Mater, **30**(1), 488-494, 2012.

[17] A. Izaguirre, J. Lanas, J.I Alvarez, "Effect of a polypropylene fibre on the behaviour of aerial lime-based mortars" Constr. Build. Mater, **25**(2), 992-1000, 2011. https://doi.org/10.36816/shedet.006.12

[18] A. Everett, Materials (4th ed.). London, Harlow: Longman, 1986.

[19] A.M. Neville, Properties of Concrete (4th ed.). Harlow: Longman Scientific & Technical, 1995.

[20] L. Mcdonald, "Hydraulic lime mortar for the house of the future" Structure Engineer, **78** (7), 2000.

[21] K. Callebaut, J. Elsen, K. Van Balen, W. Viaene, "Nineteenth century hydraulic restoration mortars in the saint Michael's church (leuven, belgium)" Cement Concrete Res., **31**(3), 397-403, 2001.

[22] K. Elert, C. Rodriguez-Navarro, E.S. Pardo, E. Hansen, O. Cazalla, "Lime mortars for the conservation of historic buildings" Stud. Conserv., **47**, 62-75, 2002.

[23] A. Moropoulou, A.S. Cakmak, G. Biscontin, A. Bakolas, E. Zendri, "Advanced byzantine cement based composites resisting earthquake stresses: The crushed brick/lime mortars of justinian's hagia Sophia" Constr. Build. Mater, **16**(8), 543-552, 2002.

[24] M. P. Seabra, H. Paiva, J.A. Labrincha, V.M. Ferreira, "Admixtures effect on fresh state properties of aerial lime based mortars" Constr. Build. Mater, **23**(2), 1147-1153, 2009. doi: 10.36816/006.12

[25] A.M. Forster, K. Carter, "A framework for specifying natural hydraulic lime mortars for masonry construction" Structural Survey, **29**(5), 373-396, 2011.

[26] C. Ince, Y. Ozturk, M. Carter, M. Wilson, "The influence of supplementary cementing materials on water retaining characteristics of hydrated lime and cement mortars in masonry construction" Mater Struct., **47**(3), 493–501, 2013. Doi: 10.1617/s11527-013-0074-9

[27] M. Tate, "The most important property of cement-lime mortar in masonry construction" International Building Lime Symposium Orlando, Florida, March 9 -11, 2005.

[28] J. Hughes, J.E. Lindqvist, CBI Betonginstitutet AB, SP – Sveriges Tekniska Forskningsinstitut, and RISE "RILEM TC 203-RHM: Repair mortars for historic masonry: The role of mortar in masonry: An introduction to requirements for the design of repair mortars" Mater Struct, **45**(9), 1287-1294, 2012.

[29] S.A. Olaniyan, A.J. Klemm, "Current Trends in Development of Lime Based Composites" Building Physics in Theory and Practice, **7**(3), 49-54, 2015.

[30] F. Pacheco-Torgal, J. Faria, S. Jalali, "Some considerations about the use of lime–cement mortars for building conservation purposes in portugal: A reprehensible option or a lesser evil?" Construction and Building Materials, **30**(1), 488-494, 2012.

[31] C. Torney, A.M. Forster, E.M. Szadurski, "Specialist 'restoration mortars' for stone elements: A comparison of the physical properties of two stone repair materials" Heritage Science, **2**(1), 1-12, 2014.

[32] C. Torney, A.M. Forster, "Concerned with compatibility" Building Surveying Journal, **28**, 2015.

[33] M.J. Mosquera, B. Silva, B. Prieto, E. Ruiz-Herrera, "Addition of cement to lime-based mortars: effect on pore structure and vapor transport" Cement Concrete Res. **36**, 1635– 1642, 2006.

[34] M. Arandigoyen, J.I. Alvarez, "Blended pastes of cement and lime: pore structure and capillary porosity" Applied Surface Science, **252**(23), 8077-8085, 2006.

[35] P.F.G. Banfill, "The rheology of fresh mortar" Magazine of Concrete Research, **154**(43), 13–21, 1991.

[36] P.F.G. Banfill, "Rheological methods for assessing the flow properties of mortar and related materials" Constr. Build. Mater, **8**(1), 43-50, 1994.

[37] M. Arandigoyen, J. I. Alvarez, "Pore structure and mechanical properties of cement–lime mortars" Cement and Concrete Res., **37**(5), 767-775, 2007.

[38] V. Starinieri, D.C. Hughes, D. Wilk, "Influence of the combination of Roman cement and lime as the binder phase in render mortars for restoration" Constr. Build. Mater, **44**, 192–199, 2013.

[39] T.H. Park, C.K. Park, M.H. Noh, "Rheological properties of cementitious materials containing mineral admixtures" Cement and Concrete Res., **35**(5), 842-849, 2005. https://doi.org/10.1016/j.cemconres.2004.11.002

[40] R. Artis "Admixtures and Additives for Lime Mortars" An online publication of 'The Building Conservation Directory', 2014. http://www.buildingconservation.com/articles/lime-mortar-additives/lime-mortar-additives.htm, retrieved in September, 2016.
[41] J. Stark, "Recent advances in the field of cement hydration and microstructure analysis" Cement and Concrete Res., **41**(7), 666–678, 2011.
[42] Y. Sébaïbi, R.M. Dheilly, B. Beaudoin, B., M. Quéneudec, "The effect of various slaked limes on the microstructure of a lime–cement–sand mortar" Cement and Concrete Res., **36**(5), 971-978, 2006. https://doi.org/10.1016/j.cemconres.2011.03.028
[43] M. Arandigoyen, J.I. Alvarez, "Pore structure and mechanical properties of cement–lime mortars" Cement and Concrete Res., **37**(5), 767-775, 2007.
[44] British Standards Institution (BSI) BS EN 998: Masonry Mortar – Part 2: Specification for Mortar for Masonry, 2010a.
[45] A. Moropoulou, "Reverse engineering: a proper methodology for compatible restoration mortars" RILEM Conference Proceedings on Historic Mortars, Delft, 2005.
[46] British Standards Institution (BSI) BS EN 13139: Aggregates for mortar - Part 3 (PD 6682): Guidance on the use of BS EN 13139, 2013.
[47] ASTM C136 "Standard Test Method for particle size distributions", 2014
[48] ASTM C 33 "Standard Specification for mortar Aggregates", 2016
[49] British Standards Institution (BSI) BS EN 1015: Methods of Test for Mortar for Masonry – Part 3: Determination of Consistence of Fresh Mortar (by Flow Table), 2000.
[50] J. Lanas, J.I. Alvarez-Galindo, "Masonry repair lime-based mortars: Factors affecting the mechanical behaviour" Cement and Concrete Res., **33**(11), 1867-1876, 2003. https://doi.org/10.1016/S0008-8846(03)00210-2
[51] British Standards Institution (BSI) BS EN 1015: Methods of Test for Mortar for Masonry – Part 6: Determination of Bulk Density of Fresh Mortar, 1999.
[52] British Standards Institution (BSI) PD 6678: Published Document - Guide to the Specification of the Masonry Mortar, 2005.
[53] British Standards Institution (BSI) BS EN 459: Building Lime – Part 2: Test Method, 2010b.
[54] J.S. Pozo-Antonio, "Evolution of mechanical properties and drying shrinkage in lime-based and lime cement-based mortars with pure limestone aggregate" Constr. Build. Mater, **77**, 472-478, 2015.
[55] K.M. Green, M.A. Carter, M. A.D. Hoff, M.A. Wilson, "The effects of lime and admixtures on the water-retaining properties of cement mortars" Cement Concrete Res., **29**(11), 1743-1747, 1999.
[56] Y. Sébaïbi, R.M. Dheilly, M. Quéneudec, "A study of the viscosity of lime–cement paste: Influence of the physico-chemical characteristics of lime" Constr. Build. Mater, **18**(9), 653-660, 2004.
[57] K. Aligizaki, Pore Structure of Cement-Based Materials: Testing, Interpretation and Requirements. Taylor and Francis, 2006.
[58] J. Grilo, P. Faria, R. Veiga, A.S.Silva, V. Silva, A. Velosa, "New natural hydraulic lime mortars – physical and microstructural properties in different curing conditions" Constr. Build. Mater, **54**, 378-384, 2014. https://doi.org/10.1016/S0008-8846(03)00210-2
[59] G. Land, D. Stephan, D. "Controlling cement hydration with nanoparticles" Cement and Concrete Comp., **57**, 64-67, 2015.
[60] Y.C. Choi, Y. K. Cho, K. Shin, K., S. Kwon, "Development and application of microcapsule for cement hydration control" KSCE Journal of Civil Engineering, **20**(1), 282-292, 2015.
[61] A. El-Turki, M.A. Carter, M.A. Wilson, R.J. Ball, G.C. Allen, "A microbalance study of the effects of hydraulicity and sand grain size on carbonation of lime and cement" Constr. Build. Mater, **23**(3), 1423-1428, 2009. https://doi.org/10.1016/j.cemconres.2009.03.028
[62] A. El-Turki, R.J. Ball, S. Holmes, G.C. Allen, W.J. Allen, "Environmental cycling and laboratory testing to evaluate the significance of moisture control for lime mortars" Constr. Build. Mater, **24**(8), 1392-1397, 2010. https://doi.org/10.1016/ 11.002
[63] V. Pavlik, M. Uzakova, M. "Effect of curing conditions on the properties of lime, lime–metakaolin and lime–zeolite mortars" Constr. Build. Mater, **103**, 14-25, 2016
[64] A. Moropoulou, A. Bakolas, P. Moundoulas, E. Aggelakopoulou, S. Anagnostopoulou, "Strength development and lime reaction in mortars for repairing historic masonries" Cement and Concrete Comp., **27**(2), pp 289-294, 2005.
[65] B.S. Briccoli, L. Rovero, "Gli additivi per malte usati nell' antichita, Ed. Mascolo G, Materiali e tecniche per il Restauro Cassino" Idea Stampa Editore AIMAT, INISM, 1997

Appendix I. Lime and Lime-Cement Mortars' Mechanical Strength Results over 12 month curing periods

Mortar Reference I.D.	**Volumetric Ratio** (L-C-Sd)	**Mechanical Parameters** (MPa)	**Curing Period (**month(s))				
			1	**3**	**6**	**9**	**12**
L13 (0%)	1-0-3	Flexural (COV(%))	0.70 (6)	0.54 (4)	0.63 (0)	0.63 (6)	0.61 (0)
		Compressive (COV(%))	0.63 (6)	0.59 (5)	0.68 (15)	0.68 (4)	0.66 (6)
		Modulus of Elasticity (COV(%))	181.47 (5)	136.73 (16)	175.46 (30)	107.71 (11)	93.08 (28)
LC31 (25%)	3-1-12	Flexural (COV(%))	1.13 (6)	1.27 (10)	1.13 (9)	1.06 (14)	1.16 (13)
		Compressive (COV(%))	1.88 (5)	1.65 (25)	2.86 (14)	2.26 (14)	2.57 (5)
		Modulus of Elasticity (COV(%))	382.75 (6)	382.53 (18)	487.68 (6)	443.32 (6)	416.10 (14)
LC21 (33%)	2-1-9	Flexural (COV(%))	1.52 (5)	1.27 (15)	1.52 (5)	1.60 (3)	1.29 (4)
		Compressive (COV(%))	2.64 (9)	2.33 (16)	2.53 (12)	3.76 (6)	3.10 (14)
		Modulus of Elasticity (COV(%))	405.02 (9)	481.46 (9)	524.15 (12)	528.21 (3)	476.86 (6)
LC11 (50%)	1-1-6	Flexural (COV(%))	2.57 (24)	2.92 (3)	2.91 (4)	2.71 (5)	2.84 (12)
		Compressive (COV(%))	5.83 (6)	5.27 (11)	8.21 (17)	6.36 (4)	8.58 (4)
		Modulus of Elasticity (COV(%))	631.89 (2)	604.78 (6)	671.51 (5)	627.62 (2)	683.58 (1)
LC12 (66%)	1-2-9	Flexural (COV(%))	4.12 (9)	3.71 (13)	4.14 (8)	3.83 (3)	4.23 (2)
		Compressive (COV(%))	8.24 (4)	9.45 (7)	10.45 (4)	11.34 (8)	10.52 (5)
		Modulus of Elasticity (COV(%))	633.61 (5)	701.02 (1)	691.50 (3)	719.96 (2)	681.81 (2)
LC13 (75%)	1-3-12	Flexural (COV(%))	3.47 (5)	3.56 (15)	4.40 (0)	3.74 (11)	4.01 (1)
		Compressive (COV(%))	9.17 (9)	9.19 (4)	12.85 (11)	9.38 (6)	12.34 (5)
		Modulus of Elasticity (COV(%))	685.49 (2)	698.41 (3)	780.83 (2)	683.30 (3)	717.74 (3)

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1499-1510 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Comparative Analysis Spline Methods in Digital Processing of Signals

Hakimjon Zaynidinov[1], Sayfiddin Bakhromov[2], Bunyod Azimov[*,3], Sarvar Makhmudjanov[1]

[1]*Department of Information Technologies, Tashkent University of Information Technologies, Tashkent, 100200, Uzbekistan*

[2]*Department of Computational Mathematics and Information Systems, National University of Uzbekistan, Tashkent, 100174, Uzbekistan*

[3]*Department of Information Technology, Arizona State University, Andijan, 170100, Uzbekistan*

ARTICLE INFO

Article history:
Received: 14 August, 2020
Accepted: 01 December, 2020
Online: 21 December, 2020

Keywords:
Cubic spline
Spline functions defect
Clasic interpolation models
Ryabenky cubic spline
Grebbennikov cubic spline
Local interpolation cubic spline

ABSTRACT

In this article, the Grebennikov cubic spline, the Ryabenky cubic spline, and we offer the local interpolation cubic spline models are selected. The construction details of the models were given and the processes of approximating the functions were performed using selected local cubic splines. We can also say that the local interpolation cubic spline models considered in this study provide high accuracy in digital processing of signals, which helps experts to make the right decision as a result of digital processing of signals. For instance, the initial values of the gastroentrological signal were calculated after that they digitally processed, and the error results were taken. Approximation of the considered models and digital processing of the gastroenterological signal by getting error results were specific analyzed comparatively based on numerical and graphical comparisons.

1. Introduction

Digital signal processing is one of the most pressing areas of science and technology today, which is developing from a practical point of view. The correct choice of mathematical apparatus in the digital processing of signals leads to a high degree of accuracy in making the right decisions by experts in the field. As an example, we can say that in medicine and geophysics accuracy is very important [1]. In medicine, accuracy plays an important role in such processes as determining the current condition of a patient brought to the hospital in a critical condition, determining the type of disease, clarify the degree of the disease. We know that in the field of geophysics, too, a large amount of money is spent on the extraction of mineral resources, so accuracy is important in confirm the location of minerals [2].

That is, according to the anomalous changes in the geophysical signal, it is possible to make predictions for the analysis the subsequent developmental activity of the object under study. Usually, as a result of forecasting, it is possible to predict such information as the location of the most accumulated minerals, the amount of their reserves in advance. As information can be used anomalous changes in the electromagnetic, gravitational fields of the earth, anomalous changes in the ionosphere, seismic noise, various acoustic vibrations.

Today, scientists are exploring new ways to process these signals using a variety of mathematical models and algorithms [3]. One of these methods is to recover the signals received in tabular form using local interpolation spline functions [4]. This is because the accuracy level of spline functions is higher than that of classical interpolation polynomials. And the algorithms derived from them require less computation [5]. Existing classical interpolation models, their application in signal recovery and digital processing algorithms are performed depending on the node points [6], [7]. The construction of classical interpolation polynomials is based on the idea of replacing functions with that are closer to it, in a sense, and simpler in structure [8], [9]. In this case, a large number of node points are required to obtain a high degree of convergence of these models [10], [11]. In the case of node points, the order of the system equations formed in the construction of the model under consideration increases, and the process of solving this system of equations becomes more complicated, which does not ensure a high level of accuracy of the model.

In order to solve this problem, in the article, the process of digital processing of signals was carried out using high-precision

*Corresponding Author: Bunyod Azimov, Email: bunyodbekazimov@mail.ru

spline models. The increase in the degree of convergence of the models under consideration does not depend on the increase in the order of the system of equations [12], [13].

That is, in the construction of the spline function considered in this work, the following node points are constructed on the basis of $x_{i-1}, x_i, x_{i+1}, x_{i+2}$ in the $[x_i, x_{i+1}]\ (i=\overline{0,n})$ interval S_i local interpolated cubic spline.

That is, the spline function divides the [*a, b*] interval into n parts and builds a spline function in a single interval (Figure 1).

To construct a spline function, the interval [a, b] is divided into n parts, and the spline function is constructed in a single interval [14].

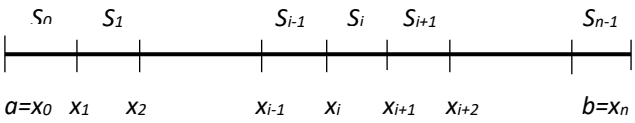


Figure 1: A view of the spaces where the cubic spline is built.

$$S_i(x_i) = S_{i+1}(x_i) \tag{1}$$

$$S_i^{(')}(x_i) = S_{i+1}^{(')}(x_i) \tag{2}$$

$$S_i^{(")}(x_i) = S_{i+1}^{(")}(x_i) \tag{3}$$

If conditions (1), (2) and (3) are met, the spline function defect is called a spline function that fully meets the actual level of demand, which is equal to 1.

The following is a study of cubic spline functions with high accuracy in digital processing of signals [15]-17].

2. Build cubic spline functions

In digital signal processing, cubic spline functions are a high-precision mathematical apparatus.

We know that the $\{x\}_{i=0}^{N}$ corresponding to the $f(x)$ function, the $S(x)$ function passing through the node points, is called the interpolation cubic spline and meets the following conditions [1],[9],[10]:

1) $S(x)$ function third-degree polynomial in each $[x_i, x_{i+1}]\ (i=\overline{0,n})$ interval;

2) The first and second-order derivatives of the $S(x)$ function must be continuous in the interval [a, b];

3) $S(x_i) = f(x_i), i=\overline{0,n}$.

The last condition is called the interpolation condition, and the function that satisfies the three conditions is called the interpolation cubic spline.

2.1. Build a cubic spline function based on basic functions

Let us be given the following function.

$$G(x,t) = (x-t)^3 = \begin{cases} (x-t)^3, & x \ge t \\ 0, & x < t \end{cases} \tag{4}$$

Assume that the $\{x_i\}$ node points on the OX axis are defined in steps h as follows [12].

$x_{i+1} = x_i + h,\ i = 1,2,...,M$, $G(x,t)$ we enter the fourth-order divisor of the function separately for the $x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}$ node points on the variable i:

$$\phi_i(x) = G(x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}),\ i = 3,4,...,M-2$$

The fourth-order difference of the $\varphi(x)$ function is determined by the following formula:

$$f(x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}) =$$

$$= \frac{f(x_{i-1}, x_i, x_{i+1}, x_{i+2}) - f(x_{i-2}, x_{i-1}, x_i, x_{i+1})}{x_{i-2} - x_{i+2}}$$

In turn, $x_{i-1}, x_i, x_{i+1}, x_{i+2}$ and $x_{i-2}, x_{i-1}, x_i, x_{i+1}$ are also calculated sequentially as above.

After some simplification, the calculation formula for the fourth-order subtraction difference of the $G(x,t)$ function will look like this:

$$\varphi_i(x) = G(x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}) =$$
$$\frac{1}{4!h^4}[(x-x_{i+2})_+^3 - 4(x-x_{i+1})_+^3 \tag{5}$$
$$+ 6(x-x_i)_+^3 - 4(x-x_{i-1})_+^3 + (x-x_{i-2})_+^3]$$

Based on the features discussed above, the following forms the basis in the space of tertiary splines

$$S_i(x) = \frac{\varphi_i(x)}{\varphi_i(x_i)},\quad i = 3,4,5,...,M-2 \tag{6}$$

Let's look at the features.

This function forms the basis in the space of cubic splines and has the following features:

1) Smoothness

$$S_i(x) \in C^2[x_1, x_M] \tag{6^1}$$

2) Locality

$$S_i(x) > 0, \quad x \in (x_{i-2}, x_{i+2})$$
$$S_i(x) \equiv 0, \quad x \notin (x_{i-2}, x_{i+2}), \qquad (6^2)$$
$$i = 3,4,5..., \quad M-2$$

The values of the $f(x)$ function at the $f_i = f(x_i),\ i = 3, M-2$ node points are given $x_4 = a, x_{M-3} = b$.

Consider the following function.

$$S(x) = \sum_{i=3}^{M-2} f_i S_i(x), \quad x \in (x_4, x_{M-3})$$

From this function

$$S_i(x) > 0, \quad x \in (x_{i-2}, x_{i+2})$$

$$S_i(x) \equiv 0, \quad x \notin (x_{i-2}, x_{i+2})$$

the local condition is required.

In that case, the values at the node point of the $S_i(x)$ function based on the localization condition are as follows.

$$S(x_i) = \sum_{j=i-1}^{i+1} f_j S_j(x_i), \quad i = 4, M-3$$

For simplicity

$$S(x_i) = \sum_{p=-1}^{1} f_{i+p} S_{i+p}(x_i)$$

That is

$$S(x_i) = f_{i-1} S_{i-1}(x_i) + f_i S_i(x_i) + f_{i+1} S_{i+1}(x_i),$$
$$i = \overline{4, M-3}$$

The $S_i(x)$ function at values p= 0.1 is as follows

$$S_{i+p}(x_i) = S_{i-p}(x_i) = S_i(x_{i+p}) = S_i(x_{i-p}) = a_p$$

has properties

$$a_p = \begin{cases} 1 & if \quad p = 0 \\ 0,25 & if \quad p = 1 \end{cases}$$

That is, at p = 0

$$S_i(x_i) = S_i(x_i) = S_i(x_i) = S_i(x_i) = 1$$

really

$$S_i(x) = \frac{\varphi_i(x)}{\varphi_i(x_i)},$$

in this

$$S_i(x_i) = \frac{\varphi_i(x_i)}{\varphi_i(x_i)} = 1,$$

The case p = 1 is similar.
As a result

$$S(x_i) = 0,25 f_{i-1} + 1 f_i + 0,25 f_{i+1} = f + 0,25(f_{i-1} + f_{i+1}),$$
$$i = \overline{4, M-3}$$

We now express the f_{i-1}, f_{i+1} through the f_i, for which we spread them to the Taylor series

$$f_{i-1} = f_i + h f_i' + \frac{h^2}{2} f_i''$$
$$f_{i+1} = f_i - h f_i' + \frac{h^2}{2} f_i''$$

By adding these expressions we get the following

$$f_{i-1} + f_{i+1} = 2 f_i + \frac{h^2}{2}(f_i'' + f_i'')$$

$$f_{i-1} + f_{i+1} = 2 f_i + O(h^2)$$

$$0,25(f_{i-1} + f_{i+1}) = 0,5 f_i + O(h^2)$$

In that case

$$S(x_i) = f_i + \frac{1}{4}(f_{i-1} + f_{i+1}) = f_i + \frac{1}{2} f_i + O(h^2),$$

From this

$$S(x_i) = \frac{3}{2} f_i + O(h^2)$$

We will have $S(x_i) = \frac{3}{2} f_i + O(h^2)$.

The smoother the $f(x)$ function, the closer it is to f_i in $(f_{i-1} + f_{i+1})/2$ calculations.

And the $S_i(x)$ spline function is close to the $k\frac{3}{2} f(x)$ at the node points then we find the next approximate function

$$S^*(x_i) = \frac{2}{3} S(x_i) = \frac{2}{3}[\frac{3}{2} f_i + O(h^2)]$$

$$S^*(x_i) = f_i + O(h^2)$$

in which case it can be obtained

$$S^*(x) = \frac{2}{3} \sum_{i=3}^{M-2} f_i S_i(x)$$

$$S^*(x) = \frac{2}{3} S_{i-1}(x) f_{i-1} + \frac{2}{3} S_i(x) f_i +$$

$$+\frac{2}{3} S_{i+1}(x) f_{i+1} + \frac{2}{3} S_{i+2}(x) f_{i+2}$$

We enter the following notation, where $t = \frac{x - x_i}{h}, x = x_i + th$,

$$\psi_1(t) = \frac{2}{3} S_{i-1}(x); \quad \psi_2(t) = \frac{2}{3} S_i(x);$$
$$\psi_3(t) = \frac{2}{3} S_{i+1}(x); \quad \psi_4(t) = \frac{2}{3} S_{i+2}(x)$$

In that case

$$S^*(x) = \psi_1(t) f_{i-1} + \psi_2(t) f_i +$$

$$+\psi_3(t) f_{i+1} + \psi_4(t) f_{i+2}$$

It is not difficult to calculate $\psi_i(t)$.
We will see one of them count

$$\psi_1(t) = \frac{2}{3} S_{i-1}(x) = \frac{2}{3}\frac{1}{4}(1-t)^3 = \frac{1}{6}(1-t)^3$$

the rest are calculated similarly.

$$\psi_2(t) = \frac{1}{6}(3t^3 - 6t^2 + 4)$$

$$\psi_3(t) = \frac{1}{6}(1 + 3t + 3t^2 + 3t^3)$$

$$\psi_4(t) = \frac{1}{6} t^3$$

As a result, we write the general view of a tertiary spline function that is independent of node points as follows

$$S^*(x) = \sum_{j=1}^{4} \psi_j(t) f_{i-2+j}$$

Here

$$\begin{cases} \psi_1(t) = \frac{1}{6}(1-t)^3 \\ \psi_2(t) = \frac{1}{6}\left(3t^3 - 6t^2 + 4\right) \\ \psi_3(t) = \frac{1}{6}\left(1 + 3t + 3t^2 + 3t^3\right) \\ \psi_4(t) = \frac{1}{6} t^3 \end{cases}$$

2.2. Local interpolation cubic spline function based on Ryabenky operator

If the function is smooth enough, then it is advisable to approximate this function with spline functions. The use of third degree Given the extreme nature of the spline function with a high approximation accuracy, a third-order Ryabenky spline can be used to approximate the function, which approximates better than classical interpolation polynomials [10].

The $f(x)$ function belongs to the class of continuous functions up to the $f(x) \in C^q(-\infty, \infty)$, q order derivative.

Where q> 0 is the fixed number. Let the values of the $f(x)$ function be given at nodes of equal spacing

$$f(x) = f(ih), \ (i = 0, \pm 1, \pm 2, ...)$$

In Ryabenky's article, a 2p+1 level $S_{ph}(x)$ spline function with defect p+1 was constructed based on the $f(ih)$ values of the $f(x)$ function. For the $\left|f(x) - S_{ph}(x)\right|$ error in the one-dimensional case (up to the derivative of order r < q), the following is appropriate.

$$\left|f^{(r)}(x) - S^r_{ph}(x)\right| \le h^{q-r} \ k(p) \sup_x \left|f^{(q)}(x)\right| \qquad (7)$$

Where k(p) is not variable, it only depends on p.

For the optional p, the $S_{ph}(x)$ spline can be written to see the exact appearance of the $f(x)$ function via $f(ih)$ values. It is therefore very convenient to apply in various fields. This spline function can also be approximated in *C[a,b]* to the advantage of the Hermit spline.

In the works of S.L. Sobolev the following Ryabenky operator is given for the $S_{ph}(x)$ function in the $[x_i, x_{i+1}]$ interval [14].

$$S_{ph}(x) = \sum_{j=0}^{p} \sum_{k=0}^{p-j} \frac{(p+k)!}{k!\,j!\,p!} [a_{pj}(x) t^{j+k} (1-t)^{p+1}$$

$$S_{ph}(x) = \sum_{j=0}^{p} \sum_{k=0}^{p-j} \frac{(p+k)!}{k!\,j!\,p!} [a_{pj}(x) t^{j+k} (1-t)^{p+1} \quad (8)$$

here $t = \frac{x - x_i}{h}$, $a_{p0}(x_i) = f(x_i)$.

The spline function given in (8) is a 2p+1 level spline with a defect defect p+1, which in general gives a local third-order interpolation spline of the special case Ryabenky given p = 1.

When p= 1, a local cubic spline is formed

$$S_{1h}(x) = a_{10}(x_i)(1-t)^2(1+2t) + a_{10}(x_{i+1})t^2(3-2t)$$

$$+ a_{11}(x_i)t(1-t)^2 - a_{11}(x_{i+1})t^2(1-t)$$

$$a_{10}(x_i) = f(x_i), \quad \Delta f(x_i) = f(x_{i+1}) - f(x_i)$$

$$a_{11}(x_i) = 1! \sum_{k=1}^{1} \frac{\Delta f(x_i)}{1!} S(1,1) = f(x_{i+1}) - f(x_i), \quad S(1,1) = 1$$

$$a_{11}(x_{i+1}) = f(x_{i+2}) - f(x_{i+1})$$

$$S_{1h}(x) - S_3(x) = f(x_i)(1-t)^2(1+2t) +$$

$$+ f(x_{i+1})t^2(3-2t) + (f(x_{i+1}) - f(x_i))t(1-t)^2 -$$

$$-(f(x_{i+2}) - f(x_{i+1}))t^2(1-t) = ((1-t)^2(1+2t)$$

$$-t(1-t)^2)f(x_i) + (t^2(3-2t) + t(1-t)^2 +$$

$$+ t^2(1-t))f(x_{i+1}) - t^2(1-t)f(x_{i+2}) =$$

$$= \phi_1(t)f(x_i) + \phi_2(t)f(x_{i+1}) + \phi_3(t)f(x_{i+2})$$

$$S_3(x) = \varphi_1(t)f(x_i) + \varphi_2(t)f(x_{i+1}) + \varphi_3(t)f(x_{i+2}). \quad (9)$$

here

$$\varphi_1(t) = (1-t)^2(1+t), \ \varphi_2(t) = t(1+2t-2t^2),$$

$$\varphi_3(t) = -t^2(1-t)$$

This function in the $[x_i, x_{i+1}]$ interval (9) can be called the "Ryabenky" cubic spline function.

2.3. Building of a local interpolation cubic spline function

Divide the $[a,b]$ interval into n equal parts

$$a = x_0 < x_1 < \ldots < x_i < x_{i+1} < \ldots < x_n = b$$

We construct the local interpolation cubic spline model under consideration using a linear combination of two $y_1(x)$ and $y_2(x)$ parabolas in the $x \in [x_i, x_{i+1}]$ $(i = \overline{0, n-1})$ interval [9],[13].

We have the following four points $A(x_{i-1}, y_{i-1}); B(x_i, y_i); C(x_{i+2}, y_{i+2}); D(x_{i+1}, y_{i+1});$ to build a $S_3(x)$ spline in the $x \in [x_i, x_{i+1}]$ $(i = \overline{0, n-1})$ range.

It should be noted that in this method, four points are of course used to construct the interpolation cubic spline model (Figure 2).

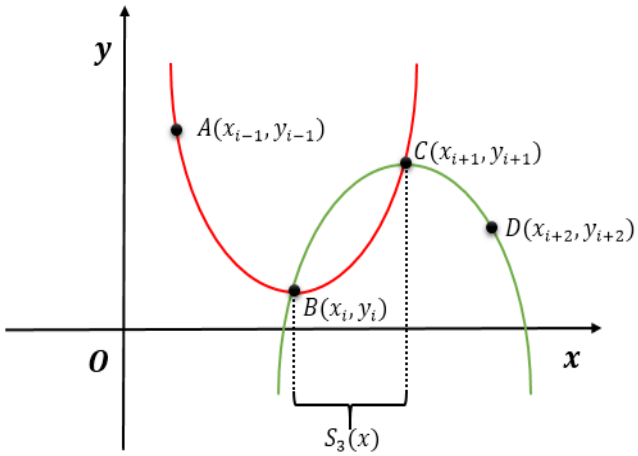


Figure 2: A view of the $S_3(x)$ local cubic spline construction range.

Using the interpolation condition to construct the following $y_1(x) = a_1x^2 + b_1x + c_1$ parabola passing through the points A, B, C, we construct the following system of equations [13], [18]:

$$\begin{cases} ax_{i-1}^2 + bx_{i-1} + c = y_{i-1} \\ ax_i^2 + bx_i + c = y_i \\ ax_{i+1}^2 + bx_{i+1} + c = y_{i+1} \end{cases} \quad (10)$$

(10) we find a, b, c from the system. And the following

$$y_1(x) = a_1x^2 + b_1x + c_1 \quad (11)$$

parabola will be formed.

We construct the following system of equations to construct the second $y_2(x) = a_2x^2 + b_2x + c_2$ parabola passing through the B, C, D points as above:

$$\begin{cases} ax_i^2 + bx_i + c = y_i \\ ax_{i+1}^2 + bx_{i+1} + c = y_{i+1} \\ ax_{i+2}^2 + bx_{i+2} + c = y_{i+2} \end{cases} \quad (12)$$

(12) we find a, b, c from the system. And the following

$$y_2(x) = a_2x^2 + b_2x + c_2 \quad (13)$$

parabola will be formed.

We perform the replacement to make it easier to estimate the error here $h_i = x_i - x_{i-1}$.

$$y_i(x) = y_i(t) = -0{,}5t(1-t)f_{i-1} + (1-t^2)f_i + 0{,}5(1+t)f_{i+1} \quad (14)$$

$$\begin{aligned} y_{i+1}(x) = y_{i+1}(t) = 0{,}5t(1-t)(2-t)f_i + \\ + t(2-t)f_{i+1} - 0{,}5t(2-t)f_{i+2} \end{aligned} \quad (15)$$

Through the combination of the above parabolas, we get the following view

$$S_i(x) = S_i(t) = (\alpha_1 + \alpha_2 t)y_i(x) + (\alpha_3 + \alpha_4 t)y_{i+1}(x).$$

From the interpolation condition

$$S_i(x_i) = f_i, \qquad S_i(x_{i+1}) = f_{i+1}$$

To determine the $\alpha_1, \alpha_2, \alpha_3$ *and* α_4, we construct equations (16):

$$\alpha_1 + \alpha_2 = 1, \quad \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1, \quad (16)$$

where α_3 *and* α_4 are found $\alpha_3 = 1 - \alpha_1$, $\alpha_4 = -\alpha_2$

The system of equations (17) and (18) is formed after some simplification of the interpolation conditions of the first $S_i'(x)$, $S_{i+1}'(x)$ and the second $S_i''(x)$, $S_{i+1}''(x)$ at the x_{i+1} node point of the spline

$$\alpha_1(\Delta^2 f_{i+1} - \Delta^2 f_{i-1}) + \alpha_2\Delta^3 f_{i-1} = \Delta^3 f \quad (17)$$

$$\alpha_1(\Delta^4 f_{i-1}) - \alpha_2(\Delta^2 f_{i-1} + \Delta f_i - \Delta f_{i+2}) = \Delta^3 f \quad (18)$$

here Δ is the difference of operators.

If we say that the system of equations (16) - (18) is solved by $\alpha_1^*, \alpha_2^*, \alpha_3^*, \alpha_4^*$,

Then

$$S_3(x) = S_3(t) = (\alpha_1^* + \alpha_2^* t)y_i(t) + (\alpha_3^* + \alpha_4^* t)y_{i+1}(t).$$

Spline will have 1 defect. But $f_{i-1}, f_i, f_{i+1}, f_{i+2}, f_{i+3}$ coefficients are complex rational functions. Therefore, such splines are inconvenient for digital processing of signals.

If we say $\alpha_1 = 1$, $\alpha_2 = 1$, then $\alpha_3 = 0$, $\alpha_4 = 1$ These values satisfy equations (16) - (17) but do not satisfy equation (18)

The defect of the interpolation spline in the $[x_i, x_{i+1}]$ interval will be 2 [1],[13].

$$S_3(f;x) = S_3(t) = (1-t)y_i(t) + ty_{i+1}(t). \quad (19)$$

We will add these following formulas

$$\varphi_1(t) = -0{,}5t(1-t)^2, \quad \varphi_2(t) = 0{,}5(1-t)(2+2t-3t^2)$$

$$\varphi_3(t) = 0{,}5t(1+4t-3t^2), \quad \varphi_4(t) = -0{,}5(1-t)t^2$$

Substituting the expressions (14) and (15) for the $y_i(t)$ and $y_{i+1}(t)$ given in (19), we obtain the view (20) for the $[x_i, x_{i+1}]$ interval.

$$S_3(f;x) = \sum_{j=0}^{3} \varphi_{j+1}(t)f(x_{i+j-1}). \quad (20)$$

Based on the theory of splines, this function in the $[x_i, x_{i+1}]$ interval (20) can be called a local interpolation cubic spline function.

3. Approximation of functions using local cubic splines

The graph of the local interpolation cubic spline function generated in Section 2 was compared with the graph of the Ryabenky local cubic spline function and the graph of the Grebennikov local cubic spline function.

Let us compare the approximation of these three local cubic splines with the given function $f(x)$.

The local interpolation cubic spline function we propose has the appearance in y $[x_i, x_{i+1}]$ cross section as studied in Section 2.3:

$$S_3(f;x) = \sum_{j=0}^{3} \varphi_{j+1}(t) f(x_{i+j-1}),$$

here

$$\varphi_1(t) = -0,5t(1-t)^2,$$

$$\varphi_2(t) = 0,5(1-t)(2+2t-3t^2),$$

$$\varphi_3(t) = 0,5t(1+4t-3t^2),$$

$$\varphi_4(t) = -0,5(1-t)t^2.$$

here $t = (x - x_i)/h$, $h = \frac{b-a}{N}$, $N = 1,2,...$

We then define this spline function with $S_3(x)$ for convenience.

The second is the Ryabenky local cubic spline function and has the following appearance in y $[x_i, x_{i+1}]$ cross section:

$$S_3(f;x) = \sum_{j=1}^{3} \psi_j(t) f(x_{i+j-1}),$$

here

$$\psi_1(t) = (1-t)^2(1+t),$$

$$\psi_2(t) = t(1+2t-2t^2),$$

$$\psi_3(t) = -t^2(1-t),$$

here $t = (x - x_i)/h$, $h = \frac{b-a}{N}$, $N = 1,2,...$

We define the Ryabenky local cubic spline function by $RS_3(x)$

The third spline is the Grebennikov local cubic spline function, which has the following appearance in the $[x_i, x_{i+1}]$ cross section:

$$S_3(f;x) = \sum_{j=0}^{3} \phi_{j+1}(t) f(x_{i+j-1}).$$

here

$$\phi_1(t) = \frac{1}{6}(1-t)^3,\ \phi_2(t) = \frac{1}{6}(4-6t^2+3t^3),$$

$$\phi_3(t) = \frac{1}{6}(1+3t+3t^2-3t^3),\ \ \phi_4(t) = \frac{1}{6}t^3.$$

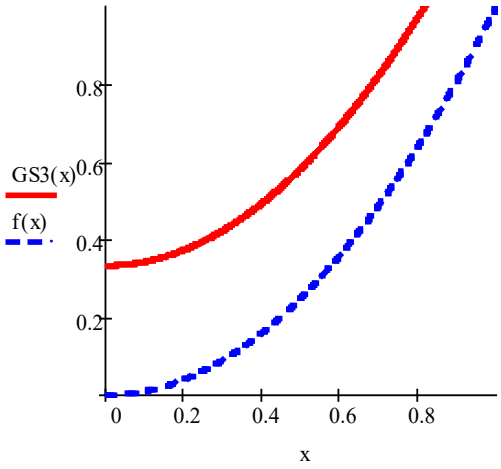
here $t = (x - x_i)/h$, $h = \frac{b-a}{N}$, $N = 1,2,...$

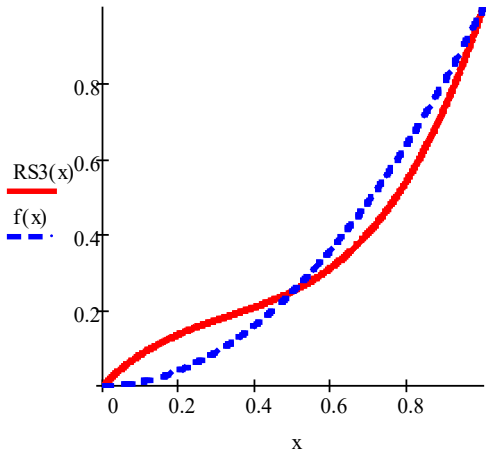
We define the Grebennikov local cubic spline function as $GS_3(x)$. In the examples, without losing generality, we only look at the situation when there is $[a,b] = [0,1]$ $N = 1$.

In this case, we consider the approximation of several functions in $t = x$ and graphs with three cubic splines (Figures 3,4,5 and 6):

1. In the case of $f(x) = x^2$ (examples given in Fig. 3)

As can be seen from the graph of the first example, the $S_3(x)$ spline graph overlaps with the $f(x) = x^2$ function graph, while the $GS_3(x)$ and $RS_3(x)$ spline graphs do not overlap.





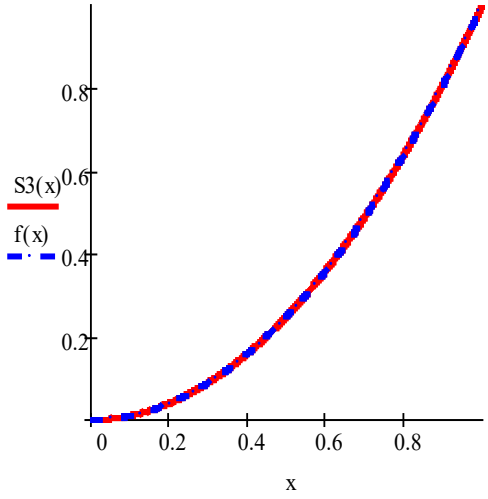


Figure 3: Results of approximation of the $f(x)=x^2$ function with three cubic splines.

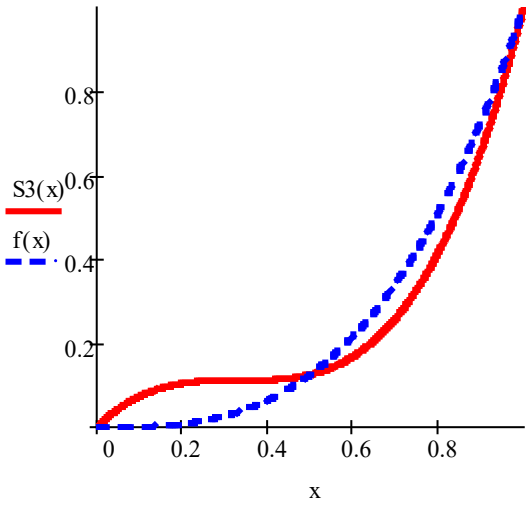
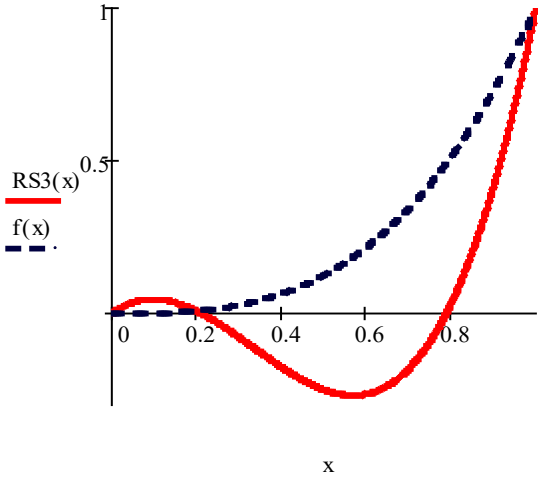
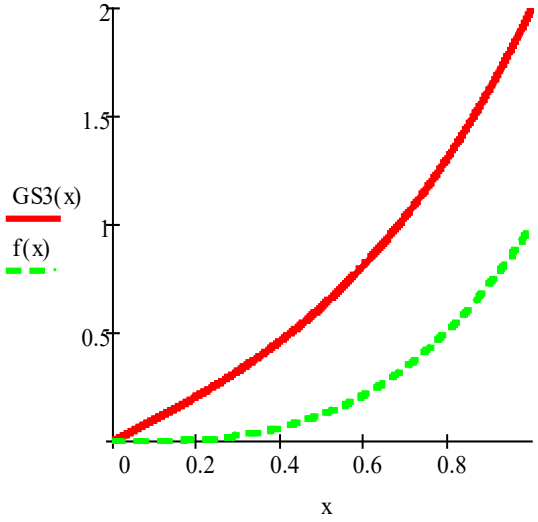


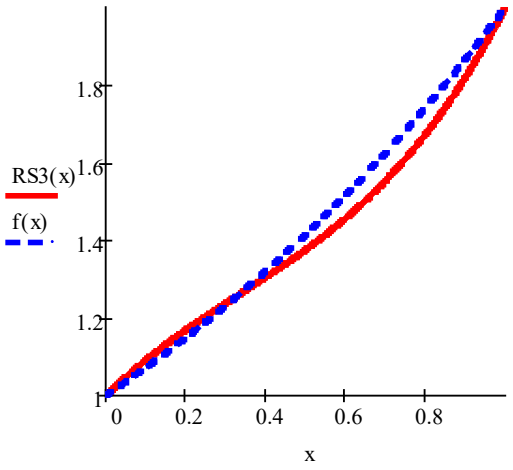
Figure 4: Results of approximation of the $f(x)=x^3$ function with three cubic splines.

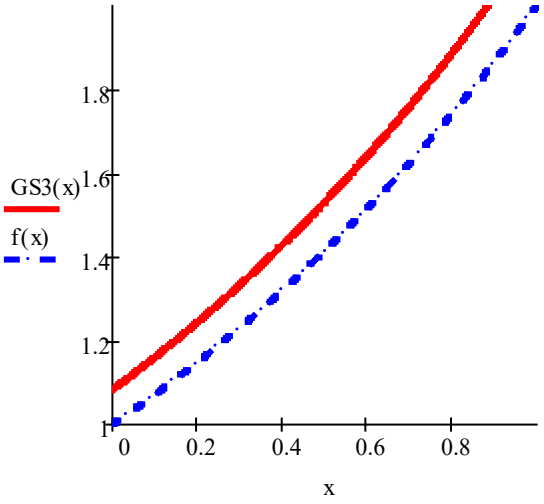
2. In the case of $f(x)=x^3$ (examples given in Fig. 4)





3. In the case of $f(x)=2^x$





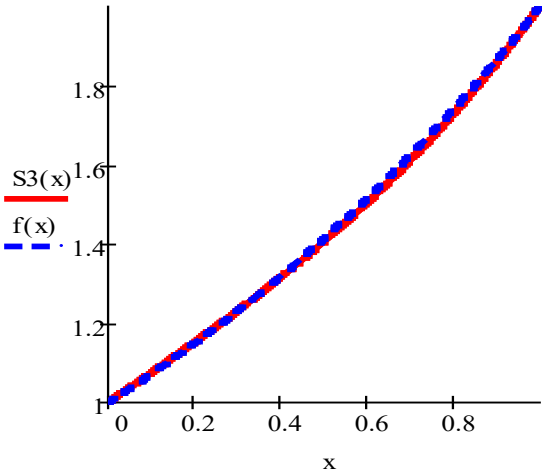
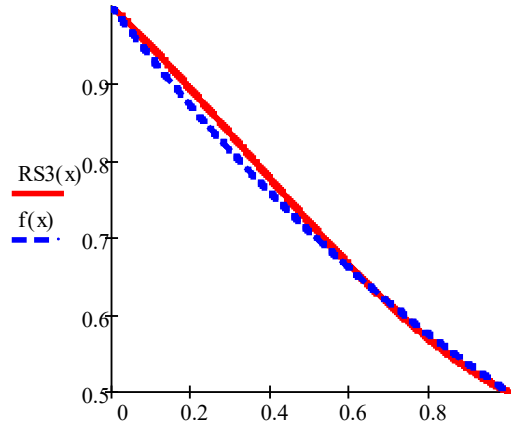
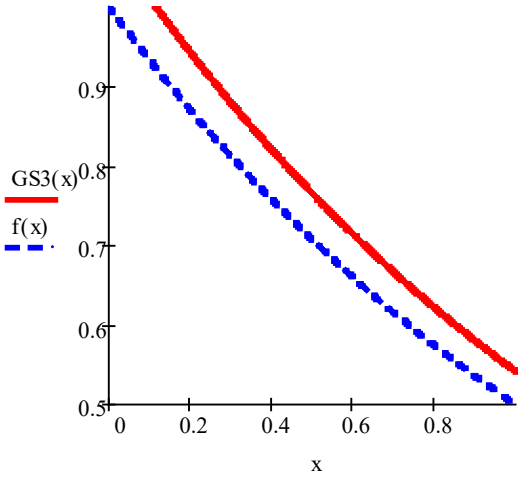


Figure 5: Results of approximation of the $f(x)=2^{x}$ function with three cubic splines

4. In the case of $f(x)=2^{-x}$





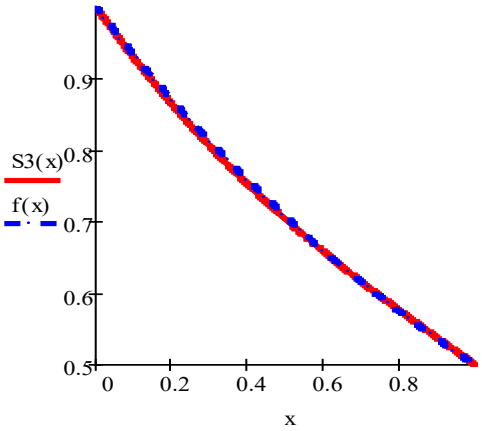


Figure 6: Results of approximation of the $f(x)=2^{-x}$ function with three cubic splines

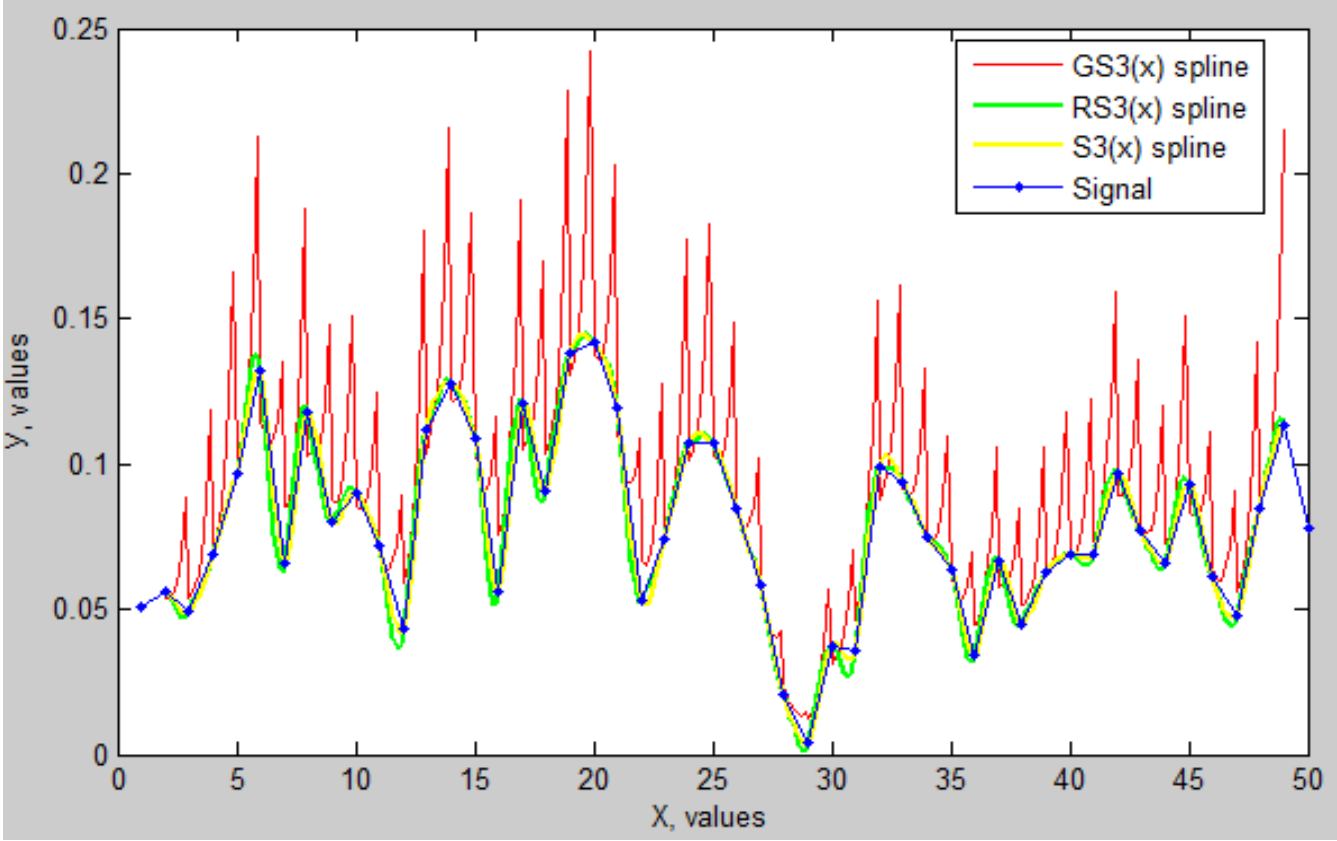
Examples 1,2,3 and 4 show that the $S_3(x)$ spline function graph approximates the $f(x)$ function graph better than the $RS_3(x)$ and $GS_3(x)$ spline function graphs.

4. Digital processing of gastroenterological signals based on developed algorithms

Using the cubic spline models considered, the restoration of the gastroenterological signal given in Table 1 was considered. Based on the above sequence, a tertiary spline construction program was developed in the MATLAB software environment and used in signal processing (Figure 7). The algorithm of this program is shown in Figure 8.

Table 1: Values of the gastroenterological signal

№	Time, s	Amplitude	№	Time, s	Amplitude
1.	1	0.051	26.	26	0.085
2.	2	0.056	27.	27	0.058
3.	3	0.049	28.	28	0.021
4.	4	0.069	29.	29	0.004
5.	5	0.097	30.	30	0.037
6.	6	0.132	31.	31	0.036
7.	7	0.066	32.	32	0.099
8.	8	0.118	33.	33	0.094
9.	9	0.080	34.	34	0.075
10.	10	0.090	35.	35	0.064
11.	11	0.072	36.	36	0.034
12.	12	0.043	37.	37	0.067
13.	13	0.112	38.	38	0.045
14.	14	0.128	39.	39	0.063
15.	15	0.109	40.	40	0.069
16.	16	0.056	41.	41	0.069
17.	17	0.121	42.	42	0.097
18.	18	0.091	43.	43	0.077
19.	19	0.138	44.	44	0.066
20.	20	0.142	45.	45	0.093
21.	21	0.119	46.	46	0.061
22.	22	0.053	47.	47	0.048
23.	23	0.074	48.	48	0.085
24.	24	0.107	49.	49	0.113
25.	25	0.107	50.	50	0.078



7. Results of gastroenterological signal recovery

Table 2: Error results in the process of digital processing of the gastroenterological signal

x_i	$f(x)$	$GS_3(x)$	$RS_3(x)$	$S_3(x)$	$\lvert GS_3(x)-f(x)\rvert$	$\lvert RS_3(x)-f(x)\rvert$	$\lvert S_3(x)-f(x)\rvert$
2	0,056	0,0562	0,056	0,056	0,0002	0	0
2.1	0,0553	0,0589	0,0551	0,0557	0,0036	0,0002	0,0004
2.2	0,0546	0,0632	0,0537	0,0549	0,0086	0,0009	0,0003
2.3	0,0539	0,0694	0,0522	0,0539	0,0155	0,0017	0
2.4	0,0532	0,0778	0,0506	0,0528	0,0246	0,0026	0,0004
2.5	0,0525	0,0887	0,0491	0,0516	0,0362	0,0034	0,0009
2.6	0,0518	0,0535	0,0479	0,0504	0,0017	0,0039	0,0014
2.7	0,0511	0,0544	0,0471	0,0495	0,0033	0,004	0,0016
2.8	0,0504	0,0559	0,0469	0,0489	0,0055	0,0035	0,0015
2.9	0,0497	0,0584	0,0475	0,0487	0,0087	0,0022	0,001
3	0,049	0,0625	0,049	0,049	0,0135	0	0
3.1	0,051	0,0684	0,0509	0,0499	0,0174	0,0001	0,0011
3.2	0,053	0,0765	0,0527	0,0511	0,0235	0,0003	0,0019
3.3	0,055	0,0872	0,0545	0,0528	0,0322	0,0005	0,0022
3.4	0,057	0,101	0,0562	0,0547	0,044	0,0008	0,0023
3.5	0,059	0,1183	0,058	0,0568	0,0593	0,001	0,0022
3.6	0,061	0,0703	0,0598	0,0591	0,0093	0,0012	0,0019
3.7	0,063	0,0729	0,0618	0,0616	0,0099	0,0012	0,0014
3.8	0,065	0,0761	0,064	0,0641	0,0111	0,001	0,0009
3.9	0,067	0,0805	0,0664	0,0666	0,0135	0,0006	0,0004
4	0,069	0,0868	0,069	0,069	0,0178	0	0
$\max_{a\le x\le b}\lvert S_3(x_i)-f(x_i)\rvert$					0,0593	0,0040	0,0023

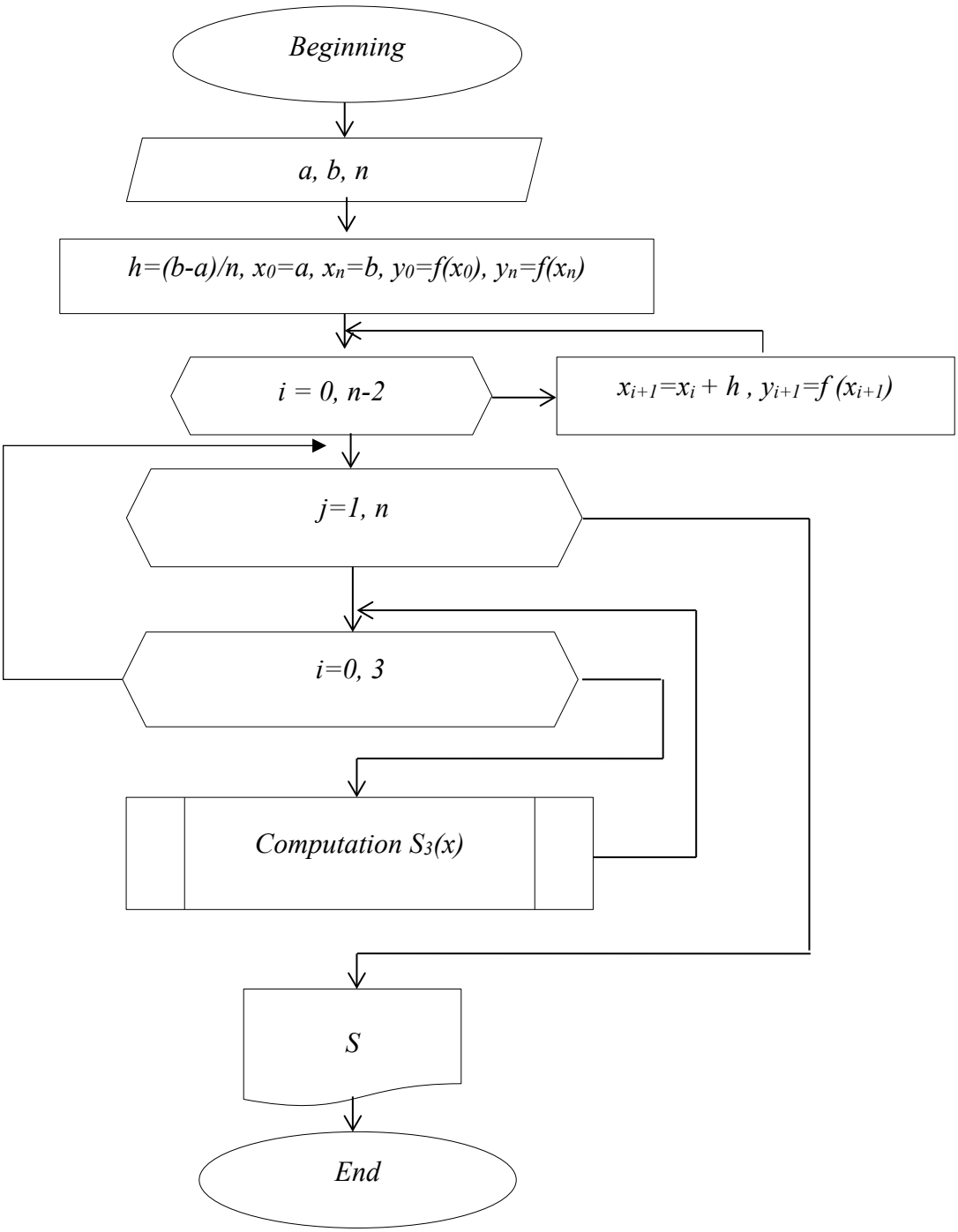


Figure 8: Algorithm of cubic spline construction program.

here, *a* and *b* are the limits of a given interval, *n* is the number of intervals of the interval.

5. Conclusion

In this article, spline functions with a high degree of accuracy in digital processing of signals which are Grebennikov cubic spline, the Ryabenky cubic spline, and the local interpolation cubic spline functions that we proposed.

The construction details selected cubic spline-functions in the equally distributed nets were given, and initially the processes of approximation functions were carried out.

Using selected cubic spline models, the gastroentrological signal was digitally processed and errors were evaluated.

According to results of approximation function using local cubic splines presented in Section 2, as well as the error results in Table 2 obtained from digital processing of the gastroenterological signal, it was found that the local interpolation cubic spline accuracy was high which .we proposed

It can be seen that the application of the mathematical model of the local interpolation cubic spline function we have considered in medicine, in geophysics, and in many areas where the level of accuracy is important in digital processing and recovery of signals leads to effective results.

References

[1] Z. Xakimjon, A. Bunyod, "Biomedical signals interpolation spline models," in International Conference on Information Science and Communications Technologies: Applications, Trends and Opportunities, ICISCT 2019, 2019, doi:10.1109/ICISCT47635.2019.9011926.

[2] J. Rajeswari, M. Jagannath, Advances in biomedical signal and image processing – A systematic review, Informatics in Medicine Unlocked, **8**, 2017, doi:10.1016/j.imu.2017.04.002.

[3] A. Subasi, Biomedical Signal Processing Techniques, 2019, doi:10.1016/b978-0-12-817444-9.00003-9.

[4] S. Patidar, T. Panigrahi, "Detection of epileptic seizure using Kraskov entropy applied on tunable-Q wavelet transform of EEG signals," Biomedical Signal Processing and Control, **34**, 2017, doi:10.1016/j.bspc.2017.01.001.

[5] M. Unser, "Splines: A Perfect Fit for Signal and Image Processing," IEEE Signal Processing Magazine, **16**(6), 1999, doi:10.1109/79.799930.

[6] K. Usman, M. Ramdhani, "Comparison of Classical Interpolation Methods and Compressive Sensing for Missing Data Reconstruction," in Proceedings - 2019 IEEE International Conference on Signals and Systems, ICSigSys 2019, 2019, doi:10.1109/ICSIGSYS.2019.8811057.

[7] D. Singh, M. Singh, Z. Hakimjon, Geophysical application for splines, 2019, doi:10.1007/978-981-13-2239-6_7.

[8] G. Lindfield, J. Penny, Numerical methods: Using MATLAB, 2018, doi:10.1016/C2016-0-00395-9.

[9] A.I. Grebennikov, "Isogeometric approximation of functions of one variable," USSR Computational Mathematics and Mathematical Physics, **22**(6), 1982, doi:10.1016/0041-5553(82)90095-7.

[10] D. Singh, M. Singh, Z. Hakimjon, B-Spline approximation for polynomial splines, 2019, doi:10.1007/978-981-13-2239-6_2.

[11] M. Singh, H. Zaynidinov, M. Zaynutdinova, D. Singh, "Bi-cubic spline based temperature measurement in the thermal field for navigation and time system," Journal of Applied Science and Engineering, **22**(3), 2019, doi:10.6180/jase.201909_22(3).0019.

[12] H.N. Zayniddinov, O.U. Mallayev, "Paralleling of calculations and vectorization of processes in digital treatment of seismic signals by cubic spline," in IOP Conference Series: Materials Science and Engineering, 2019, doi:10.1088/1757-899X/537/3/032002.

[13] H. Zaynidinov, M. Zaynutdinova, E. Nazirova, "Digital processing of two-dimensional signals in the basis of Haar wavelets," in ACM International Conference Proceeding Series, 2018, doi:10.1145/3274005.3274023.

[14] S.L. Sobolev, Selected Works of S.L. Sobolev, Springer US, Boston, MA, 2006, doi:10.1007/978-0-387-34149-1.

[15] Y. Kim, G. Lee, S. Park, B. Kim, J.O. Park, J.H. Cho, "Pressure monitoring system in gastro-intestinal tract," in Proceedings - IEEE International Conference on Robotics and Automation, 2005, doi:10.1109/ROBOT.2005.1570298.

[16] B.H. Jansen, "ANALYSIS OF BIOMEDICAL SIGNALS BY MEANS OF LINEAR MODELING.," Critical Reviews in Biomedical Engineering, **12**(4), 1985.

[17] H. Zaynidinov, J. Juraev, U. Juraev, "Digital image processing with two-dimensional haar wavelets," International Journal of Advanced Trends in Computer Science and Engineering, **9**(3), 2020, doi:10.30534/ijatcse/2020/38932020.

[18] D. Singh, M. Singh, Z. Hakimjon, Parabolic Splines based One-Dimensional Polynomial, 2019, doi:10.1007/978-981-13-2239-6_1.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1511-1516 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Impact of Salinity on Stabilized Leachate Treatment from Ozonation Process

Iva Yenis Septiariva[1], I Wayan Koko Suryawan[2], Novi Kartika Sari[3,*], Ariyanti Sarwono[2]

[1]*Sanitary Engineering Laboratory, Study Program of Civil Engineering, Faculty of Engineering, Universitas Sebelas Maret, Jalan Ir Sutami 36A, Kentingan, Surakarta, Indonesia*

[2]*Department of Environmental Engineering, Faculty of Infrastructure Planning, Universitas Pertamina, Komplek Universitas Pertamina, Jalan Sinabung II, Terusan Simprug, Jakarta 12220, Indonesia*

[3]*Environmental Engineering Study Program, Jurusan Teknologi Infrastruktur dan Kewilayahan, Institut Teknologi Sumatera, Jl. Terusan Ryacudu, Way Huwi, Kec. Jati Agung, Kabupaten Lampung Selatan, Lampung 35365, Indonesia*

ARTICLE INFO

Article history:
Received: 28 September, 2020
Accepted: 29 November, 2020
Online: 21 December, 2020

Keywords:
Salinity
Ozone pretreatment
Leachate
Biodegradability

ABSTRACT

Leachate generated from landfill Indonesia is typical has high salinity and organic matters during the dry season, potentially contaminating water and the environment. An increase in salinity might deteriorate the activity of the microorganisms on the decomposition process. This research aims to study the feasibility of the ozonation technique for treating stabilized leachate. The parameters tested were pH, COD, BOD_5 BOD_5/COD ratio, and TDS. Independent variables of this study were salinity with different concentration from ±0.4 ppt; ±5 ppt; ±10 ppt; and ±20 ppt and ozone contact time up to 60 minutes with an ozone dose of 600 mg/hour. Experimental results revealed that salinity ranging from 0.4 ppt to 20 ppt had little effect on COD removal, and the removal efficiency was around 30%. Further, the ozonation process exhibited better organic removal with an improved BOD_5 concentration from 426 mg/L to 887 mg/L in low salinity conditions (0.4 ppt). Meanwhile, it significantly affected the old leachate treatment, where the leachate biodegradability can be enhanced from 0.120 (initial condition) to 0.321 with salt addition of 20 ppt.

1. Introduction

The end of life of municipal solid waste (MSW) in landfills has always been expected because of its cost and operational efficiency [1]. Landfill leachate results from percolated rainwater and moisture passing through the layers of wastes in landfills [2]. The combination of pollutant parameters like BOD_5 (Biochemical Oxygen Demand), COD (Chemical Oxygen Demand), ammonia, and inorganic salts makes leachate a potential contamination source in river water and surface water. The characteristics and composition of leachate depend on the biological and chemical processes that occur during waste degradation, solid waste composition, rainfall density, groundwater percolation rate, and landfill age.

Government regulation requires minimizing the risk to public health, therefore all of the leachates generated from conventional landfills must be treated according to quality standards. The BOD_5/COD ratio can predict changes in the biodegradation rate in leachate. In general, the ratio of BOD_5/COD contained in leachate ranges from 0.4 - 0.6. This ratio range states that the organic compounds contained in leachate are biodegradable [3]. One of the causes of the low degradability of leachate is the high salinity value, where salinity is one of the inhibitors in the degradation process of microorganisms.

Lim et al. stated that salinity does not affect the hydrolysis and acid formation processes, but inhibits bacteria's activity carrying out the digestion process [4]. The effect of salinity on the waste degradation process can be measured through the COD content, the characteristics of the biogas produced, and methane gas production. Leachate salinity contain potassium (K^+), sodium (Na^+) and chloride (Cl^-) ions [5]. Chloride inhibits bacterial activity so that it can reduce COD content. High salinity content can reduce methane gas formation [6]. Due to the high complexity in its waste composition and characteristics, it is arduous to define a better leachate treatment strategy [7].

Apart from having a high salt concentration, leachate concentrate also contains large amounts of refractory organics [8].

*Corresponding Author: Novi Kartika Sari, Email: novi.sari@tl.itera.ac.id

In addition, old leachate consists primarily of refractory organics due to less volatile fatty acids, which is turn can lower the BOD_5/COD ratio [9]. The physical-chemical process can be applied to pretreatment, tertiary processing, and the overall leachate treatment process. The purpose of the oxidation process in leachate processing is to oxidize organic substances to the most stable form of oxidation, namely into carbon dioxide and water, and increase the biodegradability of organic pollutants suitable for economic, biological processing [10] Furthermore, in principle, physicochemical technologies are suitable for wastewater treatment with extremely low biodegradability, such as stabilized leachate [11] Ozone treatment is one of the strongest chemical processes in leachate processing, because it has the potential for oxidation to reduce stubborn organic fractions.

However, the variation of salinity in landfill leachate fluctuates, leading to further deteriorating treatment efficiency [12]. Several studies have reported on the influence of high salinity on the landfill leachate by the ozonation process [13-15]. This research was designed to examine the treatment performance subjected to different salinity by the ozonation process from stabilized landfill leachate by adopting those approaches. The scope of future research is to compare three types of leachate characteristics, namely young leachate, semi-old leachate, and stabilized leachate with variations in the concentration of salinity by processing with ozonation.

2. Material and Method

2.1. Properties of leachate

Mature leachate was obtained from an aged landfill in Batam City, Indonesia, which has been operating since 1992. This research was conducted in a research laboratory at Universal University in Batam City. Research sampling was carried out at the inlet section of the leachate treatment plant, namely the landfill in Batam City. The laboratory test results from sampling on the landfill are leachate belonging to the characteristics of old leachate which has a COD concentration of 3000-4000 mg/L.

The mature or stabilized leachate's COD concentration was 3,000-4,000 mg/L, ammonia concentration was 1,000-1,500 mg/L, and pH was 7.8. The leachate was categorized as stabilized leachate for this study due to the exceedingly low BOD_5/COD ratio (<0.15) and high salinity content. The leachate was pumped from the leachate collecting pond. After collection, the stabilized leachate was stored in a cool container (4°C) to maintain the inherent characteristics. Before subjected to the ozonation process, the leachate samples were prefiltered using a vacuum pump.

2.2. Laboratory scale reactor

An illustration of leachate treatment with ozonation can be seen in Figure 1. It is equipped with the ozone generator that can supply ozone with a flow rate of 0.25 g/hour in the frequency of 50/60 Hz, providing an ozone flow rate reaching 1.4 L/minute after calibration [9] . KI solution was prepared to absorb excess ozone gas and safety. The ozone concentration was determined according to the standard method by titration using KI solution. The processes in the reactor were run in a thoroughly complete mix condition with 10 L volume of the reactor.

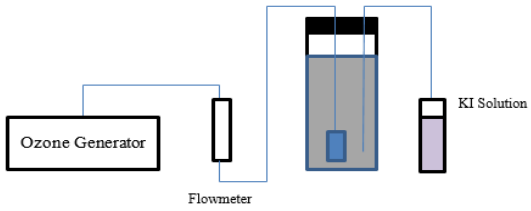


Figure 1: Schematic diagram of ozonation reactor

In Figure 1, a flow diagram of the leachate treatment process in the ozone reactor can be described. The first step is to prepare a KI (Potassium Iodide) solution by dissolving some potassium iodide powder with aquadest. Then, the solution was put into an iron tube with a tube and put into the ozone reactor. Stabilized leachate was put into the ozone reactor at a volume of 10 L. After that, prepared the flowmeter by adjusting the flowrate to 1.4 L / minute. The flowmeter was mounted on the ozone generator and connected to an ozone reactor which already contains old leachate.

2.3. Experimental procedure

The physicochemical treatment process in this study involved ozonation used in the batch operation. At the start of each experiment, the reactor was filled with 9 L of stabilized leachate. Samples were regularly taken at the specified time of 20 minutes, accounting for different applied salinity concentrations. These conditions were performed at a salt concentration of 0.4 ppt, 5 ppt, 10 ppt, and 20 ppt for 60 minutes. The ozonation for treating stabilized landfill leachate was operated for 20 minutes, 40 minutes, and 60 minutes, respectively. Different salt concentrations were adjusted by adding NaCl.

2.4. Sampling and analysis

The initial leachate quality was all measured to determine the original characteristics of the contaminants in the leachate. The initial characteristic parameters consisted of pH, temperature, conductivity, COD, BOD_5, and TDS as for leachate. The leachate used was certainly not low in salinity (as a control).

The leachate properties were analyzed for the following elements: COD, BOD_5, and TDS. The ozonation reactor was continuously monitored (online measuring of pH), and sampling was taken (COD, BOD_5, and BOD_5/COD ratio) during every cycle. Samples collected every operational cycle were filtered through a 0.45 µm filter before analysis. All the parameters were analyzed based on the procedure BOD_5 (APHA A-B 1995), COD (APHA B-C 1995), TSS (APHA N 1995), TKN (APHA C 1995), and salinity (APHA B&C 1995) [16].

3. Results and Discussion

3.1. Leachate characteristics

The collected leachates' initial characteristics depend on the municipal solid waste's waste composition and water content (MSW). The quality of stabilized leachate has not fulfilled the effluent standard, set by government regulation, and both are presented in Table 1. It can be ascertained that the salinity in leachate is only 0.4 ppt to be used as a control in research.

Table 1 shows that the untreated stabilized leachate sample possesses a high concentration of chemical parameters compared to government regulation. The leachate sample's pH value was

beyond the quality standards because anaerobic decomposition was attributed to decreased volatile fatty acid concentration. The electrical conductivity value (19.33 mS/cm) demonstrates dissolved inorganic materials in the samples. The concentration of TDS (4,100 mg/L) also fluctuates widely. The leachate's dark brown colors are mainly caused by ferrous oxide ferric form and the ferric hydroxide colloid formation and complexes with fulvic and humic substances [17]. The stabilized leachate has COD and BOD_5 concentration of 3,561 mg/L and 426 mg/L, respectively. This result reveals that organic matter in stabilized leachate is difficult to degrade biologically due to the complexity of organic compounds. The lower BOD_5/COD ratio in this leachate characteristic might be attributed to the leachate's biologically recalcitrant fraction.

Table 1: A comparison of untreated characteristics of the stabilized leachate and leachate quality standards

Parameter	Units	Stabilized Leachate	Government Regulation
pH		9.4	6.0 – 9.0
Temperature	°C	24.5	
Conductivity	mS/cm	19.3	
COD	mg/L	3,561.0	300
BOD_5	mg/L	426.0	150
BOD_5/COD		0.12	
TKN	mg/L	1,757.0	60
Salinity	ppt	0.4	
TDS	mg/L	4,100.0	

3.2. Stabilized leachate treatment: Effect of ozone contact time and salinity

The chemical oxidation process for the leachate treatment results in a better quality of stabilized leachate than the initial effluent. Ozonation with different salt concentrations contributes to the degradation of COD substances, the improvement of biodegradability ratio, the change in pH, and total dissolved solids.

3.3. The degradation of COD substances

Figure 2 compares the COD reduction in raw leachate at different salinity subjected to ozone treatment at a certain period. From the COD concentration obtained, the oxidation efficiency decreased slightly after 60 minutes. With 20 minutes of ozone exposure time, COD concentration decreased from 3,561 mg/L to 2,538 mg/L and 2,863 mg/L at salinity of 0.4 ppt and 20 ppt respectively. By adjusting the salt concentration of 5 ppt, the COD efficiency removal increases from 19.91% to 31.99% in ozonation time from 40 min to 60 min. Figure 2 shows that as ozone contact time increases, the COD content of stabilized leachate decreases faster than that of initial mature leachate. The initial COD value (3,561 mg/L) declines to 2,479 mg/L, 2,422 mg/L, 2,521 mg/L, and 2,434 mg/L respectively at the ozone contact time of 60 min and different salinity resulting in removal efficiencies of 30.38%, 31.99%, 29.21% and 31.65%.

These results suggested that COD removal efficiencies remained largely unaffected due to variations in salinity from 0.4 ppt to 20 ppt. Moreover, a minor decline in COD content removal was obtained when the ozone contact time was prolonged from 20 min to 60 min. Salinity variations showed a negligible effect on organics removal, while it influenced the nitrification and denitrification efficiency to a more considerable extent. The salt inhibition can be diminished significantly after the long acclimatization time of the biomass [18].

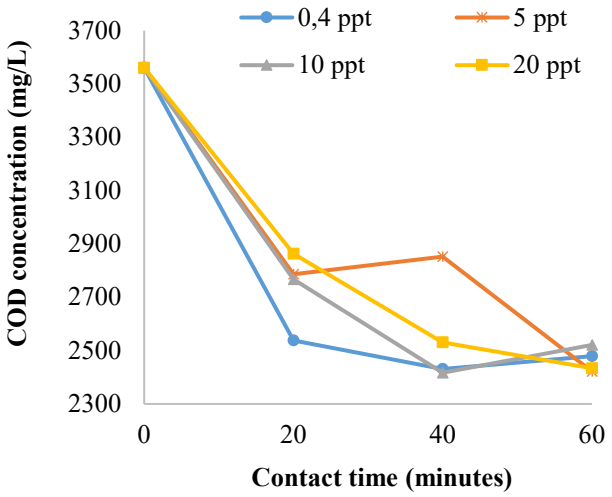


Figure 2: Effect of ozone contact time and the salinity on the COD concentration

The ozonation process and the further oxidation process by accelerating the decomposition of ozone to OH, is one of the wise efforts to be effective in treating leachate. The ozone decomposition reaction is

$$O_3 + HO^- \rightarrow HO_2^- + O_2 \quad (1)$$

$$O_3 + HO_2^- \rightarrow OH^* + O_2^* + O_2 \quad (2)$$

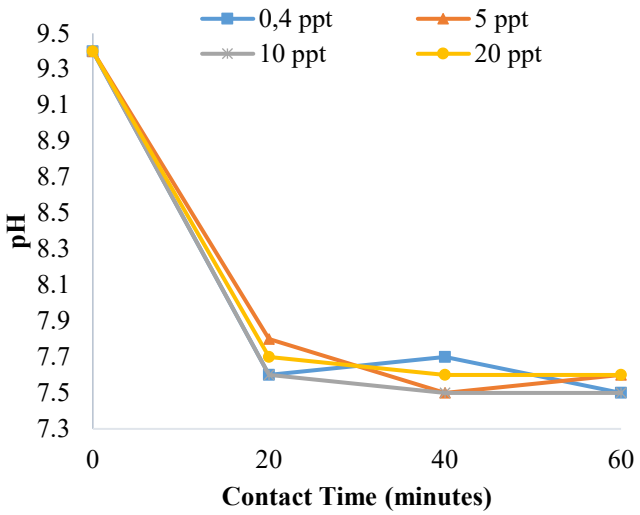


Figure 3: Effect of ozone contact time and the salinity variation on pH

3.4. The change of pH

Applying ozonation at a high leachate pH of 9.4 will most probably lead to a faster ozonation decomposition, therefore a significant enhancement of HO* production which is a more reactive compared to ozone itself, maybe achieved [19]. It is possible that at lower pH, organic pollutants are shifted toward their non-ionized forms. This promotes a lower reactivity toward

ozone compared to the ionized or dissociated form [20]. The conversions of the recalcitrant compound will also alter the occurring pH. The pH gradually decreased as ozone contact time increased, and the variation of effluent pH is depicted in Figure 3. When the salt concentration was 0.4 ppt, effluent pH slowly declined from 9.4 to 7.5 as ozonation time was increased from 0 to 60 min. Furthermore, salts of 5 ppt effluent pH gradually decreased from 9.4 to 7.6 after 60 min of ozone contact time. This might be due to the degradation of macromolecular organic content by ozone to form aliphatic acids, alcohols, and aldehydes and subsequently lead to lower pH.

3.5. The improvement of the biodegradability ratio

From Figure 4, it can be seen that ozonation leads to an increase in the BOD^5/COD ratio. On the other hand, biodegradability changes one indicator to evaluate the ozone's efficiency to treat wastewater [21].

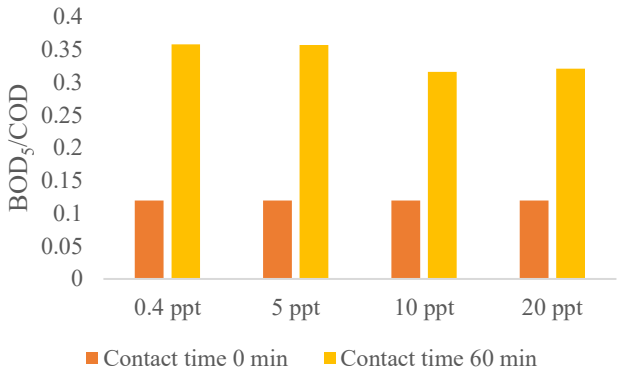


Figure 4: Effect of ozone contact time and the salinity variation on BOD_5/COD ratio of stabilized leachate

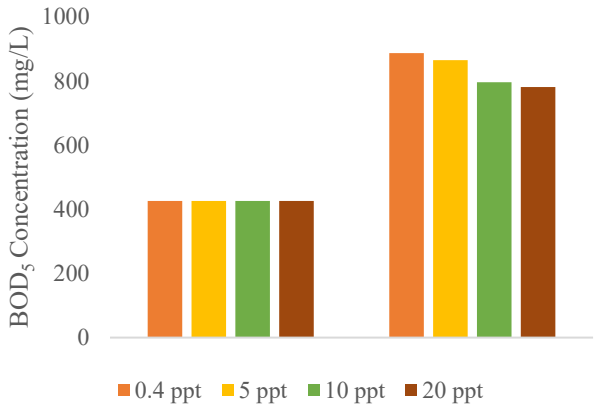


Figure 5: Effect of ozone contact time and the salinity variation on BOD_5 concentration of stabilized leachate

The improvement of biodegradability as measured by the BOD_5/COD ratio are shown in Figure 4. For a 60 min ozone contact time, COD concentration gradually decreased, and BOD_5 concentration slowly increased from 426 mg/L to 865 mg/L at a salinity concentration of 5 ppt (Figure 5). When salinity concentration is adjusted to 20 ppt, biodegradability rises from 0.12 to 0.321. Furthermore, the effect of ozonation is visible at the salinity of 0.4 ppt, and the biodegradability is improved from 0.12 to 0.358. These results indicate that ozonation technology can destruct macromolecular organics into lower molecular weight substances [22, 23] Consequently, the ozonation process yielded a more favorable effluent for biological treatment with a BOD_5/COD ratio of more than 0.3 [24].

Advanced treatment with ozone/persulfate oxidation lowered pH from herbal decoction pieces wastewater and improved biodegradability from 0.16 to 0.55 [25]. Another study with ozone/hydrogen peroxide also enhanced the biodegradability of textile wastewater [22, 23] and color removal [26], as well as the efficiency of biological wastewater treatment [27]. Figure 6 shows that there is a close relationship between salinity and biodegradability values. The higher the salinity in leachate, the non-biodegradable food will be processed biologically.

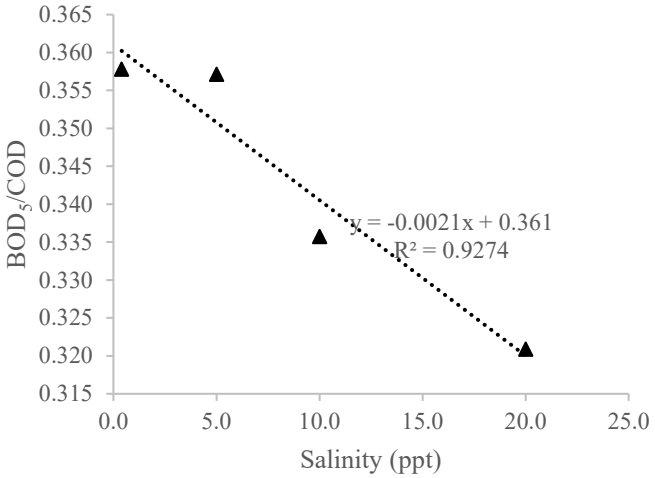


Figure 6: Correlation of ozone salinity and BOD_5/COD ratio of stabilized leachate

3.6. Total dissolved solids (TDS) analysis

Leachate generally contains high salinity and TDS. The TDS values of samples subjected to ozone treatment are presented in Figure 7. In Figure 7, the initial concentration of total dissolved solids (TDS) ranged from 4,000 – 7,000 mg/L depending on salt concentrations. TDS in leachate shows the presence of inorganic salts, organic compounds dissolved in leachate. The higher the salinity due to the higher TDS can increase leachate's toxicity because it can change the ion composition [28]. TDS can be sodium (salt), calcium, magnesium, potassium, carbonate, nitrate, bicarbonate, chloride, and sulfate [29].

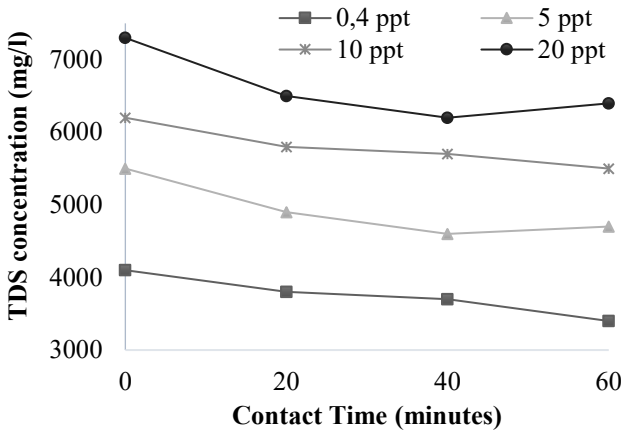


Figure 7: Effect of ozone contact time and the salinity variation on the TDS concentration

TDS primarily revealed the presence of inorganic salts and dissolved organics. The increase in salinity provokes an increase in TDS concentration. The final TDS concentrations increase for all different salinity from 0.4 ppt to 20 ppt. The amount of TDS reflects the availability of the mineralization process and becomes one of the parameters considered for the licensing discharge of landfill leachate. It can be detected in Fig.4 that following the ozone treatment for 60 min of ozone contact time at each salt concentration, the TDS values are 3,400 mg/L, 4,700 mg/L, 5,500 mg/L, and 6,400 mg/L respectively. Overall, ozonation does not significantly contribute to the reduction of TDS.

4. Conclusions

The influence of different salinity conditions on the ozonation of stabilized landfill leachate was studied for several physicochemical characteristics such as COD, BOD_5, pH, and TDS. In this project, ozonation technology was employed to treat stabilized leachate to study the COD removal efficiencies at different salinity concentrations. However, the COD value in the stabilized leachate is still above the discharge standard in conformance with the Indonesian regulation. The results indicated that significant removal of the COD parameter could achieve 31.65% for 60 minutes of ozone contact time with 20 ppt salinity. The estimated COD removal in all conditions (0.4 ppt, 5 ppt, and 10 ppt of salinity variation) was 30.38%, 31.99%, and 29.21%, respectively. Though slight difference of COD removal was obtained, biodegradability, as indicated by BOD_5/COD ratio, can be significantly enhanced by ozonation.

References

[1] Q. Xu, G. Siracusa, S. Di Gregorio and Q. Yuan, "COD removal from biologically stabilized landfill leachate using Advanced Oxidation Processes (AOPs)," Process safety and environmental protection, **120**, 278-285, 2018, https://doi.org/10.1016/j.psep.2018.09.014.

[2] R. Poblete, I. Oller and M. I. C. E. Maldonado, "Improved landfill leachate quality using ozone, UV solar radiation, hydrogen peroxide, persulfate and adsorption processes," Journal of environmental management, **232**, 45-51, 2019, https://doi.org/10.1016/j.jenvman.2018.11.030.

[3] G. Tchobanoglous, H. Theisen and S. Vigil, Integrated solid waste management: Engineering principles and management, New Jersey: McGraw-Hill, 1993.

[4] Y. G. Lim, C. Niwa, N. Nagao and T. Toda, "Solubilization and methanogenesis of blue mussels in saline mesophilic anaerobic biodegradation," International Biodeterioration & Biodegradation, **61**(3), 251-260, 2008, https://doi.org/10.1016/j.ibiod.2007.06.012.

[5] M. Loncnar, M. Zupančič, P. Bukovec and M. Z. Justin, "Fate of saline ions in a planted landfill site with leachate recirculation," Waste management, **30**(1), 110-118, 2010, https://doi.org/10.1016/j.wasman.2009.09.010.

[6] S. Supparattanapan, P. Saenjan, C. Quantin, J. L. Maeght and O. Grünberger, "Salinity and organic amendment effects on methane emission from a rain-fed saline paddy field," Soil Science and Plant Nutrition, **55**(1), 142-149, 2009, https://doi.org/10.1111/j.1747-0765.2008.00330.x.

[7] M. Y. Kılıç, K. Kestioglu and T. Yonar, "Landfill leachate treatment by the combination of physicochemical methods with adsorption process," J. Biol. Environ. Sci, **1**(1), 37-43, 2007.

[8] Wu. Chu, Chen. W, Gu. Zhepei and Li. Qibin, "A Review of the Characteristics of Fenton and Ozonation Systems in Landfill Leachate Treatment," Science of the Total Environment, Journal Pre-proof, 2020, https://doi.org/10.1016/j.scitotenv.2020.143131.

[9] I.Y. Septiariva, T. Padmi, E. Damanhuri, and Q. Helmy, "A study on municipal leachate treatment through a combination of biological processes and ozonation," MATEC Web of Conferences 276, 06030, 2019, https://doi.org/10.1051/matecconf/201927606030.

[10] X. Li, W. Zhu, Y. Wu, C. Wang, J. Zheng, K. Xu and J. Li, "Recovery of potassium from landfill leachate concentrates using a combination of cation-exchange membrane electrolysis and magnesium potassium phosphate crystallization," Separation and Purification Technology, **144**,1-7,2015,https://doi.org/10.1016/j.seppur.2015.01.035.

[11] W. Chen, Z. Gu, P. Wen and Q. Li, "Degradation of refractory organic contaminants in membrane concentrates from landfill leachate by a combined coagulation-ozonation process," Chemosphere, **217**, 411-422, 2019, https://doi.org/10.1016/j.chemosphere.2018.11.002.

[12] O. Lefebvre and R. .. Moletta, "Treatment of organic pollution in industrial saline wastewater: a literature review," Water research, **40**(20), 3671-3682, 2006, https://doi.org/10.1016/j.watres.2006.08.027.

[13] I. Monje-Ramirez and M. O. De Velasquez, "Removal and transformation of recalcitrant organic matter from stabilized saline landfill leachates by coagulation–ozonation coupling processes," Water research, **38**(9), 2359-2367, 2004, https://doi.org/10.1016/j.watres.2004.02.011.

[14] W. Chen and Q. Li, "Elimination of UV-quenching substances from MBR- and SAARB-treated mature landfill leachates in an ozonation process: A comparative study," Chemosphere, **242**, 125256, 2020, https://doi.org/10.1016/j.chemosphere.2019.125256.

[15] S. A. Amr, H. A. Aziz, M. S. Hossain and M. J. Bashir, "Simultaneous removal of COD and color from municipal landfill leachate using Ozone/Zinc sulphate oxidation process," Glob Nest J,**19** (19), 498-504, 2017.

[16] American Public Health Association, American Water Works Association, Water Pollution Control Federation, & Water Environment Federation, Standard methods for the examination of water and wastewater, **2**, USA: American Public Health Association, 1915.

[17] W. Chen, Z. Gu, P. Wen and Q. Li, "Degradation of refractory organic contaminants in membrane concentrates from landfill leachate by a combined coagulation-ozonation process," Chemosphere, **217**, 411-422, 2019, https://doi.org/10.1016/j.chemosphere.2018.11.002.

[18] E. R. Rene, S. J. Kim and H. S. Park, "Effect of COD/N ratio and salinity on the performance of sequencing batch reactors," Bioresource technology, **99**(4), 839-846, 2008, https://doi.org/10.1016/j.biortech.2007.01.037.

[19] M. Chys, V. Oloibiri, W. Audenaert, K. Demeestere and S. W. Van Hulle, "Ozonation of biologically treated landfill leachate: efficiency and insights in organic conversions," Chemical engineering journal, **277**, 104-111, 2015, https://doi.org/10.1016/j.cej.2015.04.099.

[20] W. Audenaert, D. Vandierendonck, S. H. Van Hulle and I. Nopens, "Comparison of ozone and HO induced conversion of effluent organic matter (EFOM) using ozonation and UV/H_2O_2 treatment," Water research, **47**(7), 2387-2398, 2013, https://doi.org/10.1016/j.watres.2013.02.003.

[21] P. Fu, X. Lin, G. Li, Z. Chen and H. Peng, "Degradation of thiol collectors using ozone at a low dosage: kinetics, mineralization, ozone utilization, and changes of biodegradability and water quality parameters," Minerals, **8**, (11), 477, 2018, https://doi.org/10.3390/min8110477.

[22] I. W. K. Suryawan, G. Prajati, A. S. Afifah and M. R. Apritama, "NH3-N and COD reduction in Endek (Balinese textile) wastewater by activated sludge under different DO condition with ozone pretreatment," Walailak Journal of Science and Technology (WJST), 2020, http://wjst.wu.ac.th/index.php/wjst/article/view/9127.

[23] I. Suryawan, M. J. Siregar, G. Prajati and A. S. Afifah, "Integrated ozone and anoxic-aerobic activated sludge reactor for endek (Balinese textile) wastewater treatment," Journal of Ecological Engineering, **20**(7), 2019, http://dx.doi.org/10.12911/22998993/109858.

[24] M. Á. López-Ramírez, O. P. Castellanos-Onorio, M. A. Susunaga-Miranda, F. Lango-Reynoso, M. del Refugio Castañeda-Chávez and J. Montoya-Mendoza, "Treatment of Leachates of a Controlled Landfill in Veracruz By Using the Fenton Method.," Nature environment & pollution Technology, **18**(1), 2019.

[25] G. Tang, W. Chen, Y. Wei, T. Shao, M. Zhang, Z. Jia and F. Ma, "Novel Advanced Treatment of Physically Treated Effluent from Herbal Decoction Pieces Wastewater Using a Combined Ozone/Persulfate-UBAF Process," Polish journal of environmental studies, **28**(4), 2857-2866, 2019, https://doi.org/10.15244/pjoes/92126.

[26] N. H. T. Huynh, P. H. Duong and Y. S. .. Yoon, " A New Combination Treatment System of Ozonation and Electrocoagulation for CI Acid Red

114 Dye Removal from Dyeing Wastewater," Nature environment and pollution technology, **17**(4), 1383-1390, 2018.

[27] W. Na, "Impact of Ultrasonication-Ozonation Pretreatment on Anaerobic Digestion of Sewage Sludge," Nature environment and pollution technology, **17**(2), 539-542, 2018.

[28] P. K. Weber-Scannell and L. K. Duffy, "Effects of total dissolved solids on aquatic organism: a review of literature and recommendation for salmonid species," American Journal of Environmental Sciences, **3**(1), 1-6, 2007.

[29] W. Chen, A. Zhang, G. Jiang and Q. Li, "Transformation and degradation mechanism of landfill leachates in a combined process of SAARB and ozonation," Waste management, **85**, 283-294, 2019, https://doi.org/10.1016/j.wasman.2018.12.038

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1517-1528 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Implementation of a Journalist Business Intelligence in Social Media Monitoring System

Abba Suganda Girsang*, Sani Muhamad Isa, Natasya, Megga Eunike Cristilia Ginzel

Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia

ARTICLE INFO

Article history:
Received: 12 November, 2020
Accepted: 10 December, 2020
Online: 21 December, 2020

Keywords:
Journalist
Social Media
Business Intelligence
Monitoring

ABSTRACT

The need to update the news at this time is very important. Journalists have responsibilities as their job. Apart from conveying the correct news to the public, they must also deliver it quickly. Journalists who are tasked as news seekers must always be updated about the latest news that is viral, therefore a journalist must be fast and reliable in finding news sources. With this research, it can help journalists work in finding news sources by utilizing public information from Twitter, as one of social media platforms. Social media is considered a source of public information that can involve everyone. In previous research, a system has been created to collect tweets from several news accounts on Twitter which are then grouped into various themes based on cosine similarity. Continuing from the results of previous research, this research obtained data from crawlers stored in the database server can be displayed in the form of a report on a business intelligence application dashboard using a data warehouse. Thus, there is some information that can be used to make decisions. This information includes how much news engagement is, how active users are in commenting, trending posts categories, etc. The results of the research will show how much news is generated based on users who actively like, comment and share news on news accounts that are on Twitter. These results will be used as a reference by journalists to cover live news that is viral.

1. Introduction

Along with technological developments, the distribution and exchange of information are becoming faster and kept growing. Social media is the answer towards conveying information quickly in effective and efficient ways, reaching many people. It is an open media that grants access to all people all over the world, regardless of age, occupation, or social level. Social media such as Twitter, Facebook, MySpace, YoutTube, Flickr, and others have been growing at a tremendous rate and the adoption rate of such media has been skyrocketing, which in turn, has delivered astronomical numbers of users in less than 10 years [1].

Through social media, people can create or participate in the making of every content that can be shared and discussed. The freeness and openness which social media offers were very different and more convenient, compared to other media.

In 2012, six years after Twitter got launched, Twitter had already 140 million active users and was still growing. There were 340 million tweets uploaded in a day [2], which creates a high possibility that social media will give a significant impact on gaining and reaching a lot of worthy news from a lot of people. Aware of this opportunity, we realize the worthy news that is obtained can give a huge impact to journalism. But in this case, social media will require a monitoring system to find out the latest news, which is currently discussed in society through social media. Therefore, journalists will be able to find the latest news effectively and efficiently. Monitoring directly through social media will also help journalists to understand the bigger picture of the latest trending news. The story can be told in many different points of view and also accompanied by several other people's opinions. Of course, social media also has the possibility of spreading false information or hoax news, but this is where the role of journalists is needed. Journalists can help the public to identify and inform the true version of the news or news that is trending.

Social media will become a bridge to connect journalism with public stories and opinions which contain worthy news. There is little doubt that social media has an important role to play in the future of journalism and that it has functioned to be the source of ideas and information [3].

*Corresponding Author: Abba Suganda Girsang, Email: agirsang@binus.edu

2. Related Work

Social media technology takes many forms, including internet forums, weblogs, social blogs, microblogging, wikis, podcasts, photos or images, videos, ratings and social bookmarking. By applying a set of theories in the areas of media research (social presence, media wealth) and social processes (self-presentation, self-disclosure) Kaplan and Haenlein created classification schemes for different types of social media in their article Business Horizons published in 2010. According to Kaplan and Haenlein there are six types of social media collaborative projects, blogs, content communities, social networking sites, virtual game worlds, and virtual social worlds [4].

Social media has vast growth and clearly matters for society [5]. After interviews with 53 digital journalists, some public Figureures, extra media organizations, opinion leaders, and others in social media were shown to affect news production [6] Journalists use social media to find story leads and to share their work with audiences, which has made journalism more interactive [7].

As one of social media platform, Twitter can be seen as a modern version of person-on-the-street interviews, or even a journalistic stand-in for actual polling. While this was never a reliable way of gauging public opinion, the fact that Twitter makes these vox populi searchable and embeddable vastly reduces the effort that it takes to collect and call upon them. Its use has proliferated to the point that journalists see Twitter as a reliable source of news [5]. An online survey for US journalists with participation of 212 journalists, shows that journalists who use Twitter, potentially causing them to dismiss information that many of their colleagues identify as newsworthy [8].

Investigation in 2012, presidential primary which was the first presidential election where Twitter was used as candidate and party reactions for what was broadcasted in news media, which makes Twitter become a field of new sources for journalistic content [9]. On London Olympic 2012, @London2012 account served as the official account of London Olympic and became another news source for journalistic [10]. Austria's national selection campaign proved that Twitter feeds significantly increase the parties agenda-building power, compared with the impact of parties' news releases [8].

Analyzing the content of tweets in Twitter has also been done outside the journalistic view however performing a similar function. In the context of promoting biking, hashtags become the key point of Natural Language Processing (NLP) tools implementation of the study. Generally, NLP refers to the process of understanding how each word correlates to each other to produce a certain meaning, from the text grammar [11]. In this study, selected hashtags are determined in the first place to assemble the dataset associated with bike commuting. By performing data cleaning to the raw data from Twitter, the tweet frequency could be deduced by a text mining process. To further infer the meaning behind the tweets, sentiment analysis is conducted to determine the polarity of opinions regarding biking. The results show a majority of positive remarks. Biking communities and other related instances could then benefit from the result of the analysis to generate ideas in stimulating more people to bike by understanding the past biking motivations [12]. This also further elaborates how Twitter is a major platform to gain insights on the public's interests.

Another example of tweet classification, is used to analyze user feedback and its impact on a business. The analysis was based on influential term groups that are determined specifically for each different type of business. Then the relationship between the content of tweets and real-world outcomes could be established [13].

Processing data from Twitter means relying towards short texts as tweets only consist of very brief messages. This might be one of the causes of inaccuracy when making inference or classification based on tweets. Researchers on the study of Arabic Twitter users classification also concerned about the issue when performing classification using Twitter data [14]. There are several methods that could be implemented in reducing that by selecting the appropriate method. One of them would be a support vector machine or known as SVM (Support Vector Machine) [15]. In terms of NLP, SVM is one of the machine learning models that is often used. Category classification is the main purpose of SVM implementation [11]. In relation to journalism, trending topics can be classified using SVM. Another study has implemented SVM in classifying tweets into real news or rumour. The consideration that took part in the classification includes the content of the tweet, information about the user that posted the tweet and like numbers [16].

With the same context of this paper, a research has been done on news classification [17]. The news are classified based on religion, business, entertainment, law, health, motivation, sports, government, education, politics and technology as the chosen topics. Although different machine learning approaches are implemented in the research, overall, Naive Bayes Multinomial (NBM), as one of the algorithms for classifying texts, achieved the best performance with 77.47% of accuracy. However, the result of the news data classification only focuses on segmenting the news and does not further provide insight for news trend analysis.

By making use of the data from social media, trend analysis can be done further with the implementation of business intelligence through designing the data warehouse. The process of designing the data warehouse has been introduced in a study which consists of data preprocessing with the objective to clean the data obtained from two social media, Facebook and Twitter altogether, then determining the facts and dimension [18]. Dimension tables are results of denormalization and fact tables consist of the keys that lead to each dimension, along with fact attributes [19]. In that study specifically, before facts and dimensions are determined, as there are two origins from where the data is being collected, data mapping should be performed in ensuring a valid integration of data. The next step would be the transform and load process to complete the whole Extract, Transform and Load (ETL) procedure before proceeding into the data warehouse.

Another research involving the integration of business intelligence on social media analytics has been explored in the case of brand personality analysis [20]. The details about the brand followers, tweets related to the brand, reviews from the employee of the respective brand, tweets posted by the brand

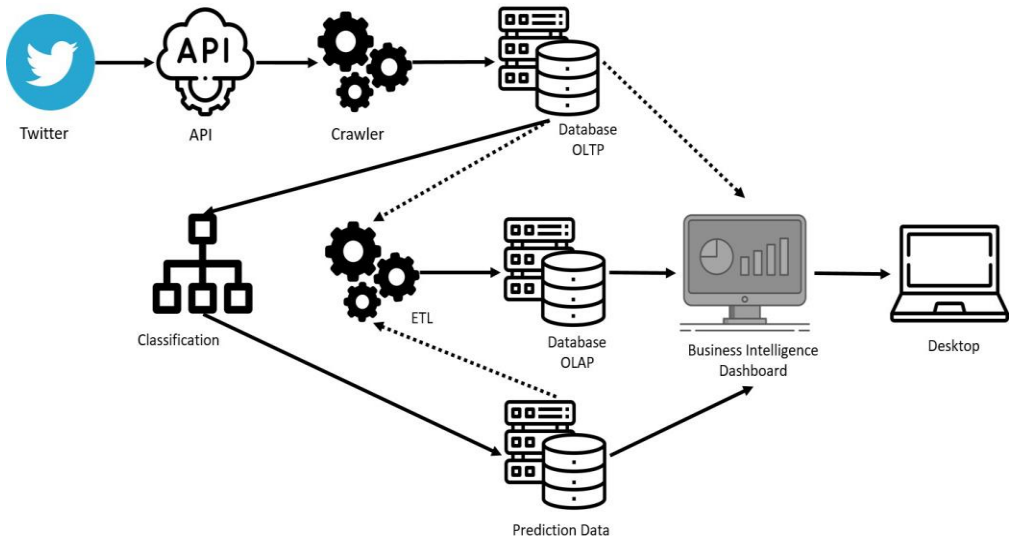


Figure 1: Business Intelligence Architecture

themselves comprise the dataset used in the research. Classification of the data into five brand personality categories supports the analysis on the chosen brands. The researchers suggested that decision making strategies can be improved with a sufficient business metrics measurement, in this context, gained from the activities associated with the brand on social media. This result could be achieved as trends are provided within social media data and useful for many types of organization [21]. Hence, classification of social media data incorporated with business intelligence could present more insight for enhancement in business strategies.

3. Proposed Method

In this research, a Business Intelligence application will be built into the architecture shown in Figure 1.

The following steps should be taken to develop the system as shown on the image above:

3.1. Data collection

At this data collection stage, there are processes including:

- Collecting data from selected Social Media Platforms such as Twitter through the Social Media Application Programming Interface (API) available on each platform.
- Data retrieval will be carried out periodically by crawlers who have been created with the Social Media API Token input which is useful for authentication and authorization in data retrieval.
- Thus, the results of data retrieval will be saved on the database as raw data.

3.2. Content Analysis

At this stage of content analysis there is an intermediate process including:

- Retrieving data for content analysis from each data on each Social Media platform.
- The data taken is then processed for text classification using SVM into 10 news categories.
- The results of text processing will be stored in the database as analysis data.

3.3. Data warehouse process

At this stage of the data warehouse process, there are processes including:

- Retrieving data for the ETL process that comes from the results of content analysis (in the content analysis database) and the results from the crawler (in the raw data database).
- The data that has been loaded will then be carried out by ETL, such as counting comments, posts, likes, and so on.
- Data resulting from the ETL process is then stored in the data warehouse.

3.4. Client Side

On the client side, there are several processes that are carried out so that the application dashboard can be accessed by users, including:

- API is used as a bridge between applications and data (both raw data and finished data in the data warehouse)
- The application will request data to the API by inputting the API Token as authorization and authentication, so that with this API people are not able to access data directly.

4. Analysis Results

4.1. Setting Variables

There are some variables used in this study as shown Table 1. In Figure 2 shown the ERD (Entity Relationship Diagram) from the transaction database of the scrapping system or can be called OLTP (Online Transaction Processing), before ETL (Extract, Transform, Load) process.

Table 1: Setting Parameter

Variable	Value	Description
Number of Twitter Account	10	Kompascom, Detikcom, BeritaSatu, MediaIndonesia, Tempo.co, BreakingNews, TMCPoldaMetro (Traffic Management Center Polda Metro), MNCNewsChannel (Media Nusantara Citra News Channel), TribunIndonesia and CnnIndonesia
Duration	20 days	Ranges from 27th November 2020 - 17th December 2020
Batch	8/days	Each batch has a three-houred range
Threshold Cosine	0.7	For measuring similarity of two comments
Threshold number comments	10	If number comments more than 10, it can be considered as top news

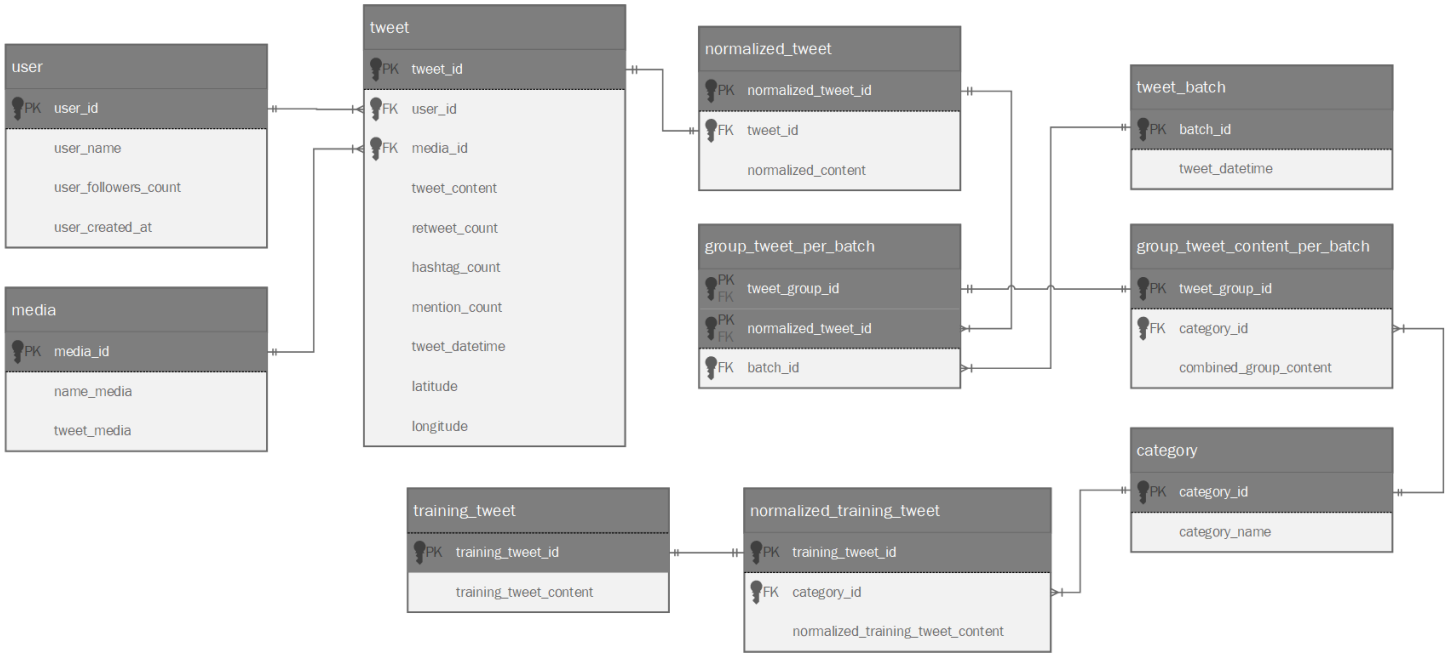


Figure 2: ERD

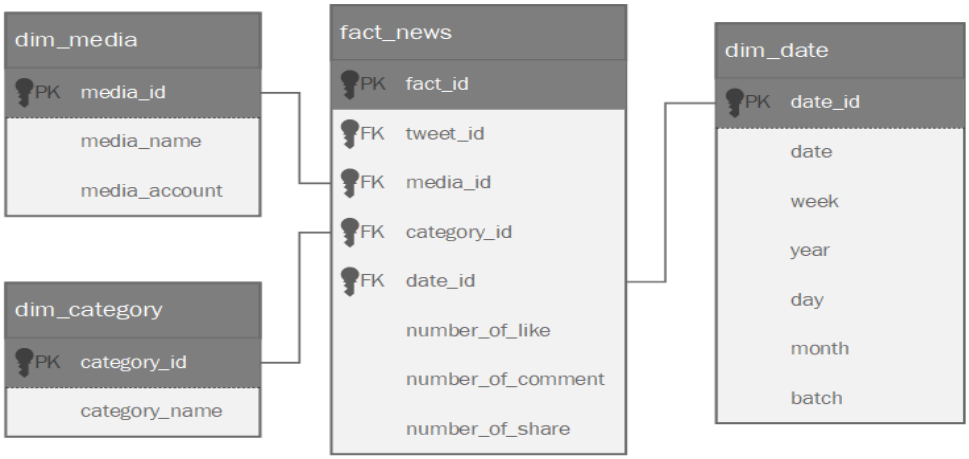


Figure 3: Star Schema

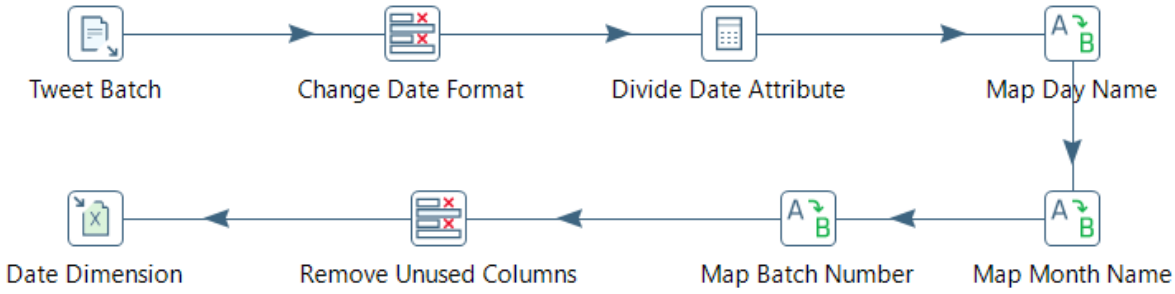


Figure 4: Date Dimension

4.2. ETL Process

In Figure 2 shown the Star Schema from OLAP (Online Analytical Processing), after ETL (Extract, Transform, Load) process. During the tweet extraction, when the time has entered the new batch, the datetime is directly inserted into the database in the server using a Python script. In order to generate the date dimension as shown in the star schema, the date obtained from the tweet extraction which started from 27th November to 17th December 2020 and shown as Tweet Batch data in Figure 3, is processed further using Pentaho. In Pentaho, the datetime format requires conversion as an error is encountered on the existing datetime from the database. Afterwards, the datetime is then divided into days, week, month, year, and batch. Week and year are represented in numbers, whereas days and month are mapped to their corresponding names. Batch itself is obtained from the hour of date, where it is then mapped into batch numbers ranging from 1 to 8. Date dimension is finally created after removing unused columns that are only used for processing purposes but not necessarily needed in the date dimension.

4.3. Performance Analysis

Each three hours, called a batch, data is evaluated. The batches are at 0-3, 3-6, 6-9, 9-12, 12-15, 15-18 and 21-24 having eight batches in total for each day as shown in Table 2. Table 3 showed tweets distribution for each category after classification with SVM. Table 4 showed the top three categories in each batch on the first day of extraction.

Table 2: Batch Time Detail

Batch	Created Time
1	00.01-03.00
2	03.01-06.00
3	06.01-09.00
4	09.01-12.00
5	12.01-15.00
6	15.01-18.00
7	18.01-21.00
8	21.01-24.00

Table 3: Tweets Distribution

Category	Number of news
Economy	11
Health	10
Entertainment	4137
Culinary	0
Lifestyle	317
Automotive	11
Politics	1991
Sport	0
Technology	284
Travel	0
Total	6761

Table 4: Distribution Top News in First Day

Day	Batch (Time)	Number top news (Based on Shares)
Day 1	Batch 1	Entertainment, Politics, Automotive
	Batch 2	Politics, Entertainment, Lifestyle
	Batch 3	Entertainment, Life, Politics
	Batch 4	Entertainment, Politics, Lifestyle

	Batch 5	Politics, Entertainment, Automotive
	Batch 6	Politics, Entertainment, Automotive
	Batch 7	Entertainment, Poliis, Technology
	Batch 8	Politics, Entertainment, Lifestyle

After all the data has been loaded to the data warehouse consisting of three dimensions, the data stored is transformed into a business intelligence dashboard. Figure 5, 6, 7, and 8 were shown as an overview and example of a business intelligence dashboard. Some insights that were taken based on chart analytics in the business intelligence dashboard in Figure 5, 6, 7, and 8 will be discussed further. Figure 5 and 6 shows the first part of the business intelligence dashboard which contains some analytics with higher quality information retrieval than Figure 7 and Figure 8, such as Total Number of Like, Comment and Share, Number of Share by Batch and Category, Number of Like, Comment, and Share by Category and Media, etc.

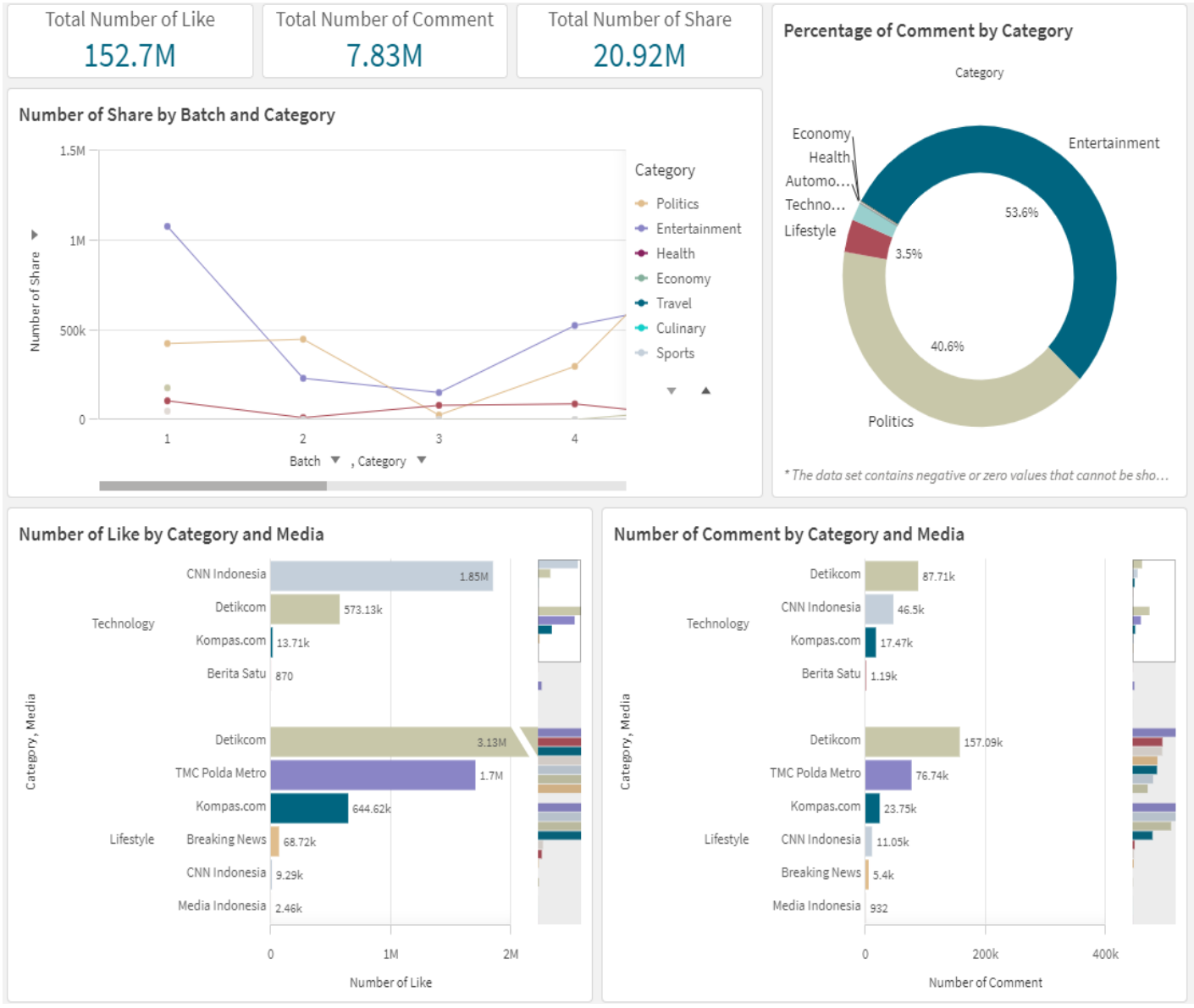


Figure 5: Business Intelligence Dashboard I

Figure 6: Business Intelligence Dashboard II

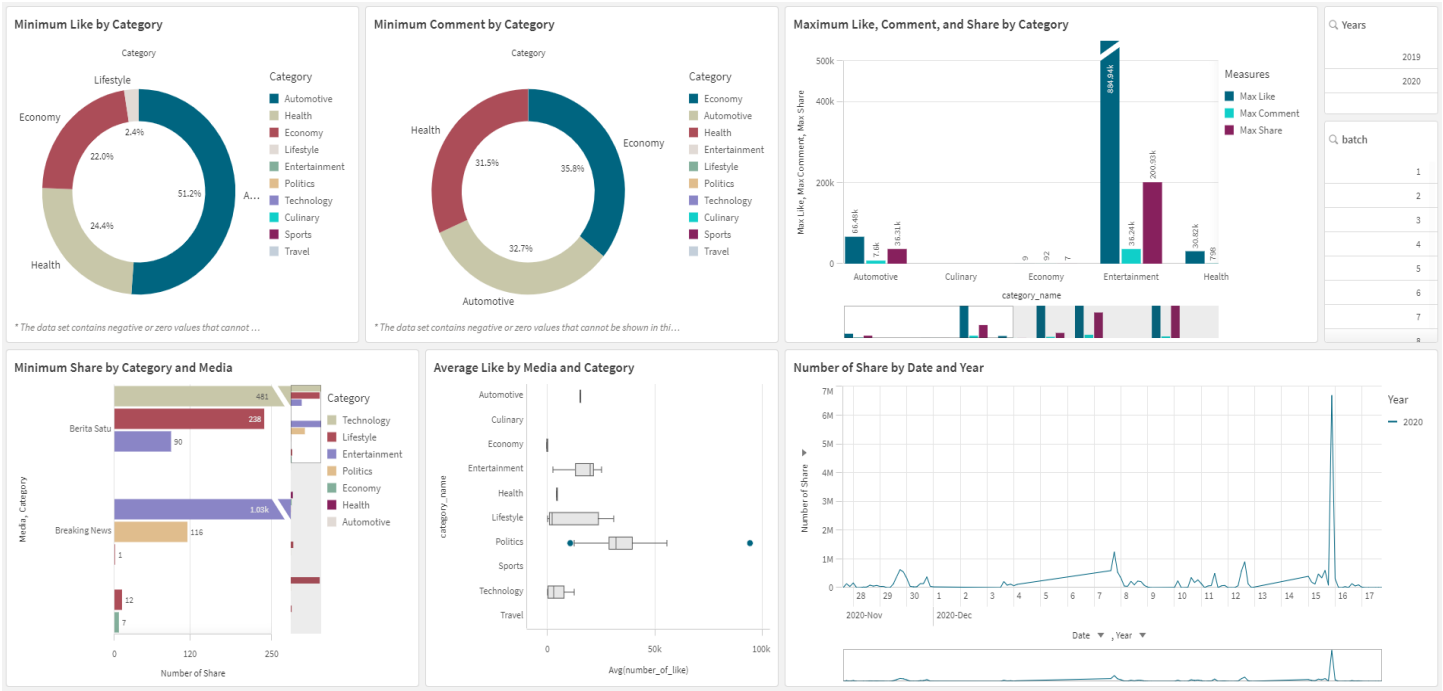


Figure 7: Business Intelligence Dashboard III

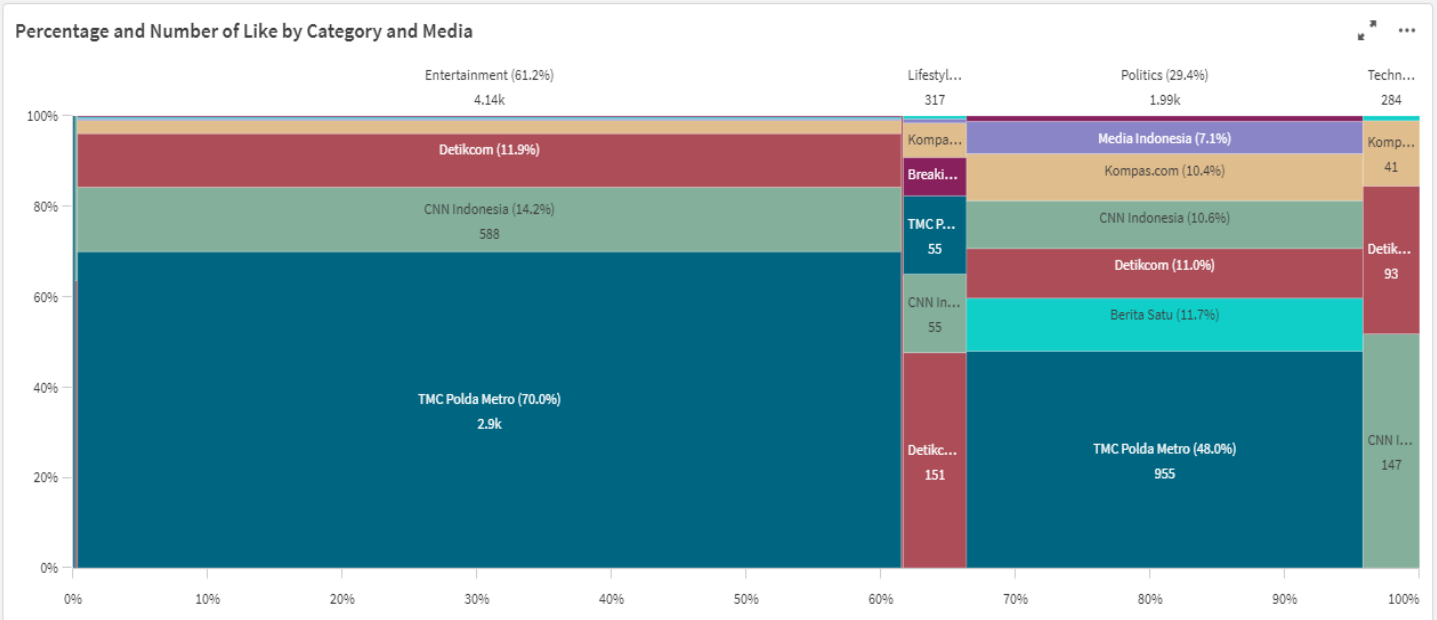


Figure 8: Business Intelligence Dashboard IV

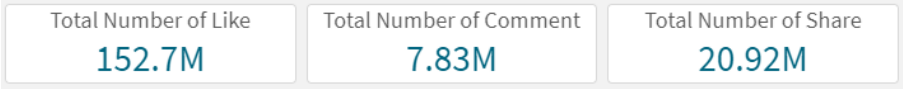


Figure 9: Total Number of Like, Comment, and Share

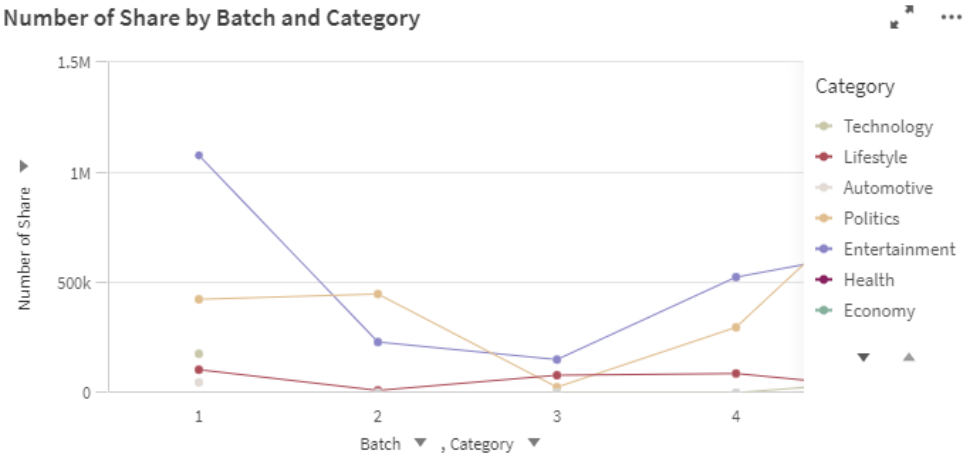


Figure 10: Number of Share by Batch and Category

Figure 9 shows a detailed screenshot on the total number of likes, comments and shares from the whole dataset. They represent the amount of engagement involved in Twitter for the specific period in 20 days. From the description in Figure 8, it could be seen that the total number of comments have the lowest value compared to the other two. This indicates how less frequently users leave replies on tweets they are interested in, whereas users mostly are prone to leave likes on a tweet as it requires less effort. Shares in another case require a moderate effort and in Figureure 10, it shows that entertainment is leading in batch 1. Data presentation in terms of batches, allows journalists to observe more about what kind of news attracts engagement of the public in specific time of day. Hence, it could provide understanding of what content type is at peak on certain hours, as well as how active users are participating in sharing information.

In terms of percentage, users have commented the most on tweets related to entertainment, politics and lifestyle. Considering the COVID-19 pandemic outbreak that started since the last month of 2019 and is still ongoing, the circumstances have impacted people's lifestyle, one of them would be staying at home more. This resulted in Twitter users behavior which is exploring more regarding something fun despite being mostly at home, as represented by the entertainment category in the top first position in terms of comment percentage shown in Figure 11. Politics category with a close comment percentage, is positioned second. During the specific chosen time period of tweet extraction, there are several political issues that users are intrigued about, leading

to higher interest in the topic. On the other side, the least percentage of comment is obtained from the economy category, indicating that the critical economic condition in the present is either quite a sensitive issue or nothing very significant, resulting in a fewer interaction between social media users.

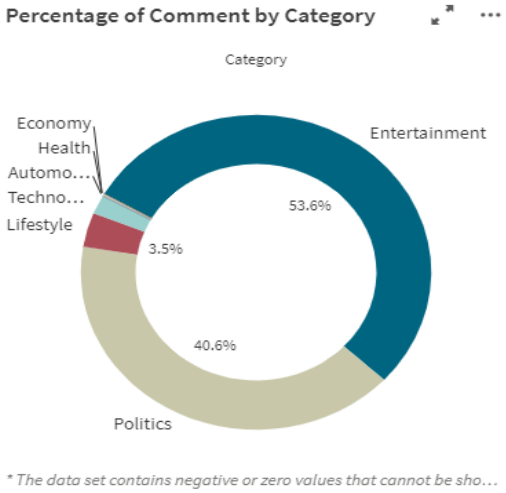


Figure 11: Percentage of Comment by Category

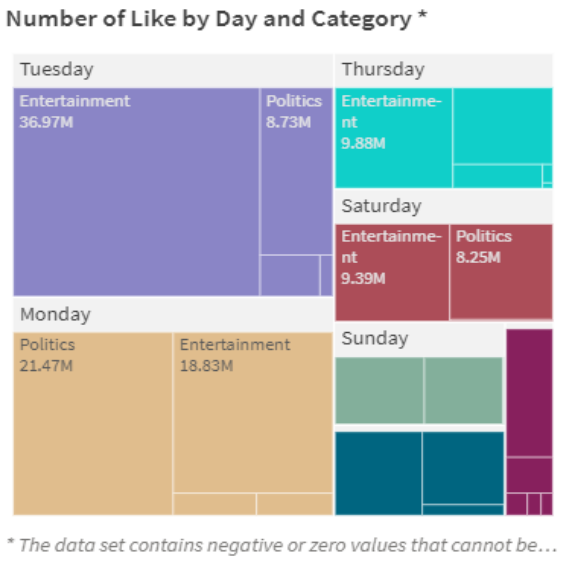


Figure 12: Number of Like by Day and Category

On a larger scale, the data can be further observed in terms of days. If batches tend to be very specific, days would display more generic data. This kind of data functions to assist and give insight on users' engagement routines. From Figure 12, on a daily basis, mostly weekdays, Twitter users frequently like tweets related to entertainment. This indicates that the users' daily interest is mostly fixed to the entertainment category. Hence, journalists can create more opportunities in searching for news that are able to raise high engagement from the public by being aware of entertainment updates during the specific period.

Number of likes within the politics category as shown in Figure 13 represented a quite unbalanced trend among the listed media. High engagement on politics category only appears to be most applicable for TMC Polda Metro, but specifically not much for Breaking News. There are two factors that allow this to occur: the number of tweets posted on that category or the attention that the tweet is able to acquire (by providing a more engaging tweet although the content is similar that triggers further response from other users, etc). Obviously if there are more tweets posted on the category, they will record a higher number of likes. However, this does not bring the possibility of having fewer tweets related to the category on the specific media, but still obtaining numerous likes down to the drain, as long as the tweets conform to the preferences of the users. In this case, Breaking News seems not to be focusing on the politics category. However, on the other side, it could be seen that most media did not cover much about the economy category. Based on this analysis, for more references regarding potential news on a specific category, journalists might be able to learn from the top media with the greatest number of likes, as the engagement towards a certain category does not apply to every media.

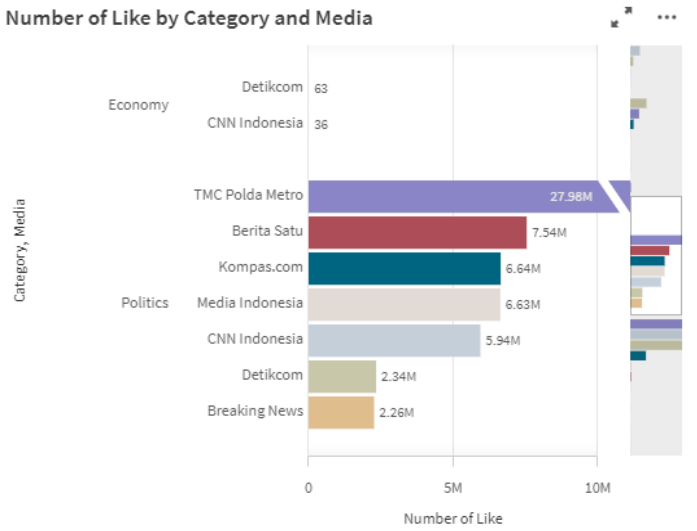


Figure 13: Number of Like by Category and Media

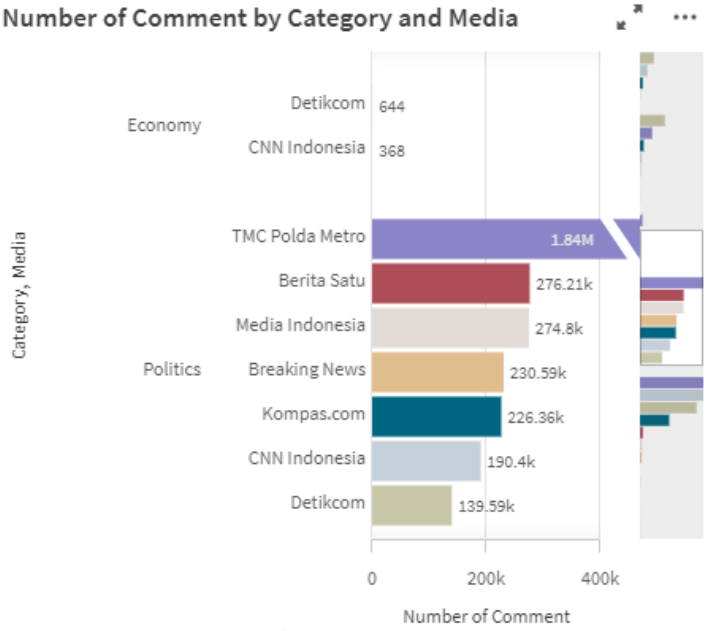


Figure 14: Number of Comment by Category and Media

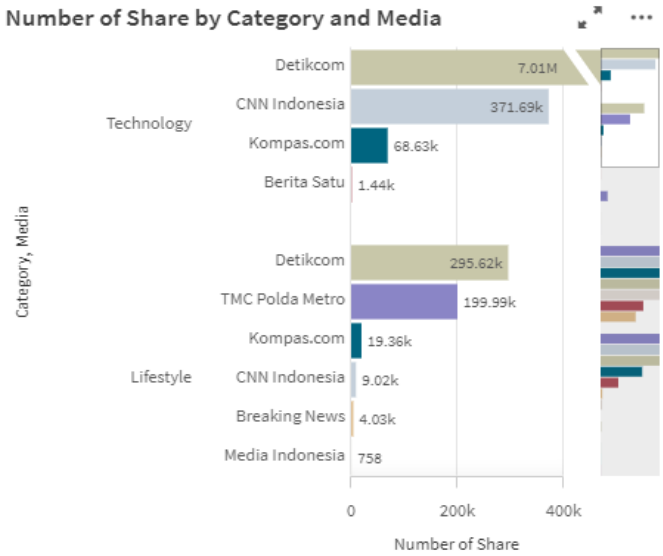


Figure 15. Number of Share by Category and Media

In Figure 14, it showed there is a significant difference between two categories for some media. The top Five Media and the others media in terms of the politics category, which was TMC Polda Metro, Berita Satu, Media Indonesia, Breaking News, and Kompas.com had the most commented tweet, which means people are more likely to reply and retweet these top five media. It also can be seen that at the certain time, there was no critical news related to the economy category. Comments can serve as discussion boards with unlimited sources and participants, limited by the need to validate information. Hence, journalists can dig and validate further information from the candidate news through comments. The comments will help provide more insight into news. If there was some big or high-quality news that occurred and uploaded to Twitter, it is speculated that the number of comments will be a balanced number for each media, either it will be a large amount of number or some range amount of number. Therefore, journalists can identify how worthy the news is but in more efficient and effective ways.

In Figure 15 can be seen the number of shares by category and media which can help decide trending news to look furthermore. Categories with a balanced number of shares in the chart have a bigger probability of containing a good quality and worthy of sharing news. With a lot of people sharing or retweeting the tweet, it concluded the news is well-known by a large number of both segmented users and other users on Twitter. Then, it will be useful for journalists to catch up nowadays trending topics in every category that is available.

Maximum like, comment, and share for each category will represent the most trending category discussed in society which can be seen in Figure 16. Showing the quantity data of the maximum engagement will be useful for journalists in defining the standard quality of news in the trending category news. At a certain time, Politics was the top trending category with 891.82k number of likes, 50.25k number of comments, and 394.71k number of shares. Therefore, when Politics hits the top trending in another certain time, the engagement's amount can be compared to gain comparison and insight. Over time, journalists can define the average of maximum engagement of the trending category which contains high-quality tweets that are worthy of sharing especially as news. Meanwhile, the lower limits of maximum engagement can also be defined by these circumstances. Hence, journalists will be able to choose the best quality of news in the category and process the news.

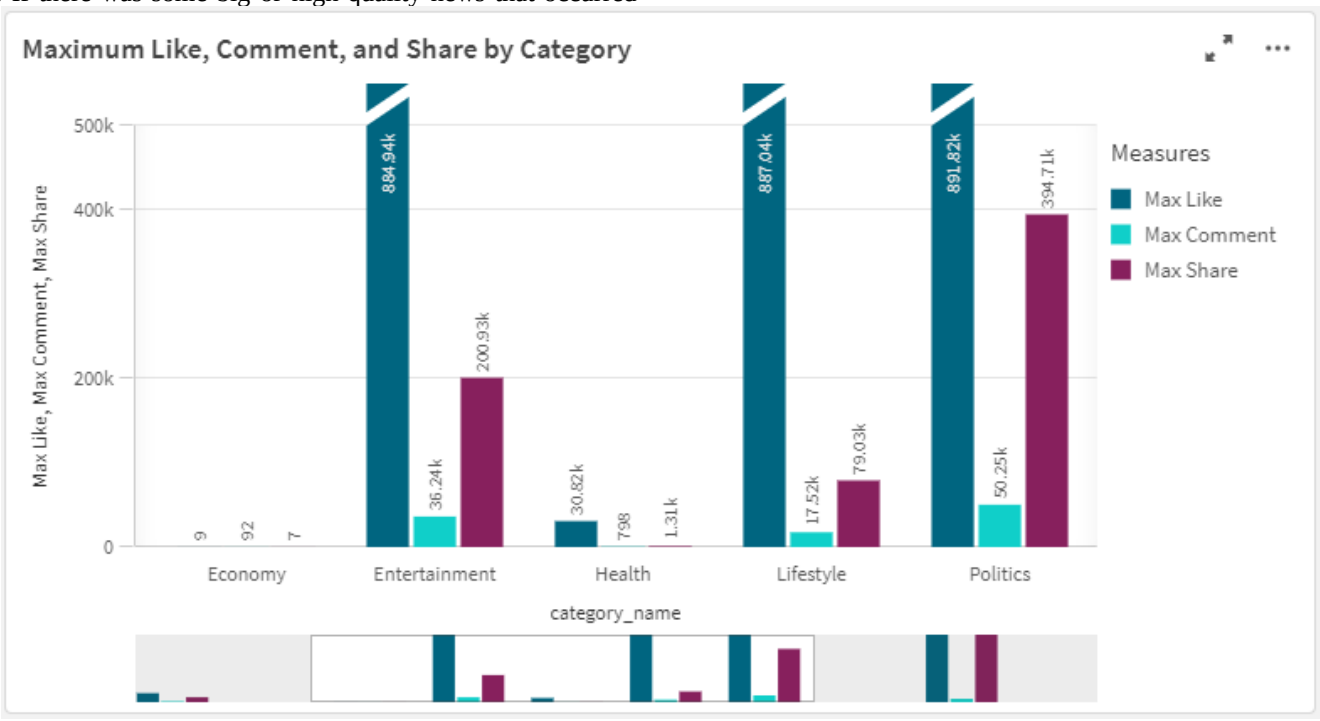


Figure 16. Maximum Like, Comment, and Share by Category

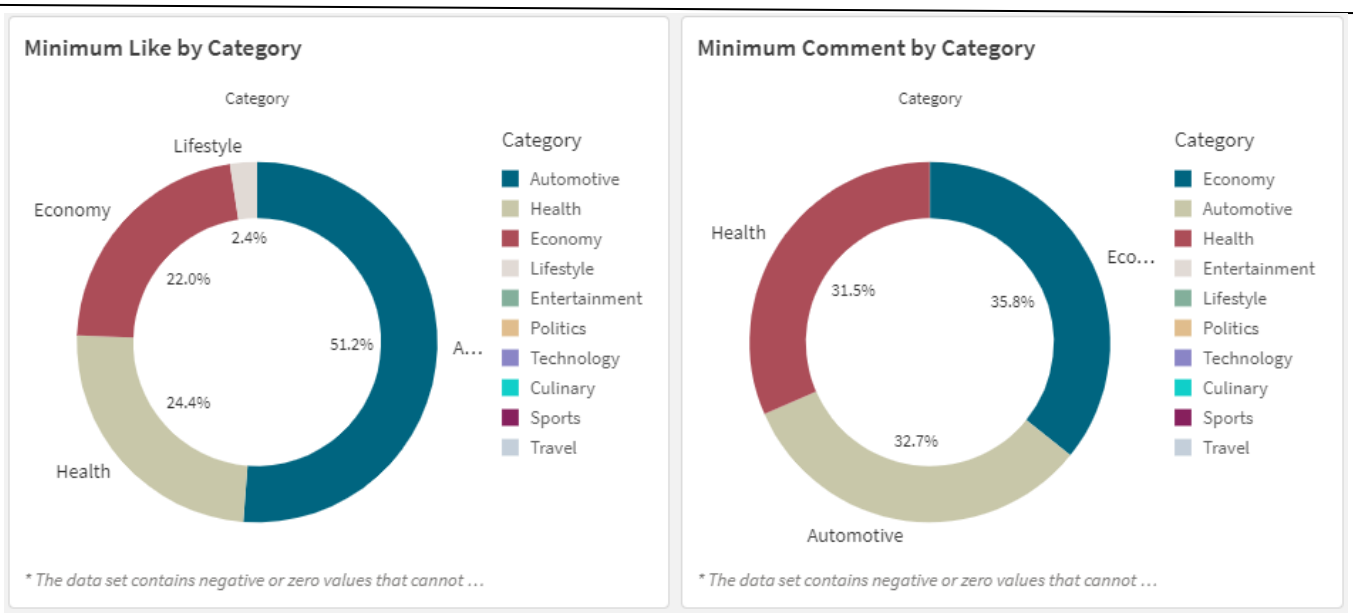


Figure 17: Minimum Like, Comment by Category

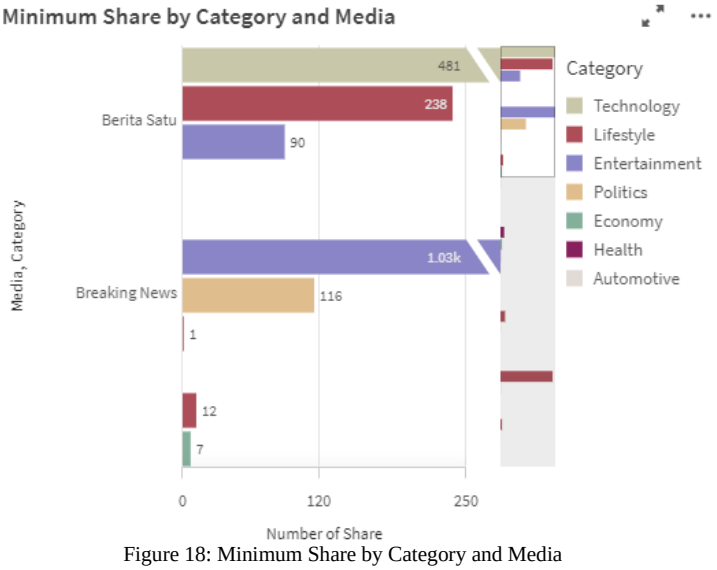


Figure 18: Minimum Share by Category and Media

Figure 17 and Figure 18 shows minimum count data of likes, comments, and shares by Category and Media which are able to help choose and exclude the certain category with the smallest percentage of likes and comments or smallest number of shares. With the standard defined, furthermore, the process of choosing the category and news will be easier. On the other hand, a big number of percentage and amount define that the certain category news was one of trending topics. In Figure 17 the entertainment category in Berita Satu minimum likes was 90, it can be seen there a large number difference to compare with the technology category. It concluded Berita Satu is not the right media to find entertainment news.

5. Conclusion

With processed and presented data at the dashboard, journalism will be helped a lot, for viral news information, through batch, date, day, news's category, news media and others. Through many data mining and other processes, Twitter was proven capable of providing and helping journalists with news sources and conveying the correct news to the public in effective and efficient ways. This process will help and give the journalism world a solution for effective, efficient, and unlimited news sources.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

This work is supported by the Directorate General of Strengthening for Research and Development, Ministry of Research, Technology, and Higher Education, Republic of Indonesia, as a part of Penelitian Dasar Unggulan Perguruan Tinggi Research Grant to Binus University titled "Sistem Penugasan Jurnalis Berdasarkan Trafik Media Sosial" or "Journalist Assignment System Based on Social Media Traffic" with contract number: 25/E1/KPT/2020, 225/SP2H/LT/DRPM /2019, 088/LL3/PG/2020, 039/VR.RTT/IV /2019.

References

[1] H.S.N. Al-Deen, J.A. Hendricks, Social Media: Usage and Impact, 307, 2012.
[2] T. Team, Twitter turns six, Twitter Blog, 21–23, 2012.
[3] W. Stassen, "Your news in 140 characters: exploring the role of social media in journalism," Global Media Journal African Edition, **4**(1), 116–131, 2011, doi:10.5789/4-1-15.
[4] S.C. McGregor, L. Molyneux, Twitter's influence on news judgment: An experiment among journalists, Journalism, **21**(5), 597–613, 2020, doi:10.1177/1464884918802975.
[5] S.C. Lewis, L. Molyneux, "A decade of research on social media and journalism: Assumptions, blind spots, and a way forward," Media and Communication, **6**(4NewsandParticipationthroughandbeyondProprietary), 11–23, 2018, doi:10.17645/mac.v6i4.1562.
[6] P. Ferrucci, Networked: Social media's impact on news production in digital newsrooms, Newspaper Research Journal, **39**(1), 6–17, 2018, doi:10.1177/0739532918761069.
[7] P. Pradhan, N. Kumari, "A study on Journalistic use of Social Media," Amity Journal of Media & Communication Studies, **8**(1), 49–59, 2018.
[8] J. Seethaler, G. Melischek, Twitter as a tool for agenda building in election

campaigns? The case of Austria, Journalism, **20**(8), 1087–1107, 2019, doi:10.1177/1464884919845460.

[9] B.A. Conway, K. Kenski, D. Wang, "The Rise of Twitter in the Political Campaign: Searching for Intermedia Agenda-Setting Effects in the Presidential Primary," Journal of Computer-Mediated Communication, **20**(4), 363–380, 2015, doi:10.1111/jcc4.12124.

[10] E.L. Frederick, L.M. Burch, M. Blaszka, A shift in set: Examining the presence of agenda setting on twitter during the 2012 london olympics, Communication and Sport, **3**(3), 312–333, 2015, doi:10.1177/2167479513508393.

[11] P.M. Nadkarni, L. Ohno-Machado, W.W. Chapman, "Natural language processing: An introduction," Journal of the American Medical Informatics Association, **18**(5), 544–551, 2011, doi:10.1136/amiajnl-2011-000464.

[12] S. Das, A. Dutta, G. Medina, L. Minjares-Kyle, Z. Elgart, "Extracting patterns from Twitter to promote biking," IATSS Research, **43**(1), 51–59, 2019, doi:10.1016/j.iatssr.2018.09.002.

[13] S. Lim, C.S. Tucker, "Mining Twitter data for causal links between tweets and real-world outcomes," Expert Systems with Applications: X, **3**(July), 100007, 2019, doi:10.1016/j.eswax.2019.100007.

[14] A. Alabdullatif, B. Shahzad, E. Alwagait, "Classification of Arabic Twitter Users: A Study Based on User Behaviour and Interests," Mobile Information Systems, **2016**(January 2016), 2016, doi:10.1155/2016/8315281.

[15] F. Kateb, J. Kalita, "Classifying Short Text in Social Media: Twitter as Case Study," International Journal of Computer Applications, **111**(9), 1–12, 2015, doi:10.5120/19563-1321.

[16] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, R. Procter, "Detection and resolution of rumours in social media: A survey," ArXiv, **51**(2), 2017.

[17] J.E. Sembodo, E.B. Setiawan, M.A. Bijaksana, "Automatic tweet classification based on news category in Indonesian language," 2018 6th International Conference on Information and Communication Technology, ICoICT 2018, (May), 389–393, 2018, doi:10.1109/ICoICT.2018.8528788.

[18] I. Moalla, A. Nabli, L. Bouzguenda, M. Hammami, "Data warehouse design from social media for opinion analysis: The case of Facebook and Twitter," Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA, **0**, 2016, doi:10.1109/AICCSA.2016.7945627.

[19] M. Golfarelli, D. Maio, S. Rizzi, "The dimensional fact model: A conceptual model for data warehouses," International Journal of Cooperative Information Systems, **7**(2–3), 215–247, 1998, doi:10.1142/s0218843098000118.

[20] Y. Hu, A. Xu, Y. Hong, D. Gal, V. Sinha, R. Akkiraju, "Generating Business Intelligence Through Social Media Analytics: Measuring Brand Personality with Consumer-, Employee-, and Firm-Generated Content," Journal of Management Information Systems, **36**(3), 893–930, 2019, doi:10.1080/07421222.2019.1628908.

[21] C.A. Pushpam, J.G. Jayanthi, "Overview on Data Mining in Social Media," International Journal of Computer Sciences and Engineering, **5**(11), 147–157, 2017, doi:10.26438/ijcse/v5i11.147157.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1529-1536 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Prophet Architecture in Normalized Meter Energy Consumption Prediction on Building

Fernando Lioexander, Abba Suganda Girsang*

Computer Science Department, BINUS Graduate Program-Master of Computer Science Bina Nusantara University, Jakarta, 11480, Indonesia

ARTICLE INFO

Article history:
Received: 11 November, 2020
Accepted: 12 December, 2020
Online: 21 December, 2020

Keywords:
Forecast
Nmec
Prophet
Time Series

ABSTRACT

Normalized Metered Energy Consumption (NMEC) is a solution for investors in determining the best energy-saving strategy for buildings. But on the other hand, investors need a fast and reliable evaluation results in measuring how effective the savings methods they use without wasting money. To address this issue, we selected Facebook's latest predictive time series method called Prophet Algorithm's which adapts the regression modular additive model with hyperparameter advantages that can be optimized based on time series parameters by automation. Although the Prophet is a relatively new method in time series predicition, but it can show promising results that offer even more easily implemented by beginners for business purposes. Furthermore, it allows gaining insight into each periodic component of the forecast separately and helping to assess energy management issues. At the end of the experiment, the Prophet model achieved an excellent result by showing the Root Mean Square Error (RMSE) below 10 points for the next two-month forecast.

1. Introduction

Energy wastage is still an enormous problem around the world, even about 40% due to inefficient city building energy management systems [1,2]. Challenges in the recent energy industry are cost-effective deficiencies, model inaccuracy, and lack of prediction procedures determining the use of energy that can be measured periodically to make significant decisions [3]. Investors have invested a lot of money to improve sustainable building's energy efficiency to reduce costs and emissions. The question is whether energy efficiency has worked success effectively reducing energy consumption to the maximum? Then the next question is how effectively the method works in the influence of various external factors? In previous research on energy efficiency, it concluded that the best strategy is not only to apply energy-saving technology. But energy management strategies can also be an effective solution in minimizing energy disposal [4].

Energy benchmarks are still a consideration for building owners to be able to measure the energy consumption performance of building with similar structures to different geographical or weather conditions. Based on the results of energy benchmark analysis, building developers can implement appropriate and proven energy-saving methods by the best practice of other buildings. The core keys to knowing the maximum energy savings are by determining the energy demand model before the development action (baseline model) and determining the energy demand prediction model after the implementation of the energy-saving method (report model) [5]. Meanwhile, the new energy benchmark introduced in California Energy Policy for verifying the success of energy-saving in 2017, namely NMEC. The NMEC considers the weather and building characteristics to normalize energy consumption data. Normalization indicates a statistical process for adjusting energy use before and after energy management improvements in a wide range of conditions related to independent variables for more sensible results that can describe actual energy consumption patterns [6].

Previously researchers successfully used the Hybrid model Artificial Neural Network (ANN) with an average Mean Absolute Percent Error (MAPE) value of 4.2% [7]. On the other hand, the model still has limitations where it can give a decent prediction for only the next 72 hours. Another research uses LSTM performs well in terms of accuracy, but it requires more data compared to other methods [8]. Until now, various algorithms have been explored widely for predicting building energy consumption. However, many external factors such as holidays or high season trends that change occasionally in many cases, which cannot capture by models. And it's hard practically to produce feasible analysis results for many business users.

In this research, the authors had the idea to implement the Prophet method to predict building energy use based on historical

*Corresponding Author: Abba Suganda Girsang, Email: agirsang@binus.edu

trends that address dynamic business needs and require time efficiency. Prophet demonstrates higher accuracy and ease of implementation to see movement trends periodically for business purposes with flexible requirements after studying the evaluation results in several previous studies such as Autoregressive Integrated Moving Average (ARIMA) [9], and Seasonal Trend Decomposition using LOESS (STL), ARIMA, Neural Network Autoregressive (NNAR), Trigonometric Exponential Smoothing State-Space model with Box-Cox transformation (TBATS), Linear, S. Naïve in [10]. Prophet can also effectively solve the problem of easy forgery using real-time prediction of the user behavior to determine the right time to switch on the security sensor on the smartphone [11].

We believe that there are still fewer applications for prophet methods, especially in predicting energy consumption in buildings. Although it is still new development methods, the prophet method can surpass various other approaches with the concept of an additive model to estimate predictions based on historical trend patterns. For further development, the authors combine affecting normalization techniques to improve the quality of the model. The purpose of the research will later contribute to producing baseline models that can minimize external effects (e.g., weather, building characteristics, etc.) as an accurate evaluation model of building energy savings in the future. And it could be a consideration for investors to prepare a strategy for the implementation of high-value prolonged energy savings.

2. Literature Review

2.1. Energy Consumption Performance Indicators

Factors that affect energy consumption in buildings generalized into 7 categories [12]:

- Weather (e.g., air temperature, wind speed, solar radiation, humidity),
- Building characteristics (e.g., building functionality, square area, number of floor levels),
- Characteristics of occupants (e.g., presence of residents on weekdays and holidays),
- Building service system (e.g., cooling/heating conditions, availability of water heaters),
- Activities and habits of residents (e.g., turning off lights during the day),
- Social and economic factors (e.g., education level, energy cost), and
- Needs for indoor air temperature.

The behavior and habits of the occupants are quite difficult parameters to identify regarding the complexity and differences of each occupant observed [12]. One of the most possibles solutions is to use real measurement data such as historical data on KWH electricity usage of the previous period. Based on four energy consumption performance methodologies on buildings explained in [13], we chose to use statistical analysis (regression model-based) as the best option to overwhelm the time limitation problem. Statistical analysis methodology refers to the measurement of Energy Performance Intensity (EPIs) in buildings derived from a combination of supportive variables with robust relationships to energy usage.

2.2. Structure Dependent Energy Usage /Loss (SDE U/L) Overview

Energy consumption estimates in buildings usually consider dry-bulb temperature, known as normalization of weather Heating and Cooling Degree Days (HDD/CDD) [14]. HDD/CDD is the standard term of measurement units for conducting weather modeling. However, this method can provide misleading solutions because unable to keep up with the dynamics of extreme weather changes by limited inputs and ratio concept. The development of this method continued to be carried out throughout until recently the SDE U/L method introduced to the public. The SDE U/L method uses the concept of linear or nonlinear models considering the coefficients of wind speed, humidity, radiation, and outdoor temperature. Moreover, several important building factors such as building size, window size, construction connection that ignored in the application of various previously normalization methods [15]. SDE U/L consists of 3 main factors, namely weather, characteristics of buildings, and behavior or habits of building occupants.

$$g_{bl}(T, W, R,H, P)=a+aT+bR+cW+dH+eP, \quad (1)$$

In steady-state, (1) represent linear models to calculate the heat balance of the building denoted by $g_{bl}(\cdot)$ function. The equation made base on heat/cool generated by building appliances that improve overall electricity use. Coefficients *a*, *b*, *c*, and *d* are parameters that depend on the building's characteristics. Parameter *a* relates to insulation, building size, etc. that affect air temperature. Parameter *b* associates to the direction of entry of sunlight, the size of the window, the material of the window that affects the intensity of radiation. While parameter *c* is related to air ventilation, the height of the building, the open space that affects wind speed. And parameter *d* relates to the infrastructure that affects the humidity of the room. Parameter *e* could consider as the number of residents of the building. While non-linear models have the same concept as linear models, but the application of methods developed with Multi-Layer Perceptron. Where each input node consists of T, R, W, H, P, and when the data do not exist, then the node input will be omitted. The model is optimally attached when the RMSE result is smaller than 10^{-10}.

2.3. Prophet Overview

The prophet method concept is adapting the Generative Additive Model (GAM) method that emphasizes regression models with non-linear smoothers on the regressor [16]. As a difference, Prophet only uses time as a regressor when there is the possibility of some linear and non-linear functions of time as essential components. Prophet can generally meet the main criteria of good time series models described further in [17] such as forecasting, modeling, and characterization. One of the advantages adapted from the GAM model is the flexibility in accepting trend changes in the data source and quickly in the process of model fittings so that users can change the parameter model interactively to improve accuracy results [18]. Prophet's model base on the concept of a decomposable time series with three main components: trend, seasonality, and holidays that represent in (2) [19]. The additive regressive model result of

combination decomposable time series denotes as $y(t)$, which will describe per movement how the Prophet model fits data distribution.

$$y(t)=g(t)+s(t)+h(t)+\varepsilon_t, \quad (2)$$

In (2), $g(t)$ describes the trend of non-periodic changes in the time series model. $s(t)$ describes periodic changes such as weekly, yearly. And $h(t)$ depicts the continued effect of holidays on a nation that has the potential to cause unusual trends and sometimes effects occurring over one day. Notation errors are distinct changes that the model cannot predict. The fundamental in a decent non-linear model is to make sure the model can describe the growth trend and how the continuation corresponds to the addition of data. With dynamic trend movements, the basic approach used is the logistic growth model follows in (3).

$$g(t)= \frac{C}{1+exp(-k\,(t-m))}, \quad (3)$$

C represents carrying capacity, k is the growth rate, and m is the offset parameter. Carrying capacity is the maximum number of individuals that can be supported resources in an ecosystem. One of the modifications implemented in the Prophet method is to define changepoints, which variable C changes to $C(t)$ defines carrying capacity expected at any given time. Then Prophet defines a vector containing a rate adjustment of $\delta \in R^S$, where δ_j is the change in the rate at the time of s_j. Then base k will be couple with adjustment at that point: $k + a(t)^T\delta$. Where vector $a(t)$ - $\{0, 1^S$ is defined in (4) to limiting that the k rate increase will only occur when the t measurement time exceeds the changepoint time occurred.

$$a_j(t)=\begin{cases} 1\ ,if\ t \geq s_j \\ 0\ ,otherwise, \end{cases} \quad (4)$$

The k rate variation is also followed by an offset parameter change to connect the endpoints of the segments. The correct adjustment at changepoint $\boldsymbol{j}$ is easily computed as and presented as follows in (5):

$$\gamma_j=\left(s_j-m-\sum_{l<j}\gamma_l\right)\left(1-\frac{k+\sum_{l<j}\delta_l}{k+\sum_{l\leq j}\delta_l}\right) \quad (5)$$

The result of modification of prophet model is presented in (6):

$$g(t)= \frac{C(t)}{1+exp(-k+a(t)^T\delta\,(t-(m+a(t)^T\gamma)))}, \quad (6)$$

In determining the changepoint on the model, the analyst can perform manually by defining a specific potential signifiant date as an example of a national holiday. Or automatic by applying the principle of historical trends to variable δ and sets the flexibility rate in the model by parameter τ. For the record, the application of historical trends in δ will not influence the central point growth rate of k. Thus, when the parameter τ value is 0, the curve of the fitting model will be linear or basic logistics. Accuracy in the prediction model becomes the final parameter as the evaluation material, but the error value of uncertainty would still occur due to the unconstant growth rate. Uncertainty calculates assuming that future predictions will have an average pattern of frequency and enormous changes in growth rates in history. To increase accuracy, we can configure the τ parameter based on the variant result of the data. The increasing value of the τ parameter will also increase the flexibility of the model fitting to historical patterns and reduce the error loss value. However, high flexibility is directly proportional to the increased interval of uncertainty.

In the Prophet model, the seasonal component $s(t)$ provides adaptability by permitting periodic changes in a different kind of seasonality. There are many cases of multi-period seasonality as a capture result of human behaviors. For instance, a 5-days workweek implies effects on a time series that repeat each week for over a month, while vacation schedules at the weekend and school breaks can generate a trend that repeats each year. To fit and forecast these effects we must specify seasonality models that are periodic functions of [time] t. Prophet relies on the Fourier series to provide a lenient model of periodic effect. P is the regular period the time series will have (e.g., P = 365.25 for yearly data or P = 7 for weekly data when time scale in days).

The impact of a particular holiday on the time series is often similar year after year, making it important incorporation into the forecast. The component $h(t)$ indicates predictable events of the year including those on irregular schedules (e.g., Halloween or Labor day). To utilize this feature, the user needs to provide a custom list of events. Fusing this list of holidays into the model is made straightforward by assuming that the effects of holidays are independent. After the model fitting process, the model will make predictions based on trends and seasonality that successfully captured previously in (1). Trend assumptions are based on the same date in the previous period. The output of the model will be yhat variables that indicate the predicted results. And another component yhat_lower and yhat_upper, which describes uncertainty intervals of prediction.

2.4. Related Works

In this section, it will explain in more detail about the methods used along with the evaluation results obtained by the previous research. Previously researchers used four approaches as a comparison to create baseline models of energy prediction such as Support Vector Machine (SVM), Long Short-Term Memory (LSTM), Nonlinear Auto-Regressive (NAR), Autoregressive Moving Average (ARMA) [8]. It could conclude that each method tested can exceed ARMA performance. While LSTM performs well in terms of accuracy, it requires more data compared to other methods. Hybrid model Artificial Neural Network (ANN)-based Short Term Load Forecasting (STLF) models with a combination of boolean metering systems to convert input values into vector bit shapes carried out in [7]. The experiment conducted using energy consumption data on 93 homes in Portugal from 2000 to 2001. The experiment proved a pretty good result with an average Mean Absolute Percent Error (MAPE) value of 4.2% as well as a maximum MAPE of 18.1%. In the evaluation results, there are limitations where the model can give a decent prediction for only the next 72 hours. They suggest further development with variable use of air temperature and weather and consider the presence of residents by distinguishing routines on weekdays and weekends to improve the accuracy of

prediction models. Other research with the concept of short-term load forecasting for the next 24 hours had a comparison between ANN and SVM methods, with an accuracy rate of 62% for ANN and 60% for SVM [20].

Another comparison method based on a data-driven approach on two buildings to get predicted results of energy consumption per hour meet the model calibration criteria defined by the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE). Various methods such as multiple linear regression, adaptive linear filter algorithms (least mean square (LMS), normalized least mean square (nLMS), and recursive least square (RLS)), and Gaussian mixture model regression (GMMR) used in this experiment [21]. Once evaluated, the GMMR method can go beyond a variety of other adaptive approaches. Cause GMMR is a non-linear method so it can better show the performance of building energy consumption. For an additional record, the data-driven method's large number of parameter inputs does not ensure better accuracy. It is due to the model's difficulty in achieving a convergence state in finding the best solution due to the combinatorial explosion problem. Table 1 shows the summary of related works, together with the dataset, method, and achieved experiment results.

Table 1: Summary of Related Works

Ref	Dataset	Method	Result
[7]	Energy Consumption on 93 Homes in Portugal	Hybrid ANN	Gives an average MAPE around 4.2 %.
[8]	Power Consumption for a Single Household.	SVM	NAR worked best for short time error variance progress while SVM best for long prediction intervals.
		LSTM	
		NAR	
		ARMA	
[20]	Power Consumption for a Single Household in Poland.	ANN	ANN worked best with an accuracy around 62 % for short-term load forecasting.
		SVM	
[21]	Buildings Energy Consumption.	MLR	GMMR give best result on fitting building energy consumption model
		LMS	
		nLMS	
		RLS	
		GMMR	

3. Experimental Design

In this section, we describe the framework that we used to perform data normalization until the implementation of prediction methods for a group of buildings with the same characteristics as the comparison with other groups. Moreover, we explain the steps and configuration parameters for the numerical assessments of Section 4.

3.1. Dataset

The data in this experiment consists of 3 datasets: weather, electrical consumption of buildings, and building characteristics. Dataset acquires from meter measurements of more than a thousand structures across the United States [22]. The data consist of measuring water, electricity, gas, hot water, and steam meters. ASHRAE's 1000-building electrical energy consumption data has an hourly interval with a measurement period from 2016 to 2017, thus meeting the regularly spaced criteria in the time series. Once the data are collected, the data going through the selection phase since some buildings do not have full sensors or there might construction conditions in certain areas, possibly making the sensor measurement value 0.

In ensuring a valid measurement where the building is inhabited, then the maximum missing data is set at 15%. Besides, missing values are replaced with a value of -999 or with interpolation automation techniques based on data of the previous and next day. To eliminate outliers on data, the plot analysis method used in looking at the trends of each district's building sensors. In this simulation, one building was randomly selected with an N/A value of less than 10 percent to provide the best results. The historical movement of energy consumption over a year with hour intervals could be seen in Figure 1.

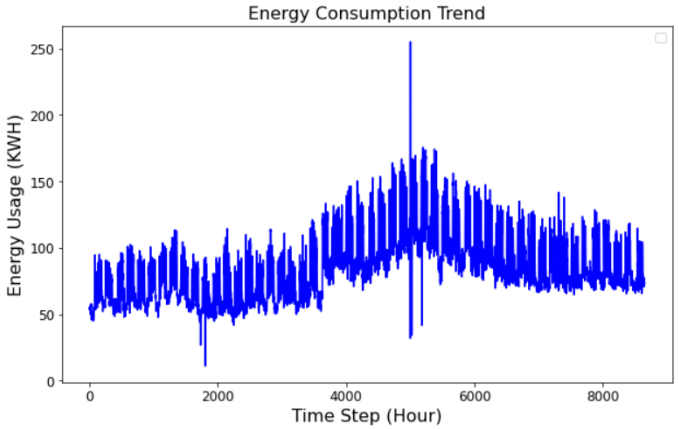


Figure 1: Electrical Energy Consumption Trend on Building with Id 1082.

3.2. Weather Normalization

Then in the stage of weather normalization, implementation is carried out by SDE U/L method to consider meter measurement with weather conditions as well as building characteristics as inputs. Normalization has done using a multi-layer perceptron with configuration four input nodes, two hidden layers, and one output node. For the normalization process, the base temperature sets to 9.6°C, which refers to the average temperature of the building area on measured year. The configuration of base variables used in normalization could be different depends on the needs of research in analyzing the most influential weather factors for the development of energy-saving methods in the future. The trend of electricity consumption after normalization could be seen in Figure 2. Differences in radiation normalization results through the same data could be seen in Figure 3. Because there is no additional information about the behavior of the occupant, thus the input node in parameter P be ignored on the model. For the

final training model configuration, we use two hidden layers with each consisting of seven nodes and a learning rate of 0.04.

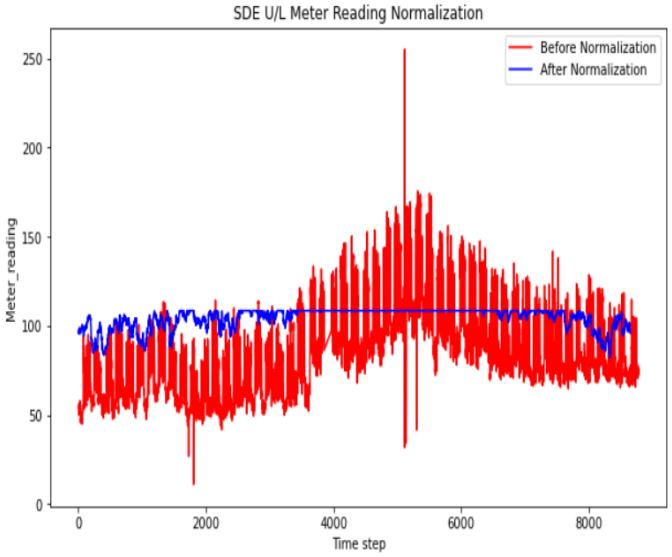


Figure 2: Temperature Base Normalization Implementation.

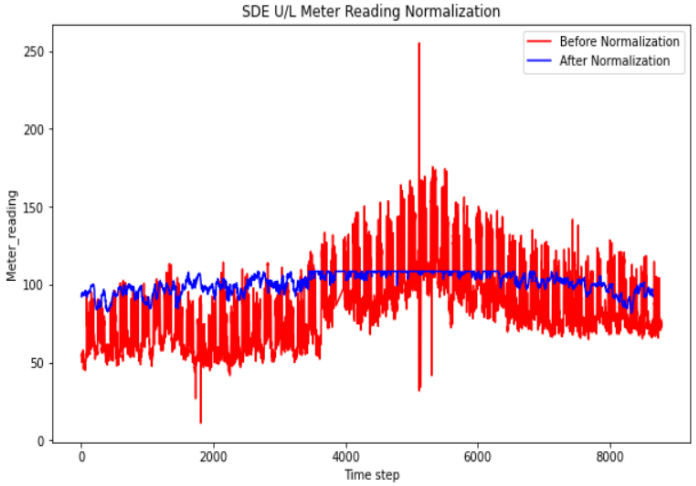


Figure 3: Radiation Base Normalization Implementation.

3.3. Prophet Framework

Preparation of the prophet prediction model carried out in the following stages:

- Split data into three sets, training set, validation set, and test set. The proportion we used in splitting the dataset into training, validation, and test sets is 80%, 10%, and 10%, respectively as a reasonable option given the sufficient amount of data [23]. Training sets were used to train models, while validation sets were used for tuning hyperparameters, and test sets were used for evaluation of models.
- Initialize the carrying capacity value for each row of data presented in the stamp column. The stamp value is not continuously adjusting to the market size at the time of measurement. The market size here refers to the maximum amount of energy consumption per hour. Therefore, carrying capacity plays a role to keep the curve of the model does not exceed the market size value specified before. For example, in the winter season (December - February), the carrying capacity value is determined based on the maximum value of energy consumption between those three months. For the implementation of carrying capacity, minimum value can be used floor variable.
- Defines specific dates such as national holidays that repeat periodically for each year in the form of data frames like Christmas, labor-day, etc.
- Add custom seasonalities such as school holidays in the long-term summer holidays for up to 3 months.
- Change the changepoint ratio on the model to set the flexibility of the model. By default, change points have automatically added to 80% of trends with the reason for creating more projection lines so that there is no overfitting at the end of the prediction. But to be aware, adding a changepoint ratio will make the model have too much flexibility and have overfitting side effects. The default changepoint ratio value in the Prophet method is *0.05*.
- Apply fouries series yearly seasonality tuning as a smoothing method when trend patterns have high fickle frequencies. For example in tropical conditions, the weather changes rapidly.
- Divide the data into 2 columns, ds for the date (timestamp) and y for the feature column which is the result of normalization of building energy consumption as input in the model.
- Repeat steps e through f to find the best model through a final evaluation using RMSE.

3.4. Evaluation Metrics

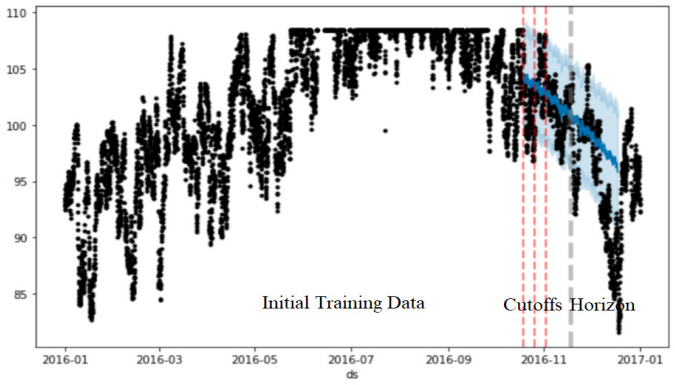


Figure 4: Cross Validation Periods.

3.5. Environment and Parameter Setting

These experiments were conducted on Google Colab using GPU accelerator with Python programming language. The models created using the Prophet forecasting model through the logistic growth model as a trend factor base. The final configuration of the Prophet model in this experiment could be seen in table 2.

Table 2: Prophet Parameter Setting

Parameter	Setting
Changepoint prior scale	0.1
Seasonality prior scale	10
Holidays prior scale	10
Seasonality mode	additive

4. Result

4.1. Analysis

An overview of the final prediction Prophet model could be seen in Figure 5. Prophet model movement trends illustrating through blue lines that can fit the pattern of data diversity well. The red vertical lines in this figure indicate where the potential changepoints placed, and by default, are uniformly placed in the first 80% of the time-series data. Looking at how the curve in fittings movement of the model, Prophet prove can handle a single outlier in the middle of the data. In the last section of Figure 5, we can also see how the prophet made predictions for the next two months.

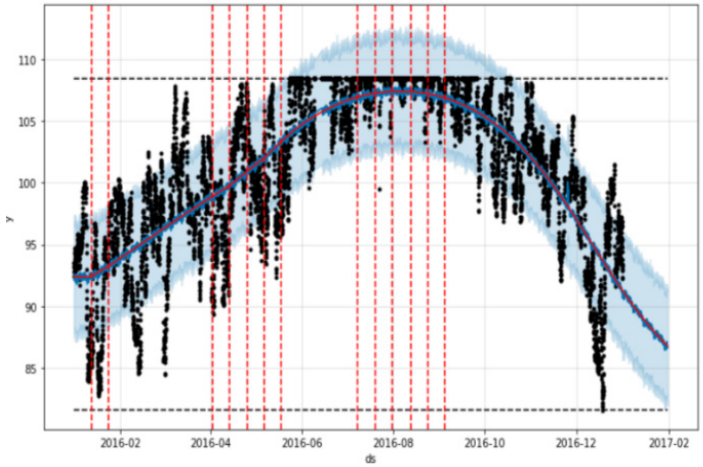


Figure 5: Forecast Plot with Prophet Built-in Method.

Observing the weekly plot provide by Prophet in Figure 6, we can conclude that the usage of electricity consumption in the building increased on weekdays following the function of the building as an office. Also, through advance analysis, we can see daily electricity usage in Figure 7 decreases at 8 p.m. and increases at 2 a.m. And the annual trend in Figure 8 shows energy usage increases in the mid-season and decreases in the late season.

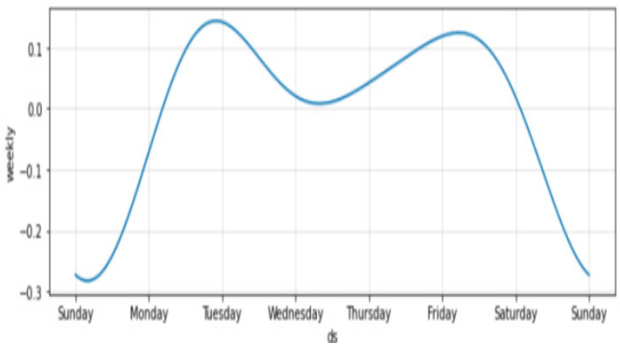


Figure 6: Weekly Plot of Energy Consumptions.

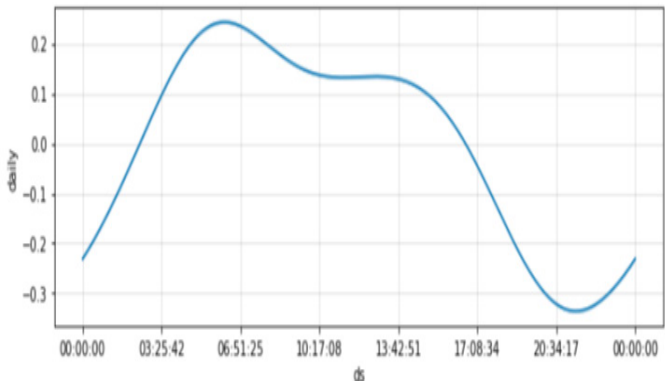


Figure 7: Daily Plot of Energy Consumptions.

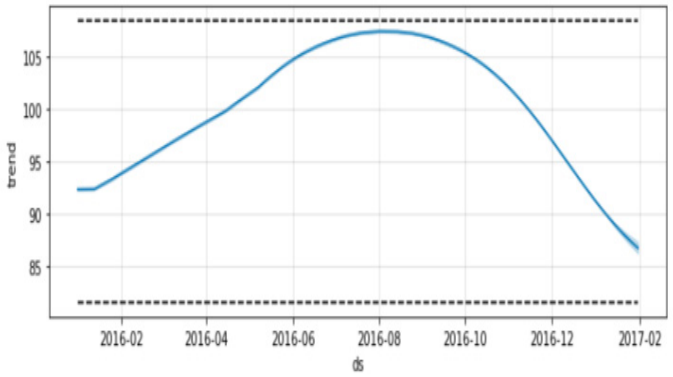


Figure 8: An Overview Plot of Energy Consumptions by Date.

4.2. Evaluation

We could investigate the trend of cross-validation errors for the next two months' predictions in Figure 9. The blue line shows the RMSE, where the mean takes over a rolling window of the dots. We realize that errors below 5 points are typical for predictions one month into the future and that errors increase up to around 8 points for predictions two months into the future.

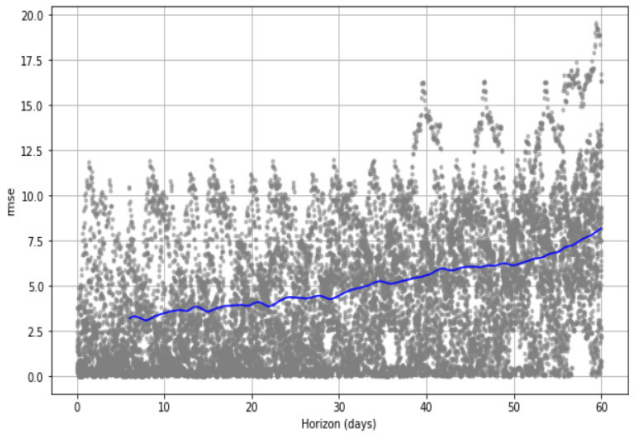


Figure 9: Cross Validation Visually Over Time.

Another experiment had conducted to further evaluate the Prophet's performance with another well-known time series algorithm, LSTM. For LSTM training model configuration, we use two hidden layers, with each consisting of 50 LSTM units,

dropout 0.25, batch size 32, and a sliding window of 720 as the accumulated number of hours in one month. And weather dimensions applied previously in the normalization process were used as inputs in the LSTM model.

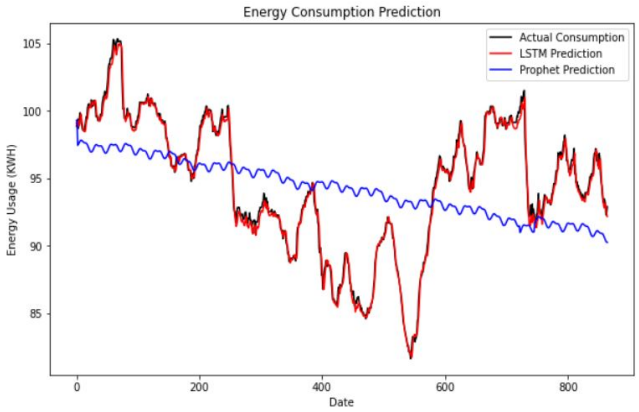


Figure 10: Prophet, LSTM Prediction and Actual Comparison.

Looking at the results on the graph in Figure 10, it appears prophet performs well in fitting data without overfitting. On the other hand, LSTM tends to shows easily overfitting results with RMSE around 0.005 points and training execution time around 330 s. Meanwhile, Prophet with just one-dimensional feature input showed a short training time of 20 s, 15 times faster than LSTM. Comprehensively LSTM still needs to apply an approach to minimize overfitting. Adding more data to LSTM maybe become the solution to have reasonable predictability results.

Cross-validation may be used for tuning hyperparameters of the model, as can be seen in Figure 11. Here parameters are evaluated on RMSE averaged over a 30-day horizon. However, different performance metrics may be acceptable for various problems. In this scenario, the changepoint prior scale is specified within [0.001; 0.01; 0.1; 0.5], while the seasonality prior scale within [0.01, 0.1, 1. 10]. Seeing the result, whereas an increase of changepoint ratio may offer minimum RMSE, nevertheless, it tends to form the model overfitting after all.

	changepoint_prior_scale	seasonality_prior_scale	rmse
0	0.001	0.01	6.932333
1	0.001	0.10	6.948724
2	0.001	1.00	6.928149
3	0.001	10.00	6.909688
4	0.010	0.01	4.614373
5	0.010	0.10	4.639601
6	0.010	1.00	4.633053
7	0.010	10.00	4.613942
8	0.100	0.01	4.261015
9	0.100	0.10	4.269093
10	0.100	1.00	4.262371
11	0.100	10.00	4.271824
12	0.500	0.01	4.261223
13	0.500	0.10	4.260438
14	0.500	1.00	4.263596
15	0.500	10.00	4.255161

Figure 11: Hyperparameter Tuning on Prophet Parameters.

5. Conclusion

With the rapid advancement of machine learning-based techniques, especially deep learning algorithms, the Prophet method is gaining popularity as alternatives way among researchers across diverse disciplines. The objective of this study was to improve the baseline model by introducing the implementation of prophet methods with its practical advantages combine with weather normalization method SDE U/L. Furthermore, we would highlight some of its ability within this experiment in addressing a variety of external factors that affect electricity consumption, such as modeling holiday effects, custom seasonalities, and additional regressor. As a result, SDE U/L – Prophet models could considered successfully predicted values without overfitting in the 30-day forecast with maximum RMSE around 5 points. In advance, more data will make predictions give better results. In the future, the Prophet can be a leading solution that offers practically insightful, flexibility, accuracy, speed, and simplicity for many non-technical users in various time series cases.

We hope that this study result could be a consideration for the researcher by using Prophet methods in various time series cases, as many advantages that Prophet have compared to other time series prediction methods. And especially for practitioners to be more efficient in choosing the best energy-saving based on the prediction analysis result of the energy savings method used. So that in the future practitioner does not have to wait until months to evaluate the final results before further modification considering the many costs incurred in the method selection process. This provides a useful tool for analysts to gain insight into their forecasting problem, besides just producing a prediction.

References

[1] Y. Ji et al., "Hybrid ventilation for low energy building design in south China. Building and Environment, **44**(11), 2245-2255, 2009. https://doi.org/10.1016/j.buildenv.2009.02.015

[2] Z. Romani et al., "Metamodeling the heating and cooling energy needs and simultaneous building envelope optimization for low energy building design in Morocco. Energy and Buildings, **102**, 139-148, 2015. https://doi.org/10.1016/j.enbuild.2015.04.014

[3] Y. Zhang et al., "A new approach, based on the inverse problem and variation method, for solving building energy and environment problems: preliminary study and illustrative examples. Building and Environment, **91**, 204-218, 2015. https://doi.org/10.1016/j.buildenv.2015.02.016

[4] I. Choi et al., "Energy consumption characteristics of high-rise apartment buildings according to building shape and mixed-use development. Energy and Buildings, **46**, 123-131, 2012. https://doi.org/10.1016/j.buildenv.2012.02.015

[5] S. Singaravel et al., "Component-based machine learning modelling approach for design stage building energy prediction: weather conditions and size. In Proceedings of the 15th IBPSA conference, 2617-2626, 2017.

[6] E. Elkind, "Powering the Savings: How California Can Tap the Energy Efficiency Potential in Existing Commercial Buildings, 2016.

[7] F. Rodrigues et al., "The daily and hourly energy consumption and load forecasting using artificial neural network method: a case study using a set of 93 households in Portugal. Energy Procedia, **62**, 220-229, 2014.

[8] R. Bonetto et al., "Machine Learning Approaches to Energy Consumption Forecasting in Households, 2017. https://arxiv.org/abs/1706.09648

[9] I. Yenidoğan et al., "Bitcoin Forecasting Using ARIMA and PROPHET. In 2018 3rd International Conference on Computer Science and Engineering, 621-624, 2018.

[10] H. Aguilera et al., "Towards flexible groundwater-level prediction for adaptive water management: using Facebook's Prophet forecasting approach. Hydrological Sciences Journal, **64**(12), 1504-1518, 2019. https://doi.org/10.1080/02626667.2019.1651933

[11] C. Mi et al., "Real-time Recognition of Smartphone User Behavior Based on Prophet Algorithms, 2019. https://arxiv.org/abs/1909.08997

[12] Z. Yu et al., "A systematic procedure to study the influence of occupant behavior on building energy consumption. Energy and buildings, **43**(6), 1409-1417, 2019.

[13] M. Ghajarkhosravi et al., "Energy benchmarking analysis of multi-unit residential buildings (MURBs) in Toronto, Canada. Journal of Building Engineering, 27, 100981, 2020.

[14] R. Zmeureanu, "A new method for evaluating the normalized energy consumption in office buildings. Energy, **17**(3), 235-246, 1992. https://doi.org/10.1016/0360-5442(92)90051-Z
[15] S. Beheshti et al., "Structure dependent weather normalization. Energy Science & Engineering, **7**(2), 338-353, 2019. https://doi.org/10.1002/ese3.272
[16] S.N. Wood, " Generalized additive models: an introduction with R. Chapman and Hall/CRC, 2017.
[17] A.S. Weigend, "Time series prediction: forecasting the future and understanding the past, Routledge, 2018.
[18] K. Larsen, "GAM: the predictive modeling silver bullet. Multithreaded. Stitch Fix, 30, 2015.
[19] S.J. Taylor et al., "Forecasting at scale. The American Statistician, **72**(1), 37-45, 2018.
[20] K. Gajowniczek, "Short term electricity forecasting using individual smart meter data. Procedia Computer Science, **35**, 589-597, 2014.
[21] L. Wang et al., "Adaptive learning based data-driven models for predicting hourly building energy use. Energy and Buildings, **159**, 454-461, 2018.
[22] Ashrae, ASHRAE - Great Energy Predictor III, Version 1. Retrieved January 25, 2020. https://www.kaggle.com/c/ashrae-energy-prediction/data.
[23] P.J. Fabri, "Measurement and Analysis in Transforming Healthcare Delivery. Springer, 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1537-1544 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

A Software-Defined Network Approach for The Best Hospital Localization Against Coronavirus (COVID-19)

Bilal Babayigit[1], Eda Nur Hascokadar[2,*]

[1]*Department of Computer Engineering, Erciyes University, Kayseri, 38039, Turkey*

[2]*Graduate School of Natural and Applied Sciences, Erciyes University, Kayseri, 38039, Turkey*

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 04 December, 2020
Online: 25 December, 2020

Keywords:
Software-Defined Network
Multi-Controller Placement
Genetic Algorithm
COVID-19
Pandemic Hospital

ABSTRACT

Traditional networks have difficulty in meeting the technological developments and the continuous increase in the size of data to be processed. Software-Defined Network (SDN) approach has emerged as an alternative to traditional networks. SDN separates the control and data planes from each other and manages the network over the control plane with flexibility and cost advantages. In networks with large data flow, SDN with multiple controllers will be useful to manage the network because a single controller may cause network interruptions and data loss. After deciding on the use of multiple controllers in the SDN approach, problems are encountered with the number of controllers that should be used and the placement of these controllers. Due to the increase in the coronavirus epidemic that has affected the whole world in recent months, the need for pandemic hospitals has increased. However, considering the establishment costs of pandemic hospitals, it will not be possible to establish these hospitals in every desired location in the impact area. For this reason, positioning pandemic hospitals at more strategic points will provide these hospitals with a wider coverage area and more functionality at lower costs. In this paper, a genetic algorithm-based SDN is presented for pandemic hospital localization. The number of coronavirus infected people of each city in Turkey is taken into consideration as well as the city distances in the ULAKNET data set within the Topology Zoo database. For the best localization of pandemic hospitals, the Dijkstra algorithm is used to best cover the cities where the coronavirus epidemic is at a minimum distance from the cities. In this paper, the number of controllers is set as 10, and the experimental results are given with maps and graphics.

1. Introduction

This paper is an extension of work originally presented in 2019 4th International Conference on Computer Science and Engineering (UBMK) [1]. In this study, firstly, the coronavirus coefficients of the cities on a certain date are added to the data file of [1] in addition to the coordinates of each city in the Ulaknet data set within the Topology Zoo database. Secondly, the fitness function values are obtained by dividing the coronavirus coefficients of the cities in the selected solution set by the arithmetic averages, as well as the algorithm used to find the minimum distances of the incoming solution clusters to other cities. In this way, while developing random solutions during Genetic Algorithm (GA) generations, solution clusters at the minimum distance to cities are found based on Dijkstra Algorithm (DA), and the solution set with cities with a high number of coronaviruses become the optimum solution set. The aim is to find the locations of pandemic hospitals that are close to the cities on the map, have a high coronavirus coefficient, and best cover cities that need hospitals.

With the continuous development of information technologies, the Internet has turned into a complex and large-scale infrastructure that profoundly affects people's working and lifestyle [2]. However, the complexity of the traditional network makes the networks difficult to manage. SDN resolves these issues based on the principle of separating the data layers. SDN enables easy management of the traditional network seperating data and control planes. By this way, SDN provides many benefits such as simplifying network management, improve network utilization efficiency, support network innovation, reduce network delays.

*Corresponding Author: Eda Nur Hascokadar, edahsckdr@gmail.com

As the network structure expands, a single controller SDN architecture may be difficult to manage a comprehensive network. To increase the scalability and reliability of the network and to avoid a single point of failure, there is a need for a logically centralized [3], physically distributed multi-controller network architecture. In multi-controller architecture, the Controller Placement Problem (CPP) is encountered. With CPP, the number and location of the controllers have a large impact on end-to-end latency reduction and network performance within the network.

2. Related Works

CPP has been intensely studied in the literature. Papers [4-16] explain the different CPP problems in detail for the proper single controller or multiple distributions of controllers and clarify the benefits of SDN.

In [4], the authors present as a solution to CPP and minimize the response delays for a large-scale network using a hybrid GA-PSO algorithm. Using the hybrid algorithm, the best position is obtained in a shorter time compared to GA.

In [5], the authors solve the distributed controllers' placement problem and apply different scenarios in CPP for the distribution of the network load.

CPP with different aspects such as transmission delay, control plane utilization, and controller workload distribution are studied in [6]. A hybrid gradient descent optimization method with GA is implemented in two different network scenarios and effectively established the balance between the use of the control plane and the network response time.

In [7], the authors examine different controller placement models and their accuracies. Other goals of the study include maximizing the controller capacity, reliability, flexibility while minimizing the network latency, deployment cost, and energy consumption. Concerning CPP, the authors present a feature selection technique adopted in the literature to illustrate the search methods used by the controller placement models.

In [8], the authors present a hierarchical K-median algorithm for CPP. Also, by dividing the large number of single-controller network SDN, they attempt to provide efficient control of the area. They also consider the load balancing.

In [9], the authors consider the packets among the controllers and switches for wide area networks, argue that it is important to reduce the propagation delay. They propose a new approach called the Cluster-Based Network Division Algorithm (CNPA). CNPA divides the network into subnets in steps to shorten the propagation delay between the controller and the controller can place the associated switches to a sub-network. They aim at minimizing the network latency, maximizing reliability and durability, minimizing distribution cost and energy consumption. The solutions are searched for CPP with different purposes. They focus on cooperation between controllers for CPP, cost awareness with the use of multiple controllers, and increasing the number of parameters to be based on optimization.

In [10], the authors review the CPP, and analyze the solution techniques and their limitations. The solution techniques are generally based on objective functions. In obtaining the optimum solutions, various factors such as propagation latency between switches and controllers and constraints such as the capacity of controllers and switches are considered.

In [11], the authors present a cost-effective real-time monitoring and recording system to combat the crime committed on the Internet using SDN architecture. In [12], the authors implement a small scale and low-cost SDN design. The performance of two different load balancing algorithms (Round Robin and Dijkstra) is investigated in [13] using round-trip times for addressing the SDN load balancing problem.

In [14], the authors examine the factor of controllers' load into the problem of the controller placement. For this, they define the Capacitated Controller Placement Problem (CCPP). The evaluation shows that the new strategy can significantly reduce the number of required controllers and reduce the load of the maximum-load controller.

In [15], the authors propose a new mathematical model for the CPP in SDNs. The goal of the model is to minimize the cost of the network while considering different constraints. The simulation results show that the model can be used to plan small scale SDN.

In [16], a new expansion model is presented to introduce the network operators plan and to update the SDN infrastructure. For a given network design and switches, the model finds how many, where and which controllers to install to (re-)design the topology of the new network.

The coronavirus also called COVID-19 or SARS-CoV-2, is one of the worst pandemics in recent history [17]. Especially, in recent months, the financial and the social impacts of the COVID-19 pandemic on our lives and economies are very destructive. To stop the spread of COVID-19, billions of people need a vaccine for coronavirus. Many research teams in companies and researchers in universities try to develop vaccines against COVID-19. But it might not be physically possible to make enough vaccine for everyone and optimistically it could be available in 2 year-time but as scientists say it will be a while before the vaccine is as useful as hoped. To this end, governments take measures such as buy blood test kits, rules for wearing masks and social-distancing, flexible working hours, online schools, severe lockdowns, etc. to slow down the spread of COVID-19. Apart from these and besides, governments need effective resource management and coordination. Especially, the key challenges where to place the pandemic hospitals and what should be the number of pandemic hospitals to be placed in a city should be effectively addressed by governments.

The main contribution of this paper is the introduction and evaluation of an SDN-based approach for the best hospital localization against COVID-19. We propose and explain a GA with the DA method that considers the number of coronavirus patients of cities to reduce the insertion end-to-end delay controller. The proposal uses the ULAKNET data set for Turkey and provides the most appropriate placement for the pandemic hospitals. We experimentally show the controllers are located at the closest distance to all other cities and close to the places where the coronavirus outbreak is common, thus reducing the end-to-end delay significantly. Our approach can be used to analyze, validate, and to evaluate the localization of the countries' pandemic hospitals' demand. If more data support is satisfied especially at the level of villages, towns, cities, and districts better prediction models can be obtained.

3. Technical Concepts

3.1. Software-Defined Network (SDN)

SDN is a new networking paradigm emerging to improve the performance of existing network architecture [18]. SDN, by separating the control plane from the data plane, the switches become simple forwarding devices. Network control of SDN is provided with the help of a controller or controllers with logical management in the control layer. Controller, applications, network policy, network configuration, topology management, finding connections, is responsible for the process flow table entry. SDN architecture data, consisting of control and application layers. The communication between the layers is carried out by North and Southbound interfaces. The basic architecture of SDN is shown in Figure 1.

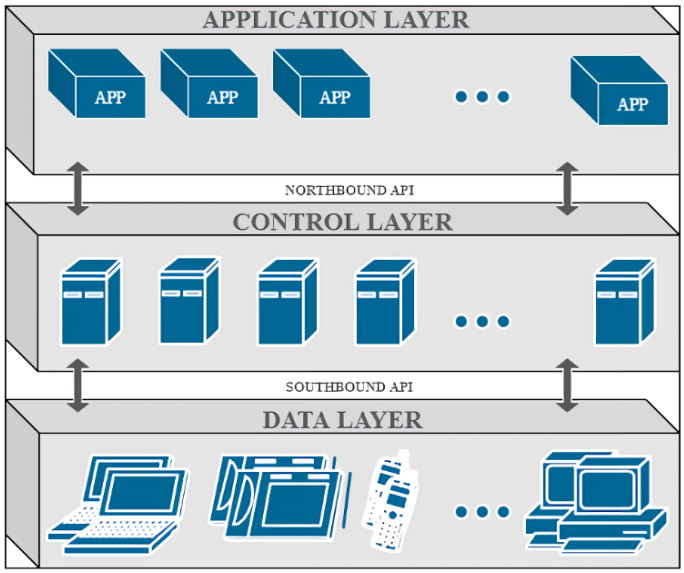


Figure 1: The Basic SDN Architecture

3.2. Multiple Controller Placement Problem

A single controller in the control layer of the SDN architecture may not be sufficient in large-scale networks. There are disadvantages such as a single checkpoint being a single point of failure, bottleneck, and delay in response time. A multi-controller structure is recommended to solve such problems in SDN. After deciding the number of controllers, the significant challenges of multi-controller SDNs to be addressed to increase performance are determining the locations of controllers and determining the number of controllers per module by evaluating the latency.

3.3. Genetic Algorithm (GA)

GA has three steps (selection, crossover, mutation) to have a successful gene. GA process starts with randomly selected individuals. After creating the initial population, it is necessary to determine the fitness function value of each individual. The value of the fitness function is directly proportional to the proximity to the solution according to the type of problem. The higher the fitness function value, the higher the chance of an individual to survive and to reproduce. Individuals that have high fitness function values are selected for crossover. In crossover step, two individuals are selected as a parent and then one or more individuals are produced using different types of crossover operators such as one-point, multi-point, uniform, etc. Then mutation is applied in a probabilistic manner to enrich the gene pool. After these steps, a new population is produced. This procedure is repeated until a specified termination condition is satisfied. The block diagram of GA is given in Figure 2.

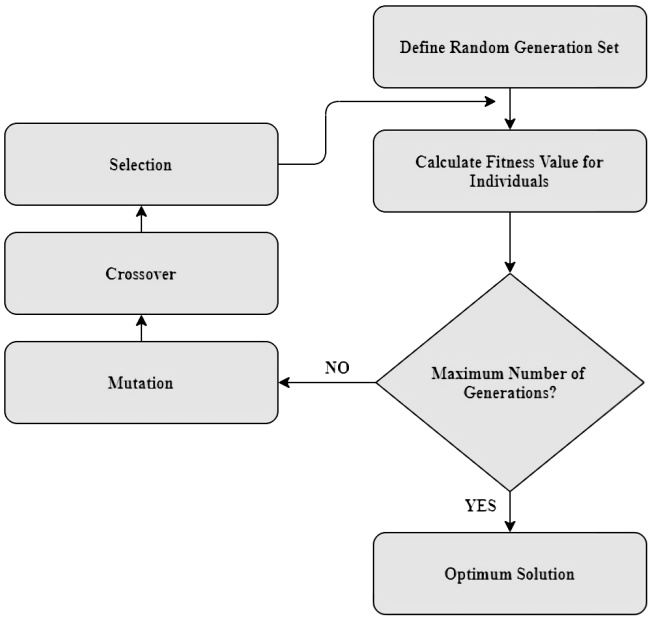


Figure 2: Genetic Algorithm Block Diagram

3.4. Dijkstra Algorithm (DA)

DA is one of the most commonly used algorithms for the shortest path searching. The purpose of DA is to find the shortest distance between nodes in a graph [19]. It uses the greedy search principle as it aims to find the most suitable solution at every step while searching for solutions to the problems it applies to. The input of the DA is the weighted graphs and its output is the shortest path from the origin node to each vertex in the graph.

4. Experiments

In this paper, the hospital placement problem for the coronavirus epidemic affecting the whole world is solved in the SDN infrastructure (see Figure 3) by applying GA and DA approach. In addition to the coordinates of the cities, the coronavirus parameter of each city is added to the ULAKNET database file used.

The ULAKNET data set for Turkey in the Topology Zoo database [20] shown in Figure 4 is used. The study is carried out with Python codes on a machine with Linux based Ubuntu 16.04 operating system, 8 GB RAM, and Core i7, 2.0 GHz processor. The Pycharm platform is used as a compiler for the Python codes.

First, a fixed node map for Turkey using latitude and longitude values in ULAKNET data sets is illustrated in Figure 5. In Figure 5, each city is also represented by a number. For example, 74 number is assigned for İstanbul. The data for İstanbul is as follows:

```
<node id="74">
<data key="d29">1</data>
```

<data key="d30">41.01384</data>
<data key="d31">Turkey</data>
<data key="d32">Red Colour</data>
<data key="d33">74</data>
<data key="d34">28.94966</data>
<data key="d35">Istanbul</data>
</node>

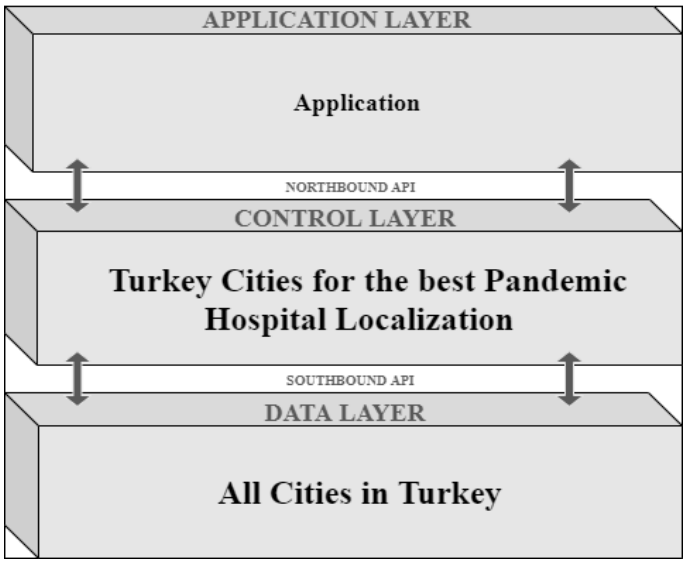


Figure 3: The SDN-based Structure Used for the Best Pandemic Hospital Localization

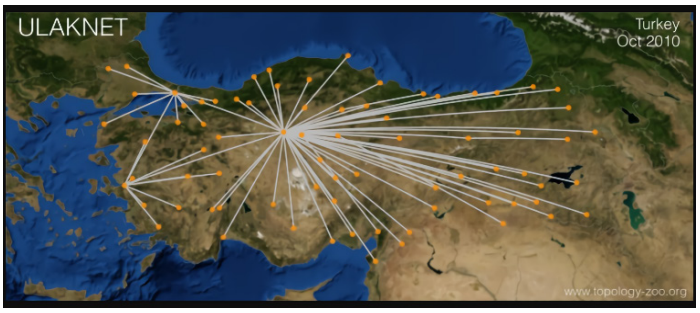


Figure 4: ULAKNET Turkey Map

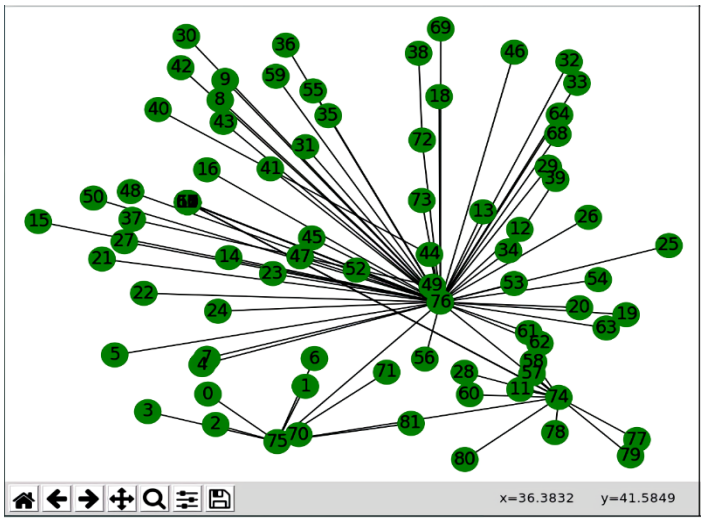


Figure 5: ULAKNET Turkey Map Before the Placement of Multi-Controllers

According to the coordinate data of the cities, the minimum distances of each city to other cities are calculated with DA. A distance matrix obtained for each city is shown in Figure 6.

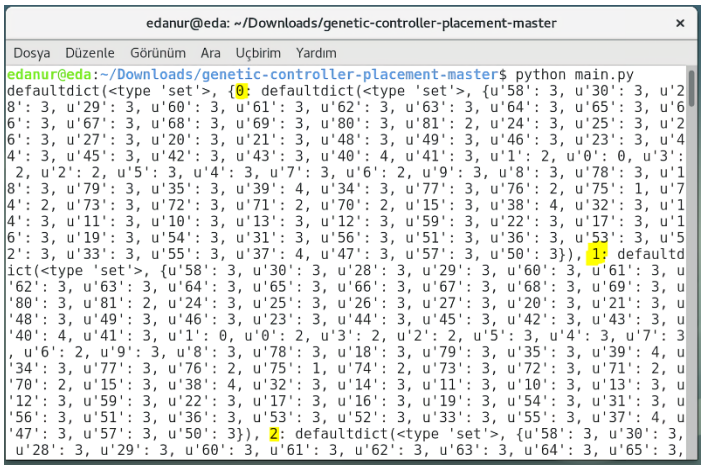


Figure 6: A City Distance Matrix

The number of coronavirus patients is assumed as the coronavirus coefficient of a city. So, the corresponding coronavirus coefficients are added for each city and are shown in Figure 7. The coronavirus patients' data of the cities of Turkey given in Figure 7 are the patients of the date 03/04/2020 [21]. For example, the city number 74 corresponds to Istanbul, and 76 corresponds to Ankara while the coronavirus coefficients of Istanbul and Ankara are 12231 and 860, respectively.

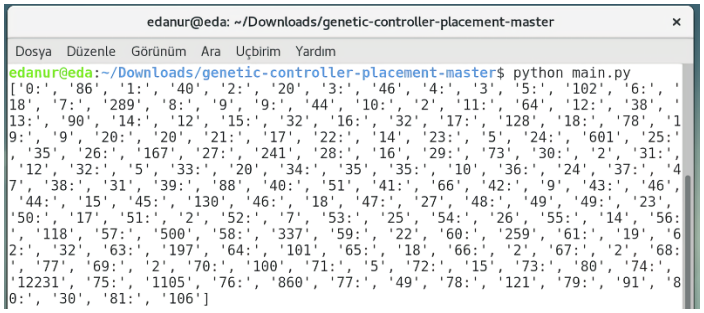


Figure 7: The Coronavirus Matrix of the Cities of Turkey

The parameter setting of GA are listed in Table 1 and the flowchart for the hospital placement problem is presented in Figure 8.

Table 1: GA parameter setting used in the simulations

Parameters	Value
Individuals in the Population	100
The Number of Controllers (Hospitals)	10
Parent Percentage	0.1
Mutation Rate	0.05
Random Selection Rate	0.01
The Number of Generations (Iterations)	1000

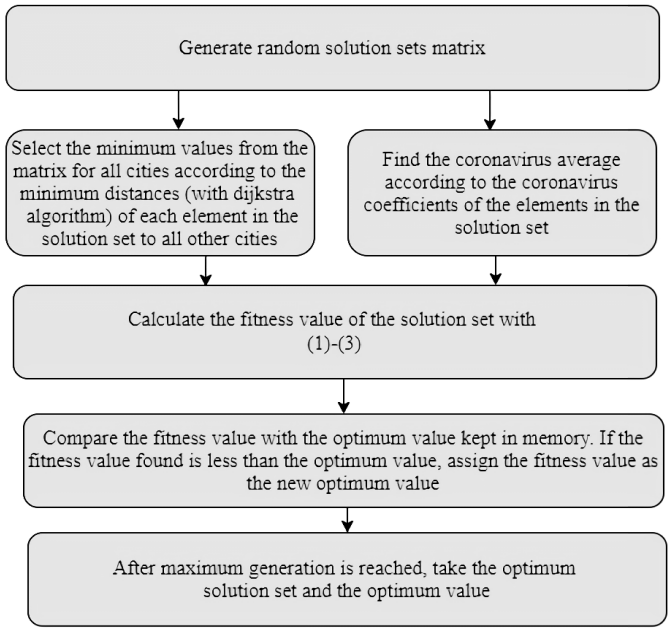


Figure 8: Flowchart for the Hospital Placement Problem

Using the parameters listed in Table 1, 100 random 1x10 solution clusters are created. The randomly generated solutions are illustrated in Figure 9.

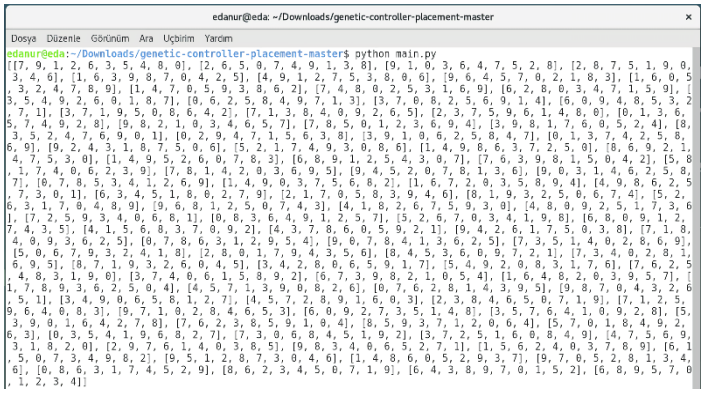


Figure 9: Random Startup Solutions

The fitness of each the random solution sets is calculated using the fitness function. The fitness function has two purposes:

- Choosing solution clusters at minimum distances to cover all cities on the map,
- Choosing solution sets that have higher coronavirus coefficients.

The fitness values of all cities are calculated according to the distance matrix created for each city according to the DA. The city in the random solution set that has a minimum distance from all other cities is selected. After calculating the optimum distance to all cities according to the minimum values in the solution set, the average of the distance from the solution set to all cities is calculated.

For the algorithm used in the simulation: n is the number of the controller (number of pandemic hospitals to be established) and set as 10. x_i is the solution set where $i = 1, 2, 3, \ldots, 10$. h is the weighted graph of cities. $f(x, h)$ is the Dijkstra function. A is the calculated value of the selected solution set with the shortest distances to all other cities. L_Opt and G_Opt are variables for Local Optimum and Global Optimum, respectively.

Equation (1) shows the arithmetic mean of the coronavirus coefficients of the elements in the solution set and (2) takes the average of the values produces according to which of the elements in the selected solution set for each city in the graph of the cities is closest. (2) calculates for the 100 solution sets. After (1) and (2) are calculated, the local optimum value of the selected solution set is obtained by dividing (1) by (2) and (3).

$$\bar{x} = \frac{\sum_{j=1}^{n} x_j}{n} \quad (1)$$

$$A = mean(min(f(x_i, h), i = 1, \ldots, n)) \quad (2)$$

$$L\_Opt = \frac{A}{\bar{x}} \quad (3)$$

Initially, a global optimum value is defined by (4). In each iteration, the global optimum value is compared with the value in (3). If the value in (3) is better than the global optimum value, the new global optimum value will now be that value (5).

$$G\_Opt = Maxint \quad (4)$$

If 'L_Opt' value is lower than 'G_Opt' then:

$$G\_Opt = L\_Opt \quad (5)$$

The aim of the study is to reach the optimum value when the average distance is minimum and the coronavirus arithmetic average is maximum. Thus, two criteria are taken into consideration for the establishment of pandemic hospitals:

- Solution clusters that are close to the cities on the map and will cover the maximum number of cities
- Solution clusters consisting of cities with a high number of coronavirus and needing hospitals more

Each experiment is run for 1000 iterations. After maximum iteration is reached, the optimum solution set and optimum value are taken. 30 experiments are conducted in the study. The best fitness values obtained of the 30 experiments are shown in Table 2. The number of the experiment and the solution set (cities) obtained during the experiment are also listed in the first and the second column of the Table 2.

The iterations and the convergence graph of the 9th experiment is shown in Figures 10 and 11, respectively. The proposed algorithm is converged faster, as iterations proceed the output gets closer to the optimum value at every iteration.

Each number in the solution set that gives the global optimum value corresponds to a city and represents the most suitable cities for the construction of pandemic hospitals. In Figure 12, the cities obtained from the 9th experiment, which gives the most optimum solution set in Table 2, are shown on the map obtained from the ULAKNET data set within the Topology Zoo database. According to the results we have obtained, the pandemic hospitals that should

be placed in the cities [74, 57, 8, 28, 0, 39, 15, 74, 18, 45]. The names of these cities obtained are as follows:

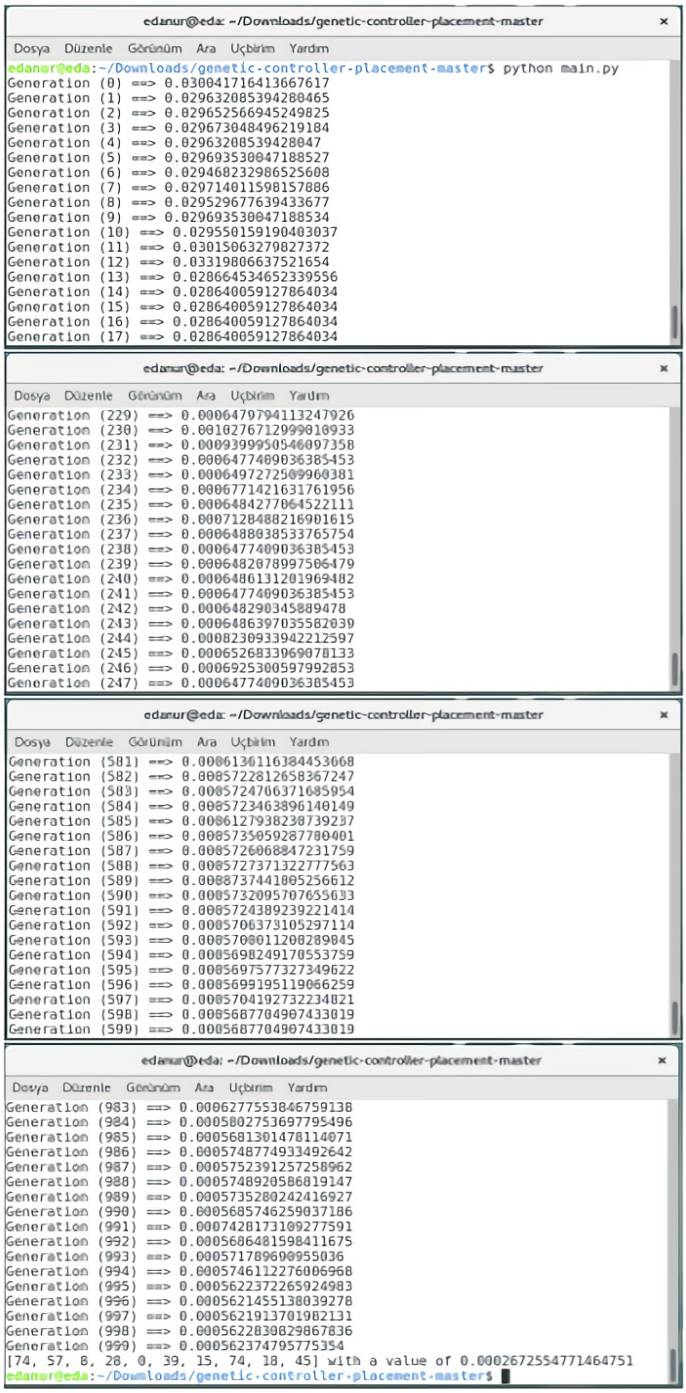


Figure 10: The Optimum Solution for the 9th Experiment

Table 2: The Experimental Results

Exp. No	Cities	Best Fitness Value
9	[74, 57, 8, 28, 0, 39, 15, 74, 18, 45]	0.0002672554771464751
12	[79, 24, 37, 76, 40, 1, 74, 75, 58, 39]	0.0003217251423851137
1	[76, 74, 57, 58, 74, 38, 28, 39, 75, 40]	0.0003219547001035476
14	[60, 42, 39, 49, 12, 40, 74, 24, 76, 33]	0.00032287621623980107
21	[65, 10, 24, 76, 74, 75, 13, 44, 63, 58]	0.0003229917768039527
11	[28, 57, 24, 43, 75, 45, 38, 19, 76, 74]	0.00032310742011823983
5	[51, 74, 52, 62, 53, 57, 58, 38, 7, 76]	0.0003235352480868282
8	[74, 24, 63, 57, 40, 3, 42, 76, 79, 59]	0.0003235708226569537
28	[74, 40, 58, 24, 76, 7, 57, 54, 75, 34]	0.0003238789470047675
17	[24, 58, 18, 75, 76, 37, 40, 7, 74, 39]	0.00032415194841603587
18	[18, 75, 7, 57, 58, 37, 44, 74, 24, 74]	0.00032422337693777505
20	[76, 74, 60, 74, 14, 2, 22, 40, 57, 78]	0.00032422337693777505
10	[44, 18, 76, 34, 58, 5, 74, 74, 30, 34]	0.0003243383497664764
22	[76, 24, 74, 74, 27, 37, 38, 58, 75, 57]	0.0003245685402205335
13	[75, 60, 74, 39, 30, 40, 57, 27, 24, 74]	0.0003247990576496674
29	[18, 40, 67, 63, 24, 81, 64, 25, 74, 76]	0.00032491443919767794
19	[64, 38, 24, 75, 76, 74, 62, 16, 39, 50]	0.0003249690401117191
15	[74, 24, 40, 49, 16, 59, 11, 2, 76, 7]	0.0003250861038869755
7	[75, 76, 39, 74, 24, 57, 46, 49, 7, 63]	0.00032526107622384904
30	[57, 24, 75, 58, 40, 74, 76, 7, 27, 74]	0.0003258893348427155
23	[74, 39, 49, 57, 68, 38, 7, 75, 76, 58]	0.00032590791780073814
2	[76, 58, 35, 74, 37, 57, 75, 60, 17, 27]	0.0003262334629682094 4

24	[7, 1, 27, 74, 76, 75, 39, 24, 58, 40]	0.0003263054392941361
16	[74, 39, 40, 60, 57, 38, 74, 75, 34, 27]	0.00032708917158037114
6	[24, 81, 0, 57, 37, 41, 74, 74, 39, 75]	0.0003281787392685552
3	[58, 75, 67, 76, 55, 31, 74, 7, 39, 60]	0.00032819733296483107
26	[76, 52, 60, 76, 24, 27, 23, 74, 58, 7]	0.00032819733296483107
4	[58, 9, 76, 57, 49, 39, 61, 79, 74, 63]	0.0003696841205081152
27	[27, 37, 39, 57, 39, 80, 74, 76, 51, 7]	0.00037254145084697076
25	[39, 58, 17, 75, 76, 75, 57, 7, 40, 24]	0.0005330035817840696

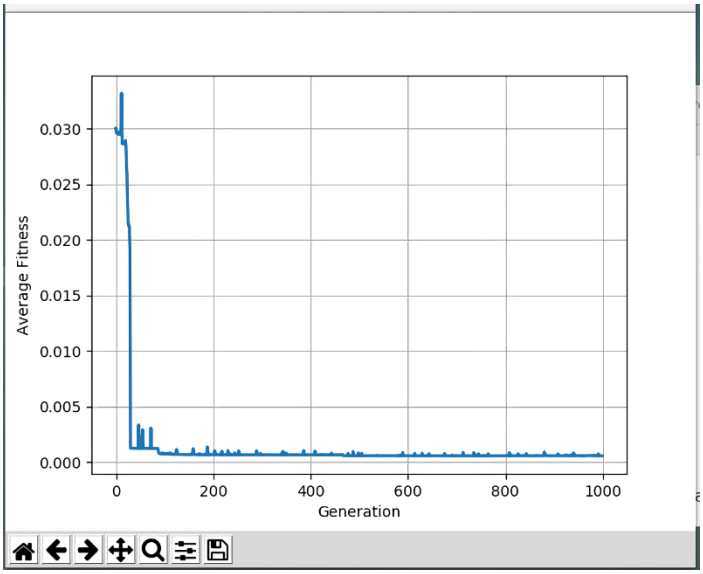


Figure 11: The Convergence Graph of the 9th Experiment

74: İstanbul;
57: Kocaeli;
8: Çanakkale;
28: Bilecik;
0: Denizli;
39: Ordu;
15: Hatay;
74: İstanbul;
18: Erzurum;
45: Kayseri

and demonstrated in the SDN structure in Figure 13. The observation is that in the solution set, there are two Istanbul cities. This corresponds Istanbul needs two pandemic hospitals. However, the capital Ankara is not in the obtained cities. But 24 of 30 experiments obtained Ankara. The cities are not obtained based on their economic, social, and cultural levels. The values obtained from the experiments should be evaluated together with the results obtained from the algorithm of all cities in the solution set, not as a single city.

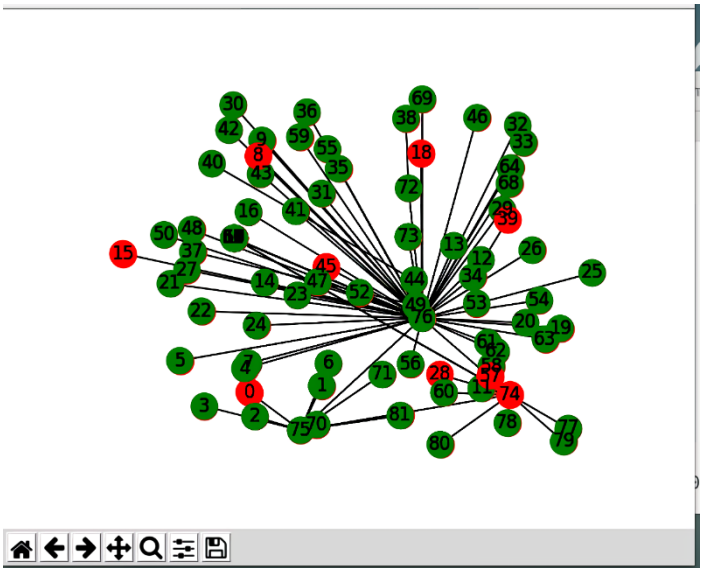


Figure 12: The Best Hospital Localization

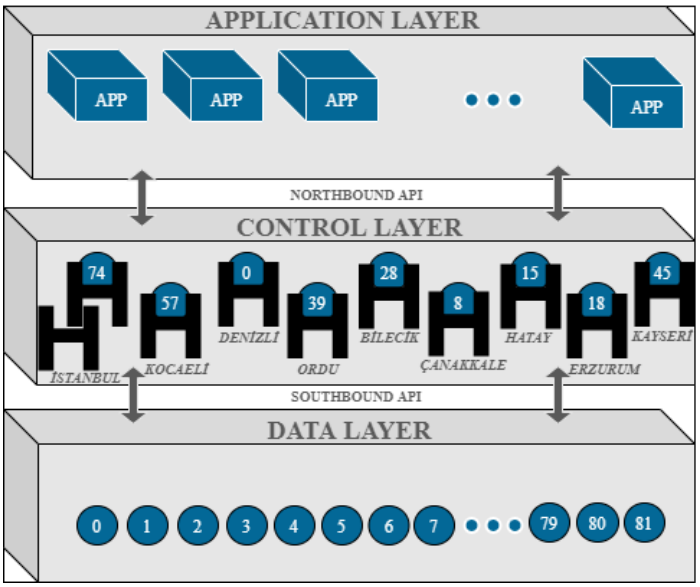


Figure 13: The Obtained Cities for the Pandemic Hospitals in the SDN Structure

5. Conclusion and Future Study

In this paper, we have studied the placement of the pandemic hospitals which should be established against the coronavirus. As it is not possible to establish the pandemic hospitals in every desired location due to the high cost of setting up the pandemic hospitals, the placement of the pandemic hospitals is of great importance. To solve the pandemic hospital placement problem, in this paper, an SDN-based solution has been proposed by applying a GA with DA method according to the distance and coronavirus parameters to the maps obtained from the ULAKNET data set.

From the experimental results, the hospital location found has the closest distance to cover all cities in the selected data set and in places where there is a high coronavirus. Ten pandemic hospitals are placed with maximum coverage.

We need more data at the level of villages, towns, cities, and districts. But now, data for coronavirus in many countries of the World are still limited and even it is very difficult to get the shared data of the countries. Such better data will help us to achieve better prediction models as future work.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We would like to thank the Erciyes University Scientific Research Projects Coordination Unit (ERU/BAP). This work is supported by the Erciyes University Scientific Research Projects Coordination Unit (ERU/BAP) under project codes FYL-2019-9299.

References

[1] B. Babayiğit, B. Ulu, E.N. Hasçokadar, "Solving multi-controller placement problem in software defined networks with a genetic algorithm," in 2019 4th International Conference on Computer Science and Engineering (UBMK), Samsun, Turkey, 666–670, 2019, doi: 10.1109/UBMK.2019.8907199.

[2] J. Lu, Z. Zhang, T. Hu, P. Yi, J. Lan, "A survey of controller placement problem in software-defined networking," IEEE Access, **7**, 24290–24307, 2019, doi: 10.1109/ACCESS.2019.2893283.

[3] Z. Guo, R. Liub, Y. Xuc, A. Gushchind, A. Walide, "STAR: Preventing flow-table overflow in software-defined networks," Computer Network, **125**, 15–25, 2017.

[4] L. Liao, V.C.M. Leung, "Genetic algorithms with particle swarm optimization based mutation for distributed controller placement in SDNs," in IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), 1-6, 2017, doi: 10.1109/NFV-SDN.2017.8169836.

[5] N. Mouawad, R. Naja, S. Tohme, "Optimal and dynamic SDN controller placement," in 2018 International Conference on Computer and Applications (ICCA), 1-9, 2018, doi: 10.1109/COMAPP.2018.8460361.

[6] V. Huang, G. Chen, Q. Fu, E. Wen, "Optimizing controller placement for software-defined networks," in International Conference on Cloud Computing and Big Data, Macau, China, 224-232, 2019.

[7] I.O. Adebayo, M.O. Adigun, P. Mudali, "Feature selection strategies for the controller placement problem in SDNs: A Review," in 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 1-5, 2018, doi: 10.1109/ICABCD.2018.8465404.

[8] H. Kuang, Y. Qiu, R. Li , X. Liu, "A hierarchical K-Means algorithm for controller placement in SDN-based WAN architecture," in 2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsha, 263–267, 2018, doi: 10.1109/ICMTMA.2018.00070.

[9] G. Wang, Y. Zhao, J. Huang, W. Wang, "The controller placement problem in software defined networking: a survey," IEEE Network, **31**(5), 21-27, 2017, doi: 10.1109/MNET.2017.1600182.

[10] A.K. Singh,S. Srivastava, "A survey and classification of controller placement problem in SDN," International Journal of Network Management, **28**(3), 1-25, 2018, doi:10.1002/nem.2018.

[11] B. Babayigit, S. Karakaya, "Cost-effective logging using SDN architecture," in International Conference on Advanced Technologies, Computer Engineering and Science , Karabuk, Turkey, 2018.

[12] B. Babayigit, S. Karakaya, B. Ulu, "An implementation of software defined network with Raspberry Pi," in 2018 IEEE 26th Signal Processing and Communications Applications Conference, Izmir, Turkey, 1-4, 2018, doi: 10.1109/SIU.2018.8404546.

[13] B. Babayigit, B. Ulu, "Load balancing on software defined networks," in 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2018 , Kizilcahamam, Ankara, Turkey, 1-4,2018, doi: 10.1109/ISMSIT.2018.8567070.

[14] G. Yao, J. Bi, Y. Li, L. Guo, "On the capacitated controller placement problem in software defined networks," IEEE Communications Letters, **18**(8).1339-1342, 2014, doi: 10.1109/LCOMM.2014.2332341.

[15] A. Sallahi, M. St-Hilaire, "Optimal model for the controller placement problem in software defined networks," IEEE Communications Letters, **19**(1), 30-33, 2015, doi: 10.1109/LCOMM.2014.2371014.

[16] A. Sallahi, M. St-Hilaire, "Expansion model for the controller placement problem in software defined networks," IEEE Communications Letters, **21**(2), 274-277, 2017, doi: 10.1109/LCOMM.2016.2621746.

[17] www.who.int/covid-19, Access Date: June 2020.

[18] Y. Zhang, L. Cui, W. Wang, Y. Zhang, "A survey on software defined networking with multiple controllers," Journal of Network and Computer Applications, **103**, 101-118, 2018.

[19] E. W. Dijkstra, "A note on two problems in connexion with graphs," Numerische Mathematik, **1**(1), 269–271, Dec 1959. [Online]. Available: https://doi.org/10.1007/BF01386390

[20] Topology Zoo. www.topology-zoo.org. Access Date: June 2020.

[21] https://ajansspor.com/haber/hangi-ilde-ne-kadar-corona-virus-vakasi-var-vefat-sayilari-362702, Access Date: June 2020.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 1545-1553 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

NemoSuite: Web-based Network Motif Analytic Suite

Wooyoung Kim*, Yi-Hsin Hsu, Zican Li, Preston Mar, Yangxiao Wang

Computing and Software Systems, School of STEM, University of Washington Bothell, WA 98011, USA.

ARTICLE INFO

Article history:
Received: 27 August, 2020
Accepted: 19 December, 2020
Online: 25 December, 2020

Keywords:
Network Motif
Web-GUI
Usability
Efficiency

ABSTRACT

Biological networks represent biological systems, and various graph analysis algorithms have been applied to solve various real-world problems. Network motif analysis, as one of network analyses, is detecting frequently and uniquely over-occurring subgraph patterns in a network. The detection process requires high computational resources, and various tools have been developed to provide efficient solutions. However, they still lack extensible output options and easy accessibility, which restricts substantial scale of experiments for many biological applications. Therefore, we provide NemoSuite (Network Motif in a Suite) as a web-interactive tool for detection and analysis of network motifs. Including both of network-centric and motif-centric approaches, it is a greatly accessible tool emphasizing on efficiency, usability, and extensibility.

1 Introduction

Systems biology focuses more on the interconnections between molecules, such as DNA, RNA, or proteins, than individual components. Biological systems are often represented as biological networks, including protein-protein interaction (PPI), metabolic, or gene-regulatory networks. Biological networks, which model the data in -ome level or in ecosystem level, show individual molecules as graph nodes and their interactions as edges. Researchers applied various graph analysis algorithms to understand biological systems in networks, and we are focusing on network motif analysis as one of graph algorithms. Here, we present an online program extending an original work presented in 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) [1], which introduces NemoMapPy as an online tool to detect network motifs.

Network motifs are defined as statistically over-recurring subgraph patterns in a network, which are assumed to have biological significance in biological systems. They are defined in line with "sequence motifs" that are playing a key role in analyzing DNA/RNA or protein sequences with the goal of, for example, controlling the translation of the encoding mRNA, characterizing protein families, detecting DNA binding sites, or controlling tumor suppressor gene expression.

Analysis of network motifs have applied to various applications such as predicting protein-protein interactions [2], determining protein functions [3], detecting breast-cancer susceptibility genes [4], investigating evolutionary conservation [5, 6], and discovering essential proteins [7, 8]. Furthermore, a broad spectrum of network motif applications has been explored: "motif clustering" [9], "motif themes" [10], "relative graphlet frequency distances" [11], "motif modes" [12], and "MotifScores" [13].

However, currently existing tools lack the easy accessibility, and flexible scalability, which restrict their range of applications, and thereupon cause some amount of faithlessness on the analysis [14]. Most tools are unable to collect instances of network motifs and lack of functionality, usability, and extensibility. Therefore, we provide NemoSuite (Network Motif in a Suite) to resolve these problems: it includes Nemo and NemoMapPy programs. Nemo program implements a network-centric approach using NemoLib [15, 16]. NemoMapPy implements NemoMap [17] algorithm which is one of the motif-centric methods.

The contributions of NemoSuite include the followings: (1) As a web-interactive graphical user interface (GUI) program, it provides easily accessible tools with great efficiency and usability; (2) It provides various output results based on users' choice, which will explore network motif analysis to various real-world applications; (3) It provides both network-centric and motif-centric tools. As far as the authors' knowledge, NemoSuite is the first program providing both network-centric and motif-centric methods as an web-interactive, and GUI format.

The rest of the paper starts providing the review of various detection algorithms and tools in Section 2. Thereafter, Section

*Corresponding Author, Computing and Software Systems, School of STEM, University of Washington Bothell, WA 98011, USA., kimw6@uw.edu

3 describes NemoSuite, followed by experiments and results in Section 4, and Section 5 respectively, and conclusion in the last section.

2 Related Works

2.1 Network Motif

Recent survey of network motifs showed that there are different types of network motifs [18]. Originally, Milo et al. [19] introduced network motifs as the "patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks" in 2002. After discovering the under-represented patterns could also delineate some features of networks, 'anti-motifs' are defined as the subgraphs with significantly less frequencies in the input network compared in a random pool [20]. On the other hand, 'colored-motifs' are introduced as the network motifs in multi-label networks such as metabolic, social or traffic networks [21]. Additionally, Parida proposed 'maximal-motif' as the network motifs with maximal nodes and edges in a given size [22].

However, as we focus on biological networks, we use the original network motifs in this work, whose mathematical definition follows. A network motif M is a connected subgraph of size k which can be found frequently and uniquely in an input network G. To determine the uniqueness, P-value (Eq. (1)) or Z-score (Eq. (2)) of the recurrence of M is calculated within a huge pool of randomly generated graphs.

$$\text{P-value}(M) = \frac{1}{N}\sum_{n=1}^{N} c(n), c(n) = \begin{cases} 1, \text{if } f_R(M) \geq f_G(M). \\ 0, \text{otherwise} \end{cases} \quad (1)$$

$$\text{Z-score}(M) = \frac{f_G(M) - \mu(f_R(M))}{\sigma(f_R(M))} \quad (2)$$

Here, N refers to the number of random graphs, and $f_G(M)$ is the frequency of M in the input G while $f_R(M)$is the frequency of M in the random network R. $\mu(f_R(M))$ is the mean of frequencies of M in the random networks and $\sigma(f_R(M))$ is the standard deviation of frequencies of M in the random networks. For a fair comparison, the random graphs are generated given the same degree sequences as the original graph. In general, a subgraph with P-value less than 0.01 or Z-score greater than 2.0 is considered as a network motif after compared with $1,000$ random networks [23].

2.2 Algorithms

Grow and Kellis classified various algorithms of discovering network motifs into network-centric and motif-centric approaches [24]. The former method first searches all possible patterns from the input graph, then determines which patterns are network motifs through a statistical testing, whereas the latter approach outputs the number of occurrences of the given query pattern in the input network without statistical testing results.

Various algorithms and software programs implement network-centric approach, which include MAVisto [25], MFinder [26], Kavosh [27], NemoFinder [28], NetMODE [29], QuateXelero [30], Motif-Discovery[31], MotifNet [32], ESU algorithm[33], and FANMOD program [34]. Network-centric approach, requiring an input network and the size of motif to search, involves enumerating non-isomorphic k-size subgraph patterns in the original network. It then compares the frequency of each pattern in a number of generated random graphs. The uniqueness of each pattern is determined based on the P-value or Z-score. This approach is specifically useful if the goal is to determine which pattern is a network motif. However, enumeration process limits the size of detectable motifs due to the time-consuming process of scanning for every nodes and edges of the network. Consequently, network-centric algorithms using exhaustive enumeration of k-size subgraphs in the network does not exceed size 8 nodes [24]. Larger size of network motifs can be detected if searched approximately .

Vertices	Number of non-isomorphic subgraphs	
	Undirected	Directed
1	1	1
2	1	2
3	2	13
4	6	199
5	21	9 364
6	112	1 530 843
7	853	880 471 142
8	11 117	1 792 473 955 306
9	261 080	13 026 161 682 466 200
10	11 716 571	341 247 400 399 400 000 000

Table 1: It shows the number of non-isomorphic patterns in undirected or directed graphs increases exponentially with the size of pattern, courtesy of [23].

Focusing on detecting larger size of motifs, Grochow and Kellis introduced a motif-centric approach (GK method), which finds the frequency of each pattern in a given query set by mapping each query graph to all possible location in the input graph [24]. GK method, MODA[35], FASCIA [36], ISMAGS[31], ParaMODA[37], and NemoMap [17] are implementing this motif-centric approach. This approach was able to find motifs of up to 15 nodes with symmetry-breaking technique to significantly reduce isomorphic testing. One disadvantage is that it assumes that the user has pre-knowledge of what patterns are network motifs in the target graph as a set of query patterns have to be provided. Otherwise, this approach is not straightly applicable to finding network motifs. In fact, no algorithms or tools have described how to provide the query set. Therefore, if the goal is to find which pattern is significantly frequent than others, the user has to provide all possible k-size subgraph patterns as a query set to compare, which can suffer heavily since full list of all possible variations of k-size subgraphs grows exponentially with k. Table 1 shows the exponential growth of pattern variations of k-size subgraphs up to 10 nodes, and in the worst-case scenario the application would have to iterate over $341,247,400,399,400,000,000$ of 10-node patterns many of which might not even exist in the target graph.

Table 2 summarizes various aspects of these two approaches. Although both the network-centric and motif-centric approaches have

Table 2: Summary of network-centric and motif-centric approaches.

	Network-centric	Motif-centric
Input	An input network, a motif size k, the number of random graphs to generate	An input network and (a) query pattern(s)
Output	All discovered subgraph pattern's frequency in the input network, Z-score and/or P-values	Each query pattern's frequency in the input network
Processes	Enumerate or sample all possible non-isomorphic subgraphs, and each pattern's frequency is compared in a huge random pool	Search the instances of each query graph in the input network, and its frequency or the set of instances of the pattern is provided
Tools	FANMOD, MAVisto, MFinder, Kavosh, NemoFinder, NetMODE, QuateXelero, MotifNet, Motif Discovery	GK, MODA, FASCIA, ISMAGS, ParaMODA, NemoMap
Benefits	Can determine which patterns are network motifs. Non-existing patterns in the network will not be searched.	Can search large size (more than 8) of a pattern. Relatively fast to collect all instances of a given query graph.
Drawbacks	Exponential computational time growth with increasing size of motifs and networks. Statistical analysis adds more computational burden.	Exponential computing time increase on the number of possible query patterns with increasing size of motifs. It cannot be directly applied for discovering network motifs.

drawbacks, each approach can be respectively applied to different problems with appropriate goals. If the goal is determining which patterns will be network motifs in a target graph, then network-centric approach is appropriate. Motif-centric method would work better for a pattern matching. Therefore, we aimed to provide a tool with both approaches to maximize the usability and efficiency.

2.3 Tools

There are various tools that follow network-centric or motif-centric approaches. After careful review of the tools [18, 38, 39], we found out that most tools are out-dated, dependent on operating systems or other external tools, restricted output options, inaccessible, or lack of usability. As shown in Table 3, we tested the following tools such as mFinder, FANMOD, Kavosh, QuateXelero, Fascia, ISMAGS, MotifNet, and Motif-Discovery, because others are unable to locate, unable to install, out-dated, incompatible with current security setting, or inexecutable.

mFinder, FANMOD, Kavosh, and QuateXelero are network-centric based tools that accept a text file as an input graph, the size of network motifs, and the number of random graphs to generate. While others require a list of edges in the text graph file, Kavosh and QuateXelero additionally require the number of nodes at the first line in the file. Before running these programs, however, since each node should be labelled as an integer, non-integer labels should be converted to integer labels. They provide outputs consisting of each pattern's frequency, and its Z-score or P-value so that users can determine which patterns are network motifs. When we design our program, NemoSuite, we set FANMOD program as an example, since it is a GUI-based program, efficient, with additional filtering options, and can provide the set of instances as well. MotifNet, as a web-program, also succeeds FANMOD, taking two files (edge list and node list), the size of network motif, and the number of random graphs. However, it requires the user to submit the work in a queue, which seemed discontinued the service currently, as we never got the results back. Motif-Discovery, as a cytoscape[31] plug-in, can read string-labelled nodes, visualize the network, and provide the frequency and Z-score for each pattern. Unfortunately, since the Z-scores are significantly different from the ones by other tools, it is unclear how to determine network motifs.

Fascia and ISMAGS are motif-centric based motif discovery tools. Fascia requires two text files. One is the input graph and the other is a query graph. The input graph file should contain the number of nodes, the number of edges, and the list of edges where the node ID's are integers. A query text file is a list of edges. As a command-based program, it provides the number of occurrence of the query pattern in the input network. On the other hand, ISMAGS is cytoscape plug-in, which can visualize the input network. In this work, we excluded other tools including MODA, since we failed to install or run them properly.

Table 3 lists the tools, approach, year of publication, the input options, the output options, interface, and shortcomings. In general, current available tools lack of visualization, usability, easy accessibility, and comprehensive results. To overcome the shortcomings of other tools, we provide NemoSuite as a comprehensive and efficient web-interactive program.

3 Methods: NemoSuite

We developed 'NemoSuite' (Network Motif in a suite) to provide a unified access to network motif analysis including network-centric and motif-centric approaches. NemoSuite consists of two separate programs: 'Nemo' and 'NemoMapPy.' Nemo program employs a network-centric approach using a NemoLib [15, 16] that improves ESU (Enumerate SUbgraphs) algorithm [33]. NemoMapPy program follows a motif-centric method by implementing NemoMap [17] algorithm in Python and Flask framework. Figure 5 shows the NemoSuite homepage with Nemo and NemoMapPy program pages. The program is available at https://bioresearch.css.uwb.edu/biores/NemoSuite/.

Table 3: Network motif detection tools are compared based on their approach (network-centric vs. motif-centric), shortcoming, year, input, output, and interface options.

Tool	Approach	Year	Input	Output	Interface	Shortcoming
mFinder	network	2002	edge list file (integer ID)	text file: freq, z-score, p-value	command-line	restricted input and output options, command-line
FANMOD	network	2006	edge list file (integer ID)	html: freq, z-score, p-value, visual. Instance set	GUI	restricted input and output options
Kavosh	network	2009	#n,edge list file (integer ID)	text file: freq, z-score, motif matrix	command-line	restricted input and output options, command-line
QuateXelero	network	2013	#n,edge list file (integer ID)	text file: freq, z-score, motif matrix	command-line	restricted input and output options, command-line
Fascia	motif	2013	#n,#m,edge list file (integer ID)	count	command-line	restricted input and output options, command-line
ISMAGS	motif	2016	cytoscape input	count	cytoscape plug-in	exception error, plug-in
MotifNet	network	2017	edgelist, nodelist	html: freq, z-score, p-value, visual. Instance set	web-interactive	queue-based, no-prompt result
Motif-Discovery	network	2018	cytoscape input	text file: freq, z-score	cytoscape plug-in	restricted output options, significantly different Z-scores from others, cytoscape-dependent

3.1 *Nemo*

Nemo, as shown in Figure 5(b), follows a network-centric approach which includes the steps of enumeration, random graph generation, and statistical analysis. The user should select an input file (see Fig 1 (a) for the file format), along with the directness, the size of motif and the number of random graphs to generate. Nemo program works on both directed and undirected graphs, and the input graph's node can be represented as string or integer. There are three output options: NemoCount, NemoProfile, and NemoCollection, whose formats are introduced by Kim and Haukap [40], and illustrated in Figure 1 (b).

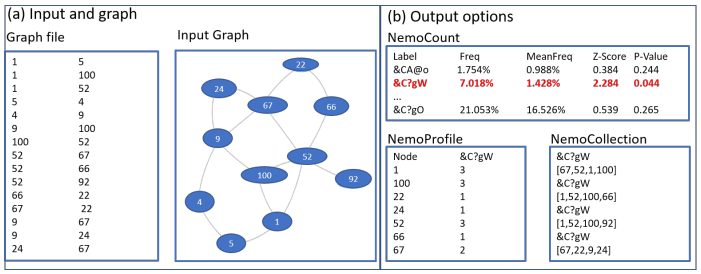


Figure 1: Nemo: (a) shows the input file format with visualized graph on the right; (b) shows the three output options including NemoCount, NemoProfile, and NemoCollection.

'NemoCount' is a default output format which lists the patterns of each detected subgraphs, along with its frequency in the input graph, average frequency in the random graphs, then finally shows its statistical results so that the user can determine which pattern is a network motif. Each pattern is labeled with g6 format which is used to identify isomorphic graphs in the Nauty program [41]. 'NemoProfile' option returns two files. One is a file of *NemoProfile* that contains two dimension of matrix of N rows and M columns, where N is the number of nodes and M is the number of network motifs. For example, from the Figure 1 (b), the node '100' is found in three instances of the motif **&C?gW**, and the node '67' belongs to two instances of the same motif. Wang and Kim showed that NemoProfile can be used as a data feature to detect essential genes [42]. Another file is a *SubgraphProfile*, which is an extended NemoProfile containing not only motifs but also non-motifs in the column. If none of patterns are network motifs, this option provides SubgraphPrifle only. Likewise, 'NemoCollection' also provides two files: *NemoCollection* and *SubgraphCollection*. While NemoCollection provides the set of network motif instances as a file, SubgraphCollection provides instances of all subgraph patterns.

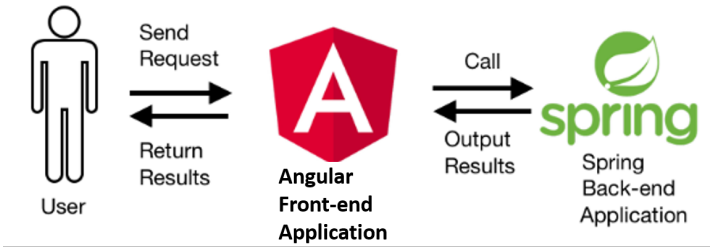


Figure 2: Block diagram for Nemo program.

Nemo program is using an angular application for the front-end, and a spring application for the back-end as shown in Figure 2. The front-end uses the angular framework, which takes advantage of

the Model-View-Controller (MVC) pattern. Developers can easily display the data and the model from the back-end and control the model with the 'appcomponent.' Also, it can be easily deployed with angular. The back-end uses the Spring framework, which produces RESTful APIs. With this simple and intuitive design, developers can easily update modules for additional purposes. Nemo program also has high maintainability with the REST architecture style. Developers can fix any defected part if any, and restart the Spring application on the server while the client application remains untouched. Figure 3 shows the overall architecture of the program.

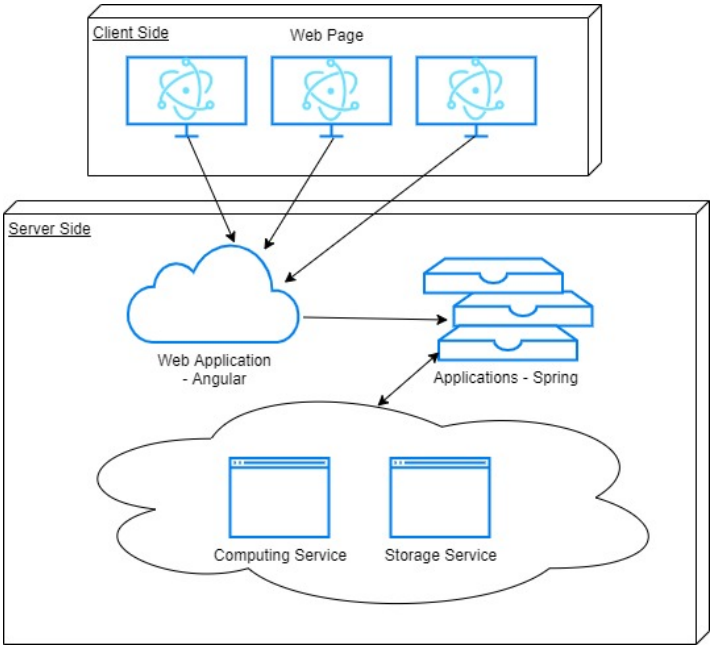


Figure 3: Descriptive architecture of Nemo program.

3.2 *NemoMapPy*

NemoMapPy implements a motif-centric approach which requires an input file, and a query pattern. While other motif-centric tools do not provide statistical results, NemoMapPy provides the frequency and statistical results of the query pattern in the input graph if the number of random graphs are entered, so that whether this pattern is a network motif can be determined with its P-value or Z-score.

NemoMapPy, as an online tool, refined and upgraded the features of NemoMap [17] that improved the performance of ParaMODA [37], which also increased the efficiency of the motif-centric algorithm introduced by Grochow and Kellis (GK) [24], and MODA program [35]. In fact, NemoMap improved the performance of ParaMODA, MODA and GK method by modifying the symmetry-breaking method. It removed the need of maintaining a list of all mappings to prevent duplicate occurrences. It also reduced some of symmetry-breaking conditions, thereafter, contributes on optimizing space and computational time. NemoMap has been proven to be more efficient than ParaMODA, MODA and GK in most cases, exception with some simple query patterns.

NemoMapPy ported NemoMap with additional features on the web [1], and was implemented in Python 3 with the Flask web framework [43]. As a consequence, NemoMapPy can accept streaming text inputs as well as file uploading. The query pattern can be provided as a text input, file uploading, or g6 format. Since Nemo program provides g6 format for detected pattern, the g6 format helps a quick transition from Nemo to NemoMapPy.

Figure 4 illustrates the overall architecture of NemoMapPy. Nginx [44] is used in the first layer, which is a reverse proxy for Gunicorn [45] that redirects incoming requests to Gunicorn with remarkable performance. Gunicorn is a WSGI server which allows multiple applications to run side-by-side in the same process. It handles the load balancing for a program, distributes incoming requests, and communicates with the Flask services accordingly. The back-end of NemoMapPy program is developed by python Flask framework, which takes advantages of simplicity and flexibility. It is convenient for developers to deploy or undeploy it from the server, and the simplified usability make Flask program faster to debug. Here we focused on providing an easy and accessible interface with multiple input options so that users can use NemoMapPy in a variety of situations.

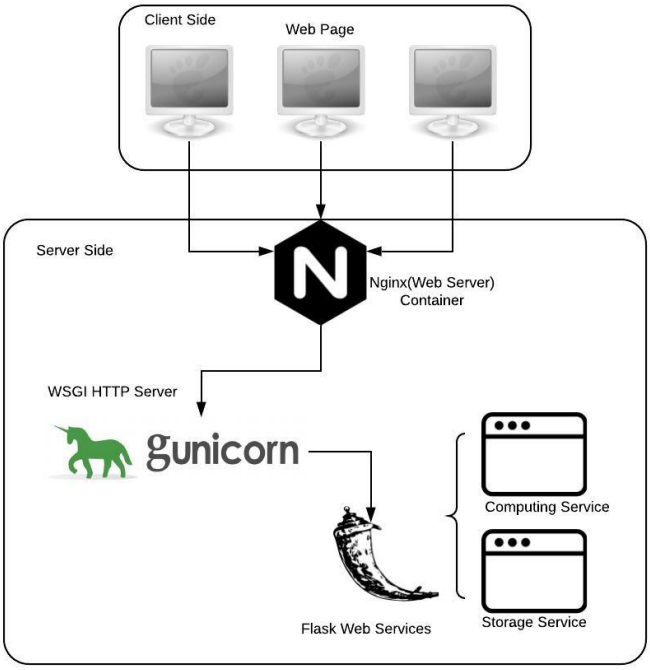


Figure 4: Descriptive architecture of NemoMapPy program

Initially, NemoMapPy only worked on undirected graph and the input file should have integer-labeled vertices [1]. Also, it only provided the frequency of the query pattern as a result. Here, we improved the initial NemoMapPy by adding the following features. (1) NemoMapPy now works both on directed and undirected graphs; (2) the vertices can be integer or string; (3) the query pattern's statistical significance is provided along with its frequency; (4) the input format of query graph pattern has three options. Input text, a file, or g6 format; and (5) the instances of the query pattern is provided as a file.

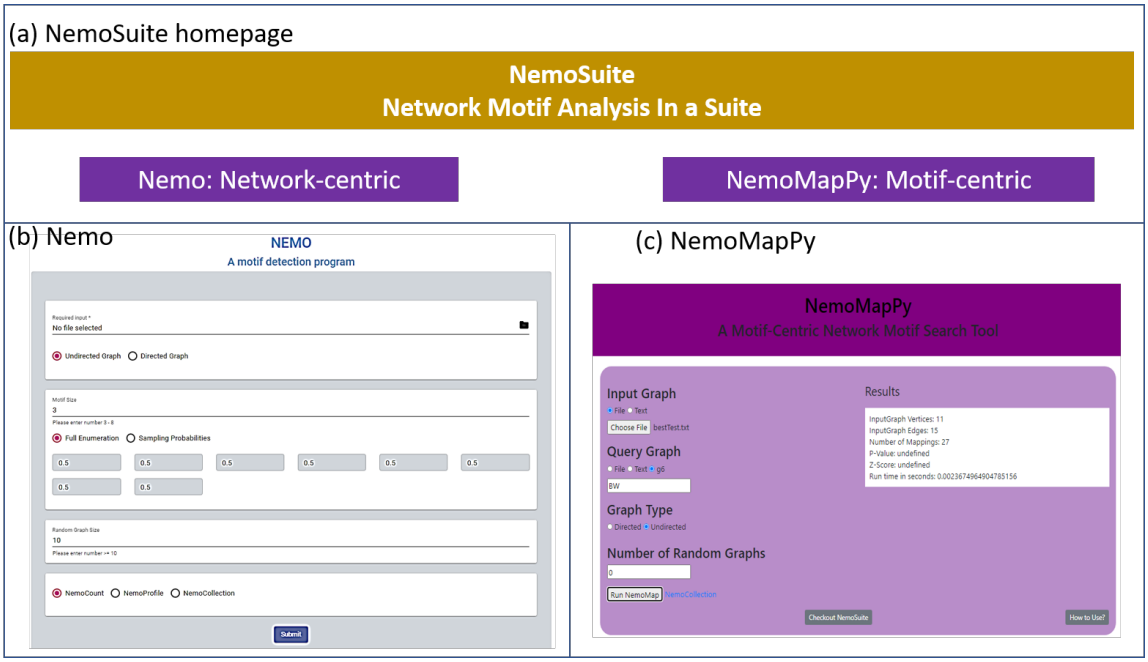


Figure 5: NemoSuite: (a) is the homepage view of NemoSuite program; (b) is the homepage of Nemo program with network-centric method; (c) shows the homepage for NemoMapPy with motif-centric method.

4 Experiments

NemoSuite is an online network motif detection program with both of network-centric and motif-centric methods. It is designed to satisfy the following functional and non-functional requirements.

Functional Requirements

- The user can use the program with any web browser.
- The user can upload the data file, which is an undirected or directed graph.
- The user can use the data file with string or integer type of nodes.
- The user can use both programs to detect the frequency of network motifs (NemoCount).
- The user can use both programs to obtain the instances of network motifs (NemoCollection).
- The user can use the Nemo program to obtain the network motif profile (NemoProfile).

Non-functional Requirements

- The returned results should be clear and easy to understand.
- The running time should be reasonably fast based on the size of the input data.
- The program works for any devices that have access to the internet and have a web browser.

In other words, as a web program, NemoSuite provides easy accessibility for users to detect network motifs, and apply them for various real-world problems. With additional features, the program provides good usability, functionality and computational efficiency. In the following section, we provide results of the comparison of NemoSuite with other programs to demonstrate its computational efficiency and good usability.

We compared Nemo with FANMOD [34] program, since both follow the motif-centric ESU algorithm in essence. Also, after reviewing all other tools, we found FANMOD is efficient and the most usable program among all others. Therefore, we compared Nemo program with FANMOD to evaluate the accessibility, usability, functionality and computational efficiency.

NemoMapPy is implementing a motif-centric approach, which is provided by Grochow and Kellis (GK), called GK method. Unlike network-centric methods, few programs are available that follow GK method, such as, MODA [35], ParaMODA, Fascia [36], ISMAGS[31] and NemoMap. They are available as source codes, cytoscape plug-in, or executable programs. However, the standalone MODA program is out-dated and had many restrictions on running the program, ISMAGS is cytoscape plug-in, and Fascia is a parallel and command-based program. Therefore, we compared NemoMapPy with NemoMap for its efficiency, assuming its easy accessibility and good usability are self-demonstrated with its web-interactive features.

5 Results and Analysis

5.1 Nemo vs. FANMOD

As a web program, Nemo program can be launched with any web-browser. Therefore it is easy to access, and easy to use. FANMOD

program is a stand-alone GUI (graphical-user-interface) program that can be downloaded from a website and it can be installed in Windows, Mac, and Linux operating systems. While both programs have good accessibility, Nemo has a better accessibility as the user can use it anytime and anywhere as long as there is an internet connection.

Nemo and FANMOD programs have good usability too. Usability measures "the capacity of a system to provide an environment for the users to perform the tasks safely, effectively, and efficiently [46]." With Nemo and FANMOD, network motifs are easily detected given an input graph, the size of motif, and the number of random graphs to generate. FANMOD has more options for the random graphs, and the user can see the progress of the task. However, Nemo program performs faster than FANMOD and has more input and output options.

Table 4: Input graph files used for testing

Type	**biological**			**collaboration**		**social**
File	Y2k	Y11k	MIPS	CA-GrQc	CA-HepTH	Face-book
#vertex	988	2401	4545	5241	9875	4545
#edge	2455	11000	12317	14484	25973	12317

We tested the performance of Nemo and FANMOD program with three biological networks, two collaboration networks, and one social network. The detail features of testing networks are described in Table 4. When we tested the run-time for network motif detection with different graphs, the size of network motif is set as four, and the number of random graphs as 100. All tests are performed in Windows OS. Table 5 and Figure 6 provide the comparison results.

Table 5: Nemo vs FANMOD run-time (rounded in seconds) comparison on the size of input graph

	Y2k	**Face-book**	**CA-GrQc**	**CA-HepTH**	**Y11k**	**MIPS**
Nemo	24	54	201	205	517	551
FANMOD	72	248	607	561	2,189	2,859

We can observe that Nemo runs relatively stable as the network size grows, while the run-time of FANMOD program increases abruptly. For the Y11k and MIPS networks, Nemo program runs roughly 5 times faster than FANMOD program.

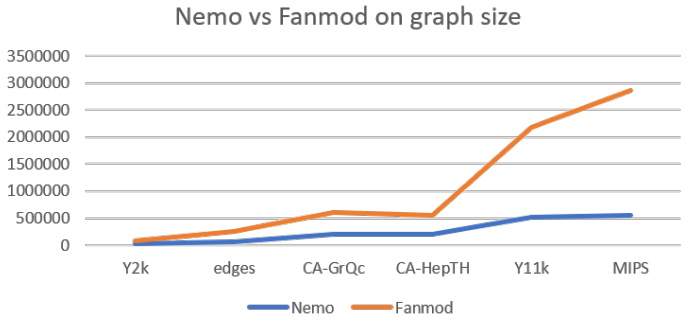


Figure 6: Graph of runtime vs. input graph size between Nemo and FANMOD in millisecond(ms)

We also tested the performance of two programs based on the size of motifs. To save time, we used Y2k data with only 10 number of random graphs, and used Windows OS.

Table 6: Nemo vs FANMOD run-time (in seconds) comparison on the size of motif

	3	**4**	**5**	**6**
Nemo	0.1	2.4	60.8	1,580.5
FANMOD	0.3	7.6	208.5	6,946.9

Table 6 and Figure 7 provide the results. We can observe that Nemo program runs consistently faster than FANMOD program, although the run-time of both programs still increases steeply with the increment of the size of motifs.

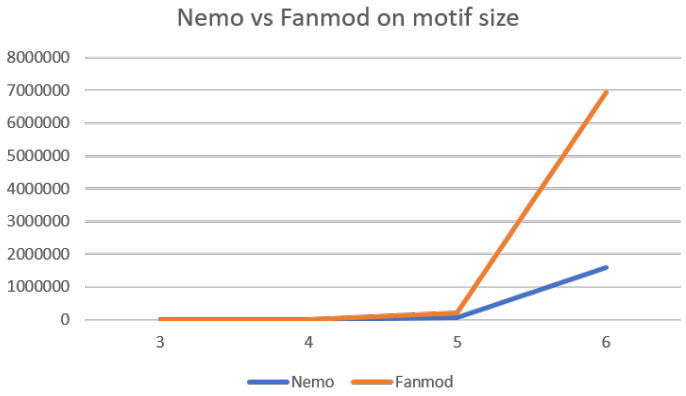


Figure 7: Graph of runtime vs. motif size between Nemo and FANMOD in millisecond(ms)

Through various experiments, we could see that Nemo program provides better accessibility, functionality, usability and computational efficiency than FANMOD program.

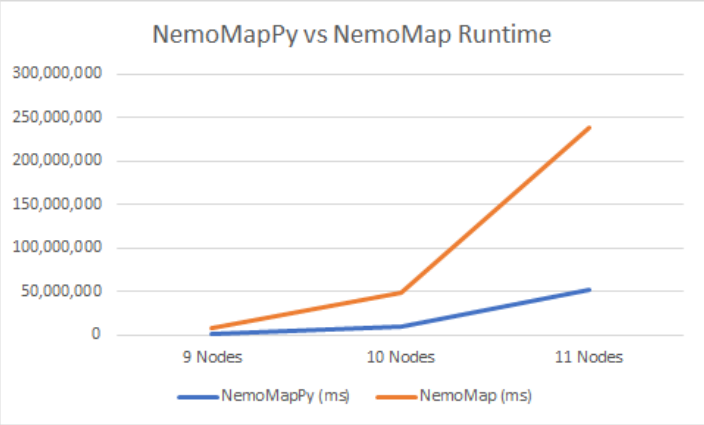


Figure 8: Graph of runtime vs. query graph size between NemoMapPy and NemoMap

5.2 *NemoMapPy vs. NemoMap*

We tested the efficiency of NemoMapPy by comparing with the NemoMap, as NemoMap is known the fastest program with motif-centric approach [17] in most of cases, when not considering parallelization. We tested NemoMapPy with a graph of 1,018 nodes and 1,331 edges, the query graphs that has 9, 10, 11 nodes with average complexity. The complexity of query graph was measured in node-to-edge ratio. We obtained NemoMap source code from the authors of [17] and compiled with Visual Studio 2017. We compared the performance with run-time.

The testing environment was on a desktop computer in Windows OS. In summary, NemoMapPy is approximately 70% faster than NemoMap on the graphs with average complexity and average density, as shown in Table 7, and Fig. 8.

Table 7: NemoMapPy vs NemoMap run-time (in seconds) Comparison

Query Graph	9 Nodes	10 Nodes	11 Nodes
NemoMapPy	1,728,213	9,923,137	51,904,930
NemoMap	8,599,994	48,674,827	239,088,982
% Reduction	79.90%	79.61%	78.29%
# Mappings	13,882,459	251,584,995	1,070,991,776

6 Conclusion

In this paper, we present NemoSuite that provides a combination of easy accessibility, usability, functionality, extensibility, and efficiency that the current network motif detection software programs lack. More specifically, we believe NemoSuite contribute greatly to development of network motif detection programs in the following reasons.

As a web-interactive program, it is easily accessible, and highly usable tool with great efficiency. The various input and output options enable for the users to apply the results to various real-world problems. It is also the first tool that provides both of network-centric and motif-centric approaches. We also demonstrated that NemoSuite is substantially designed to meet the functional and non-functional requirements, and showed its computational efficiency by comparing with other available tools.

Currently, NemoSuite includes Nemo and NemoMapPy programs to represent network-centric and motif-centric methods. Source codes of both programs are available in public github: Nemo is available at https://github.com/lytbfml/Nemolib_App and NemoMapPy at https://github.com/zicanl/NemoMapPy. We plan to add more online tools implementing recent algorithms such as QuateXelero [30], and utilize distributed computing to improve its efficiency further. Additional future works include adding more utility programs such as motif-visualization, graph file conversion, and generating random networks with various methods.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment We thank the Computing and Software Systems, School of STEM, at the University of Washington Bothell for the support of this research.

References

[1] P. Mar, W. Kim, "NemoMapPy: Motif-centric network motif search on a web," in 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 1226–1228, 2019, doi:10.1128/mBio.01377-14.

[2] I. Albert, R. Albert, "Conserved network motifs allow protein-protein interaction prediction." Bioinformatics, **20**(18), 3346–3352, 2004.

[3] J. Chen, W. Hsu, M. L. Lee, S.-K. Ng, "Labeling network motifs in protein interactomes for protein function prediction," Data Engineering, International Conference on, **0**, 546–555, 2007.

[4] Y. Zhang, J. Xuan, B. G. de los Reyes, R. Clarke, H. W. Ressom, "Network motif-based identification of breast cancer susceptibility genes," in 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 5696–5699, IEEE, 2008.

[5] G. C. Conant, A. Wagner, "Convergent evolution of gene circuits," Nature genetics, **34**(3), 264–266, 2003, doi:http://dx.doi.org/10.1038/ng1181.

[6] S. Wuchty, Z. N. Oltvai, A.-L. Barabasi, "Evolutionary conservation of motif constituents in the yeast protein interaction network," Nature genetics, **35**(2), 176–179, 2003.

[7] W. Kim, "Prediction of essential proteins using topological properties in GO-pruned PPI network based on machine learning methods," Tsinghua Science and Technology, **17**(6), 645–658, 2012.

[8] W. Kim, M. Li, J. Wang, Y. Pan, "Essential Protein Discovery based on Network Motif and Gene Ontology," in Proceedings of IEEE Bioinformatics and Biomedicine, 470–475, 2011.

[9] R. Dobrin, Q. K. Beg, A.-L. Barabasi, Z. N. Oltvai, "Aggregation of topological motifs in the Escherichia coli transcriptional regulatory network," BMC Bioinformatics, **5**, 10, 2004.

[10] L. Zhang, O. King, S. Wong, D. Goldberg, A. Tong, G. Lesage, B. Andrews, H. Bussey, C. Boone, F. Roth, "Motifs, themes and thematic maps of an integrated Saccharomyces cerevisiae interaction network," Journal of Biology, **4**(2), 6, 2005.

[11] N. Przulj, D. G. Corneil, I. Jurisica, "Modeling interactome: scale-free or geometric?" Bioinformatics, **20**(18), 3508–3515, 2004.

[12] W.-P. Lee, B.-C. Jeng, T.-W. Pai, C.-P. Tsai, C.-Y. Yu, W.-S. Tzou, "Differential evolutionary conservation of motif modes in the yeast protein interaction network," BMC Genomics, **7**(1), 89, 2006.

[13] Z.-R. Xie, M.-J. Hwang, "An interaction-motif-based scoring function for protein-ligand docking," BMC Bioinformatics, **11**(1), 298, 2010.

[14] A. S. Konagurthu, A. M. Lesk, "On the origin of distribution patterns of motifs in biological networks," BMC Systems Biology, **2**(1), 1–8, 2008.

[15] A. Andersen, W. Kim, NemoLib: A Java Library for Efficient Network Motif Detection, 403–407, Bioinformatics Research and Applications: 13th International Symposium, ISBRA 2017, Honolulu, HI, USA, May 29 – June 2, 2017, Proceedings, Springer International Publishing, Cham, 2017, doi: 10.1007/978-3-319-59575-7_42.

[16] W. Kim, Z. A. Brader, "NemoLib: Network Motif Libraries for Network Motif Detection and Analysis," in Z. Cai, I. Mandoiu, G. Narasimhan, P. Skums, X. Guo, editors, Bioinformatics Research and Applications, 327–334, Springer International Publishing, Cham, 2020.

[17] T. Huynh, S. Mbadiwe, W. Kim, "NemoMap : Improved Motif-centric Network Motif Discovery Algorithm," Advances in Science, Technology and Engineering Systems Journal, **3**(5), 186–199, 2018, doi:10.25046/aj030523.

[18] S. Yu, Y. Feng, D. Zhang, H. D. Bedru, B. Xu, F. Xia, "Motif discovery in networks: A survey," Computer Science Review, **37**, 100267, 2020, doi: https://doi.org/10.1016/j.cosrev.2020.100267.

[19] M. R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, "Network Motifs: Simple Building Blocks of Complex Networks," Science, **298**(5594), 824–827, 2002.

[20] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, U. Alon, "Superfamilies of Evolved and Designed Networks," Science, **303**(5663), 1538, 2004, doi:10.1126/science.1089167.

[21] V. Lacroix, C. G. Fernandes, M. F. Sagot, "Motif search in graphs: application to metabolic networks," IEEE/ACM transactions on computational biology and bioinformatics, **3**(4), 360–368, 2006, doi:10.1109/TCBB.2006.55[doi].

[22] L. Parida, "Discovering Topological Motifs Using a Compact Notation," Journal of Computational Biology, **14**(3), 300–323, 2007.

[23] B. H. Junker, F. Schreiber, Analysis of Biological Networks, Wiley, 2008.

[24] J. A. Grochow, M. Kellis, "Network motif discovery using subgraph enumeration and symmetry-breaking," in Proceedings of the 11th annual international conference on Research in computational molecular biology, RECOMB'07, 92–106, Springer-Verlag, Berlin, Heidelberg, 2007.

[25] F. Schreiber, H. Schwobbermeyer, "Mavisto: a tool for the exploration of network motifs," Bioinformatics, **21**, 3572–3574, 2005.

[26] N. Kashtan, S. Itzkovitz, R. Milo, U. Alon, "Efficient sampling algorithm for estimating sub-graph concentrations and detecting network motifs," Bioinformatics, **20**, 1746–1758, 2004.

[27] Z. Kashani, H. Ahrabian, E. Elahi, A. Nowzari-Dalini, E. Ansari, S. Asadi, S. Mohammadi, F. Schreiber, A. Masoudi-Nejad, "Kavosh: a new algorithm for finding network motifs," BMC Bioinformatics, **10**(1), 318, 2009.

[28] J. Chen, W. Hsu, M. Lee, S. Ng, "NeMoFinder: Dissecting genome-wide protein-protein interactions with meso-scale network motifs," Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, 106–115, 2006.

[29] X. Li, D. S. Stones, H. Wang, H. Deng, X. Liu, G. Wang, "NetMODE: Network Motif Detection without Nauty," PLoS ONE, **7**(12), e50093, 2012, doi: 10.1371/journal.pone.0050093.

[30] S. Khakabimamaghani, I. Sharafuddin, N. Dichter, I. Koch, A. Masoudi-Nejad, "QuateXelero: An Accelerated Exact Network Motif Detection Algorithm," PLOS ONE, **8**(7), 1–15, 2013, doi:10.1371/journal.pone.0068073.

[31] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks," Genome research, **13**(11), 2498–2504, 2003.

[32] I. Y. Smoly, E. Lerman, M. Ziv-Ukelson, E. Yeger-Lotem, "MotifNet: a web-server for network motif analysis," Bioinformatics, **33**(12), 1907–1909, 2017.

[33] S. Wernicke, "Efficient Detection of Network Motifs," IEEE/ACM Trans.Comput.Biol.Bioinformatics, **3**(4), 347–359, 2006.

[34] S. Wernicke, F. Rasche, "FANMOD: a tool for fast network motif detection," Bioinformatics, **22**, 1152–1153, 2006.

[35] S. Omidi, F. Schreiber, A. Masoudi-Nejad, "MODA: An efficient algorithm for network motif discovery in biological networks," Genes and Genetic Systems, **84**(5), 385–395, 2009.

[36] G. M. Slota, K. Madduri, "Fast Approximate Subgraph Counting and Enumeration," in 2013 42nd International Conference on Parallel Processing, 210–219, 2013, doi:10.1109/ICPP.2013.30.

[37] S. Mbadiwe, W. Kim, "ParaMODA: Improving motif-centric subgraph pattern search in PPI networks," in 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 1723–1730, 2017, doi:10.1109/BIBM.2017.8217920.

[38] A. Masoudi-Nejad, F. Schreiber, Z. R. Kashani, "Building blocks of biological networks: a review on major network motif discovery algorithms," IET systems biology, **6**(5), 164–174, 2012, doi:10.1049/iet-syb.2011.0011[doi].

[39] D. Jain, R. Patgiri, "Network Motifs: A Survey," in M. Singh, P. Gupta, V. Tyagi, J. Flusser, T. Ören, R. Kashyap, editors, Advances in Computing and Data Sciences, 80–91, Springer Singapore, Singapore, 2019.

[40] W. Kim, L. Haukap, "NemoProfile: Effective representation for network motif and their instances," in 12th International Symposium on Bioinformatics Research and Applications (ISBRA), 2016.

[41] B. McKay, "Practical graph isomorphism," Congr Numer, **30**, 45–87, 1981.

[42] Y. Wang, W. Kim, "Identification of Essential Genes with NemoProfile and Various Machine Learning Models," in Z. Cai, I. Mandoiu, G. Narasimhan, P. Skums, X. Guo, editors, Bioinformatics Research and Applications, 319–326, Springer International Publishing, Cham, 2020.

[43] A. Ronacher, "Flask," https://flask.palletsprojects.com, 2010.

[44] W. Reese, "Nginx: The high-performance web server and reverse proxy," Linux J., **2008**(173), 2008.

[45] B. Chesneau, "gunicorn," https://gunicorn.org/, 2009.

[46] J. Y. Lee, J. Y. Kim, S. J. You, Y. S. Kim, H. Y. Koo, J. H. Kim, S. Kim, J. H. Park, J. S. Han, S. Kil, H. Kim, Y. S. Yang, K. M. Lee, "Development and Usability of a Life-Logging Behavior Monitoring Application for Obese Patients," Journal of Obesity & Metabolic Syndrome, **28**(3), 194–202, 2019, doi:10.7570/jomes.2019.28.3.194.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1554-1559 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Application of Deep Belief Network in Forest Type Identification using Hyperspectral Data

Xianxian Luo[1,2], Songya Xu[*,3], Hong Yan[4]

[1]*Faculty of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou, 362000, China*

[2]*Fujian Provincial Key Laboratory of Data Intensive Computing, Quanzhou, 362000, China*

[3]*Faculty of Educational Science, Quanzhou Normal University, Quanzhou, 362000, China*

[4]*Fujian Forest Inventory and Planning Institute, Fuzhou, 350000, China*

ARTICLE INFO

Article history:
Received: 23 August, 2020
Accepted: 08 December, 2020
Online: 25 December, 2020

Keywords:
Deep Learning
Deep Belief Network
Forest Mapping
Hyperspectral Data

ABSTRACT

Forest mapping by remote sensing is a hot topics in forestry. At present, many researchers focus on the research of forest type classification or tree species identification using different machine learning methods and try to improve the accuracy of classification of satellite image. However, forest type classification using deep belief network (DBN) is still limited in previous literatures. Our research focuses on forest mapping in the western part of Dehua county in southern China. Most important objective was to assess the feasibility of forest mapping from hyperspectral data using deep learning. The HJ-1A hyperspectral data was adopted in this paper. We applied deep belief network and got a thematic map of four forest types, such as coniferous forest, broad-leaved forest, mixed forest and non-forest. Our finding shows that optimal network depth of DBN model is 3 and best node in each layer is 256 in our experiment. Overall accuracy is 85.8% and kappa coefficient is 0.785 with best-fit parameters in DBN model, while for SVM is 73% and 0.6447 respectively. DBN obtain better performance compared with support vector machine. Furthermore, network depth and number of nodes in each hidden layer in DBN model has a significant effect on overall accuracy and Kappa coefficient. In general, DBN is promised to be dominant method of forest mapping by hyperspectral data.

1. Introduction

It is very important to obtain the forest information timely and correctly for forest management and land cover mapping in forest ecosystems. For example, forest mapping plays an important role in the issues of forest restoration and forest degradation assessment [1]-[4]. Nowadays, it is an effective way to obtain forest structure from satellite image [5], [6]. Hyperspectral image is a typical data of satellite image and widely used in forest mapping because there are lots of spectral bands of this image and beneficial for forest type recognition. What's more, A series of achievements have been made based on hyperspectral remote sensing technology [7], [8]. However, it leads to the Hughes phenomenon because of redundant spectrum of hyperspectral image [9]. Normally, it needed to be reduced by different methods before image classification [10]. After dimension reduction, traditional classifiers were adopted, such as random forests [6] and support vector machines [7], SVM is the most popular and nonparametric supervised learning method. It is widely used in classifying forest ecosystems [11]. Fortunately, deep learning can extract feature automatically, past studies have showed that deep learning can achieve higher accuracy of image processing of remote sensing avoiding human intervention [12]. Deep belief network is typical model of deep learning. There is an input layer, one output layer and several restricted in Boltzmann machines [13]. Pre-training and fine-tuning are two important steps in DBN model. Pre-training is an unsupervised learning from bottom to up and output result with a classifier. Backpropagation was coined by Rumbelhart [14] and used to update the parameter of model for better result.

There are lots of references about image classification using DBN. It seems that DBN was first adopted by Lü et al. in 2014 year [15]. The experiments showed that the optimal network depth was three hidden layers and best nodes of each hidden layer were

*Corresponding Author: Songya Xu, 29974817@qq.com

64. In year 2015, Chen et al. created DBN model combing with principal component analysis and logistic regression. They confirmed better effect of DBN than that of SVM [12] adopting two Hypersperctral datasets, such as Indian Pines and University of Pavia. On the whole, there are a lot of application situation of deep learning [16], but few literatures is reported about forest mapping using deep learning from Hyperspectral image.

Therefore, one of our research objectives is to confirm the performance of forest mapping by DBN using hyperspectral data. Another objective of this study is to compare the classification performance of DBN model and SVM algorithm. Third objective is to investigate how optimal parameters of DBN affect overall accuracy and Kappa coefficient of forest type identification. The major contribution is to determine the optimal parameters of DBN for forest mapping from Hyperspectral data.

2. Methods

2.1. Restricted Boltzmann Machines (RBM)

RBM is a directionless, full connective model [17]. A RBM consists of two layers. One of them is an input layer, or visible units *v*. Output layer, or hidden units *h* is another layer of RBM. There are no connections within each layer. From Figure 1, we know that the units between visible and hidden are full connected by weight W. The energy of a joint configuration with *v* visible units and *h* hidden units is

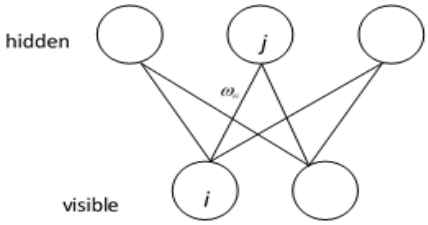


Figure 1: The model of an RBM

$$E(v,h|\theta)=-\sum_{i=1}^{I}a_i v_i-\sum_{j=1}^{J}b_j h_j-\sum_{i=1}^{I}\sum_{j=1}^{J}v_i h_j \omega_{ij} \quad (1)$$

Where $\theta=\{a_i,b_j,\omega_{ij}\}$ represents parameters of RBM. ω_{ij} is the weight between input layers *i* and hidden layers *j*. a_i is the bias of v_i, b_j is the bias of h_j. *I* and *J* are total numbers of input and output layers respectively. The probability of the joint configuration over both layers is

$$P(v,h|\theta)=\frac{e^{-E(v,h|\theta)}}{Z(\theta)} \quad (2)$$

where $Z(\theta)$ is the normalized factor, equals to $\sum_{v}\sum_{h}e^{-E(v,h|\theta)}$.

In the binary case where $h_j\in\{0,1\}$, given input layer, a conditional probability of an output layer being 1 is given as:

$$P(h_j=1|v,\theta)=\sigma\left(b_j+\sum_{i=1}^{I}v_i\omega_{ij}\right) \quad (3)$$

In the same way, the conditional probability of input layer given the output layer:

$$P(v_i=1|h,\theta)=\sigma\left(a_i+\sum_{j=1}^{J}h_j\omega_{ij}\right) \quad (4)$$

where $\sigma(x)$ is a function, such as a sigmoid function,

$$\sigma(x)=\frac{1}{1+e^{-x}}.$$

Normally, total error *E* is minimized to acquire the most favorable parameters θ, RBM.

$$E=\frac{1}{2}\sum_{k}^{K}(d_k-o_k)^2 \quad (5)$$

where the d_k stands for the true outputs, o_k corresponds real outputs. In general, there are *K* outputs. Gradient descent is a good method to minimize total error, it is ecessary to compute the partial derivative of *E* with respect to each weight in the network. The parameters, such as weights and biases will be iterated by stochastic gradient ascent which can be formulated as [18]:

$$\Delta\omega_{ij}=\varepsilon\left(\langle v_i h_j\rangle_{data}-\langle v_i h_j\rangle_{recon}\right) \quad (6)$$

$$\Delta a_i=\varepsilon\left(\langle v_i\rangle_{data}-\langle v_i\rangle_{recon}\right) \quad (7)$$

$$\Delta b_j=\varepsilon\left(\langle h_j\rangle_{data}-\langle h_j\rangle_{recon}\right) \quad (8)$$

where ε is a key parameter, that is learning rate. $\langle\bullet\rangle_{data}$ is expectation for data distribution and $\langle\bullet\rangle_{recon}$ is the expected value of the model distribution.

2.2. Deep Belief Network

Several RBMs stacked into a classical model of deep belief network. There are two important steps in the DBN [19]. One is pre-training and the other is fine-tuning, shown in the Figure 2. Initial parameters are produced in the pre-treatment by an unsupervised method, such as gradient descent method to avoid local optimum. Back-propagation is normally used to update model parameters according to error, such as squared error or cross entropy error.

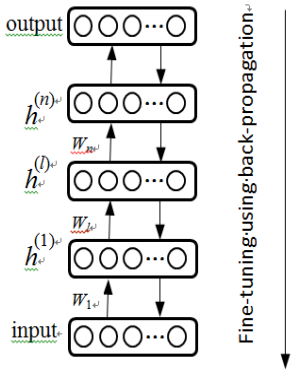


Figure 2: A generative model of deep belief network

2.2.1. Pre-training

Pre-training of DBN model is an unsupervised procedure. At the beginning, the input layer is initialized to a training vector. Then the subsequent hidden layer of RBM is trained through the output of the previous layer. Moreover, this whole process can be repeated until the final hidden layer. By doing so can the DBN extract more and deep feature from the original input data.

2.2.2. Fine-Tuning using Backpropagation algorithm

Backpropagation was coined by Rumbelhart [20]. Backpropagation update model parameters using cross entropy error because there is four kinds of outputs in this paper.

$$\delta_k = o_k(1-o_k)(d_k - o_k) \tag{9}$$

For the *l*th hidden layer,

$$\delta_i^l = (1-o_i)o_i \sum_j^n \omega_{ij}^l \delta_j^{l+1} \tag{10}$$

The batch update rules are

$$\omega_{ij}^l = \omega_{ij}^l + \varepsilon_{fine-tuning} o_i \delta_j^{l+1} \tag{11}$$

$$b_j^l = b_j^l + \varepsilon_{fine-tuning} \delta_j^{l+1} \tag{12}$$

2.3. Indicators of Evaluation of Classification

- *Overall accuracy*

The overall accuracy is total assessment of classified quality, which equals the total pixels sum divided by correctly classified pixels by total pixels. The formula for calculating the overall accuracy based on the confusion matrix can be listed as following:

$$p = \sum_{i=1}^{c} m_{ii} / N \tag{13}$$

here, *C* represents the number of categories, and m_{ii} represents the elements on the diagonal of the confusion matrix, and *N* represents the total number of test samples.

- *Kappa coefficient*

Kappa coefficient adopts a multivariate discrete analysis technique to reflect the consistency between classification results and reference data. It considers all factors of the confusion matrix, and it is a more objective evaluation index, which is defined as:

$$k = \frac{N\sum_{i=1}^{c} m_{ii} - \sum_{i=1}^{c}(m_{i+}m_{+i})}{N^2 - \sum_{i=1}^{c}(m_{i+}m_{+i})} \tag{14}$$

Among them, m_{i+} and m_{+i} represent the sum of the line *i* of the confusion matrix and the sum of the column *i* respectively. The higher the Kappa coefficient, the higher the classification precision.

3. Data and Process

3.1. Study site and hyperspectral data

Hyper Spectral Imager (HSI) is the first Chinese space-borne hyperspectral sensor aboard the HJ-1A satellite. In September 6, 2008, China launched the environmental and disaster monitoring and forecasting small satellite constellation A (HJ-1A), and the HJ-1A satellite was equipped with CCD camera and hyperspectral imager (HSI). With HSI interference imaging spectroscopy, the cutting width is 50km, the ground pixel resolution is 100 meters, 110-128 spectral bands, the spectral resolution of up to 2.08 nm, in the continuous spectral image of earth observation and available features to achieve the direct identification from the space object surface. HSI often produces certain stripe noise on the image. Satellite data was HJ-1A star HSI data 2 level products, imaging time is August 24, 2011, a total of 115 bands with working spectrum range 459 ~ 956 nm.

The research area has 8 downtowns in the west of Dehua county, Quanzhou City, Fujian Province. The administrative area of study site and its pseudo color synthetic imaging (105th, 7th, 40th band for pseudo colour synthesis) are shown in Figure 3. The image data of the product is geometrically corrected by previously corrected image with polynomial transform, and then orthorectified and outputted onto a fixed uniform spatial grid using nearest neighbour resampling. The correction error is not less than one pixel. The corrected image is unified into the specified map projection coordinate system (Xi'an 1980 coordinate system). There is obvious stripe noise in the part of the HSI image data, mainly in bands 1-29. Therefore, these 29 bands are eliminated from the HSI image. The remaining 86 bands are used for research within the wavelength range (529.6350-951.54 nm).

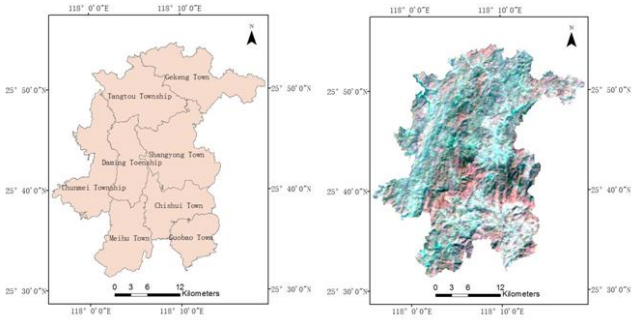


Figure 3: Study area and pseudo colour synthetic image

3.2. Experimental processing flow

According to the field data, the labelled samples were selected in the experiment. Then 86 bands of data were used as input of DBN. Before the training, these data were normalized by min-max normalization, which was mapped into the value between 0 and 1. Then, these data were shuffled and split into training and testing data randomly. After pre-training and fine tuning, categorical data was represented by one-hot coding, as shown in Table 1, 1 was Coniferous forest, 2 was Broad-leaved forest, 3 was Mixed forest, 4 was Non-forest.

The experiment is based on Windows 10 with 64-bit operating system. The processor is Intel (R) Core (TM) i5-8250U CPU @1.60GHz. The experiment is carried out in PyCharm 2018.3x64

editor to code and tune parameters. TensorFlow 1.11.0 framework was first set up as development environment. Also, The Python Extension Library is loaded, including Numpy, Pandas, Matplotlib, etc.

Then the selected samples were converted into a CSV file that is easy to be processed by Python program. All these samples were shuffled into training and testing data. Moreover, training samples were thrown into the DBN model. The trained parameters of model were stored into Tensorboar. Finally, the classified map of was outputted.

3.3. Distribution of samples

There are total 97258 pixels in the pseudo color synthetic image, shown in Table 1. The samples of four types in study site are selected manually. Different allocation of training samples and test samples are selected by experiments again and again. At last, 28000 pixels of known categories were selected as total samples. Among them, 51989 pixels pertain to the coniferous forest, 6142 pixels are broad-leaved forest, 16283 pixels are mixed forest, and the remaining 28986 pixels are non-forest. In the training process, there are 10,000 training samples for coniferous forest and 6,000 training samples for other forest type.

Table 1: Class code of dataset

Code of label	Forest type	Number of training samples	Number of testing samples	Number of all samples
1	Coniferous forest	10000	4000	45848
2	Broad-leaved forest	6000	4000	6142
3	Mixed forest	6000	4000	16283
4	Non-forest	6000	4000	28985
	total	28000	16000	97258

4. Experimental Results

4.1. Selection of super parameters

In the experiment, the dimension of the input data is 86. At the moment, network structure of DBN model is only designed by experience without theoretical basis. For simplicity, we assume that each hidden layer has the same number of nodes. The number of layers of the DBN is selected from {2, 3, 4, 5, 6, 7}, and the number of hidden layer nodes is selected from {16, 32, 64, 128, 256, 512}. According to the pre-experiments and references, several parameters are set as following: pre-training and fine-tuning of the learning rate is 0.001, the size of mini-batch is 100. Note that 10000 iteration times are run due to stability. Next, we discuss the influence of different network depth and the number of hidden layer nodes on the classification effect by fixing all these super parameters.

4.2. Network depth

Overall accuracy (OA) and Kappa coefficient were normally selected as the evaluation of image classification of remote sensing. The higher OA and Kappa coefficient, the better effect of classification precision. As we all know, network depth is key parameter of DBN. It is the number of hidden layers. Reference 21 summarized the range of optimal network depth is from 2 to 3 based on previous researches.

In experiment, we keep super parameters fixed and set the number of nodes in each hidden layer 256. Six different numbers (2, 3, 4, 5, 6, 7) of network depth were tested. Figure 4 shows that DBN with 3 hidden layers performs best, which both overall accuracy and Kappa coefficient are the largest. There is no obvious feature that how network depth affects the OA and Kappa coefficient. Reference [21] pointed out that there is no perfect theoretical basis for network structure selection. The optimal parameters should be given by experiments in different applications.

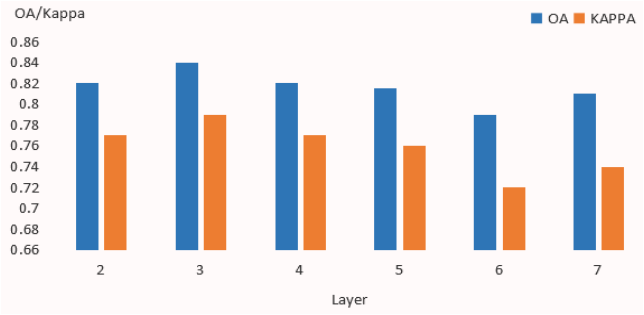


Figure 4: Effect of network depth on OA and Kappa coefficient

4.3. Number of nodes in each hidden layer

Number of nodes in the hidden layer is also import parameter in DBN model. When nodes are too large, it may cause the over-fitting problem. Oppositely, when nodes are too small, perhaps it will fail to extract deep information and gain high classification accuracy. Summarized by reference 21, the range of number of nodes in the hidden layer is from 50 to 500 according to previous research.

In experiment, we set super parameters unchanged and keep the network depth 3. Six various nodes (16, 32, 64, 128, 256, 512) were selected one by one and tested. Effect of number of nodes in each hidden layer on OA and Kappa coefficient is showed in Figure 5. It demonstrates that DBN with 256 nodes performs best, because it got the biggest overall accuracy and Kappa coefficient.

There is no obvious feature that how network depth affects the OA and Kappa coefficient. Reference [21-24] pointed out that there is no perfect theoretical basis for network structure selection. The optimal parameters should be given by experiments in different applications. Meanwhile, OA and Kappa coefficient keep relatively stable with changing nodes.

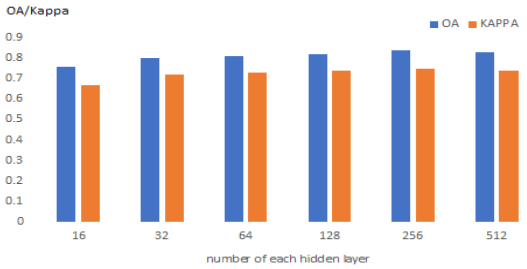


Figure 5: Effect of number of nodes in each hidden layer on OA and Kappa coefficient

4.4. Comparative analysis

SVM is a supervised and non-parametric machine learning algorithm. There is different type of kernels. A radial basis function (RBF) was adopted and has been used in some former works concerning forest type classification. The penalty factor C is selected from [1, 0.1, 0.001]. In order to keep the equit, the same samples are adopted between SVM and DBN method. Furthermore, to optimize of SVM method, five-fold cross validation and network search are adopted. When the C value is 1, the maximum overall accuracy is 73%, and the Kappa coefficient is 0.6447. For deep belief network, the optimal number of network depth is 3 and number of each node is 256.The classification process resulted in a thematic map of four kinds of forest types in Dehua county, as shown in Figure 6. As can be shown, the coniferous forest is green, broad-leaved forest is yellow, mixed forest is pink, and non-forest is blue.

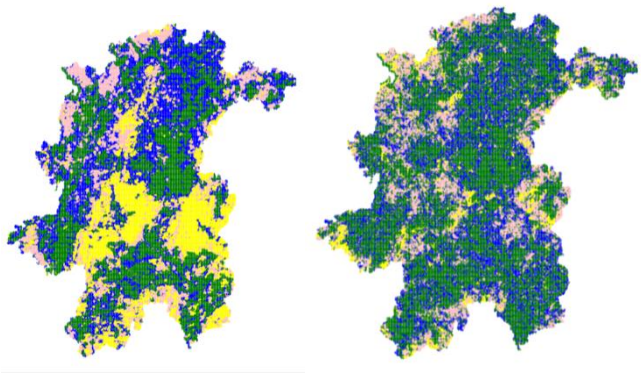
Table 2 reports the kappa values and the accuracy of forest types for both DBN and SVM classification algorithms. The accuracies of four kinds of forest types for SVM oscillated between 0.64 and 0.74 while DBN method varied from 0.82 to 0.89. Hyperspectral image classification using DBN algorithm yields the higher accuracy than that of SVM classifier for each forest type. Broad-leaved forest was generally better recognized than coniferous forest. Broad-leaved forest is classified with highest accuracy for both methods, 0.92 for DBN, while for SVM it is 0.84. The second top-classified forest type tends to be non-forest for DBN, while for SVM it is coniferous forest. Slightly lower result is obtained for mixed forest by DBN, while for non-forest by SVM. Furthermore, the worst effect performed by DBN algorithm is coniferous forest, compared to mixed forest for SVM. As the result, DBN model tends to attain better capability than SVM method from accuracy. There are two reasons due to it. One is that all feasible spectral features are thrown into input of DBN model, yet three bands or several components are selected as input for SVM method. The other is that there is large amount of data when all hyperspectal bands are considered. This is good for training of DBN model.

5. Conclusion and Discussion

This paper examined the capability DBN model in forest type classification with HJ/1A hyperspectral image. Many experiments are tested to obtain optimal parameter in DBN model. Conclusions can be drawing out as followed. At the beginning, the results showed that DBN model outperform SVM algorithm. Then, optimal network depth is 3 and best node in each layer is 256 in our experiment. Overall accuracy is 85.8% and kappa coefficient is 0.785 with best-fit parameters in DBN model.

Table 2: Comparison with SVM

Forest type	*SVM*	*DBN*
coniferous forest	0.74	0.82
broad-leaved forest	0.84	0.92
mixed forest	0.64	0.87
non-forest	0.70	0.89
OA	0.73	0.875
Kappa	0.6447	0.835



(a) SVM (b) DBN

Figure 6: Comparison of classification results based on different machine learning techniques

What makes our study unique is to make use of deep belief network to classify the forest type. Regardless of the spatial resolution of 100 m of HJ/1A hyperspectral image, our results provide satisfied recognition of four kinds of forest types. There are several open questions and future research directions remain worthy of investigation in the future. Firstly, in this paper, it is novative to adopt deep belief network to identify forest types. The classification accuracy of this method are better than that of support vector machine. However, it is still unclear that the mechanism of how this method can solve the traditional problem of "same object with different spectra" and "same spectra with different object ". Secondly, our results confirm the capability to map forest by deep belief network. Our experience in the study reveals that network depth and width affect accuracy of classification. The optimal structure needed to be adjusted according to different remote sensing data and further identification, such as tree species mapping. Finally, combining with field survey data, nearly one third of the samples are selected as training and testing samples to improve the classification effect. It is uncertain that how can we obtain the best result with so high number of total samples. It is a good idea to expand the sample by Generative Adversarial Network. Last but not least, optimal network structure is only for forest type recognition. The optimal network structure of specific forest type recognition needs further study. At the same time, it is urgent to establish the criteria and norms of forest classification using deep learning and remote sensing image.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The work was supported by the Project of Natural Science Foundation of Fujian Province under Grant No. 2020J01785, the project from the department of Quanzhou Science and Technology under Grant No. 2016N057 and the project from the department of education under Grant No. JAT170475. The author gratefully acknowledges the support of K.C. Wong Education and DAAD (No. 91551268). The authors also acknowledge the support by Fujian Provincial Key Laboratory of Data-Intensive Computing, Fujian University Laboratory of Intelligent Computing and Information Processing, and Fujian Provincial Big Data Research Institute of Intelligent Manufacturing.

References

[1] G. Modica, A. Merlino, F. Solano, R. Mercurio, "An index for the assessment of degraded Mediterranean forest ecosystems", Forest Systems, **24**, 1-3, 2015, doi : 10.5424/fs/2015243-07855.

[2] X.G. Kang, "Forest Management". Beijing: China Forestry Publishing House, 10-28, 2011.

[3] I.D. Thompson, M.R. Guariguata, K. Okabe, et al, "An operational framework for defining and monitoring forest degradation", Ecology and Society, **18**(20), 2013, doi: 10.5751/ES-05443-180220.

[4] L. Kurvonen, M.T. Hallikainen, "Textural information of multi-temporal ERS-1 and JERS-1 SAR images with applications to land and forest type classification in boreal zone", IEEE Transactions on Geoscience and Remote Sensing, **37**, 680-689, 1999, doi:10.1109/36.752185.

[5] Mei, A. X., Peng, W. L., Qin, Q.M. et al. "Introduction to remote sensing". Beijing: Higher education press, 225-286, 2006.

[6] A. Ghosh, F.E. Fassnacht, P.K. Joshi, B. Koch, "A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales". International Journal of Applied Earth Observation and Geoinformation, **26**, 49-63, 2014, doi:10.1016/j.jag.2013.05.017.

[7] Y.Y. Dian, Z.Y. Li, Y. Pang, "Spectral and Texture Features Combined for Forest Tree species Classification with Airborne Hyperspectral Imagery". J Indian Soc Remote Sens, **43**(1), 101-107, 2015, doi:10.1007/s12524-014-0392-6..

[8] N. Puletti, N. Camarretta, P. Corona, "Evaluating EO1-Hyperion capability for mapping conifer and broadleaved forests". European journal of remote sensing, 157-169, 2016doi:10.5721/eujrs20164909.

[9] G. Hughes, "On the mean accuracy of statistical pattern recognizers". IEEE transactions on information theory, **14**(1), 55-63, 1968, doi:10.1109/TIT.1968.1054102.

[10] R.E. Faßnacht, "Assessing the potential of imaging spectroscopy data to map tree species composition and bark beetle-related tree mortality". Freiburg: Freiburg university, 1-20, 2013, doi: 10.1016/j.rse.2013.09.014.

[11] M. Mustafa, A.R. James, S.J. Karl, R. David, M. Charles, "Remote Distinction of A Noxious Weed (Musk Thistle: CarduusNutans) Using Airborne Hyperspectral Imagery and the Support Vector Machine Classifier". Remote Sensing, **5**(2), 612-630, 2013, doi: 10.3390/rs5020612.

[12] Y.S. Chen, X. Zhao, X.P. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network". IEEE Journal of selected topics in applied earth observations and remote sensing. 1-12, 2015, doi:10.1109/JSTARS.2015.2388577.

[13] E. Alpaydin, Introduction to machine learning(third edition), The MIT press, 267-311, 2014.

[14] L. Deng, S.S. Fu, R.X. Zhang, "Application of deep belief network in polarimetric SAR image classification". Journal of image and graphics, **21**, 933-941, 2016, doi: 10.1109/IGARSS.2015.7326284.

[15] Q. Lü, Y. Dou, X. Niu, J.Q. Xu, F. Xia, "Remote sensing image classification based on DBN model". Journal of computer research and development. 51, 1911-1918, 2014, doi:10.7544/issn1000-1239.2014.20140199.

[16] P. Zhong, Z.Q. Gong, S.T. Li, et al., "Learning to Diversify Deep Belief Networks for Hyperspectral Image Classification". IEEE Transactions on Geoscience and Remote Sensing, 1-15, 2017, doi:10.1109/TGRS.2017.2675902.

[17] G.E. Hinton, J.S. Terrance, Learning and relearning in Boltzmann machines. *MIT Press, Cambridge*, 282-317, 1986.

[18] H.M. Xie, S. Wang, K. Liu, S.P. Lin, B. Hou. "Multilayer feature learning for polarimetric synthetic radar data classification". in IEEE International Geosicence and Remote Sensing Symposium. 2818-2821, 2014, doi: 10.1109/IGARSS.2014.6947062.

[19] G.E. Hinton, R.R. Salakhutdinov. "Reducing the dimensionality data with neural networks", Science, **313**, 504-507, 2006, doi: 10.1126/science.1127647.

[20] L. Deng, S.S. Fu, R.X. Zhang. "Application of deep belief network in polarimetric SAR image classification". Journal of image and graphics, **21**, 933-941, 2016, doi:10.11834/jig.20160711.

[21] C.C. Yao, X.X. Luo, Y.D. Zhao, et al, "A review on image classification of remote sensing using deep learning" in Proceedings of 2017 3rd IEEE Internaitonal Conference on Computer and Commmocaotpms, 1947-1955, 2017, doi: 10.1109/CompComm.2017.8322878.

[22] S.J. Ge, J. Lu, H. Gu, et al., "Polarimetrie SAR image classification based on deep belief network and superpixel segmentation." in 2017 3rd International Conference on Frontiers of Signal Processing, 115-119, 2017, doi: 10.1109/ICFSP.2017.8097153.

[23] M.M. Lau, J.T.S. Phang, K.H. Lim, "Convolutional Deep Feedforward Network for Image Classification." in 2019 7th International Conference on Smart Computing & Communications, 2019, doi:

[24] A. Mughees, L.Tao, "Multiple Deep-Belief-Network-Based Spectral-Spatial Classification of Hyperspectral Images." Tsinghua Science and Technology, **24**(2), 183-194, 2019, doi: 10.26599/TST.2018.9010043.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1560-1568 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

A Computational Modelling and Algorithmic Design Approach of Digital Watermarking in Deep Neural Networks

Revanna Sidamma Kavitha[1 *], Uppara Eranna[2], Mahendra Nanjappa Giriprasad[1]

[1]*Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University Anantapur, Ananthapuramu, 515002, India*

[2]*Department of Electronics and Communication Engineering, Ballari Institute of Technology and Management, Ballari-583104, India*

ARTICLE INFO

Article history:
Received: 21 October, 2020
Accepted: 12 December, 2020
Online: 25 December, 2020

Keywords:
Digital Water Marking
Neural Networks
Algorithms
Convolution Neural Network

ABSTRACT

In this paper we propose an algorithmic approach for Convolutional Neural Network (CNN) for digital watermarking which outperforms the existing frequency domain techniques in all aspects including security along with the criteria in the neural networks such as conditions embedded, and types of watermarking attack. This research addresses digital watermarking in deep neural networks and with comprehensive experiments through computational modeling and algorithm design, we examine the performance of the built system to demonstrate the potential of watermarking neural networks. The inability of intruder towards the retrieval of data without the knowledge of architecture and keys is also discussed and results of the proposed method are compared with the state of the art methods at different noises and attacks.

1. Introduction

The research presented here is the extension of the work originally presented at International Conference on Artificial Intelligence and Signal Processing (AISP),2020 [1]. The digital revolution and the internet have paved a way to the creation of massive digital information containing images, videos, transactions, intellectual properties. The ease with which this digital data can be copied and reproduced has created avenues for copyright infringements. The massive explosion of digital multimedia devices has resulted in the creation of a large chunk of data and increased demand (and role) of data hiding techniques. Digital watermarking is employed for various applications such as copyright protection (ownership assertion), broadcast monitoring (really broadcasted or not), tamper detection (persistent item identification, forgery detection), data authentication and verification (integrity verification), fingerprinting (transaction tracking and privacy control), content description (labelling and captioning), publication monitoring and copy control (unauthorized distribution) covert communication (data hiding), Medical applications (Annotation and privacy control), and Legacy system enhancement (backward compatibility) [2]. All information hiding techniques revolve around 3 parameters: Imperceptibility, robustness and payload capacity [3]. While payload capacity is important in steganography, in digital watermarking the trade-off between imperceptibility and robustness is needed for the excellent quality of data hiding [4]. The ability of a data hiding technique to remain unchanged to human perception is called imperceptibility i.e. an unintended user should not be able to make out whether the image has undergone watermarking [5]. Robustness is the ability of watermarked data to be immune to attacks and threats. Other goals of digital watermarking are Security (watermark should be secret and undetectable by an unauthorized parties), effectiveness (ease of detection immediately after embedding), uniqueness (multiple watermarks to coexist), cost effectiveness (in terms of hardware and computational speed), and scalability [6]. Several Artificial Intelligent techniques such as neural network, evolutionary computation, fuzzy logic, swarm intelligence, Probabilistic reasoning and multi-agent systems were proposed to secure watermark in digital media [7]. These techniques will be employed at either the transmitter side during embedding the watermark or during the extraction stage at receiver. Recent approaches have also attempted to incorporate AI methods in the pre-processing stage [8].

With such diverse approaches and their unparalleled capabilities, the hiding of data in the digital image will be far more superior in terms of imperceptibility and robustness as compared to the existing ones [9].

*Corresponding Author: R.S Kavitha, drkavithakavana@gmail.com

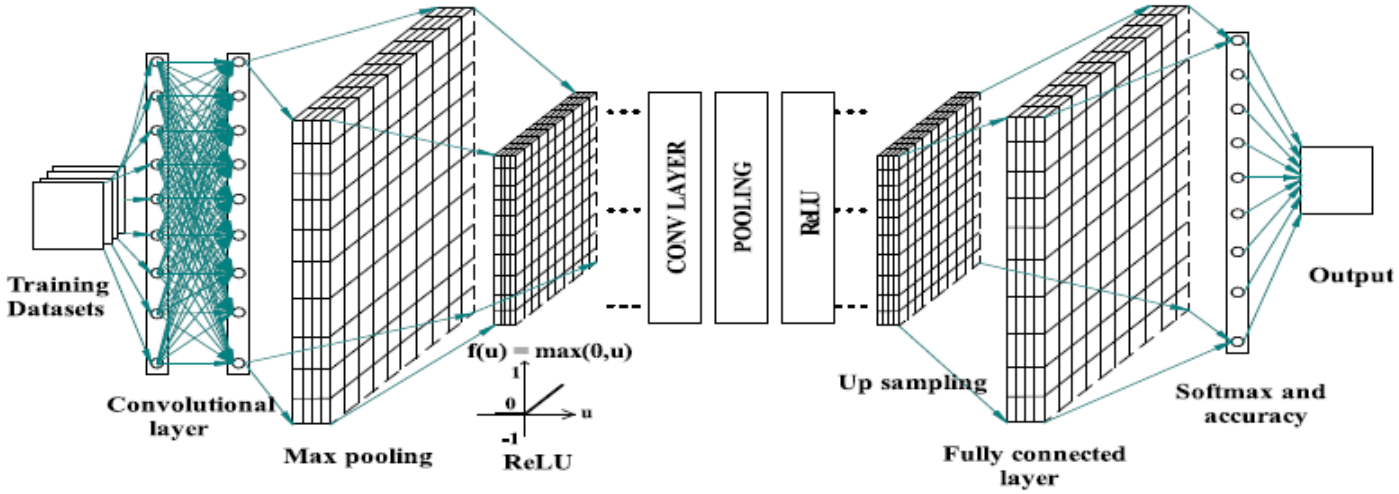


Figure 1: A typical digital watermarking process with Deep learning and its associated architectures

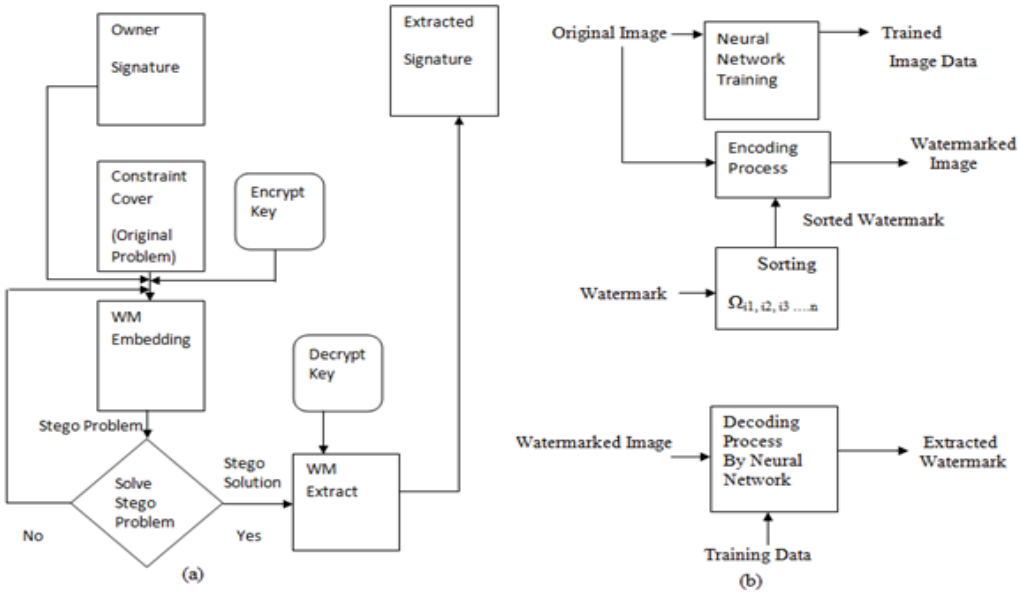


Figure 2: Experimental setup block diagram for digital watermarking process

A typical digital watermarking mechanism with embedding and extracting stages is depicted in Figure 1. It enables owners of the digital documents to embedded their copyright information for information security.

Digital watermarking can be done on text, image, audio, video and graphics in spatial or frequency domain. The watermark can be of noise type (pseudo noise, Gaussian random and chaotic sequences) or image type (binary image, stamp, logo and label) [10]. Based on the deployment conditions various watermarking techniques can be used. For public use, visible watermarks are preferred while for private applications and to arrest unauthorized copying invisible watermarking can be used [11]. Fragile watermarks are used in tamper-proof applications whereas robust watermarking is used in applications where the watermark should remain intact even after modification or tampered with [12]. According to the detection stage, visual watermarking is more robust which needs the original media and the embedded watermark for detection while blind watermarking does not require any of these. This is the most demanding type of watermarking as the watermarking is generated and embedded at the transmitter, while detection and extraction will happen at receiver as illustrated in Figure 2. Watermarking is also done even in prepossessing stage and the approach presented here does not impact the efficiency of the network in which a watermark is inserted as the embedded watermark while it is training the host network [13, 14].

The robustness of hard and soft decision detectors can be measured using Receiver Operating Characteristic graphs while Bit Error Rate is used for the detector response with bit sequence.

2. Literature Review

Many standard quantitative measures and metrics have been proposed to evaluate digital watermarking while comparing with their counterparts. Here we report all the standard methods found in the literature as illustrated in Table 1 and those used in this work to measure imperceptibility and robustness (along with capacity and computational cost) [34]. We use the methods for performance evaluation by ensuring

Table 1: Summary of literature review on Algorithm Transforms and Classifications

Reference		Survey/Tutorial	Parameter(s)
Frequency Domain + Back Propagation NN			
[15]		Discrete Wavelet Transform	Robustness
[16]		Multiwavelet Transform	Imperceptibility
[17]		Discrete Wavelet Transform	Robustness
[18]		Fourier Transform	Fidelity
[19]		Discrete wavelet Transform	Imperceptibility
[20]		Discrete wavelet Transform	Imperceptibility
Frequency Domain + Radial Basis NN			
[21]		Discrete Cosine Transform	Robustness
[22]		Discrete wavelet Transform	Imperceptible
[23]		Discrete wavelet Transform	Invisible and Robust
[24]		Discrete wavelet Transform	Robustness
Frequency Domain + Hopfield NN			
[25]			Capacity
[26]			Imperceptibility
[27]			Image Quality
[28]		Discrete Cosine Transform	Invisible and Robust
Frequency Domain + Full Counter Propagation NN			
[29]			Robustness, Imperceptibility
[30]		Discrete Cosine Transform	Robustness
[31]		Discrete Cosine Transform	Complexity, Capacity, PSNR
[32]		Discrete Cosine Transform	Imperceptibility, and robustness
Frequency Domain + Synergetic NN			
[33]		Discrete Wavelet Transform	Robustness and Imperceptibility

i) model and sources of distortion remain uniform
ii) All test images are 8-bit grey-scale images and are defined in same color space.

Quality assessment can also be done by comparing the original watermark and extracted watermark. This alternate metric is called a Normalized Correlation which exploits the correlation between the original watermark and the extracted one. The value lies between [0 1], any value nearer to 1 assures better quality. Another way of defining the similarity between the original watermark and the extracted one is accuracy ratio. It is the ratio of correct bits to the total bits. The architecture of CNN is different, unlike neural network where all layers are fully connected, here the layers are recognized in 3D: height, width, and depth. Further neurons in one layer are connected to only a small region of the next layer. Finally, the output is a single vector of probability scores, organized along the depth dimension [35]. The CNN consists of series of convolutional, pooling/sub-sampling layers followed by a fully connected layer. To implement the act of recognition in machines we need to show an algorithm of millions of images before it makes a pattern by generalizing the input and start making predictions for images it has never seen before [36]. The aim is to evolve a more robust and imperceptible watermarking scheme that can cater to the needs of content protection [37].

3. Algorithm Design

Watermarking can be achieved in any of the following two methods: either by changing the pixel values (least significant) of the image or by changing the coefficient values [38]. The quality of watermark depends on the method used. The first method where bits (representing the pixel values) are manipulated refers to a spatial domain which is very simple but not robust and can be easily perceived. In simple terms, spatial domain techniques refer to replacing pixels of the original image by watermark image [39]. In a given image first, the target pixels are identified and are replaced by pixels of watermark image. The spatial domain techniques are simple, fast and with less computational complexity [40]. They are immune to cropping and noising but are sensitive to signal processing attacks. In pursuit of enhancing robustness, if more pixels are manipulated we may end up with visible watermarks. These algorithms should carefully achieve a trade-off between robustness and imperceptibility [41]. In this section the experimental setup of watermarking a digital image is described with Algorithm 1 and Algorithm 2. At the various stages of digital watermarking from encoding of original image to decoding process and finally up to extracting the watermark and obtaining the high-resolution output decoded image. The aim is to evolve a more robust and imperceptible watermarking scheme that can cater to the needs of content protection and piracy prevention [42]. All the transform (and few hybrid approaches with NN) methods are good for watermarking but lack learning and adaptability [43]. We propose a digital watermarking method using deep learning methods which exploit the expressiveness of deep NN to securely embed invisible, imperceptible, attack-resilient binary signatures into the cover images. Coming to the

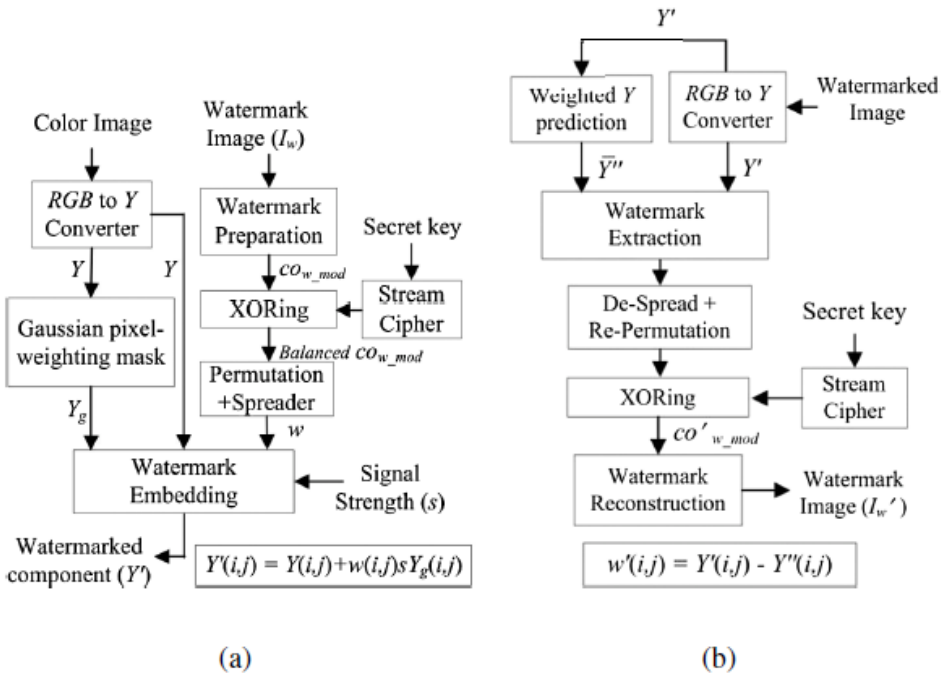


Figure 3: Block diagram representation of a typical digital watermarking process (a)Watermark Embedding and (b)Watermark Extraction.

decoder, we adopt adversarial model techniques to cause disorders to decode the desired signature [44]. We perform extended gradient descent under the Expectation over Transformation framework. In training the decoder network an Expectation-Maximization (EM) framework is employed to learn feature transformations that are more resilient to the attacks [45]. Experimental results indicate that our model achieves robustness across different transformations (all transformations, including scaling, rotation, adding noise, blurring, random cropping, and more) [46]. The aim is to evolve a more robust and imperceptible watermarking scheme that can cater to the needs of content protection and piracy prevention [47]. All the transform (and few hybrid approaches with NN) methods are good for watermarking but lack learning and adaptability [48]. We propose a digital watermarking method using deep learning methods which exploit the expressiveness of deep NN to securely embed an invisible, imperceptible, attack resilient binary signature into the cover images [49]. Coming to the decoder, we adopt adversarial model techniques to cause disorders to decode the desired signature.

Algorithm 1: Watermarking using Deep NN(Part-I)

Result: Encoding
Initialization;
Input_IMG: Image for Watermarking
Compute α transparency;
Construct an overlay;
Add Watermark to the overlay
while *setup* **do**
  Start encoding model;
  Image Encoding Process Starts;
  return encoded images;
  if *Given the model and targets* **then**
    compute the cross entropy loss;
  Else
    Given number of Iterations;
    Encode the set of images with specified binary targets;
    Image Encoding Process Ends;
  End Encoding Model
End

We perform extended gradient descent under the Expectation over Transformation framework. In training the decoder network an Expectation-Maximization (EM) framework is employed to learn feature transformations that are more resilient to the attacks. To implement the act of recognition in machines we need to show an algorithm of millions of images before it makes a pattern by generalizing the input and start making predictions for images it has never seen before [50]. The architecture of CNN is different, unlike NN where all layers are fully connected, here the layers are recognized in 3D: height, width, and depth. Further neurons in one layer are connected to only a small region of the next layer. Finally, the output is a single vector of probability scores, organized along the depth dimension. The CNN consists of series of convolutional, pooling/sub-sampling layers followed by a fully connected layer. The Figure 3 elucidates the block diagram representation of a typical digital watermarking process with Figure 3(a) as Watermark embedding process and Figure 3(b) watermark extraction phase. In this the input of the deep learning algorithm is the colour image and watermark to embed inside the colour image to watermark the deep Neural Network (DNN). The Algorithm 1 and Algorithm 2 are formulated based on the Figure 3(a) and 3(b).

Algorithm 2: Watermarking using Deep NN(Part-II)

Result: Decoding
Initialization;
While *setup* **do**
  Start the Decoding model;
  if *decoding network* **then**
    Initialize Decoding;

Extract Watermark;
Obtain High Resolution
Decoded Image;
Else
End of Decoded Network;
Return predictions;
Return decoding model;
Image Decoding Process Ends;
End Decoding Model
End

The Algorithm 1 illustrates the encoding process of digital watermarking using deep neural network and the Algorithm 2 presents the decoding process of digital watermarking using deep neural network.

4. Performance Evaluation

Fair performance evaluation of any system is fundamental for its acceptance and accreditation. Many standard quantitative measures and metrics have been proposed to evaluate digital watermarking while comparing with their counterparts. Here we report all the standard methods found in the literature and those used in this work to measure imperceptibility and robustness (along with capacity and computational cost). We use the methods for performance evaluation by ensuring i) model and sources of distortion remain uniform ii) all test images are 8-bit gray-scale images and are defined in same color space. For an image size of M x N pixels, with a pixel value of O for original image (without watermark) and W for watermarked image, the performance metrics can be calculated as below:

MSRE:

$$MSRE(O,W) = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{J=0}^{N-1}[\mathrm{O(i,j)} - W(i,j)]e^2$$

PSNR:

$PSNR(O,W) = 10\log 10[\{255\}\,e^2/\mathrm{MSRE\ (O,W)}$

4.1. Robustness Metrics

Robustness Metrics: Robustness of a watermark is a measure of resistance to attacks. The detectors dictate the evaluation method to measure robustness. The three types of detector responses are i) hard decisions (true or false for the presence and absence of watermark respectively), ii) soft decisions (correlation or similarity coefficients in terms of real numbers) iii) bit sequence (if the embedded watermark is in form of a message). The robustness of hard and soft decision detectors can be measured using Receiver Operating Characteristic (ROC) graphs while Bit Error Rate (BER) is used for the detector response with bit sequence.

ROC:

True Positive Fraction (TPF): $\frac{Number\ of\ True\ Positive\ Results}{True\ Positive+False\ Negative\ Results}$

False Positive Fraction (FPF): $\frac{Number\ of\ False\ Positive\ Results}{Fasle\ \ Positive+True\ Negative\ Results}$

Quality assessment can also be done by comparing the original watermark and extracted watermark. This alternate metric is called a Normalized Correlation which exploits the correlation between the original watermark and the extracted one. The value lies between [0 1], any value nearer to 1 assures better quality. Another way of defining the similarity between the original watermark and the extracted one is accuracy ratio. It is the ratio of correct bits to the total bits. When the attack is not going to affect the commercial value then it is better to consider only the watermark-to-noise ratio. This ratio signifies the power of watermark signal against the noise introduced by such attacks. For a watermark of size m x n pixels, the performance metrics can be calculated as below:

$$NC(OW,EW) = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{J=0}^{n-1}[\delta[\mathrm{OW(i,j)} - EW(i,j)]$$

where,

δ (X,Y) = 1 if X=Y and 0 otherwise;

5. Results Obtained

In this section we discuss the watermarking in the presence of various attacks at different noise levels. We present the results in terms of robustness of the watermark (BER and NC). The CNN is trained using standard descent back propagation algorithm. The performance consistency of the proposed method is verified by considering 2 different cover images: Lena and Camera man as presented in Figure 4 shows the watermarking process with different transforms (scaling, rotation, adding noise, blurring, random cropping, and more)and also confirms that that the proposed method is applicable to any cover image and with any watermark. BER is a measure of noise injected when the

signal is received after transmission channel. The BER shows the signal loss and fading in a wireless channel. The BER for various noise levels is shown in Figure 5(a). Normalized (cross) correlation is a template-matching method in digital watermarking. The template will be an image that shows a critical feature; by repeatedly computing a statistic between the watermarked image and corresponding pixels of a subset of an original image presents the noise correlation for various noise levels of 2,5,10 and 15. The important parameter in watermarking is the loss of original information and the accuracy with which the watermark is hidden in the cover image. There is a fine balance between the robustness and imperceptibility. The trade-off is to achieve high robustness without showing any trace of watermark. The accuracy of our model is gradually increasing as the epochs increases, on the contrary loss is gradually decreasing. The models training and testing results are shown in Figure 6 with the over fit, under fit and good fit are shown in Figures 6(a), 6(b) and 6(c) respectively. Our model Adam performs well and is consistent as compared to other models as shown in Figure 7. The Figure 8 shows the noise injected/present and the model performance. In the Figure 8(a) noise in hidden layer is shown while Figure 8(b) depicts the noise in input layer.

Figure 4: Typical digital watermarking process for Lena and Cameraman Images (a)original image (b)watermarked image (c)initial decoded image with watermark (d) extracted watermark (e)high resolution output image

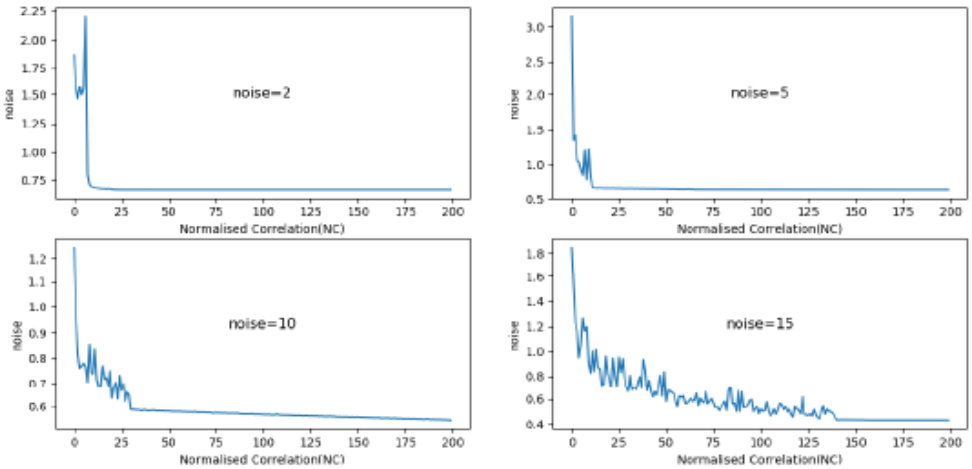


Figure 5: Verification of Robustness of the digital watermark against various type of transformations, shows the BER of watermark for various noise levels 2, 5, 10 and 15

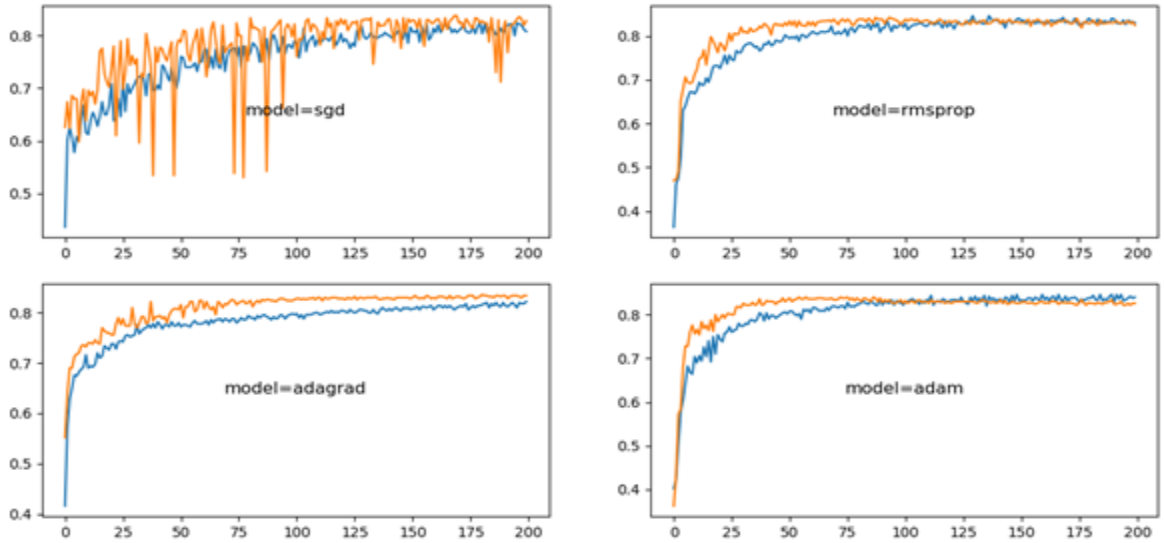


Figure 6: Validation of Neural Network Model

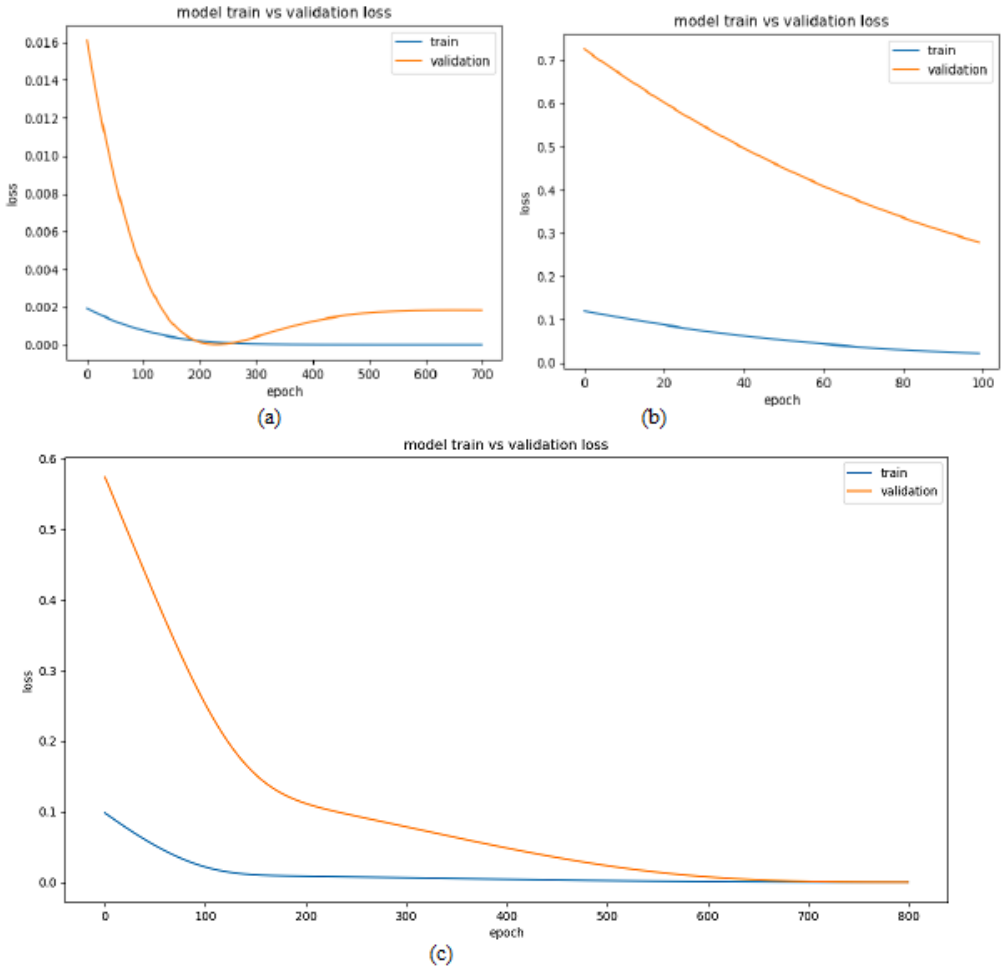


Figure 7: Fitting the model for validation (a)over fit (b) under fit and (c) good fit.

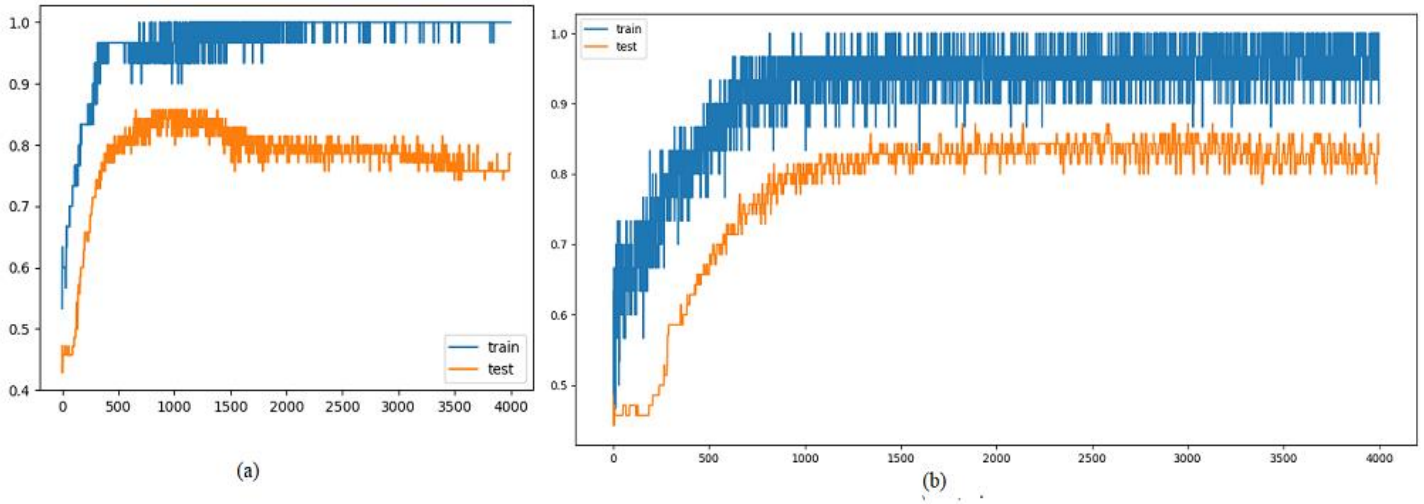


Figure 8: Injected/present noises in(a)hidden layers and (b)input layer of the proposed convolutional neural network.

Table 2: Comparison of Results

S.No	Noise/Attack	Parameter	[51]	[52]	[53]	[54]	[55]	[56]	[57]	[58]	Our Work
1	No Noise	PSNR(db)	58.91	48.29	48.47	51.45	34.33	56.47	35.6	47	**62.3**
		NC	1	1	1	1	1	1	0.99	0.98	**1.00**
2	With Noise	PSNR(db)	55.10	43.04	32.74	31.66	22.43	40.18	29.1	41.72	**58.78**
		NC	1	0.91	0.81	0.92	0.91	0.93	0.95	0.91	**1.00**

The graph shows that training and test results are in good accordance and validate the model used for training and optimization in digital watermarking. The proposed model is suitable for video too as the training is done off-line hence can be used in real-time applications. Hence the use of deep neural networks will enable us to realize complex features like learning weight sharing and update mechanism, noise-resilience and immunity towards attacks, and scaling. The proposed method has low loss and the accuracy goes on increasing as the number of epochs increases. The accuracy is 85% for 15 epochs and will slightly increase for more epochs. The level of high accuracy shows that the original image and watermarked image are indistinguishable.

This high performance of the model can be owed to the learning feature embedded in the watermarking. A digital watermarking task comprises embedding a signal into an image in accordance with robustness and quality constraints, it can be said that it is in essence a multi-objective optimization problem. The faster convergence of the algorithm (more accuracy with less epochs) can be achieved by introducing reinforcement learning or transfer learning method which has got huge attention in recent times. The current digital watermarking can also be validated by applying it to protect the copyrights of trained neural networks where ownership protection and piracy prevention is of utmost priority. The traditional approaches in digital watermarking are not fit for the digital data that can be stored efficiently and with very high quality and manipulated easily using computers [59, 60]. The aim is to evolve a secure digital communication that remarkably pushes forward the limits of legacy digital watermarking schemes across all dimensions of performance metrics. As this research space is ever increasing and innovations have spurred at all levels of communication, considerable progress is required in understanding the deep learning approaches for digital watermarking. The convergence of technological, economic and environmental forces is driving the digital watermarking and deep learning simultaneously, then each drives the other forward. Be it the deep neural networks driving digital watermarking or vice-versa, the continued expansion of each is good for the other. The complete suitability of digital watermarking for securing deep neural networks dataset is yet to be conducted. The trained models can be viewed as intellectual property, and it is a worthy challenge to provide copyright protection for trained models. We emphasis on how the copyrights of trained models can be protected computationally and propose for neural networks a digital watermarking technology. We propose a conceptual framework for integrating a watermark into models of deep neural networks to safeguard copyrights and identify violation of trained models of intellectual property. The Table 2 illustrates the comparison of obtained results with the previous works and it is noteworthy that the proposed work has shown an improvement of 5.75% in PSNR without noise and 6.68% in PSNR with noise = 5.

6. Conclusions

In this paper, we proposed a learning framework for robust digital image watermarking technique based on deep neural network. As observed, previous efforts in this space focused on optimization of embedding parameters with use of evolutionary computing. Demonstration of the watermarking under various noise and attacks is performed. The detailed experiments were carried out and we analyzed the performance of designed system. We have shown that our model could embed a watermark without impairing a deep neural network's efficiency. In future the research would continue in the direction towards digital watermarking based on intelligence.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] R.S. Kavitha, U. Eranna, M.N. Giriprasad, “DCT-DWT Based Digital Watermarking and Extraction using Neural Networks,” in 2020 International Conference on Artificial Intelligence and Signal Processing, AISP 2020, 2020, doi:10.1109/AISP48273.2020.9073104.

[2] N. Memon, P.W. Wong, “A buyer-seller watermarking protocol,” IEEE Transactions on Image Proc., **10**(4), 643-649, 2001, doi:10.1109/83.913598.

[3] I.A.T. Hashem, I. Yaqoob, N.B. Anuar, S. Mokhtar, A. Gani, S. U Khan, "The rise of “big data” on cloud computing: Review open research issues." Information systems, **47**, 98-115, 2015, doi:10.1016/j.is.2014.07.006.

[4] N.F. Johnson, Z. Duric, S. Jajodia, N. Memon, “Information Hiding: Steganography and Watermarking—Attacks and Countermeasures,” Journal of Electronic Imaging, 2001, doi:10.1117/1.1388610.

[5] J.C. Ingemar, M.L. Miller, A.B. Jeffrey, J. Fridrich, T. Kalker, Digital Watermarking and Steganography, 2008, doi:10.1016/B978-0-12-372585-1.X5001-3.

[6] Y.S. Singh, B.P. Devi, K.M. Singh, “Image comp using Multilayer Feed Forward Artificial Neural Network with Conjugate Gradient,” in Proceedings of the 2012 World Congress on Information and CommunicationTechn,976-980,2012, doi:10.1109/WICT.2012.6409216.

[7] A.Z. Tirkel, G.A. Rankin, R.M. Van Schyndel, W.J. Ho, N.R.A. Mee, C.F. Osborne, “Electronic Water Mark,” Digital Image Computing, Technology and Applications (DICTA’93), 1993.

[8] S. Wu, S. Zhong, Y. Liu, “Deep residual learning for image steganalysis,” Multimedia Tools Apps, **77**(9), 10437-10453, 2017, doi:10.1007/s11042-017-4440-4.

[9] S.J. Lee, S.H. Jung, “A survey of watermarking techniques applied to multimedia,” in IEEE International Symposium on Industrial Electronics,1,272-277, 2001, doi:10.1109/isie.2001.931796.

[10] R. Warkar, P. More, D. Waghole, “Digital audio watermarking and image watermarking for information security,” in 2015 International Conference on Pervasive Computing: Advance Comm. Tech. and Application for Society, ICPC2015, doi:10.1109/PERVASIVE.2015.7086980.

[11] A.K. Singh, B. Kumar, M. Dave, A. Mohan, “Robust and Imperceptible Spread-Spectrum Watermarking for Telemedicine Applications,” Proceedings of the National Academy of Sciences India Section A - Physical Sciences, **85**(2),295-301, 2015, doi:10.1007/s40010-014-0197-6.

[12] M. Barni, F. Bartolini, V. Cappellini, A. Piva, “A DCT-domain system for robust image watermarking,” Signal Processing, 1998, doi:10.1016/S0165-1684(98)00015-2.

[13] T. Ashkan et al.,"A modified dual watermarking scheme for digital images with tamper localization/detection and recovery capabilities." 2012 9th International ISC Conference on Information Security and Cryptology. IEEE, 2012.

[14] P.K. Dhar, J.M. Kim, “Digital watermarking scheme based on fast Fourier transformation for audio copyright protection,” International Journal of Security and Its Applications, **5**(2),33-48, 2011. doi: 10.1016/j.future.2018.07.029.

[15] I.J. Cox, J. Kilian, F.T. Leighton, T. Shamoon, “Secure spread spectrum watermarking for multimedia,” IEEE Transactions on Image Processing, 1997, doi:10.1109/83.650120.

[16] D. Kundur, D. Hatzinakos, “Digital watermarking for telltale tamper proofing and authentication,” Proceedings of the IEEE, 1999, doi:10.1109/5.771070.

[17] A. Tiwari, M. Sharma, R.K. Tamrakar, “Watermarking based image authentication and tamper detection algorithm using vector quantization approach,” AEU - International Journal of Electronics and Communications, **78(5)** ,114-123, 2017, doi: 10.1016/j.aeue.2017.05.027.

[18] Z. Ma, M. Jiang, H. Gao, Z. Wang, “Block chain for digital rights management,” Future Generation Computer Systems, **89**(7), 746-764, 2018, doi: 10.1016/j.future.2018.07.029.

[19] J. Kishigami, S. Fujimura, H. Watanabe, A. Nakadaira, A. Akutsu, "The Block Chain-Based Digital Content Distribution System," in Proceedings - 2015 IEEE 5th International Conference on Big Data and Cloud Computing, B D Cloud 2015, 2015, doi:10.1109/BDCloud.2015.60.

[20] Stefik, Mark J., et al. "System for controlling the distribution and use of rendered digital works through watermaking." U.S. Patent No. 6,233,684. 15 May 2001.

[21] T. Jitha Raj, E.T. Sivadasan, "A survey paper on various reversible data hiding techniques in encrypted images," in Souvenir of the 2015 IEEE International Advance Computing Conference, IACC 2015, 2015, doi:10.1109/IADCC.2015.7154881.

[22] A.A. Philip, P.R. Geetharanjin, "Fingerprint Encryption and Dual Watermarking to Verify the Security Issues in Teleradiology," in Proceedings of the 3rd International Conference on Communication and Electronics Systems, ICCES 2018, 2018, doi:10.1109/CESYS.2018.8724112.

[23] I.J. Cox, M.L. Miller, "The first 50 years of electronic watermarking," EURASIP Journal on Applied Signal Processing, **2**(2)**,** 820-936, 2002, doi:10.1155/s1110865702000525.

[24] J. Bajpai, A. Kaur, "A literature survey - Various audio watermarking techniques and their challenges," in Proceedings of the 2016 6th International Conference - Cloud System and Big Data Engineering, Confluence 2016, 2016, doi:10.1109/CONFLUENCE.2016.7508162.

[25] C. Kumar, A.K. Singh, P. Kumar, "A recent survey on image watermarking techniques and its application in e-governance," Multimedia Tools and Applications, **77(3)**, 3597-3622, 2018, doi:10.1007/s11042-017-5222-8.

[26] M.H. Alkawaz, G. Sulong, T. Saba, A.S. Almazyad, A. Rehman, "Concise analysis of current text automation and watermarking approaches," Security and Communication Networks, **9**(18), 6365-6378, 2016, doi:10.1002/sec.1738.

[27] A. Świerkosz, "Digital watermarking in telemedicine an example from ECG - Review of challenges, methods and applications," Advances in Intelligent Systems and Computing, 248-255, 2017, doi:10.1007/978-3-319-47154-929.

[28] S. Mirghasemi, P. Andreae, M. Zhang, "Domain-independent severely noisy image segmentation via adaptive wavelet shrinkage using particle swarm optimization and fuzzy C-means," Expert Systems with Applications, **133**(4), 126-150, 2019, doi: 10.1016/j.eswa.2019.04.050.

[29] Q.D. La, M. V. Ngo, T.Q. Dinh, T.Q.S. Quek, H. Shin, "Enabling intelligence in fog computing to achieve energy and latency reduction," Digital Communications Networks, 2019, doi: 10.1016/j.dcan.2018.10.008.

[30] H. Kandi, D. Mishra, S.R.K.S. Gorthi, "Exploring the learning capabilities of convolutional neural networks for robust image watermarking," Computers Security ,**65**(3), 247-268, 2017, doi: 10.1016/j.cose.2016.11.016.

[31] M.S. Hsieh, D.C. Tseng, Y.H. Huang, "Hiding digital watermarks using multiresolution wavelet transform," IEEE Transactions on Industrial Electronics, **48**(5) ,875-882, 2001, doi:10.1109/41.954550.

[32] N.M. Makbol, B.E. Khoo, "Robust blind image watermarking scheme based on Redundant Discrete Wavelet Transform and Singular Value Decomposition," AEU - International Journal of Electronics and Communications,**67(2)**,102-112, 2013, doi: 10.1016/j.aeue.2012.06.008.

[33] H.T. Hu, L.Y. Hsu, J. Garcia-Alfaro, "Exploring DWT-SVD-DCT feature parameters for robust multiple watermarking against JPEG and JPEG2000 compression," Computers and Electrical Engineering, **41**(8), 52-63, 2015, doi: 10.1016/j.compeleceng.2014.08.001.

[34] T.T. Takore, P.R. Kumar, G.L. Devi, "A modified blind image watermarking scheme based on DWT, DCT and SVD domain using GA to optimize robustness," in International Conference on Electrical, Electronics, and Optimization Techniques, 2016, doi:10.1109/ICEEOT.2016.7755190.

[35] A.K. Singh, "Improved hybrid algorithm for robust and imperceptible multiple watermarking using digital images," Multimedia Tools and Applications, **76(6)**,8881-8900, 2017, doi:10.1007/s11042-016-3514-z.

[36] B.J. Saha, Arun, K.K. Kabi, C. Pradhan, "Non-blind watermarking technique using enhanced one-time pad in DWT domain," in 5th International Conference on Computing Communication and Networking Technologies, ICCCNT 2014, 2014, doi:10.1109/ICCCNT.2014.6963061.

[37] N.M. Kumar, T. Manikandan, V. Sapthagirivasan, "Non-blind image watermarking based on similarity in contourlet domain," in International Conference on Recent Trends in Information Technology, ICRTIT 2011, 2011, doi:10.1109/ICRTIT.2011.5972280.

[38] G.H. Yann LeCun, Yoshua Bengio, "Deep learning,", Nature, **521**(7553)**,** 436–444, 2015.

[39] R. Namba, J. Sakuma, "Robust watermarking of neural network with exponential weighting," in Asia CCS 2019 - Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security, 2019, doi:10.1145/3321705.3329808.

[40] E. Le Merrer, P. Pérez, G. Trédan, "Adversarial frontier stitching for remote neural network watermarking," Neural Computing and Applications, **32**(13), 9233-9244, 2020, doi:10.1007/s00521-019-04434-z.

[41] B. Prashanth, "Design and Implementation of Radar Cross-Section Models on a Virtex-6 FPGA," Journal of Engineering (United Kingdom), 2014, doi:10.1155/2014/489765.

[42] L. Pang, S. Zhu, C.W. Ngo, "Deep Multimodal Learning for Affective Analysis and Retrieval," IEEE Transactions on Multimedia, **17**(11), 2008-2020, 2015, doi:10.1109/TMM.2015.2482228.

[43] B. Prashanth, P.A. Kumar, G. Sreenivasulu, " Design and Implementation of floating point ALU on a FPGA processor," in 2012 International Conference on Computing, Electronics and Electrical Technologies, ICCEET 2012, 2012, doi:10.1109/ICCEET.2012.6203790.

[44] B. Vijay Krishna, B. Venkata Prashanth, P. Sujatha, "Design and implementation of DPFC for multi-bus power system," International Journal of Engineering and Technology (UAE), **7**(2), 2018, doi:10.14419/ijet.v7i2.8.10318.

[45] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015, doi:10.1109/CVPR.2015.7298594.

[46] S.R. Qasim, H. Mahmood, F. Shafait, " Rethinking table recognition using graph neural networks," in Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, 2019, doi:10.1109/ICDAR.2019.00031.

[47] J. Li, P. He, J. Zhu, M.R. Lyu, "Software defect prediction via convolutional neural network," in Proceedings-IEEE International Conference on Software Quality, Reliability and Security, 2017, doi:10.1109/QRS.2017.42.

[48] Y. Uchida, Y. Nagai, S. Sakazawa, S. Satoh, "Embedding watermarks into deep neural networks," in ICMR 2017 - Proceedings of the 2017 ACM International Conference on Multimedia Retrieval, 2017, doi:10.1145/3078971.3078974.

[49] A. Van Den Oord, S. Dieleman, B. Schrauwen, "Deep content-based music recommendation," in Advances in Neural Information Processing Systems, **26(1)**, 2643-2651, 2013, doi:10.1145/2500098.2500109.

[50] A. Karatzoglou, B. Hidasi, "Deep learning for recommender systems," in RecSys 2017 - Proceedings of the 11th ACM Conference on Recommender Systems, 2017, doi:10.1145/3109859.3109933.

[51] A. Karpathy, L. Fei-Fei, "Deep Visual-Semantic Alignments for Generating Image Descriptions," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, doi:10.1109/TPAMI.2016.2598339.

[52] H.O. Song, Y. Xiang, S. Jegelka, S. Savarese, "Deep Metric Learning via Lifted Structured Feature Embedding," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, doi:10.1109/CVPR.2016.434.

[53] B. Kim, H. Kim, K. Kim, S. Kim, J. Kim, "Learning not to learn: Training deep neural networks with biased data," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019, doi:10.1109/CVPR.2019.00922.

[54] L. Shao, D. Wu, X. Li, "Learning deep and wide: A spectral method for learning deep networks," IEEE Transactions on Neural Networks and Learning , **25**(12), 2303-2308, 2014, doi:10.1109/TNNLS.2014.2308519.

[55] A.B. Duque, L.L.J. Santos, D. Macêdo, C. Zanchettin, "Squeezed Very Deep Convolutional Neural Networks for Text Classification," in Lecture Notes in Computer Science, 2019, doi:10.1007/978-3-030-30487-4_16.

[56] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision Pattern Recognition, 2016, doi:10.1109/CVPR.2016.90.

[57] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," IEEE Transactions on Image Processing, **26**(7), 3142-3155, 2017, doi:10.1109/TIP.2017.2662206.

[58] B.D. Rouhani, H. Chen, F. Koushanfar, "Deep Signs: An End-to-End Watermarking Framework for Ownership Protection of Deep Neural Networks," in International Conference on Architectural Support for Programming Languages and Operating Systems - ASPLOS, 2019, doi:10.1145/3297858.3304051.

[59] A.I. Orhean, F. Pop, I. Raicu, "New scheduling approach using reinforcement learning for heterogeneous distributed systems," Journal of Parallel and Distributed Computing, **117**(1)**,** 292-302, 2018, doi:10.1016/j.jpdc.2017.05.001.

[60] Q. Guo, X. Wang, Y. Wu, Z. Yu, D. Liang, X. Hu, P. Luo, "Online knowledge distillation via collaborative learning," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, doi:10.1109/CVPR42600.2020.01103.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1569-1576 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

ESP2: Embedded Smart Parking Prototype

Tarek Frikha[1,*], Hedi Choura[1], Najmeddine Abdennour[1], Oussama Ghorbel[1,2], Mohamed Abid[1]

[1]*CES-Laboratory, Sfax University, Sfax, 3029, Tunisia*

[2]*Computer Sciences Department, Jouf University, Skaka, 2014, Kingdom of Saudi Arabia*

ARTICLE INFO

Article history:
Received: 16 October, 2020
Accepted: 16 December, 2020
Online: 25 December, 2020

Keywords:
Embedded System
Smart City
Smart Parking
Wireless Sensor Network
Actuator
Sensor
Mobile application
Web Application

ABSTRACT

The technological evolution and the frequent use of the Internet as a technology within the framework of the Internet of Things and Robotics has affected various fields such as Industry 4.0, smart agriculture, smart cities, home automation and autonomous cars. These advances applied to everyday life have facilitated the implementation of new concepts to minimize pollution, traffic jams and accidents while making life easier. In this article, we propose an intelligent parking concept based on embedded systems. In order to solve the problem of permanent traffic jams and their repercussions on large cities we have implemented this system. It is a complete platform allowing remote reservation from different platforms. It will also collect the various data obtained by parking sensors. These data will be processed and saved if necessary. This concept was first developed to highlight the feasibility and adaptability of this system in the context of smart cities. It has been realized using wireless sensor networks connected to embedded platforms. These platforms are not only connected to motors, sensors and actuators but also to servers to save data and deploy reservations. These reservations are made via web and mobile application.

1. Introduction

The robotics automation, wireless sensors, artificial intelligence and embedded systems are famous scientific research fields with a wide margin for innovation and growth. Although these are areas that are literally different and divergent. However, they remain compatible. One can thus join these different axes in order to carry out applications that make life easier for mankind.

In this framework, we can mention several types of uses such as smart houses [1], smart agriculture [2] and smart parking [3], which are smart city [4] applications. We can also talk about e-health [5] and Industry 4.0 [6] which are applications in constant development.

In this work, we're going to touch on smart parking. The automated systems for vehicle parking management [7] are widely available in the recent years. However, it is still an exclusive product that is not largely used for a couple of reasons including being very cost expensive and complex. These difficulties also contribute to other problems such as hard maintenance and resource management issues [8]. Nowadays, android applications and Arduino based technologies [9] are the mainstream and the most popular solution in numerous domains due to being open sourced. This attribute makes them easy to implement and heavily supported with the majority of hardware and software. For these reasons we used android based application for our embedded smart parking and management system that outperformed the other available options in resource optimization and performance.

In our previous work [10] we presented a fully embedded architecture for an automated intelligent parking system that communicates through an Android application, which delivered an intelligent parking management and a top-notch costumer service for smart parking.

In this paper, we describe in details our previous work [10] while we expand on the practical implementation and the communication interface. With the objective of having a functional prototype that develops the research work and can be integrated in the field of smart cities, we will present two platforms:

- an embedded platform that has the role of data acquisition and processing in our parking lot. The use and architecture will be described later on.
- a client/server platform with Android and Web applications: these will describe the reservation, confirmation and payment parts of the proposed service.

*Corresponding Author: Tarek Frikha, tarek.frikha@enis.tn

Our paper will be presented as follows. We start by an overview on the smart parking state of the art. In the theoretical analysis section, we will introduce the architectural design of the car parks on one hand and the developed computer application to be realized on the other. Then, we present the practical implementation section where we demonstrate the developed applications as well as the implemented embedded architecture. Finally, we compare between the implemented Intelligent Parking results before we declare the conclusion and perspectives.

2. State of Art

Several works have presented smart parking. In [11], they present a flexible and real time information system. This parking is easy to use for the client via web connexion. A CMOS cam is used to control the parking. The disadvantages of this system are the need to inquire for the available slot, the high implementation cost. The most important disadvantage is the bottlenecks due to an overload of incoming informations.

In [12], they use RFID technology. It permits to have not only secured but also an easy and fast checking out of the vehicles in less time. The data security is provided by ZigBee Wireless Technology. The data of the arrival and departure of the vehicle are stored in database. To inform the driver informations are displayed on the LCD proposed screens in the parking. Despite all the advantages, the system has many disadvantages. It allows one by one. That's why it consumes a lot of time and prevent multiple check out and check in at the same time. The drive can't make a reservation via mobile device. Another disadvantage of RFID is the increase of cost and of the node to node implementation time.

In [13], if the most important disadvantage is that the parking does not alarm the user about the car theft, the proposed smart parking use ultrasonic sensors and RFID. Needing for deployment ZigBee and 3G (or 4G, 5G…), the most important advantages of the proposed work are the use of the Car Parking Framework that gives two performance classes in term of financial and consumption cost deployment. In addition, the heterogeneous wireless sensor network cause energy fluctuations. To remedy they use LIBP gathering protocol.

In [14], they use Zigbee for deployment and ferromagnetic sensor. Suryady and al introduced another concept of smart parking based on cloud-based platforms as a service (PaaS). This technology gives the smart parking an interesting quality. The smart-parking system is faster to use. It reduces time, cost and system maintenance. The platform has the advantage of being flexible to collaborate with other platforms. This system has also disadvantaged. It does not provide neither guidance for the drive nor payment due to the application insecurity to protect user privacy. The user has no possibility to display spaces for user on smartphone or parking LCD.

In [15], the smart parking is more complete. It has many qualities like his flexibility and security increase. This security appears for the user because of less of human interaction for booking slots. The drive can find the availability of parking space and then books slots. The IoT's devices provide data storage. The data are processed and collected to be saved. The drive has another security aspect to preserve the user from steals by theft management.

As we already declared, there are numerous automated parking management systems, yet only three methods which are Park Assist [16], Hectronic [17] and Parking plus [18] that stand out and achieve respectable results. These intelligent car parks are available on the market but have some missing features that need to be addressed.

2.1. Park Assist smart parking system

the TKH group developed this smart parking system that have various traits including guiding the client to the empty parking space while providing access to a mobile application with the vehicle geolocation and the reliance on machine learning analysis for the parking spaces to insure the right exploitation.

2.2. Hectronic intelligent parking

Contrary to Park assist, this intelligent parking system delivered by the German international company Hectronic has a separate entrance and exit terminals. The main feature of this system is their powerful and well built security alongside other adequately constructed services.

2.3. Parking Plus mobile application

This parking system is based on a mobile application that delivers a central parking service by gathering information from a cluster of parking zones to provide the driver with the details about the nearest empty parking place and the possibility to reserve it and pay for it by the phone.

When we analyze these various methods, we found that some of them have security shortcomings both in capturing information for reservation and payment or in securing vehicles in case of theft. Therefore, in this work we present an intelligent car park allowing the capture of information on the different vehicles when they enter, exit and during their presence in our car park. An offline backup is made throughout the day before being transmitted to the cloud in order to maintain a trace if necessary.

Reservation, payment and placement of the car in the car park are done in a secure way via not only web but also mobile applications compatible on different types of terminals.

Given the lack of space proposed for parking lots and the permanent increase in the number of cars in large cities generally and particularly in Tunisia, our goal is to set up an intelligent system to give the user the choice, if possible, to choose his parking space, if not to allocate an automatic space to optimize the space.

Indeed, thanks to the automatic parking system chosen, we will have the possibility:

- to secure the parking lot by minimizing human intervention: cars will be closed and parked at the entrance of the parking lot. Their placement will be carried out automatically afterwards.
- to optimize space and minimize accidents: the cars will be parked automatically and therefore no space or accidents due to bad maneuvers will occur.
- to minimize pollution and traffic jams: by allowing the reservation system to not search for parking spaces in a

random way and therefore optimize fuel consumption and minimize pollution.

The innovation proposed in our paper is to create a fully automated parking system that solves the limitations described above.

This design will present not only the architecture of the realized prototype but also the technologies based on the IoT to allow the automatic placement of cars in the proposed concept.

There will be a design part, an IoT part and a hardware implementation part which will be described in parts 3 and 4.

3. Theoretical Analysis

Looking at the systems available on the market, we can initially assume that depending on their points of strength we can choose from them a tolerable approach [19]. However, these methods have many apparent concerns such as the inefficient space exploitation with Park Assist and Hectronic by the need for a driving pavement to connect the vehicle to the driving spot and the capacity problem caused by the lack of long term parking strategy [20]. We can also point out the high service cost and the absence of security plan for Parking Plus. In this section we will describe the different axis of our own parking architecture that fix these issues.

We begin with our proposed parking architecture for car parks that is set up in narrow plots of land. We present a cylinder shape architecture as shown in figure 1. The second part of this theoretical analysis introduce the proposed embedded system to collect, process and update the information before transmission step. The last part describes the frontend and the backend of the application architecture. It will describe different actors and secured reservation and payment service to be used by the driver.

3.1. Smart parking architecture

3.1.1. Proposed smart parking architecture

In order to solve the parking management concerns that the previous methods failed to answer, we conceptualized a structure with a cylinder-shaped architecture for vehicle parking as revealed in figure 1.

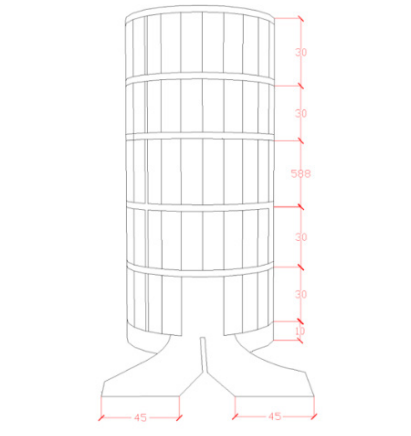


Figure 1: The front face display of the cylinder-shaped architecture for our parking

We choose using this architectural shape in an attempt to reduce the space spent on pavement allowing the vehicle movement within the structure and switching to a unique vehicle area, which is the mechanical component that act as an elevator and the selector for the parking place simultaneously.

3.1.2. Detailed units of the architecture

To simplify and optimize the driver's access to the proposed architecture, the parking place selection is done via the android mobile application. Afterwards, the driver has to station his vehicle on the entrance elevator panel that has the same size as standard parking place as shown in figure 2.

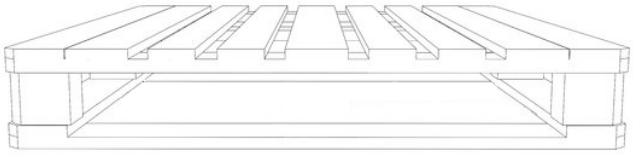


Figure 2: A side view of the entrance elevator panel

Having reserved a parking space via web or mobile application, the customer will provide the information for the car registration number. This number will be the reference to allow the car to enter the parking place.

Once the vehicle is checked to the entrance panel, a two-armed cart raise the panel as an elevator. Figure 3 present an example of this activity.

Afterwards, the vehicle will be directed directly to the parking space relying on the elevator's ability to rotate 360 degrees. This feature allows the system to set the vehicle in the appropriate location selected in advance by the customer. The exit process is also similar to the entry procedure but backwards. The customer has to do his check out and pay for the service via the mobile application and once he does the elevator selects the vehicle corresponding to the car registration number and bring it down to the exit of the parking lot.

The full automation of elevator with the minimization of manpower and human intervention in the parking process will make the process secure by avoiding human errors and effective by efficient exploitation of the parking facility.

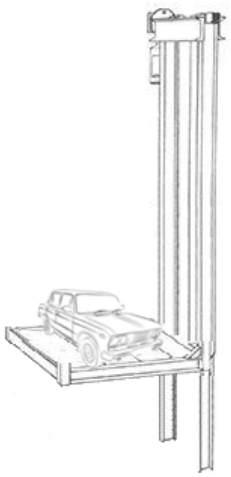


Figure 3: A sketch example of a vehicle elevator

3.2. Hardware embedded architecture

Having chosen a suitable architectural form for our intelligent car park, we will present the embedded implementation of our system in this section. It will, in fact, consist of three parts. Firstly, we have chosen a set of sensors and actuators that will be installed in order to obtain information about the parking spaces if they are available or occupied. Second, a system will collect this information from the installed sensors to update and save the occupancy in the parking lot. This data will be saved locally and subsequently updated in the cloud. In part C, we will present the software backup platform implemented.

3.2.1. Used set of sensors and actuators (SSA)

As described in the state-of-the-art section, our intelligent car park will not only make the driver's life easier by offering him an available space but will also provide him with a secure space for both him and his vehicle.

Thus, a surveillance camera at the entrance but also on each floor of the car park will detect the license plate of the car in order to validate its access to the car park but also to the reserved space.

As soon as the car reaches its space a sensor will transfer the information and cancel the space for any other reservation. At the exit of the car park, the information is transmitted directly to the driver and the space will be directly released for another reservation.

The billing for this operation will be proportional to the time occupied in our car park.

To sum up, we'll need:

- Surveillance cameras at the entrance and exit of the car park but also at the entrance and exit of each floor of our car park.

- Barrier level detection sensors to validate entry and exit to the car park.

- Infrared sensors to transmit information on the presence and absence of the car from the reserved space.

In parallel with the sensors, the need for actuators is equally important. Indeed, these will allow us to manage the different engines of our car park remotely. We can therefore mention the following:

- The motors of the car park barriers
- The engines of the car lift
- Vehicle placement jacks as shown in Figure 3.

Due to the importance of precision in the realization of our system. Different electrical voltage values must be received from the sensors and sent to the different motors. The accuracy of these data requires a system that can directly manipulate the analog data by minimizing the use of digital-to-analog and analog-to-digital converters. This will be detailed in the section sensor and actuators control and collection platform.

3.2.2. Sensor and actuators control and collection platform (SAC2P)

In order to control the various sensors and actuators present in our system, we will use an on-board platform that can receive and transmit analog data. In the prototype that will be described in section IV, part B, we will use an embedded Arduino platform that will allow us to validate the proposed approach.

This platform will be connected via Bluetooth to both actuators and sensors. It will thus control entry to the car park, the placement of the car and the sending of information to the information backup and transfer platform. Figure 4 describes sensor and actuators control and collection platform.

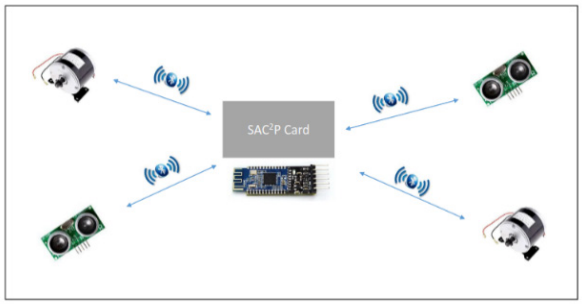


Figure 4: Sensor and actuators control and collection platform

3.2.3. Information backup and transfer platform (IBTP)

The Information backup and transfer platform part offers a digital embedded platform.

Several types of platforms can be proposed such as GPUs, FPGAs, ASICs. This platform will have four main tasks with the following priorities:

a) Acquisition and processing of information from surveillance cameras.

b) The administration of the Sensor and actuators control and collection platform via Bluetooth.

c) The offline backup and management of the various data recovered.

d) Updating information and saving it on the cloud server via wifi or 4G

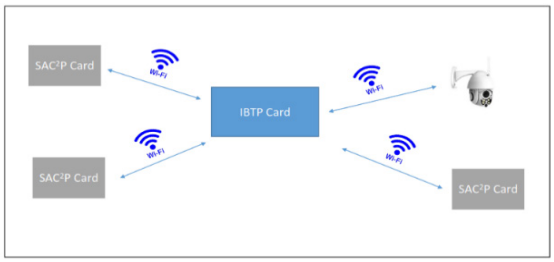


Figure 5: Information backup and transfer platform

Since, for our application, obtaining a real-time result is not paramount, it is not necessary to use GPUs that are quite expensive in terms of hardware and energy costs. Thus, a simple multiprocessor architecture based on 4 processors working in parallel is largely enough.

For our application, we decided to use a Raspberry Pi as a platform to meet the needs of our Smart Parking prototype. In part 4, we will describe in more detail the system implemented.

In figure 5, there is a description of information backup and transfer platform.

3.3. Software reservation platform architecture

After having described and set up the architecture part 3.1 and the hardware part 3.2, we will highlight in part 3.3 the software part and particularly the online booking platform.

In the reservation part we will have several interfaces.

Firstly, the administrator will access the different functionalities via the web. He will be able to modify the terms of payment, accept complaints, validate possible subscriptions from members and deal with disputes.

It is the administrator who will be able to manipulate the number of spaces, the availability as well as the occupation of the car park, the rates per hourly session, penalties for late payment.

The customer will access the online booking platform either via mobile application or via web.

Each customer must first register on the system either via mobile application or via the web. He will indicate his personal data (surname, first name, telephone number, e-mail, license plate, make and colour of the car, ...). He will then choose the method of payment either by credit card or in cash.

As a nominal scenario, a customer will be able to reserve a place on one or the other of the applications. He will indicate the duration of his parking period. At the time of reservation, the user will indicate the method of payment. At 15 minutes from the end of the indicated parking period, a SMS and an e-mail will be sent to inform the customer. He will be able to extend his reservation by reply or maintain the duration indicated at the time of reservation.

On leaving the car park, the parking space is released and is therefore open for a new reservation.

If the customer leaves the car park on time and has paid by credit card, they will leave the car park directly: the camera at the car park entrance will identify the car and open the entrance barrier. If he is going to pay in cash, a machine will take the coins or notes and give him the rest of the amount.

If the customer is late without replying to the SMS received, he will have to pay a penalty. The amount charged for leaving the car park will be the equivalent of a full day's parking.

If the customer is late and this delay has been indicated by message, the customer will pay the difference when leaving the car park.

4. Practical implementation

In this part we will present the parts describing the implementation and setup of our demonstrator. We will start with the modeling part including the class diagram and the use case diagram.

Then we propose the mobile SW part (Arduino application) and finally the implementation of the embedded part implemented on an Arduino platform.

4.1. Uses case diagram

In this section we will present the actors of our platform. The administrator, the client as a physical actor and the system as a logical actor.

0As shown in Figure 6, the administrator has the possibility to access to the different functionalities via the web. He will be able to accept claims, check the number of seats and set rates per hourly session.

The administrator can also be able to manipulate the penalties for late payment, validate the possible subscription of the member and change the terms of payment, etc.

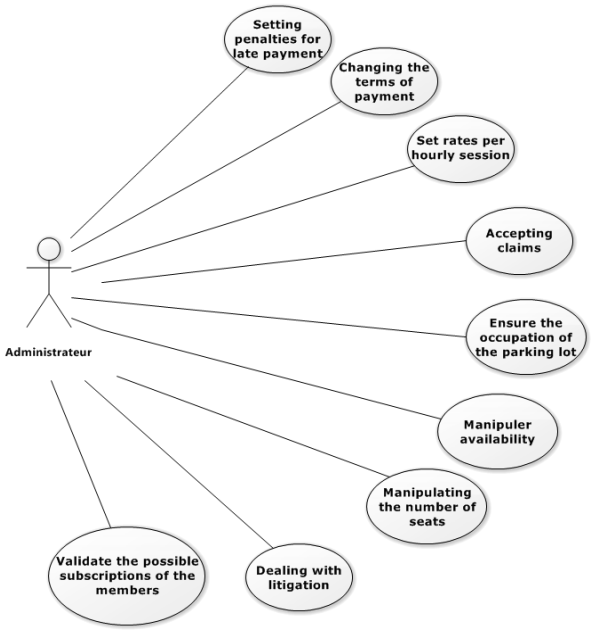


Figure 6: Admin Use Class Diagram

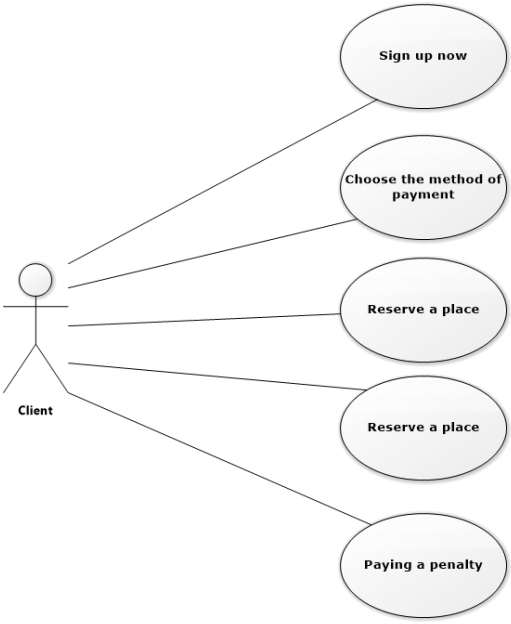


Figure 7: Client Use Class Diagram

The client, as mentioned in Figure 7, must start by registering on the system either via mobile application or via the web. He I indicate his personal data such as surname, first name, phone number, e-mail, license plate, make and color of the car, etc. After that, he chooses the method of payment either by credit card or cash. He will be able to reserve a place by indicate the duration of his parking period.

At the time of his reservation, the client will indicate the method of payment. He will be able to extend his reservation by reply or maintain the duration indicated at the time of reservation.

The last actor is the system, as represented in Figure 8, it must be able to identity the car, take coins or bills, remit the rest of the amount, open the entry barrier and before 15 minutes of the end of the period subscription, it sends a SMS to client to inform him; if he wants to extend the period or he free the space.

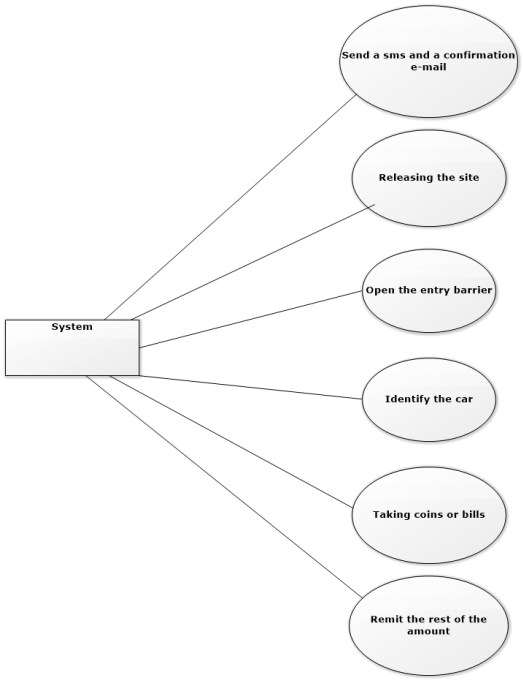


Figure 8: System Use Cas Diagram

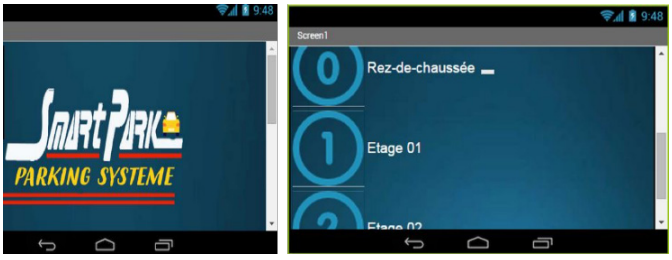


Figure 9: Our smart park application main menu

4.2. Android application

The Android operating system is by far the largest used operating system for smartphones in the world for the recent years. To communicate with our automated smart system, we developed an Android application to exploit the ease of implementation, compatibility and support [21]. We used the App Inventor development tool generate by Google to create our Android web application for smartphones [22]. Our application contains a main menu for the user to select the desired parking floor from the parking structure as shown in figure 9.

From this menu the driver first selects the floor and then searches for the empty parking location. Finally, the application creates a security code for the specific location to be used by the user when he exits the parking facility. It is also possible for the user to let the parking system to choose automatically which place to allocate.

4.3. Arduino technology embedded system:

To control the various parts of our system we choose to work with a simple Arduino board [23] and we have decided on the Arduino mega board. This embedded bored is responsible for the communication with the client commands coming from the Android application. It is also responsible for checking and monitoring the sensors that provide the information about empty locations and vehicle presence, while managing the mechanical elements of the elevator and the rotating panel.

For the communication part, the Arduino mega board needs the Arduino WIFI shield in order to connect to the network via WIFI [24]. As displayed in figure 10, this wifi shield is pin by pin compatible with the Arduino mega board, which makes it easy to work with.

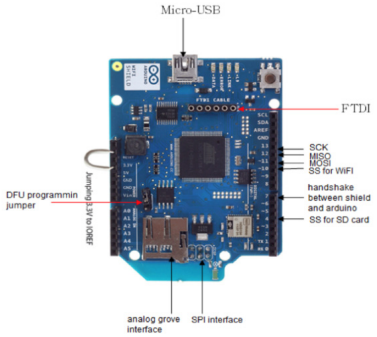


Figure 10: The Arduino WIFI shield

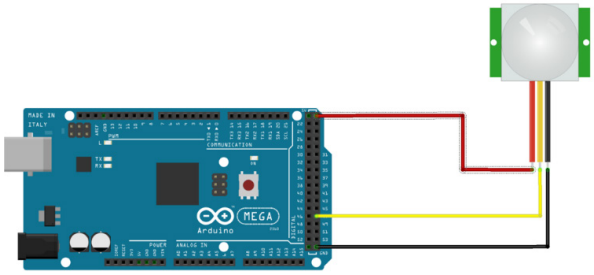


Figure 11: Arduino mega board and the PIR sensor wiring diagram

In order to collect the information about the availability of the parking locations and the elevator panel a cluster of spread out sensors is crucial for our system. Therefore, we decided to lay out a collection of the PIR (Passive Infrared Sensor) in every parking location, the entrance, the exit and in the elevator panel. As shown in figure 11, connecting the PIR sensor to our Arduino mega board is a very simple operation that demands the use of only three pins. We also considered using the HC-SR04 ultrasonic sensor for the same operation mainly for compatibility reasons since it is also an

Arduino technologies component. However, considering the low expensive PIR sensor was the optimal choice since it does not compromise the proficiency of our system [25].

To manage the mechanical part of our system, which are the elevator panel motor and the rotation motor, we also relied on our Arduino mega board. For our prototype , we sufficed with the use of simple direct current motor driver to emulate a real life situation where our Arduino mega board can also control the motor of the elevator using two relays and terminal levers for the 230V alternative current motor [26]. Our choice for a simple DC motor driver is the L298N circuit that grants our Arduino board full authority over the speed and the rotation direction of the direct current motor [27]. For the panel rotation motor, we used a servomotor connected to the Arduino mega. The servomotor delivers an immense management and control to every detail of the panel rotation and angular rotation manipulation [28]. With the use of the servomotor, we guarantee that our smart parking system is capable to reach all parking positions throughout the facility thanks to the 360 degrees of freedom available to the rotating panel. In figure 12 we present the wiring diagram for both the DC motor and the servomotor.

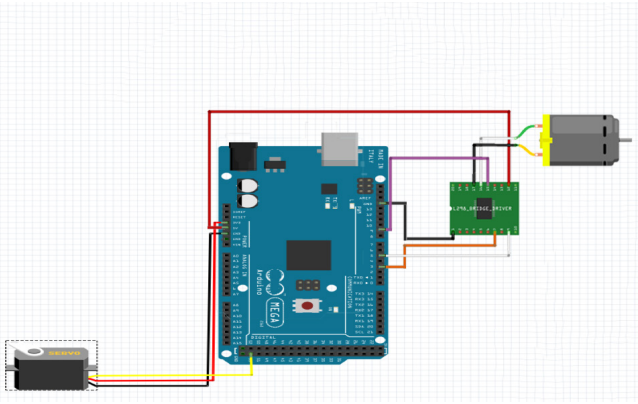


Figure 12: The Arduino mega board wiring diagram with the L298N motor driver, the DC motor and the servo motor.

4.4. Result comparison

Comparing the system we have set up with the different intelligent parking lots existing in the literature we notice that our system has several points in common with them but has advantages in terms of space optimization and automation of the system.

Table 1 illustrates this comparison.

Table 1: ESP2 parking comparison

	Mobile App	Automatic	Secured	Smart
Park Assist	Yes	No	*	No
Hectronic	Yes	No	***	No
Parking +	Yes	No	***	Yes
ESP2	Yes	Yes	*****	Yes

As shown in Table 1, it can be seen that all the systems described in the literature offer a mobile application for booking.

However, the car parking requires the intervention of the man (driver or parking manager) to park the car unlike our system which is fully automated based on sensors, IoT and embedded systems.

The minimization of human intervention makes it possible to secure the system against theft, minimize human errors and save energy consumption.

Finally, as Parking+, our system has an intelligent part to update the status of the parking lot and therefore make it easier to park the car while minimizing the waiting time for parking.

5. Conclusion

Throughout this paper we introduced a smart vehicle parking system that effectively manage the storage and automates vehicle parking using an Android application as the front-end user communication platform with our embedded control system. Our embedded control unit includes the main Arduino mega embedded board and the Arduino WIFI shield that ensures WIFI connection. The information for the presence of vehicles and the availability of empty parking location is acquired by the collection of PIR sensors and provided to the Arduino mega board to take the appropriate decision. This main control unit is also responsible for the management and control of the elevator panel and the rotation engine in charge of accessing the empty locations and stored vehicles.

Our novel smart parking system with the Android customer communication unit has delivered a better storage management capability than the previously mentioned commercially available competitors. The dominance of our developed system over Park Assist and Hectronic parking system is clearly manifested in the long term cost gained from the smart management of the allocated space that our competitors sacrifice in order to provide the driving pavement needed to connect the vehicle to the driving spot. The Parking Plus application also miss the security quality that our solution provides compared to their nonexistent answer to this problem.

Although our system is a great solution for smart parking, we do not deny the fact that it has a few shortcomings such as the need for more flexibility when it comes to the architectural structure of the facility and the power management of the mechanical parts (elevator panel and rotation engine). These limitations can be the subject of future research that can include introducing a sustainable energy resources and other architectural designs. we can also ensure a better security by implementing a decentralized Blockchain based security control system in case of an expansion to a multiple units of smart parking systems.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. Li, W. Gu, W. Chen, Y. He, Yannian Wu, Y. Zhang, “Smart Home: Architecture, Technologies and Systems”, Procedia Computer Science, 131, 393-400, ISSN 1877-0509, 2018, doi:10.1016/j.procs.2018.04.219.

[2] J. Muangprathub, N. Boonnam, S. Kajornkasirat, N. Lekbangpong, A. Wanichsombat, P. Nillaor, "IoT and agriculture data analysis for smart farm", "Computers and Electronics in Agriculture", 156, 467-474, ISSN 0168-1699, 2019, doi: 10.1016/j.compag.2018.12.011.

[3] F. Al-Turjman, A. Malekloo, "Smart parking in IoT-enabled cities: A survey", Sustainable Cities and Society, 49, 101608, ISSN 2210-6707, 2019

[4] A. Mathijssen and A. J. Pretorius, "Specification, Analysis and Verification of an Automated Parking Garage",. 40, 2005.

[5] K. Ghidouche Aït-Yahia, F. Ghidouche, G. N'Goala, "Smart city of Algiers: defining its context", Leonidas Anthopoulos, Smart City Emergence, Elsevier, 391-405, 2019. https://doi.org/10.1016/B978-0-12-816169-2.00019-5

[6] I. Chiuchisan and O. Geman, "Trends in embedded systems for e-Health and biomedical applications," 2016 International Conference and Exposition on Electrical and Power Engineering, 304-308, 2016, doi: 10.1109/ICEPE.2016.7781352

[7] R. Y. Zhong, X. Xu, E. Klotz, S. T. Newman, "Intelligent Manufacturing in the Context of Industry 4.0: A Review," Engineering, **3**(5), 616-630, ISSN 2095-8099, 2017, doi: 10.1016/J.ENG.2017.05.015.

[8] Howard, K. Charles, K. Cayetano, and O. Omojola. "Parking management systems." U.S. Patent No. 6,885,311. 26 Apr. 2005.

[9] A. D'Ausilio "Arduino: A low-cost multipurpose lab equipment." Behavior research methods, **44**(2), 305-313, 2012. doi:10.1016/2012.04.219.

[10] https://www.parkassist.com

[11] A. Najmeddine and T. Frikha, "A fully automated intelligent system for vehicle storage and parking management using android and arduino hardware and applications," 2019 IEEE International Conference on Design & Test of Integrated Micro & Nano-Systems (DTS), Gammarth-Tunis, Tunisia, 2019, 1-4, doi: 10.1109/DTSS.2019.8915278.

[12] P. Dharma Reddy, A. Rajeshwar Rao, A.S Musthak. "An Intelligent Parking Guidance and Information System by using image processing technique". International Journal of Advanced Research in Computer and Communication Engineering. **2**(10), 2013.

[13] M. Patil, V.N Bhonge, "Wireless sensor network and RFID for smart parking system". International Journal of Emerging Technology and Advanced Engineering, **3**(4), 188-192, 2013

[14] E. Karbab, D. Djenouri, S. Boulkaboul, & A. Bagula, "Car park management with networked wireless sensors and active RFID" In 2015 IEEE International Conference on Electro/Information Technology (EIT), 373-378, 2015.

[15] Z. Suryady, G.R Sinniah, S. Haseeb, M.T Siddique, M.F.M Ezani, "Rapid development of smart parking system with cloud-based platforms." In Information and Communication Technology for The Muslim World (ICT4M), 2014, 1-6

[16] B.M.K Gandhi, M.K Rao. "A Prototype for IoT based Car Parking Management system for Smart cities", Indian Journal of Science and Technology, **9**(17), 2016.

[17] https://hectronic.com/en/solutions/parking/

[18] https://www.parkingplus.ro

[19] T. Miroslav, S. Gutta, and V. Philomin. "Smart parking advisor." U.S. Patent No. 6, 426, 708. 30 Jul. 2002.

[20] J. Yatin, & al. "Understanding smart and automated parking technology." Int. J. u-e-Serv. Sci. Technol, **8**, 251-262, 2015.

[21] M. Goadrich, M. P. Rogers. "Smart smartphone development: iOS versus Android." Proceedings of the 42nd ACM technical symposium on Computer science education. ACM, 2011.

[22] D. Wolber, &al. App Inventor. " O'Reilly Media, Inc.", 2011.

[23] Arduino, S. A. "Arduino." Arduino LLC 2015.

[24] A. Javed, "Building Arduino Projects for the Internet of Things." Experiments with Real-World Applications." United States of America: Apress Media, LLC,177- 193, 2016.

[25] J. Xin, et al. "Target detection and classification using seismic and PIR sensors." IEEE Sensors Journal **12**(6), 1709-1718, 2012.

[26] https://www.instructables.com/id/3-floor-elevator-arduino/

[27] J.D Warren, J. Adams, and H. Molle. "Arduino for robotics." Arduino robotics. Apress, Berkeley, CA, 51-82, 2011.

[28] J. Blum, "Exploring Arduino: tools and techniques for engineering wizardry" John Wiley & Sons, 80-85, 2013

[29] http://www.ee.ic.ac.uk/pcheung/teaching/DE1_EE/stores/sg90_datasheet.pdf

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 1577-1592 (2020)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

A Framework for Adoption and Diffusion of Mobile Applications in Africa

Chinedu Wilfred Okonkwo[*,1,2], Magda Huisman[2], Estelle Taylor[2]

[1]*School of Consumer Intelligence and Information Systems, University of Johannesburg, Johannesburg, 2001, South Africa*

[2]*School of Computer, Statistical and Mathematical Sciences, North-West University, Potchefstroom Campus, 2520, South Africa*

ARTICLE INFO

Article history:
Received: 11 October, 2020
Accepted: 28 November, 2020
Online: 25 December, 2020

Keywords:
Mobile Applications
Adoption
Diffusion
Framework
Africa

ABSTRACT

The adoption and diffusion of mobile applications (mobile apps) has become the base of modern activities in Africa owing to the services and values that are obtained through mobile apps innovations. More emphasis has been on the development of mobile apps whereas the adoption and diffusion process as well as their predictors are ignored or given less attention. This study explores and develops a framework for the adoption and diffusion of mobile apps in Africa. A survey was conducted on the basis of the diffusion of innovation framework in five-selected African nations. A total of 1285 of the 2300 distributed questionnaires were returned, giving a response rate of 55.87%. The results indicated that many factors/predictors drive the acceptance and use of mobile apps and these factors were structured to develop a common framework for mobile apps adoption and diffusion in Africa.

1 Introduction

The advancement in mobile technology has had a major impact on the manner in which business and personal activities are performed. It could be said, in particular, that a radical evolution started with the advent of handheld devices such as mobile phones that caused the creation of distinct kinds of mobile apps that are used in separate activities [1, 2]. A number of mobile technology start-ups have been created in Africa to encourage mobile app innovation [2, 3]. Advances in digital technology have enhanced communication systems in Africa, leading to a growing amount of African communities linked to digital communications [4, 5]. The use of mobile apps are changing the African society, socially, economically and politically [1], [6, 7]. Despite the penetration and growth of mobile services, some mobile apps are rejected and discontinued to be used. Moreover, there are some non-users of mobile apps in Africa [8]. The process of acquiring and disseminating mobile app services in Africa must therefore be improved, particularly for those with local content.

It is imperative to information and communication technology (ICT) experts and scholars to know the flow of information technologies [3]. This can be achieved by conducting a research on adoption and diffusion, which implies accepting to use [2, 9, 10] and continuing the use of a new innovation [11, 12]. Various models have been proposed to aid this process including frameworks for mobile health [13], mobile banking [14] and e-government [15]. These models dealt on a particular mobile app in a specific country and cannot be generalised.

However, African mobile consumers are growing interest in home-grown mobile apps. Several mobile apps have originated in Africa, especially in South Africa, Nigeria and Kenya including social networking mobile apps like Mxit, 2go and How Far, entertainment mobile apps like Simfy Africa, Spinlet, iRokoTV and SOLO, and financial mobile apps like M-pesa [2]. These mobile apps are mostly used in the country where they are created whereas in the other countries of the continent, there is little or no adoption of the mobile apps. In other words, there is a significant difference between the rate of adoption of each respective mobile app between the country where it is created and the other countries in Africa.

This study proposes a general framework for mobile apps adoption and diffusion in African social system. Drawing from the literature on the adoption of mobile app innovation and deliveries, this research intends to address the following questions:

1. What are the factors that influences mobile apps adoption in

[*]Corresponding Author: Chinedu Wilfred Okonkwo, School of Consumer Intelligence and Information Systems, University of Johannesburg, Johannesburg, 2001, South Africa., +27 83 579 6557 & Email: chinedung@yahoo.com

Africa and

2. How can we structure these factors to form a general model to enhance the acceptance and use of mobile app innovation in Africa.

In order to achieve these goals, we conducted a survey in the five countries representing the five geographical locations in Africa to assess the user's view of mobile app use in order to explore the factors predicting the acceptance and use of mobile apps in Africa. These factors were structured to develop a framework. With leverage on the report that there is a need to know the diffusion of mobile technologies and the important of mobile technological services on developing nations such as African countries, this study would therefore make the following contributions. We have:

1. Examined and discovered the acceptance and use of mobile apps in Africa,
2. determined the factors/predictors that influence users' intention to adopt and use mobile apps innovations in Africa, and
3. developed a general framework that can improve the adoption and use of different kinds of mobile apps across African nations.

The remainder of this research work is organised as follows: Section 2 discusses the background information of the research work. Section 3 describes the related works and the gap. Section 4 shows the research conceptual model and Section 5 explains the research methodology. Section 6 describes the methods of analysis and Section 7 zooms into the survey results. Section 8 presents the regression analysis and Section 9 discusses analysis of variance (ANOVA) results. Section 10 interprets the results to validate the proposed framework. Section 11 concludes the study.

2 Background information

A brief theoretical overview of the study is presented in this part.

2.1 What are mobile apps?

Mobile app creativity has recently attracted scholars from different fields of study to carry out studies in their respective fields. As a consequence, mobile apps have been described by several writers, write-ups and web pages in many perspectives including Technology [16, 17], Educational [18], Context-aware Computing [19], Internet of Things [20], Business [21], Cloud Computing [22] and Social Services [23]. In this study, mobile apps are considered to be mobile technological innovations of a mobile computer software that allows users to carry out various activities including political, social, economic, and educational regardless of the user's location and background.

2.2 Features of mobile apps

There are five main characteristics of mobile apps, according to Rogers [9]. They are relative advantage,compatibility, complexity, observability and trialability. Such characteristics help to decrease the level of misunderstanding and increase the adoption rate of mobile apps. Technology reliability and cost are conceptualized in this study as important features of mobile apps [8].

2.3 Innovation decision process

This is a process of decision-making which includes an information-seeking and information-processing practices in which mobile app adopters are driven to reduce the confusion surrounding the effect of acceptance and use of a new mobile app innovation. Rogers's diffusion framework consists of five stages of the innovation decision process including knowledge, persuasion, decision-making, implementation and confirmation [9]. For this study implementation and confirmation are merged because they describe diffusion (putting into use and continue to use) and referred to as the diffusion stage. Thus, the innovation decision process of this research entails four phases including knowledge, persuasion, decision-making and diffusion. These stages are briefly discussed below.

Knowledge stage:The mobile apps adoption and diffusion begins with the knowledge about mobile apps. This stage involves acquisition, method and sharing of information about mobile apps. Having the right knowledge about mobile apps requires information sources, learning methods and information sharing. In Africa, there are two distinct sources of information about mobile apps including external initiated actions and personal initiated actions, learning methods are interpersonal communications, the Internet, operational manual and informal training, whereas information sharing involves organizations and events, and individual activities [8].

Persuasion stag: is shaping the minds of adopters towards adopting and using mobile apps. The main factor under this stage is the social influence [9]. Decision-making stage: This includes the adoption and rejection of mobile apps. In this stage, an individual/adopter of mobile apps chooses to adopt or reject the mobile apps. The adoption covers the adopter's category whereas rejection covers the reasons for non-adoption [9]. There are five categories of adopters in Africa including late majority and laggards, non-adopters, innovators, early adopters and early majority [8]. Also, the reasons for non-adoption of mobile apps in Africa are using mobile apps is time-consuming, using mobile apps takes a lot of phone memory, using mobile apps distracts attention from other activities and using mobile is expensive [8]. Diffusion stage: This involves the implementation and confirmation. That is putting mobile apps into use and continuing to use them. It covers the use of mobile apps, ways of using mobile apps and reasons for discontinuance of the use of mobile apps. There are two distinct ways of using mobile apps in Africa including business use and individual use, and the reasons for discontinuance include a better replacement, unsatisfactory performance, High complexity and lack of support [8].

2.4 Adopting Unit

The adopting unit consists of features of individual adoption, comprising personality variables, socioeconomic and communication behavior. [9].

Personality variables: Personality variables entail the individual characteristics (demographic) that are part of the factors influenc-

ing the outcome of the innovation decision process [9]. Adopters' attitudes, anxiety (degree of nervousness) and self-efficacy are conceptualized as part the personality variables. Attitude is measured by the feelings and reactions of the participants on the importance and contributions of mobile apps in Africa [24].

Socioeconomic characteristics: Socioeconomic characteristics involve the external factors that may influence individuals' adoption and diffusion of mobile apps. These characteristics include the contributions of mobile apps to the development of Africa [25], facilitating conditions and employment status[8] .

Communication behaviour: This covers the primary purpose of using and disseminating the information about a newly developed mobile app [9]. It is one of the primary goals that encourages a person to embrace and disseminate mobile apps. In this study, It deals with the function of information sharing as specified in the abstraction model of the research. This research deals with two key reasons of spreading the information about a newly created mobile app including: [8].

1. to facilitates both individual and business activities, and
2. to improve and make communications easier.

2.5 Adoption

According to Rogers [9], adoption is the absolute willingness to embrace and use new advanced technologies (mobile apps). GSMA [2] explained that adoption is embracing and using a new technology. It is the acceptance [10] and use [11] of an emerging product such as mobile apps by the consumer. Innovation is new when it is first introduced to a person. Therefore, we describe the adoption as the users' acceptance and use of a new invention.

2.6 Diffusion

Diffusion is a means of spreading the information and use of an innovation among the individuals of a social structure across certain networks [9, 3]. It is a mechanism by which a commodity is spread from one customer to another [11]. For this research, with regards to Rogers framework, we combine the implementation and confirmation phases to form diffusion. Thus, we defined diffusion as the method of sharing, uses and discontinued the uses of a product.

2.7 Usage of mobile applications in Africa

GSMA [26] recorded that, there are more than five billion mobile subscribers across the globe and the number is expected to rise to 5.9 billion by the year 2025. The spread of mobile technology is wider than any other technology [27]. Mobile service penetration is expanding at an extraordinary rate in Africa, and mobile apps are part of the main foundations of all mobile services. The introduction of mobile services in Africa is rapidly growing such that the mobile app is the driving force behind some of the emerging developments in African nations.[16, 7]. The African continent currently ranks among the world's largest adopters of mobile technology services, introducing new mobile services to the population [28, 29].

The beginning of mobile app development can be linked to the program required to operate the emerged mobile devices [30]. Over the years, telecommunications in Africa was primarily through a fixed-line system and the networks were regulated by the respective countries' central governments [31, 32]. In the late 1990s, the advent of mobile telecommunications brought significant changes to the African continent, so that countries like Nigeria, Cote d'Ivoire, and Uganda had more mobile phones than fixed lines [31]. In Africa, the mobile telecommunications system has a faster and broader penetration and deployment than fixed lines, and some nations, such as the South Africa, Comoros, Uganda, Kenya, Mauritius, Malawi, and the Seychelles, have increased their rural mobile network coverage as of 2010 [33]. The advancement of mobile technology has contributed to the creation, acceptance and dissemination of mobile apps to promote mobile services, leading to the production and rise in use of different kinds of mobile devices such as smartphones.

The procurement and service of a fixed-line telephone system is highly expensive. As a result, much of the African population in particular rural dwellers and low-income earners cannot afford or have access to it [4, 32]. A broad range of network coverage is provided by mobile technology enabling quick access to the African population to telecommunications services [34, 35]. Communication services are currently accessible and affordable and more than 70% of the African people is linked to the digital telecommunications network[27]. Thus, mobile technology services enable digital inclusion.

The development, delivery and use of mobile technology services included a group of interconnected network entities called the mobile ecosystem [16, 36, 37]. The African mobile ecosystem comprises of various entities that are involve in the development and delivery of mobile innovation services including mobile network operators, sales agencies, government and other stakeholders [16, 7]. Several facets of the mobile ecosystem in conjunction with a defined system development method are involved in the process of creating, delivering, implementing and disseminating mobile apps [38]. With regard to the stakeholder's priorities, aspirations, operational services, and some predicting factors [39] mobile apps are developed for different services. There are several types of mobile apps that offer unique services to their users, including: Mobile Learning [40], Mobile Commerce [41], Mobile Health [42], Mobile Agriculture [43], Mobile Banking [44], and Mobile Governance [45] used to offer different online services respectively. In this study, these various kinds of mobile apps are categorised as the classification of mobile apps. The use of mobile apps in Africa has added tremendously to the growth of Africa [1]. That is, mobile services are rising the living standards of the people of Africa, resulting in the promotion of the economic, social, political, and educational development of all African nations. [46, 47]. Thus, the ecosystem promotes mobile apps innovations which accelerate economic growth, though there are some issues to be considered.

While mobile app developments are growing in Africa, some areas, especially rural locations, are still experiencing a slow penetration of this advanced technology due to certain factors or issues of influence [29]. Around 60% of Africans resides in remote areas

where a large number of potential adopters are limited by the poor network connectivity, low income, unstable economic situations, and social and political instability [29]. In addition, some scholars have established numerous important factors that need to be addressed, including relative advantage, complexity, compatibility, trialability, observability, [9], [48]–[49], social influence, user perception, perceived technology reliability [49, 50], facilitating conditions [51], demographic, mobile technology self-efficacy, ICT anxiety, perceived financial cost [49, 52, 53] and culture [54], for greater acceptance and distribution of mobile apps in Africa.

However, in the African background, it is important to understand the whole procedures of adopting and disseminating mobile apps and to create a well-defined structure for this process. The goal of this study is to address these issues by developing a well-defined framework that can encourage the wider and more efficient use of home-made mobile apps, not only in the country of origin but also in other African nations. Through a survey conducted in five selected nations representing five geographical regions in Africa the process of mobile apps adoption and diffusion was examined. These countries include south (South Africa), west (Nigeria), north (Morocco), east (Kenya) and central (Democratic Republic of Congo).
These countries were selected because:

1. Each country is situated in one of Africa's five regions.
2. There is a significant population in each of the countries chosen.
3. More references were made to them in the literature review than to those in their respective regions.
4. Each region is presumed to have a similar lifestyle type.

2.8 What is a Framework?

A framework is a blueprint or guide consisting of principles, constructs and concept of a theory [55]. According to Fulton, Krainovich-Miller [55, 56], conceptual framework is a structure of concepts that offers the foundation for the development of an abstract idea. For this study, a framework is an abstraction for a successful mobile apps adoption and diffusion among entire African social system. This study seeks to develop a conceptual framework for mobile apps adoption and diffusion in Africa. Applying regression analysis to the research model, the relationships between the research constructs can be revealed.

3 Related Works

Although there is no general framework for the adoption and diffusion of mobile apps in all African geographical areas but some related works have been carried out. The existing frameworks apply to a particular category of mobile app and to a unique environment or country. A model for the adoption of mobile health in developing countries was proposed by Kiongo [13] using Kenya as a case study. In the framework, social influence, facilitating conditions, reliability, cost and simplicity are considered as determining factors of the intention to adopt and use mobile health. In a model for mobile banking adoption in Ghana, Crabbe, Standing [14] highlighted that usefulness, reliability and culture are factors shaping the attitudes and behavioural intention to use mobile banking. Bwalya and Healy [15] proposed a conceptual model for e-government adoption in the SADC region with consideration to the following factors; culture, user support, usefulness, ease of use and appropriate ICT infrastructure. Social influence, facilitating conditions, cost, culture and reliability are the predicting factors considered in a model for the adoption of e-commerce in Nigeria [57]. According to Moore and Benbasat [58], relative advantage and complexity are equivalent to usefulness and ease of use. In the light of the above-proposed frameworks, there is a need for a general model that can be applied to different domains and different regions/countries in Africa to increase the use of African-based mobile apps across all African nations.

3.1 The Gap

The development of mobile apps in Africa is growing rapidly [7]. A variety of mobile app start-ups have been introduced across several African countries to promote mobile app technologies. [2]. The African mobile ecosystem has promoted the creation of various kinds of mobile apps that are available on the African technology market. As a result, a large number of mobile apps originated in Africa [2]. From the literature, all the existing suggested frameworks are for specific kinds of mobile apps [13]–[15], [57] and there is no framework for the general mobile apps adoption and diffusion in Africa. In practice, most of the mobile apps created in Africa are actively used in the country where they are originated, while there is low or no acceptance of these mobile apps in other nations of the continent. It is necessary to have a common structure that will enable a wider adoption of mobile apps (especially local mobile apps) across African nations. This research will therefore assist to close the gap by suggesting an overall framework for mobile apps adoption and diffusion in Africa.

4 Research Conceptual Model (RCM)

Drawing from the diffusion of innovation framework, previous related studies on Africa context regarding the framework for adoption of mobile apps and their findings, a conceptual model for this study is, therefore, proposed as displayed in figure 1. The RCM demonstrates the methodological approach to accomplish the research aim. The model consists of three components including mobile apps innovations, innovation decision process and adopting unit. The mobile apps innovations component describes mobile apps and their features; innovation decision processes component deals with the acceptance, rejection, use, discontinuance and continues the use of mobile apps whereas adopting unit contains the influencing factors of mobile apps adoption and diffusion. Achieving the primary goals of this research work would validate this model.

The Mobile apps section describes that: as an innovation, a newly created mobile app should be considered to be better than the previous similar apps that it replaces, fits well with the intended potential users, easy to understand and use, tested before implemen-

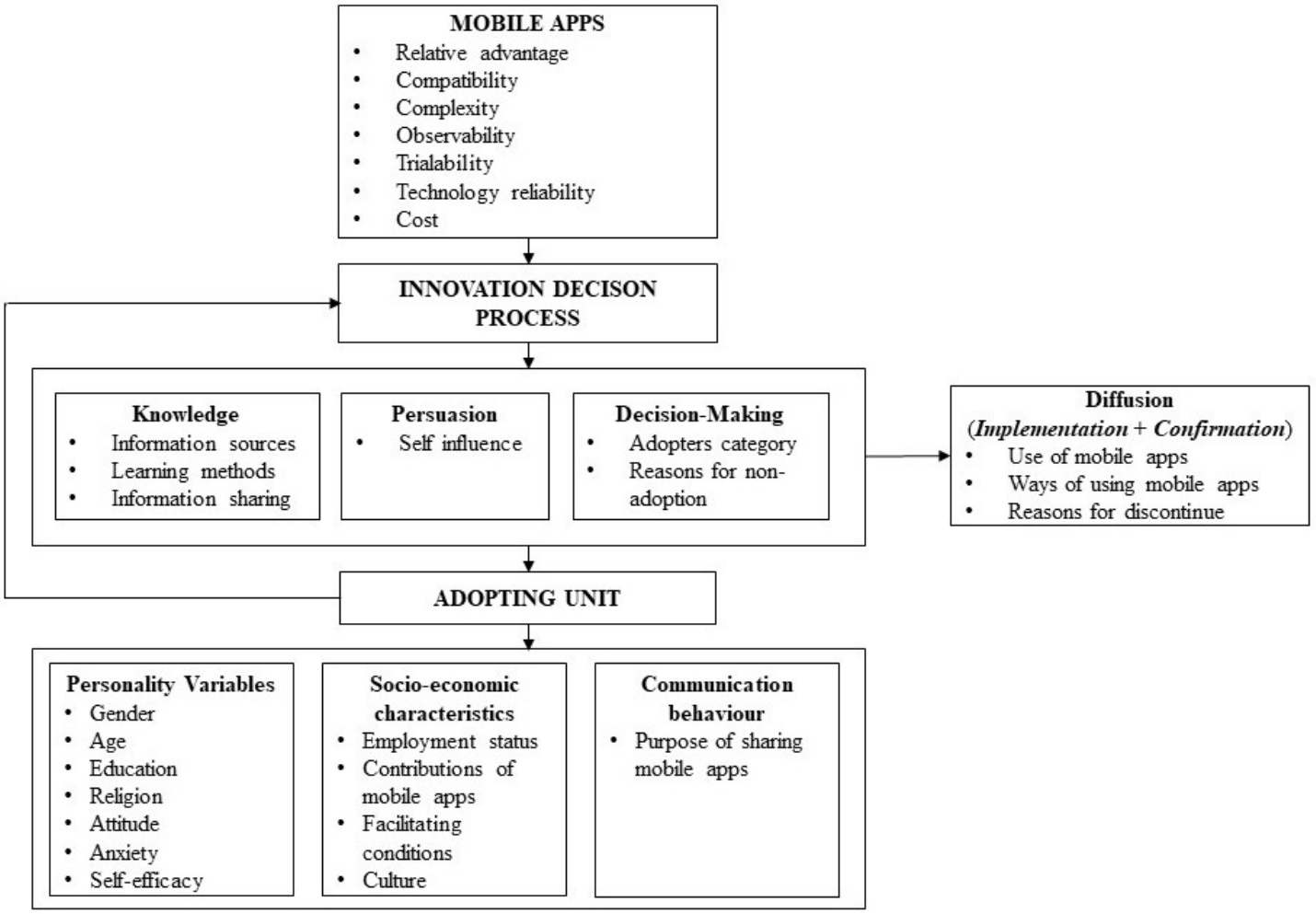


Figure 1: Research conceptual model

tation, accessible to people, and be credible. The innovation decision process explains the need for the public to be aware about the nature of a new mobile app, including source of information, the method of using the mobile app, and how to exchange information about it. Given these will help to influence the decision-making process of the adopters, either to accept or reject the mobile app. The adopting unit deals directly with the individual adopters, whereby certain personalized factors, such as personality variables, socio-economic characteristics, and communication behaviour, can affect the process of adoption. Overall, with the influence of good innovation characteristics, personalised factors and good decision-making process, a newly developed mobile app may gain widespread acceptance within the African social system.

5 Research Methodology

This research requires quantitative data from mobile apps users residing in five selected countries representing various African geographical regions. Hence, the positivism paradigm was adopted whereby a scientific method was used to gather self-determined data from the participants without the researcher's influence. The non-probabilistic technique of sampling was introduced since it is not easy in Africa to get the correct amount of mobile app users. In choosing the sample population for data gathering, purposive and convenience methods were considered. Purposive sampling involves the selection of a target audience based on a specific purpose. The study aims at the mobile apps adoption and diffusion and the population was selected from the area where there are large number of users of mobile apps for various activities. Convenience sampling is a particular type of data collection approach that entails the gathering of data from the participants that can be conveniently reached within the chosen population.

As the entire likely sample could not be covered by the researcher, therefore it was appropriate to pick a sample population containing all the desired participants. Thus, the purposive method was needed. The researcher was unable to reach everyone in the selected sample except those who were easily accessible; the convenience technique became essential at this juncture and was also applied. Educational institutions (universities) were selected with the purposeful method, and data was obtained from the participants with the convenience method. In order to achieve the research goals, the RCM formed the basis of the survey performed in the selected

nations.

5.1 The Research Instrument – The Questionnaire

The questionnaire was used as a data collection tool. It consists of seven sections containing close-ended questions with a 5-point scale rating of 1–5 (1 strongly disagreed and 5 strongly agreed). These sections are:

1. Background information – collects individuals (participants) details.
2. Mobile applications – examines the extent of participants' awareness of mobile app technologies.
3. mobile apps innovation Adoption and diffusion – examine Africa's acceptance and use of mobile apps.
4. Innovation features – examine the participants' perceptions of the importance of the characteristics of mobile apps.
5. Impact of mobile apps – examine the contributions of mobile apps to African growth.
6. Predicting factors – examine the factors influencing the acceptance and use of mobile apps in Africa.
7. Adopter's category – to determine the types of adopters of mobile apps existing in Africa.

A pilot or experimental test was conducted to make sure that the questionnaire can resloved the important parts of the study. All the recommendations were properly considered and effected. Also, the questionnaire was vetted and approved by a Statistical Consultant. The target population is the mobile apps users from the chosen nations in Africa. In the delivery of the study questionnaires, drop off and electronic/online approaches were applied. Participants were drawn from various age groups and job status to ensure that general views are covered by the collected data. A contact person was hired to help us in collecting responses from the respondents in each of the respective counties. The person was informed of the research process and the aim of the study was clarified. A total of 1285 of the 2300 distributed questionnaires were returned, providing a response rate of 55.87%. The responses comprise of 957 drop off and 328 electronically (online google forms) completed questionnaires. The responses obtained from the participants of each country was shown in Table 1.

Table 1: Participants' responses from each country

Country	**Amount disseminated**	**Amount received**
South Africa	500	335
Nigeria	500	370
Morocco	400	116
Kenya	400	212
DRC	500	252
Total	2300	1,285

Table 2: Reliability

Research variables	Mean	Std dev.	Cronbach Alpha Coefficient
Business interactions.	3.19	0.99	0.76
Individual interaction.	3.83	0.86	0.69
External initiated actions.	3.21	0.98	0.66
Personal initiated actions.	3.61	0.99	0.60
Organizations and events.	3.36	0.99	0.61
Individual activities.	3.68	0.86	0.60
Relative advantage and compatibility	3.59	0.84	0.79
Observability	3.44	0.87	0.60
Trialability	3.48	0.79	0.69
Perceived technology reliability	3.38	0.83	0.61
Self-efficacy	3.58	0.79	0.64
Social influence	3.56	0.96	0.64
Facilitating conditions	3.36	0.99	0.65
Financial cost	3.44	0.98	0.64
Cultural effect	3.31	0.88	0.63
Contributions to the healthcare system.	3.52	0.96	0.75
Contributions to education and social activities.	3.65	0.75	0.61
Contributions to the economy.	3.62	0.85	0.61
Contributions to politics.	3.60	0.86	0.60
Contributions to finance.	3.79	0.87	0.61
Contributions to agriculture.	3.39	0.95	0.69
Late Majority and Laggards	3.25	0.95	0.63
Non-Adopters	3.17	1.02	0.78
Innovators	3.19	0.88	0.80
Early adopters	3.33	0.89	0.72
Early majority	3.34	0.88	0.71

6 Analysis

To explain the frequency and response percentage of the participants, statistical analysis was performed on all the study variables. Cronbach alpha reliability analysis was used to measure reliability coefficients of the investigating variables (criteria 0.6). In order to assess the difference in the groups of dependent variables, analysis of variance (ANOVA) was conducted on the study variables, involving individual profile data (criteria, $p < 0.05$). Finally, regression analysis was performed to measure the relationship among the dependent and the independent variables. The results obtained from the data analysis will be interpreted on the basis of the RCM components. See section 7

6.1 Reliability

Using Cronbach alpha coefficient analysis, the reliability of the study variables was checked and it was found that the factors were internally consistent ($\alpha = 0.67$). Also, the correlation scores of all the elements are the same (r=1), meaning that the factor elements

associate with each other. The reliability results are displayed in Table 2.

7 Survey Results

The main objective of this research is to develop a framework for mobile app adoption and diffusion in Africa. The quantitative data was obtained from 1,285 participants from five chosen nations representing the geographical regions of Africa. The data collected was carefully extracted, statistically analysed and the results were obtained. The outcomes of the survey are presented according to the mean values of the responses obtained from the participants with respect to scale rating except the participants' profiles which is presented on percentages.

7.1 Participants' profiles

The research sample population consists of 703 (54.71%) of males and 582 (45.29%) of females. This suggests a reasonably equal distribution of the gender categories (male and female). Mobile apps are designed to be used equally by all, regardless of gender [56]. It shows a good representation of the African population. Participants were of various age groups, from less than 20 to 70 years of age and above. The results revealed that the majority of participants were between the age groups of 20-30 and 31-40 with the proportions of 409 (31.8%) and 379 (29.5%) respectively. That is, >60% of the participating population was between the ages of 20 and 40 years. All levels of education, including Matrics, Diploma, University, Masters, Doctorate and others were involved in this study. 'Others' means the set of people that are deemed to be uneducated and could not fall into these classifications. 93.3% of respondents are educated, while 3.7% are uneducated. This shows that the responses were collected from individuals who had acquired some knowledge of mobile apps innovations and services. A large proportion of the participants are working class and students with 615 (47.9%) and 402 (31.3%) respectively. Such people have the capacity of embracing and using mobile apps, even though some financial costs are involved. In addition, students and workers are also more involved in many things than elderly and unemployed people, allowing more use of mobile apps for different tasks.

7.2 Mobile apps Component

This section investigated the participants' perceptions of the features of mobile apps.

Relative advantage - is the degree to which users of a specific mobile app considered it to be more improved than another equivalent mobile app. The mean value of the participants' responses is 3.74 meaning that relative advantage is an influencing feature of mobile apps.

Compatibility – is the extent to which mobile apps suit with the exiting values, practices and demands of the prospective adopters. The obtained mean value of the participants' responses is 3.78 suggesting that compatibility is a significant feature of mobile apps.

Complexity – the degree to which participants believe that the use of mobile apps is affected by how simple it is to use the mobile app. The obtained mean value of the participants' responses is 3.78 suggesting that complexity is an influencing characteristic of mobile apps.

Observability – is the degree to which participants believe that seeing the impact or what can be achieved with the mobile app will convince people to accept and use the mobile app. The obtained mean value is 3.52 suggesting that observability is an influencing feature of mobile apps.

Trialability – is the degree to which the participants believe that testing a mobile app before adopting can influence the adoption of the mobile app. The obtained mean value of the participants' responses is 3.72 suggesting that trialability is an influencing feature of mobile apps.

Perceived technology reliability - deals with the belief, danger, and credibility involved in using a new mobile app. The obtained mean value of the participants' responses is 3.38 suggesting that perceived technology reliability is an influencing characteristics of mobile apps.

Cost – are the financial costs involved with the adoption and usage of mobile app services. The obtained mean value of the participants' responses is 3.44 suggesting that cost is an influencing feature of mobile apps.

7.3 Innovation Decision Process Component

This section investigated the innovation decision process of mobile apps including the knowledge, persuasion, decision-making, and how the information about mobile apps are being spread in Africa.

Information sources – sources of disseminating information about the existence of mobile apps. The obtained mean values are external initiated actions, 3.21 and personal initiated actions, 3.61 indicating that the sources are important to the acceptance and use of mobile apps.

Learning methods – are the methods of learning how to use mobile apps correctly. The obtained mean values from the participants' responses are: interpersonal communications, 3.6; the internet, 3.5; operational manual, 3.3; informal training, 3.3 and formal training, 2.8. This implies that all the methods are significant except formal training method.

Information sharing – are the communication channels through which the participants share the information about mobile apps with one another. The obtained mean values from the participant's responses are: organizations and events, 3.36 and personal activities, 3.68 indicating that these are significant communication channels of mobile apps adoption diffusion.

Social influence – is the level to which the adopters of mobile apps are influence by others (family, friends, government, etc.). The obtained mean value is 3.56 indicating that social influence has a significant effect on the consumers intention to adopt and use mobile apps.

Late majority and Laggards category – are the most skeptical (wait and see) and most localized set of adopters of mobile apps. The obtained mean value is 3.25 indicating a significant group of mobile apps adopters in Africa.

Non-adopters category – are the set of participants that do not adopt and use mobile apps at all. The obtained mean value from the participants' responses is 3.18 indicating that there are non-users of

mobile services in Africa.

Innovators category – involve a group of users eager to experience or practice new innovation, regardless of the advantages or losses associated with it. The obtained mean value from the participants' responses is 3.19 indicating a significant group of mobile apps adopters in Africa.

Early adopters category – they are social system leaders and advisors and their acts are confined within the limits of the social system. The obtained mean value from the participants' responses is 3.34 indicating a significant group of mobile apps adopters in Africa.

Early majority category – they are non-leaders with strong social interactions who actively promote invention and they are neither the first nor the last to adopt the innovation. The obtained mean value from the participants' responses is 3.35 indicating a significant group of mobile apps adopters in Africa.

Reasons for non-adoption – are the factors that cause the rejection of mobile apps. The obtained mean values from the responses of the participants are: mobile apps occupies a large memory space in a phone, 3.4; use of mobile apps takes time, 3.3; use of mobile apps distracts focus from other tasks, 3.3; use of mobile apps is costly, 3.3. The results indicate that these reasons may cause the rejection of mobile apps.

Use of mobile apps – Ask to decide whether participants are mobile app adopters/users. The obtained mean value from the participants' responses is 4.0 indicating that most of the sample population are mobile apps users.

Ways of using mobile apps – determine the different aims of using mobile apps. From the responses of the participants, the obtained mean values are: business use, 3.20 and individual use, 3.83 indicating that these are the important ways of using mobile apps in Africa.

Reasons for discontinuance – are the causes of stopping the use mobile apps. The obtained mean values from the participants' answers are: a new improved one to replace the existing one, 3.5; unsatisfactory results, 3.5; not easy to use, 3.3; lack of support, 3.2. The results imply that these are significant motives that may cause users to stop using a particular mobile app.

7.4 Adopting Unit Component

Attitude – is the reaction (positive or negative) of adopters towards new mobile apps. The obtained mean value from the the answers of the participants is 3.39 suggesting that attitude has significant effect on the acceptance and use of mobile apps.

Nervousness/anxiety – is the degree of nervousness whenever a new mobile app is introduced. The obtained mean value from the participants' responses is 3.20 suggesting that anxiety has significant effect on the acceptance and use of mobile apps.

Predicting factors – are the factors that influence mobile apps adoption and diffusion. The obtained mean values from the participants' responses are: self-efficacy, 3.58; facilitating conditions, 3.56; culture, 3.31. The results imply that these are important factors that may impact mobile apps adoption and diffusion.

Contribution of mobile apps – are the effect of the use of mobile apps services on the growth of Africa. The obtained mean values from the participants' responses are: contributions to healthcare system, 3.52; contributions to education and social activities, 3.65; contributions to economy, 3.62; contributions to politics, 3.60; contributions to finance, 3.80; contributions to agriculture, 3.40.

Purpose of information sharing – are the motives for spreading the news about mobile apps. The obtained mean values from the participants' responses are: to enable easy communications, 3.88 and to facilitate ones activities, 3.55 indicating that these are significant purposes of information sharing regarding mobile apps.

8 Regression Analysis

To analyse the association among the dependent and independent variables, multiple regression analysis was conducted. Two regression techniques were used to ensure that more accurate results are produced, including Enter and Stepwise techniques. Because the proposed framework measures the adoption and diffusion of mobile apps, the variables in the adoption and diffusion sections of the RCM are the dependent variables, while the remaining research variables are the independent variables. Regression analysis determines how the variations in the independent variables are associated to the variations in the dependents variables. These variations are represented by the coefficients while the p-values show the significant level of the variations. P <0.05 is the appropriate significant level. It suggests a 95% confidence interval (the possibility of obtaining the accurate effect). Any of the variables (predictor) that has p <0.05 is potential to be included into the model. The R^2 shows the proportion of the variance accounted for by the independent variables in the dependent variable. In Table 3, the effect sizes and the significant level of R^2 are shown, according to Ellis and Steyn [59].

Table 3: Effect sizes and significant levels of R^2

Effect size	Effect	Values of R^2	Conclusions of R^2
<0.15	Small	<0.13	Non-significant
0.15-0.35	Medium	0.13-0.25	Significant
>0.35	Large	>0.25	Practically important

The dependent variables are; the Use of mobile apps, Business uses, Individual uses, and the five categories of adopters.

The independent variables are; External initiated actions, Personal initiated actions, Organizations and events, Individual activities, relative advantage and compatibility, Observability, Trialability, Perceived technology reliability, Self-efficacy, Social influence, Facilitating conditions, Cost, Cultural effect, Contributions to the healthcare system, Contributions to education and social activities, Contributions to the economy, Contributions to politics, Contributions to finance and Contributions to agriculture. Also included to the independent variables are Informal training, Operational manual, The Internet, Interpersonal communications, Good innovations and Add to development. The significant results obtained from the regression analysis performed are shown in Tables 4 – 11.

8.1 General Use of Mobile App

This model indicates that the R^2 is 0.291, thus, the important independent variables accounted for 29.1% of the dependent variable variance (*General use of mobile apps*). R^2 >0.25 which is practically significant. Eight variables/predictors entered into the model which accounted for 27.3% of the entire dependent variable variance. These variables are Add to development, Good innovations, Relative advantage and compatibility, Perceived technology reliability, Contributions to healthcare, Occupy phone memory, Satisfy user expectations and Social influence. Thus, R^2 = 28.2% suggesting that the variables added in this model are practically significant for the proposed framework. See Table 4

8.2 Business Use

It indicates that the R^2 is 0.380, thus, the important independent variables accounted for 38.0% of the dependent variable variance (*Business use*). R^2 >0.25 which is practically significant. Eighteen variables that accounted for 37.0% of the overall dependent variable variance entered the model. These variables are Enable work activities, Relative advantage and compatibility, Culture, Add to development, Anxiety, Informal training, Internet, Contributions to finance. Satisfy User expectations, Working manual, Interpersonal communications, Advantageous, Contributions to agriculture, Contributions to politics, Observability, Social influence, Contributions to educations and social activities and Enable easy communications. Thus, R^2 >25.0% showing that the variables added in this model are practically significant for the proposed framework. See Table 5

8.3 Individual use

This model indicates that the R^2 is 0.447, so the significant independent variables accounted for 44.7% of the dependent variable variance (*Individual use*). R^2 >0.25 which is practically significant. Also, 15 variables that accounted for 43.8% of the overall dependent variable were added into the model. These variables are Add to development, Self-efficacy, Good innovations, Trialability, Internet, Interpersonal communications, Time consuming, Contributions to finance, Tested correctly, Observability, Contributions to agriculture, Contributions to education and social activities, Individual activities, Fits well, and Contributions to healthcare. Thus, R^2 >25.0% suggesting that the variables added in this model are practically significant for the intended framework. See Table 6

8.4 Late Majority and Laggards

It indicates that the R^2 is 0.256, so the important independent variables accounted for 25.6% of the dependent variable variance (*Late majority and laggard*). R^2 >0.25, which is practically significant. Also, 12 variables that accounted for 26.7% of the dependent variables variance were added into the model. These variables are Contributions to agriculture, Contributions to politics, Formal training, Used by others, Self-efficacy, Contributions to economy, Satisfy user expectations, Organizations and events, High complexity, Social influence, Perceived technology reliability, A better replacement, Occupy phone memory, and Informal training. Thus, R^2 >25.0% suggesting that the variables added in this model are practically significant for the intended framework. See Table 7.

8.5 Non-adopters Category

This model indicates that the R^2 is 0.312, so the significant independent variables accounted for 31.2% of the dependent variable variance (*Non-adopters*). R^2 >0.25, which is practically significant. Also, 15 variables that accounted for 30.1% of the dependent variable variance were added into the model. These variables are Financial cost, Formal training, High complexity, Use by others, Culture, Satisfy user expectations, Facilitating conditions, Lack of support, Distraction of attention, Contributions to educations and social activities, Contributions to economy, Trialability, Observability, Good innovations, and Perceived technology reliability. Thus, R^2 >25.0% suggesting that the variables added in this model are practically significant for the intended framework. See Table 8.

8.6 Innovators Category

It suggest that the R^2 is 0.326, thus, the important independent variables accounted for 32.6% of the dependent variable variance (*Innovators*). R^2 >0.25, indicating practically significant. Also, 14 variables that accounted for 29.7% of the overall dependent variable variance were added into the model. These variables are Relative advantage and compatibility, Informal training, Contributions to healthcare, Perceived technology reliability, Formal training, Fits well, Add to development, Facilitating conditions, High complexity, Contributions to economy, Culture, Unsatisfactory performance, Tested correctly and Satisfy user expectations. Thus, R^2 >25.0% demonstrating that the variables added in this model are practically significant for the intended framework. See Table 9

8.7 Early Adopters Category

It indicates that the R^2 is 0.316, thus, the important independent variables accounted for 31.6% of the dependent variable variance (*Early Adopters*). R^2 >0.25, indicating practically significant. Also, 9 variables that accounted for 29.7% of the overall dependent variable variance were added into the model. These variables are Relative advantage and compatibility, Self-efficacy, Contributions to agriculture, Contributions to politics, Culture, Fits well, Contributions to education and social activities, Satisfy user expectations, and Facilitating conditions. Thus, R^2 >25.0% suggesting that the variables added in this model are practically significant for the suggested framework. See Table 10.

8.8 Early Majority Category

This model indicates that the R^2 is 0.271, thus, the important independent variables accounted for 27.1% of the dependent variable variance (*Early majority*). R^2 >0.25, which is practically significant. Also, 14 variables entered into the model which accounted for 25.7% of the total dependent variable variance. These variables are Self-efficacy, Financial cost, Contributions to education and social activities, Occupy phone memory, Informal training, Tested correctly, A better replacement, Advantageous, Contributions to

Table 4: Model summary of the General use of mobile apps

Model Summary									
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**Change Statistics**				
1	.539a	0,291	0,270	0,967	**R Square Change**	**F Change**	**df1**	**df2**	**Sig. F Change**
					0,291	13,746	37	1241	0,000

Table 5: Model summary of Business use

Model Summary									
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**Change Statistics**				
1	.617a	0,380	0,362	0,79478	**R Square Change**	**F Change**	**df1**	**df2**	**Sig. F Change**
					0,380	20,574	37	1241	0,000

Table 6: Model summary of Individual use

Model Summary									
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**Change Statistics**				
1	.669a	0,447	0,431	0,64418	**R Square Change**	**F Change**	**df1**	**df2**	**Sig. F Change**
					0,447	27,126	37	1241	0,000

Table 7: Model summary for the Late majority and Laggard category

Model Summary									
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**Change Statistics**				
1	.506a	0,256	0,234	0,83069	**R Square Change**	**F Change**	**df1**	**df2**	**Sig. F Change**
					0,256	11,555	37	1241	0,000

Table 8: Model summary for Non-adopters category

Model Summary									
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**Change Statistics**				
1	.559a	0,312	0,292	0,86144	**R Square Change**	**F Change**	**df1**	**df2**	**Sig. F Change**
					0,312	15,234	37	1241	0,000

Table 9: Model summary for Innovators category

Model Summary									
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**Change Statistics**				
1	.571a	0,326	0,306		**R Square Change**	**F Change**	**df1**	**df2**	**Sig. F Change**
				0,73808	0,326	16,203	37	1241	0,000

Table 10: Model summary of Early adopters category

Model Summary									
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**Change Statistics**				
1	.562a	0,316	0,296	0,75119	**R Square Change**	**F Change**	**df1**	**df2**	**Sig. F Change**
					0,316	15,521	37	1241	0,000

Table 11: Model summary of Early majority category

Model Summary									
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**Change Statistics**				
1	.509a	0,271	0,237	0,76663	**R Square Change**	**F Change**	**df1**	**df2**	**Sig. F Change**
					0,259	11,741	37	1241	0,000

politics, Internet, External initiated actions, Culture, Satisfy user expectations and Trialability. Thus, R^2 >25.0% suggesting that the variables added in this model are practically significant for the proposed framework. See Table 11.

9 ANOVA Test

In order to examine the correlation among the profiles of the participants and the dependent variables, the ANOVA test was used. It helps to assess the impact or statistical importance of the participants' age, educational level, employment position and religion on the acceptance and distribution of mobile apps in Africa. The ANOVA results obtained are depicted in the Table 12 with regard to the five selected countries. This suggests that with the exception of innovators and early adopters, age has relevant impact on all the dependent variables. Education level indicated an important effect on Business use and Individual use. Employment status showed an important influence on business use and non-adopter's category while religion has a substantial impact on business use, individual use, non-adopter's and innovators categories concerning the acceptance and use of mobile apps within the African social system.

10 Interpretations and Validation of the Proposed Framework

The RCM was developed and used in this research as the basic principles. The model entails three parts or components, which include:

1. Mobile apps Innovations,
2. Innovation decision process, and
3. Adopting unit.

Each of these part contains a variety of factors that constituted the study variables. Various statistical analyses were conducted to measure the importance of the study variables from the perspectives of the participants. The obtained mean values from the responses of the participants show that the research variables are essential for mobile apps to be adopted and disseminated in Africa. With regard to the proposed framework, several analyses were also, carried out to confirm if the variable fits well in the model.

To examine the relationship between the model variables, multiple linear regression analysis was carried out. For dependent variables, eight factors were considered, including the Use of Mobile apps, Business uses, and Individual uses, Late majority and Laggards, Non-adopters, Innovators, Early adopters, and Early majority. These variables are derived form the adoption and diffusion component of the model. Regarding independent variables, the remaining factors were considered. The association among the variables was tested by regression analysis. After the analysis, the eight dependent variables resulted in the eight models. These models accounted over 25.0% of the variance (R^2 >25.0% <50.0%). With regards to the criteria for significant association namely the p-values, R^2, normality, tolerance and VIF, the output from regression analysis on the test of the relationships between the dependent and independent variables are interpreted below.

Model for the general use of mobile apps: R^2 = 0.273 suggests that the model is practically significant to the general use of mobile apps. This means that the model explained 27.3% of the variance concerning the general use of mobile apps in Africa. The model included a total of eight positive predictors: Add to development, Good innovations, Relative advantage and compatibility, Perceived technology reliability, Contributions to health care, Occupies phone memory, Satisfies user expectations and Social influence. All these variables have a positive impact on Africa's overall use of mobile apps.

The model therefore suggests that the dependent variable (general use of mobile apps) increases as predictive variables increase. this implies that that the adopters' attitude towards adoption and use of mobile apps can be shaped by good mobile apps that can contribute to Africa's development. Competitive benefit and usability increase the opportunity for mobile apps to be adopted and used. Safety assurance and decreased risk of using mobile apps, as well as the conviction of close friends and family increase the general use of mobile apps, even though the use of mobile apps takes up a great portion of phone storage.

Model for Business use: R^2 = 0.370 suggests that the model is practically significant to the use of mobile apps for business purposes. This means that the model explained 37.0% of the variance in the use of mobile apps in Africa for business purposes. In the model, 14 predictors, namely Enables work activities, Relative advantage and compatibility, Culture, Adds to development, Anxiety, Informal training, Internet, Satisfies user expectations, Working manual, Interpersonal communication, Contributions to agriculture, Observability, Social influence and Enables easy communication have a positive relationship with the use of mobile apps for business purposes, while four predictors, Contributions to finance, Advantageous, Contributions to politics, and Contributions to education and social activities have a negative relationship with the use of mobile apps for business purposes.

This means that the use of mobile apps for business purposes requires mobile apps to be good, beneficial and business-friendly in order to improve their meaningful business activities. Enables work activities and easy communications improve the use of mobile apps for business interactions. Organizations, such as banks are sharing the information about their mobile banking apps and other service-related mobile apps, such as FNB eWallet with their clients to improve their services. Increasing these predictors will increase the rate of adoption and diffusion of these mobile apps meant for business purposes.

African countries are developing nations, and developing mobile apps with local content containing social behaviour and norms will attract more adopters, especially rural dwellers and the less educated. Individuals can easily observe mobile apps created with local content, such as local language, and be convinced to use them. M-commerce apps in a local language can easily be adopted and used by people in remote areas to improve their lives in that area. Interestingly, anxiety about the use of mobile apps for business purposes has been positive. It is possible because most mobile app users were nervous when using mobile apps, especially for business transactions. Using the banking apps, as well as M-commerce apps

Table 12: ANOVA on Participants' profiles

Variables	p-values (ANOVA)			
	Age	Education level	Employment status	Religion
Use of mobile apps	0.023	0.598	0.442	0.066
Business use	0.000	0.000	0.000	0.000
Individual use	0.000	0.000	0.227	0.022
Late majority and laggards	0.000	0.746	0.990	0.264
Non-adopters	0.000	0.080	0.004	0.012
Innovators	0.176	0.835	0.777	0.023
Early adopters	0.063	0.607	0.197	0.065
Early majority	0.010	0.110	0.572	0.313

attracts some degree of fear and nervousness, but they facilitate easy, quick and comfortable transactions. Although people are anxious about using such apps, the number of users increases rapidly. Users need to learn how to use mobile apps adequately for business. Increasing the learning methods of how to use mobile apps will improve the adoption and diffusion of mobile apps. Unfortunately, the use of mobile apps for business is not profitable, and it attracts some costs. The more banking apps you use to carry out banking transactions, the more fees you pay. The use of mobile apps for business is also not affected by political, educational and social advancement.

Model for Individual use: $R^2 = 0.438$ suggests that the model is practically significant to the individual use of mobile apps. This shows that the model explained 43.8% of variance for individual use of mobile apps in Africa. The model included a total of 15 predictors, including 13 positive predictors (Adds to development, Self-efficacy, Good innovations, Trialability, Internet, Interpersonal communication, Time-consuming, Contributions to finance, Tested correctly, Observability, Contributions to education and social activities, Fits well, and Individual activities) and two negative predictors (Contributions to agriculture and Contributions to health care). Positive predictors have a positive relationship with the individual use of mobile apps, while negative predictors have a negative relationship with the use of mobile apps for individual purposes.

The model suggests that the attitude of mobile app adopters and their use thereof can be shaped by good mobile apps that can contribute to the development in Africa. The advantages of using mobile apps, including financial, educational and social benefits, affect the adoption and diffusion of mobile apps for individual purposes. The benefits of mobile social networking apps, for example have improved the adoption and use of mobile apps by individuals. Individual use of mobile apps involves the technical knowledge of how to use mobile technology, and the trial of new technology, as well as seeing others using it increases the adoption and diffusion of mobile apps. Users need to learn how to use mobile apps. The availability of learning methods, such as the Internet and interpersonal communication increase the rate of adoption and diffusion of mobile apps for individual use. The model also shows that, due to time and costs involved, there is less interest in using mobile apps to perform agricultural and health activities. In other words, the time and cost involved in using mobile apps reduces mobile app adoption and dissemination in Africa.

Model for Late majority and Laggards category: $R^2 = 0.265$ indicates that the model is practically significant to the Late and Laggards category of mobile app adopters. It implies that 26.5% of the variance can be explained by the adopters of mobile apps in Africa's Late majority and Laggards category model. The model included a total of 12 predictors, including nine positive (Contributions to agriculture, Contributions to politics, Formal training, Used by others, Self-efficacy, Contributions to the economy, Organizations and events, High complexity, Perceived technology reliability, Occupies phone memory, and Informal training) and three negative predictors (A better replacement, Satisfies user expectations and Social influence). Any increase in positive predictors leads to an increase in the Late majority and Laggards category, while any increase in negative predictors leads to a reduction in the category of the Late majority and Laggards adopters in Africa.

The Late majority and Laggards category of adopters is the set of adopters that is sceptical about adopting new mobile apps and they adopt a "wait and see" attitude towards adopting new technology. For this category to adopt and use mobile apps, they need proper information sources, such as organization and events, as well as seeing others using them and knowing the benefits of using mobile apps. In addition, they require training and reliability assurance on the adoption and diffusion of mobile apps. Since they are the local set in the social system, they become more discouraged by their close peers and are more afraid to adopt new mobile apps.

Model for Non-adopters category: $R^2 = 0.301$ indicates that the model is practically significant to the non-adopters' category of mobile apps adopters. This means that the model explained 30.1% of the variance. The model included a total of 15 significant predictors, containing 11 positive (Financial cost, Formal training, High complexity, Use by others, Culture, Facilitating conditions, Lack of support, Distraction of attention, Contributions to the economy, Observability, and Perceived technology reliability) and four negative predictors (Good innovations, Trialability, Satisfies user expectations and Contributions to education and social activities). The 11 predictors have a positive relationship with the category of non-adopters, while the four negative predictors have a negative relationship with the category of mobile app users in Africa who are not adopters.

The model explained that the high complexity, lack of support and the financial cost involved in the adoption and diffusion of mobile apps cause some adopters in the social system of Africa not to adopt and use mobile apps. The cultural beliefs, as well as the extent of support from government and other stakeholders, and dis-

traction from other activities increase the existence of non-adopters in Africa. It is expected that the perception of mobile apps as good innovations will change the attitudes of non-adopters. Providing a trial opportunity, satisfying users' expectations and the benefits of the use of mobile apps, especially in education and social activities, will reduce the number of non-adopters in Africa.

Model for Innovators category: $R^2 = 0.317$ indicates that the model is practically significant to the innovators' category of mobile apps adopters. It implies that the innovator category model explained 31.7% of the variance. The model is related to 14 important variables/predictors, including 12 positive variables (Relative advantage and compatibility, Informal training, Contributions to health care, Perceived technology reliability, Formal training, Fits well, Facilitating conditions, Contributions to the economy, Culture, Unsatisfactory performance, Tested correctly and Satisfies user expectations) and two negative variables (Adds to development and High complexity). The positive variables increase the category of innovators as they increase, while the negative variables reduce the category as they increase.

The model explained that the challenge of competitive advantage, availability of necessary environment from government and other stakeholders, as well as trying to satisfy the consumers' demands results in the increase of the innovators' category of adopters in Africa. The advancement of technology increases users' experience and requirements, thus there is a need for innovators to create mobile apps that will suit the new improvements. For purposes of improving the economy, poor performance of mobile apps and social behaviour, more improved mobile apps are needed. This encourages or raises more innovators in the social system. On the other hand, the higher the development progress and higher complication of mobile apps in Africa, the lesser innovators are required.

Model for Early adopters category: R2 = 0.297 suggests that the model is practically significant to the early adopters' category of mobile app adopters. It means that the model described 29.7% of the variance. The model is connected to nine predictors, including eight positive predictors (Relative advantage and compatibility, Self-efficacy, Contributions to agriculture, Contributions to politics, Culture, Fits well, Contributions to education and social activities and Satisfies user expectations) and one negative predictor (Facilitating conditions). Any rise in positive predictors leads to an increase in the early adopt category, while any increase in negative predictors leads to a decrease in the early adopt category.

The model shows that the ability to use a mobile app with advanced features will be of good interest and increase the potential adopters in the early adopter's category. In addition, the ability to use mobile apps to conduct educational, agricultural and political activities, as well as to fulfill consumer needs, is increasing the category of early adopters. Conversely, the availability of a suitable setting decreases the size of the group of early adopters.

Model for the Early majority category: $R^2 = 0.257$ shows that the model is practically important for mobile apps adopters in the early majority category. This means that 25.7% of the variance was explained by the model. The model is associated with 14 predictors, including 11 positive (Self-efficacy, Financial cost, Contributions to education and social activities, Occupies phone memory, Informal training, Tested correctly, Advantageous, Contributions to politics, Internet, Culture, and Trialability) and three negative predictors (A better replacement, External initiated actions, and Satisfies user expectations). A rise in positive predictors results to an increase to the size of early majority adopter category while an increase in negative predictors reduces the size of the category.

This shows that testing a new mobile app and finding it useful will increase the interest of consumers to embrace and use the mobile app without considering the impact on the phone's memory. The availability of methods of learning, such as the Internet and informal training, increases the user's skill and knowledge. The creation of better mobile apps and the acquisition of information from external sources, such as TV and radio, on the other hand increases consumer demand, thereby decreasing the acceptance of prior mobile apps.

Regarding the personality variables (age, education level, employment status and religion), ANOVA was conducted to examine the association among personality variables and the dependable variables. The findings showed that age has a statistically significant impact on all dependent variables, with the exception of the categories of Innovators and Early Adopters. The level of education showed a major effect on business and individual uses. A significant impact from the employment status was noticed on the business use and the non-adopters categories. Concerning religion, the impact was much on business use, individual use, non-adopters' and innovators' categories. Therefore, age, level of education, employment status and religion have an important impact on the acceptance and diffusion of mobile apps in Africa in various ways.

To explore the influence of gender on the acceptance and use of mobile apps, a t-test study was carried out. No noticeable effect was shown by the result. There are no variances among males and females in the adoption and delivery of mobile apps. Mobile apps are created for everyone regardless of the gender.

Overall, various substantial portions of the variance were explained by the eight models and all R^2 scores were over 25.0 points, (R^2 >25.0% <50.0%). This demonstrates that these models are practically important for the adoption and diffusion of mobile applications in Africa, and all the factors/predictors of the RCM can be applied to the proposed system. Some personality profiles have also been shown to have an effect on the adoption and distribution of mobile apps in Africa, including age, education, employment status and religion. As a consequence, a correlation occurs among them and the dependent variables. The association among the study variables typically shows that they are valuable predictors of mobile apps adoption and diffusion in Africa.

10.1 The Framework

Based on the findings obtained from various statistical studies and recommendations, Figure 2 presents the final proposed framework for the effective adoption and diffusion of mobile apps in Africa.

11 Conclusions and Future Work

11.1 Conclusion

This research aims to create a general framework for mobile apps adoption and diffusion in Africa. A three components model was conceptualized consisting of the research variables. The positivism

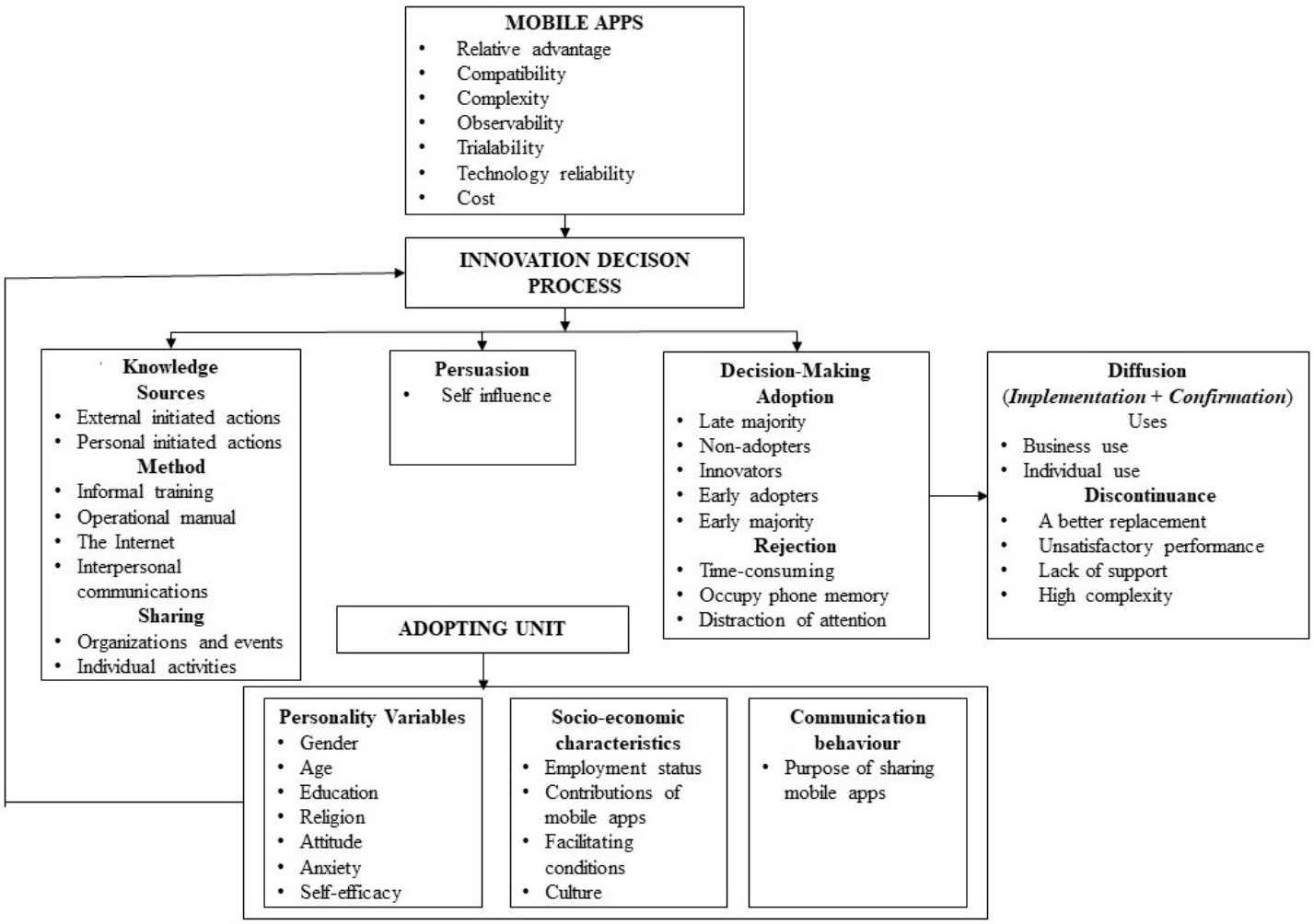


Figure 2: Validated Proposed Framework

research approach was applied. Through a survey, quantitative data were collected from the mobile apps users living in the five distinct nations representing the various regions in Africa. Various statistics analyses were carried out and the results show that the participants perceived these variables to be significant factors for mobile apps adoption and diffusion in Africa. Furthermore, a regression analysis was conducted to measure the relationships among these variables in association with mobile apps adoption and diffusion. The variable in the adoption and diffusion component of the RCM are the dependable variables whereas the remaining variables are the independent variables. The findings indicated that variables (predictors) explained more than 25.0% of the variance (R^2 >25.0% <50.0%) with appropriate level of significance ($p < 0.05$). These results suggest that the research variables are practically significant and they are potential or meaningful predictors of mobile apps adoption and diffusion in Africa. Thus, the variables fits well in the proposed framework. Also, the study revealed that some personality profiles such as age, education level, employment status , and religion have certain degree of influences on mobile apps adoption and diffusion in Africa. Therefore, there is a relationship among the personality variables and the dependent variables. Overall, the relationships existing among the research variables and their significant impact on mobile apps adoption and diffusion suggest that they are potential predictors of mobile apps adoption and diffusion in Africa. Consequently, the study proposed a three components framework for mobile apps adoption and diffusion in Africa.

11.2 Limitations

Africa is a region made up of multilingual nations and a variety of political challenges. Language gaps, cultural disparities and political uncertainty in most nations posed some serious limitations to this research work especially in Kenya, Morocco, and DRC. Also, due insufficient funding, it was not possible to reach all target audiences, particularly some rural dwellers, instead the convenient participants were used.

11.3 Future work

African nations are growing with potential population that are ready to embrace and use mobile technology innovations such as mobile apps. In order to further encourage mobile technology adoption, further research is needed on information awareness, the expansion of mobile services to unconnected areas, and how to bridge the gap

between educated and uneducated users of mobile apps in Africa.

References

[1] J. Hellström, P.-E. Tröften, The innovative use of mobile applications in East Africa, Swedish international development cooperation agency (Sida), 2010. https://www.sida.se/contentassets/b55f10e43c874c9da41203d991a7eaa0/14993.pdf.

[2] GSMA, The mobile economy Sub-Saharan Africa 2015., GSMA Intelligence. Regional report. https://gsmaintelligence. com/research. Accessed on 23/03/2016., 2015.

[3] R. Nickerson, M. Austreich, J. Eng, "Mobile technology and smartphone apps: A Diffusion of innovations analysis," in 20th Americas Conference on Information Systems, AMCIS 2014, 2014, 1–12.

[4] G. Van Zyl, Why Kenya's fixed line market is shrinking., itweb. https: //itweb.africa/content/KjlyrvwdnK57k6am. Accessed on 24/06/2016., 2013.

[5] S. Lee, M. Marcu, S. Lee, "An empirical analysis of fixed and mobile broadband diffusion," Information Economics and Policy, **23**(3-4), 227–233, 2011, doi:10.1016/j.infoecopol.2011.05.001.

[6] A. A. Shaikh, H. Karjaluoto, "Mobile banking adoption: A literature review," Telematics and informatics, **32**(1), 129–142, 2015, doi:10.1016/j.tele.2014.05.003.

[7] S. Murugesan, "Mobile apps in Africa," IT Professional, **15**(5), 8–11, 2013, doi:10.1109/MITP.2013.83.

[8] C. W. Okonkwo, M. Huisman, E. Taylor, "An investigation on the adoption and diffusion of mobile apps in Africa." Thesis submitted in fulfilment of the requirements for the degree Doctor of Philosophy in Computer Science at the North-West University, Potchefstroom campus, South Africa., 2019., 2019.

[9] E. Rogers, Diffusion of Innovations (5th ed)., New York: Free Press., 2003.

[10] I. Lule, T. K. Omwansa, T. M. Waema, "Application of technology acceptance model (TAM) in m-banking adoption in Kenya." International Journal of Computing & ICT Research, **6**(1), 2012.

[11] D. G. Taylor, T. A. Voelker, I. Pentina, "Mobile application adoption by young adults: A social network perspective," 2011. http://digitalcommons.sacredheart.edu/wcob_fac/1.

[12] D. W. Stacks, M. B. Salwen, An integrated approach to communication theory and research, Routledge, 2014.

[13] N. Kiongo, "A framework for mobile health adoption in developing countries: Case study Kenya," A framework for mobile health adoption in developing countries: Case study Kenya, 2014. http://hdl.handle.net/11295/74353.

[14] M. Crabbe, C. Standing, S. Standing, H. Karjaluoto, "An adoption model for mobile banking in Ghana," International Journal of Mobile Communications, **7**(5), 515–543, 2009. DOI: https://doi.org/10.1504/IJMC.2009.024391.

[15] K. J. Bwalya, M. Healy, "Harnessing e-government adoption in the SADC region: a conceptual underpinning," Electronic Journal of E-government, **8**(1), 23, 2010.

[16] C. Z. Qiang, S. C. Kuek, A. Dymond, S. Esselaar, "Mobile applications for agriculture and rural development," 2012. http://hdl.handle.net/10986/21892.

[17] K. S. McNamara, "Mobile applications in agriculture and rural development: Framing the Topic, and Learning from Experience," in World Bank Workshop on Mobile Innovations for Social and Economic Transformation, 2009.

[18] M. O. M. El-Hussein, J. C. Cronje, "Defining mobile learning in the higher education landscape," Journal of Educational Technology & Society, **13**(3), 12–21, 2010. https://www.jstor.org/stable/jeductechsoci.13.3.12.

[19] C. Perera, A. Zaslavsky, P. Christen, D. Georgakopoulos, "Context aware computing for the internet of things: A survey," IEEE communications surveys & tutorials, **16**(1), 414–454, 2013, doi:10.1109/SURV.2013.042313.00197.

[20] E. Oriwoh, P. Sant, G. Epiphaniou, "Guidelines for internet of things deployment approaches–the thing commandments," Procedia Computer Science, **21**, 122–131, 2013, doi:10.1016/j.procs.2013.09.018.

[21] A. Giessmann, K. Stanoevska-Slabeva, B. De Visser, "Mobile enterprise applications–current state and future directions," in 2012 45th Hawaii International Conference on System Sciences, 1363–1372, IEEE, 2012. https://doi.org/10.1109/HICSS.2012.435.

[22] H. T. Dinh, C. Lee, D. Niyato, P. Wang, "A survey of mobile cloud computing: architecture, applications, and approaches," Wireless communications and mobile computing, **13**(18), 1587–1611, 2013, doi:10.1002/wcm.1203.

[23] J. Donner, C. A. Tellez, "Mobile banking and economic development: Linking adoption, impact, and use," Asian journal of communication, **18**(4), 318–332, 2008, doi:10.1080/01292980802344190.

[24] C. W. Okonkwo, M. Huisman, E. Taylor, "The adoption of m-commerce applications: rural dwellers perspectives," in 12th, IADIS, International conference. Information systems., pages:99–106, 2019. DOI: 10.33965/is2019_201905L013.

[25] C. W. Okonkwo, M. Huisman, E. Taylor, "Socio-Economic Contributions of Mobile Applications in Africa: Impact of Local Mobile Applications," in 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC), 1–8, IEEE, 2019. DOI: https://doi.org/10.1109/IMITEC45504.2019.9015912.

[26] G. Intelligence, "The Mobile Economy. 2020," 2020.

[27] G. Association, et al., "The mobile economy 2018," London: GSM Association, 2018.

[28] L. Hosman, E. Fife, "The use of mobile phones for development in Africa: Top-down-meets-bottom-up partnering," The Journal of Community Informatics, **8**(3), 2012.

[29] G. Association, et al., "The Mobile Economy Sub-Saharan Africa," London: GSM Association, 2017.

[30] B. Simon, "A History of Mobile Application Development," https://manifesto.co.uk/history-mobile-application-development/. Accesses on 23/06/2016. 2014., 2017.

[31] I. T. Union, World Telecommunication/ICT Development Report, International Telecommunication Union, 2006.

[32] H. K. Chavula, "Telecommunications development and economic growth in Africa," Information Technology for Development, **19**(1), 5–23, 2013. https://doi.org/10.1080/02681102.2012.694794.

[33] M. E. Batuo, "The role of telecommunications infrastructure in the regional economic growth of Africa," The Journal of Developing Areas, **49**(1), 313–330, 2015. https://www.jstor.org/stable/24241299.

[34] S. Etzo, G. Collender, "The mobile phone 'revolution'in Africa: rhetoric or reality?" African affairs, **109**(437), 659–668, 2010. DOI: https://doi.org/10.1093/afraf/adq045.

[35] S. H. Lee, J. Levendis, L. Gutierrez, "Telecommunications and economic growth: an empirical analysis of sub-Saharan Africa," Applied Economics, **44**(4), 461–469, 2012. DOI: https://doi.org/10.1080/00036846.2010.508730.

[36] J. Zhang, X.-J. Liang, "Business ecosystem strategies of mobile network operators in the 3G era: The case of China Mobile," Telecommunications policy, **35**(2), 156–171, 2011. DOI: https://doi.org/10.1016/j.telpol.2010.12.009.

[37] R. Lindgren, O. Eriksson, K. Lyytinen, "Managing identity tensions during mobile ecosystem evolution," Journal of Information Technology, **30**(3), 229–244, 2015. DOI: https://doi.org/10.1057/jit.2015.8.

[38] W. C. Okonkwo, M. Huisman, "The use of system development methodologies in the development of mobile applications: Are they worthy of use?" in 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), 2, 278–283, IEEE, 2018. DOI: https://doi.org/10.1109/COMPSAC.2018.10243.

[39] C. W. Okonkwo, M. Huisman, "The predicting success factors of mobile applications development: organizational perspectives," in 12th, IADIS, International conference. Information systems., 11–18, 2019. DOI: 10.33965/is2019_201905L002.

[40] W.-H. Wu, Y.-C. J. Wu, C.-Y. Chen, H.-Y. Kao, C.-H. Lin, S.-H. Huang, "Review of trends from mobile learning studies: A meta-analysis," Computers & Education, **59**(2), 817–827, 2012. DOI: https://doi.org/10.1016/j.compedu.2012.03.016.

[41] S. Okazaki, F. Mendez, "Exploring convenience in mobile commerce: Moderating effects of gender," Computers in Human Behavior, **29**(3), 1234–1242, 2013. DOI: https://doi.org/10.1016/j.chb.2012.10.019.

[42] C. Free, G. Phillips, L. Watson, L. Galli, L. Felix, P. Edwards, V. Pa-tel, A. Haines, "The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis," PLoS medicine, **10**(1), e1001363, 2013. DOI: https://doi.org/10.1371/journal.pmed.1001363.

[43] A. Gichamba, I. A. Lukandu, "A model for designing M-agriculture applications for dairy farming," The African Journal of Information Systems, **4**(4), 1, 2012. https://digitalcommons.kennesaw.edu/ajis/vol4/iss4/1.

[44] P. Van der Boor, P. Oliveira, F. Veloso, "Users as innovators in develop-ing countries: The global sources of innovation and diffusion in mobile banking services," Research Policy, **43**(9), 1594–1607, 2014. DOI: https://doi.org/10.1016/j.respol.2014.05.003.

[45] J. Hellström, "Mobile Governance: Applications, Challenges and Scaling-up in Poblet, M.(Ed.),"Mobile Technologies for Conflict Management: Online Dispute Resolution, Governance, Participation"," 2012. DOI: https://doi.org/10.1007/978-94-007-1384-0_12.

[46] D. Baelden, L. Van Audenhove, "Participative ICT4D and living lab research: The case study of a mobile social media application in a rural Tanzanian University setting," Telematics and Informatics, **32**(4), 842–852, 2015. DOI: https://doi.org/10.1016/j.tele.2015.04.012.

[47] J. C. Aker, I. M. Mbiti, "Mobile phones and economic development in Africa," Journal of economic Perspectives, **24**(3), 207–32, 2010. DOI: 10.1257/jep.24.3.207.

[48] S. Tunmibi, A. Aregbesola, E. Asani, "Factors influencing the adoption of smart phones by university students–A cross-border approach," African Journal of Computing & ICT, **8**(1), 1–16, 2015.

[49] B. M. Achieng, B. K. Ingari, "Factors influencing the adoption of mobile banking in Kenya's commercial banks: A case of Kenya Commercial Bank (KCB) Kilindini branch," International Journal of Scientific and Research Publications, **5**(10), 1–14, 2015.

[50] R. A. Oluoch, Factor affecting adoption of mobile banking technology in Kenya: A case of bank customers within Nakuru Municipality, Ph.D. thesis, 2012.

[51] J. Mtebe, R. Raisamo, "Investigating students' behavioural intention to adopt and use mobile learning in higher education in East Africa," International Journal of Education and Development using ICT, **10**(3), 2014. https://www.learntechlib.org/p/148476/.

[52] W. P. Mtega, R. Bernard, A. C. Msungu, R. Sanare, "Using mobile phones for teaching and learning purposes in higher learning institutions: The case of Sokoine University of Agriculture in Tanzania," 2012. https://repository.ubuntunet.net/handle/10.20374/130.

[53] F. Iddris, "Barriers to adoption of mobile banking: Evidence from Ghana," International Journal of Academic Research in Business and Social Sciences, **3**(7), 356–370, 2013. DOI: https://doi.org/10.6007/IJARBSS/v3-i7/59.

[54] F. O. Bankole, O. O. Bankole, I. Brown, "Mobile banking adoption in Nigeria," The Electronic Journal of Information Systems in Developing Countries, **47**(1), 1–23, 2011. DOI: https://doi.org/10.1002/j.1681-4835.2011.tb00330.x.

[55] A. Osanloo, C. Grant, "Understanding, selecting, and integrating a theoretical framework in dissertation research: Creating the blueprint for your "house"," Administrative issues journal: connecting education, practice, and research, **4**(2), 7, 2016. DOI: 10.5929/2014.4.2.9.

[56] S. Fulton, B. Krainovich-Miller, C. Cameron, "Finding and appraising the literature," Nursing Research in Canada-E-Book: Methods, Critical Appraisal, and Utilization, 90, 2014.

[57] S. Chiemeke, A. Evwiekpaefe, "A conceptual framework of a modified unified theory of acceptance and use of technology (UTAUT) Model with Nigerian factors in E-commerce adoption," Educational Research, **2**(12), 1719–1726, 2011. http://www.interesjournals.org/ER.

[58] G. C. Moore, I. Benbasat, "Development of an instrument to measure the perceptions of adopting an information technology innovation," Information systems research, **2**(3), 192–222, 1991. DOI: https://doi.org/10.1287/isre.2.3.192.

[59] S. Ellis, H. Steyn, "Practical significance (effect sizes) versus or in combination with statistical significance (p-values): research note," Management dynamics: journal of the southern african institute for management scientists, **12**(4), 51–53, 2003.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1593-1597 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Features Preference using Conjoint Analysis Method for E-marketplace Social Care System

Angelina Ervina Jeanette Egeten[*,1], Harjanto Prabowo[2], Ford Lumban Gaol[3], Meyliana[4]

[1]*Information Systems Department, Doctor of Computer Science, Bina Nusantara University, Jakarta, 11530, Indonesia*

[2]*Management Department, BINUS Business School Undergraduate Program, Bina Nusantara University, Jakarta, 11530, Indonesia*

[3]*Computer Science Department, BINUS Graduate Program, Bina Nusantara University, Jakarta, 11530, Indonesia*

[4]*Information Systems Department, School of Information System, Bina Nusantara University, Jakarta, 11530, Indonesia*

ARTICLE INFO

Article history:
Received: 16 September, 2020
Accepted: 10 December, 2020
Online: 25 December, 2020

Keywords:
Features Priority
E-marketplace Social Care
Intention to Use
Customer Satisfaction
Conjoint Analysis

ABSTRACT

With the features of the application system make it easier for users to transact, make the user feel comfortable in using the application system. The problem is that the intensity of application usage decreases judging by the application rating given by the user each time they perform a transaction, this is because the company initially make the features based on their urge without further investigating on user satisfaction level on similar social care e-marketplace, this causes users feel dissatisfied in using the application system on the same business system, and this is the reason why it is important to decide the feature preferences in the application development to meet user needs. The purposes of this study are to find out the feature preferences that would be used in the social care e-marketplace and to find out the actual result of user opinion related to the feature preferences found. The method used in this study are predictive accuracy conjoint analysis that can be measured by Pearson correlation to find out the correlation between respondent opinion and actual opinion. Respondents in this study are 521 users of the social care e-market, who filled out a distributed questionnaire to determine the feature priority in each functional category. The result is that six feature preferences have positive value namely Photo, Service Assistance Chat, Special Request, Worker Performance Rating, Application Rating, Tips & Articles. As well as Pearson's R Value is 0.558 with a significance rate of 0.002 and Kendall's Tau correlation value is 0.320 with significance level of 0.012. This indicates a correlation between the respondents opinion pattern (estimates part-worth) and the actual opinion (actual preferences) that answer the problem by determining the priority feature preferences from the result of user satisfaction level can help increase the intensity of using the e-marketplace social care.

1. Introduction

E-marketplace Social Care is an online business using the services of domestic workers who help users to meet their needs. One of user needs is the support of features which make it easier for user to transact [1]. The problem or the gap in this research is that the intensity of application usage decreases judging by the application rating given by the user each time they perform a transaction activity by the user [2], this is because of the company initially make the features based on their urge without further investigating on user satisfaction level and user experience [3] on similar social care e-marketplace, this causes users feel dissatisfied in using the company application system because they are considered not focus on improving the feature functionality in the same business, and this led a reason why it is important to decide the feature preferences in the application development of e-marketplace social care to meet their needs and help companies to increase user satisfaction on used feature functionality. Previous research supports the importance of assessing user satisfaction rating for information technology artifacts in this case is a feature of the whole system as an incentive for consumers to use the system intensely or loyal [2]. Other studies also support the assessment of user satisfaction with user experience in using the features of application system, where by the result of feature assessment provided by user can help companies to improve feature functionality on the company's application system [3]. The purposes of this study are to find out the feature preferences that will be used in the e-marketplace social care and to find out the

[*]Corresponding Author Angelina Egeten, Email: angelina.egeten@binus.ac.id

actual result of user opinion related to the feature preferences found. Thus, this research will answer the research questions about:

- Research Question 1 : what are the feature preferences in e-marketplace social care?
- Research Question 2 : are the result of the respondent opinion pattern much different from the result of the respondent actual opinion?

2. Theoretical Background

2.1. E-marketplace Social Care

E-marketplace is a place where sellers and buyers are brought together in the market demand and supply with various types of products and services that available and provide added value to companies [4]. E-marketplace acts as a third-party that brings together sellers and buyers to get the desired choices and at the same time reduces transaction cost in terms of products and services used [5].

E-marketplace social care provides a household services or health social care which aims to help users to meet with workers in helping meet the needs of the users themselves [6], where these workers are under the distribution agencies who cooperate with e-marketplace companies.

2.2. Customer Satisfaction

In general, customer satisfaction depends on the services used by users, as well as the purchased products. If users are satisfied with the products and services, user satisfaction is fulfilled, or the service exceeds user expectations, the user is very happy and satisfied. Likewise, if the service does not match the user's expectations, the user is not satisfied [7].

Customer Satisfaction is not only based on the user's assessment of the products and services used, but also on the user experience in using service application features and the system's ability to convey complete service information [3] and has an innovative feature [2].

2.3. Intention to Use

It is very important to maintain the intensity of reuse the services used in meeting user needs, because when users need the service they will always predict or surely use the service when they need assistance in meeting their needs [1]. Not only in terms of the offered services, but the system support of the services is also the main reason for users to use the services because it can be used anywhere, anytime and are supported with innovative features [2].

3. Research Method

This study uses a conjoint analysis method to determine the features preference used by users in the e-marketplace social care system. This method was chosen to evaluate the user's features preference of the services used [8] [9] and be able to produce accurate answer predictions that did not much different between the respondent's opinion pattern and the respondent's actual opinion. Conjoint analysis can be expressed in the following model as a part of the multivariate dependence method, namely [10]:

$$Y_1 = X_1 + X_2 + X_3 + ... + X_n$$

where:

X = Independence variables are often referred as factors and a non-metric data part of the factor called levels.

Y_1 = The overall opinion of respondents on the various factor and levels of a product or service.

The conjoint process is carried out to categorize the e-marketplace social care features against the functional attributes of the e-marketplace social care, which consist of Identity, Conversation, Information and Service, Reputation, dan Sharing. The process produces preference features for each functionality, so that each functionality has a feature that represent in need.

The population in this study was users of e-marketplace social care services in Indonesia where the data is collected from January-February 2020. 521 respondents are used as sample where the questionnaire is distributed using google docs and paper directly filled by users who already had experience in using similar e-marketplace social care applications. Users are asked to rate the 25 preferences that have been set out in the questionnaire based on the results of random features of the orthoplan using SPSS version 23. The questionnaire result that has been filled in by the user is collected and tabulated, then the data tested for validity and conjoint analysis to answer the research questions in accordance with the purpose of this study.

4. Result and Discussion

The following are the steps in conjoint analysis, namely:

4.1. E-marketplace Social Care Feature Preferences Based on Similar Application

Based on the feature functionality equation of 4 other similar e-marketplace social care applications that have been selected, they are mapped into the social care application functionality, which will be used for conjoint analysis test to identify the preferred features of the e-marketplace social care system.

Table 1: Feature Preferences of E-marketplace Social Care

Feature	**Functionality Social *Care* Application**	***App A***	***App B***	***App C***	***App D***
Photo	Identity	✓	-	-	-
Sign Up (*Sign Up*)	Identity	✓	✓	✓	✓
Account Profile	Identity	✓	✓	✓	✓
Forgot Password	Identity	✓	✓	✓	✓
Services Assistance Chat	Conversation	✓	✓	✓	✓
Testimonial	Conversation	✓	✓	✓	✓
Like (*Like)*	Conversation	✓	-	-	-
Special Request	Service	✓	-	-	-
List of Services & Detail of Services	Service	✓	✓	✓	✓
Order and Order History	Service	✓	✓	✓	✓

Feature	Functionality Social *Care* Application	*App A*	*App B*	*App C*	*App D*
Search (*Search*)	Service	✓	✓	-	-
Worker Performance Rating	Reputation	✓	-	-	-
Application Rating	Reputation	-	✓	✓	✓
Best Dealer	Reputation	✓			
Tips & Articles	*Sharing*	✓	✓	✓	✓
Social Media	*Sharing*	-	-	-	-

4.2. Mapping of Features to E-marketplace Social Care Functionality

The following table is features mapping based on category or functionality attributes in e-marketplace social care.

Table 2: Mapping of Features to E-marketplace Functionality

Identity	Conversation	Information & Service	Reputation	Sharing
Photo	Chat Assistance Services	Special Request	Worker Performance Rating	Tips & Articles
Sign Up	Testimonial	List of Services & Detail of Services	Application Rating	*Social Media*
Account Profile	Like	Order and Order History	Best Dealer	
Forgot Password		Search		
4	3	4	3	2

4.3. Orthoplan

In obtaining the minimum number of combinations stimuli that orthogonal respondent must evaluate, the orthoplan process or orthogonal array is used. Orthogonal is the absence of correlation between levels on an attribute [11]. Based on the orthoplan formula, the result of the calculation is:
N_Identity * N_Conversation * N_Information&Service * N_Reputation * N_Sharing

Orthoplan = 4 * 3 * 4 * 3 * 2 = 288 combination. The result of each component multiplication is 288 combination. Therefore, a randomized orthoplan was conduct in this study because it was not possible to test 288 combination.

The following is a table of random orthoplan results generated by the system using SPSS version 23 used in the questionnaire.

Table 3: Result of Random Feature Based on Orthoplan

Identity	Conversation	Information & Service	Reputation	Sharing
Photo	Chat Assistance Services	Special Request	Application Rating	Social Media
Forgot Password	Chat Assistance Services	Order and Order History	Application Rating	Tips & Articles
Photo	Chat Assistance Services	Special Request	Worker Performance Rating	Social Media
Photo	Like	Special Request	Application Rating	Social Media
Account Profile	Chat Assistance Services	List of Services & Detail of Services	Worker Performance Rating	Tips & Articles
Photo	Testimonial	Order and Order History	Worker Performance Rating	Tips & Articles
Sign Up	Chat Assistance Services	List of Services & Detail of Services	Application Rating	Tips & Articles
Account Profile	Testimonial	Special Request	Worker Performance Rating	Tips & Articles
Photo	Like	List of Services & Detail of Services	Best Dealer	Tips & Articles
Account Profile	Chat Assistance Services	Order and Order History	Best Dealer	Social Media
Forgot Password	Chat Assistance Services	Search	Worker Performance Rating	Tips & Articles
Photo	Chat Assistance Services	Search	Best Dealer	Tips & Articles
Sign Up	Like	Order and Order History	Worker Performance Rating	Social Media
Account Profile	Like	Search	Application Rating	Tips & Articles
Forgot Password	Testimonial	List of Services & Detail of Services	Application Rating	Social Media
Sign Up	Testimonial	Search	Worker Performance Rating	Social Media
Sign Up	Testimonial	Special Request	Best Dealer	Tips & Articles
Photo	Testimonial	List of Services & Detail of Services	Worker Performance Rating	Tips & Articles
Photo	Chat Assistance Services	Special Request	Worker Performance Rating	Tips & Articles
Forgot Password	Like	Special Request	Worker Performance Rating	Tips & Articles
Sign Up	Services Assistance Chat	Special Request	Application Rating	Tips & Articles

Identity	Conversation	Information & Service	Reputation	Sharing
Account Profile	Testimonial	Special Request	Application Rating	Tips & Articles
Photo	Testimonial	Order and Order History	Application Rating	Tips & Articles
Forgot Password	Testimonial	Special Request	Best Dealer	Social Media
Photo	Testimonial	Search	Application Rating	Social Media

Based on the random orthoplan result table above, the user must rate 1-25 of each row in the table, where the user can only fill in 1 rating number on each row and this number appear once on each line.

4.4. E-marketplace Social Care Feature Preferences

The following is the result of e-marketplace social care feature preferences based on conjoint analysis using SPSS version 23:

Table 4: The Result of E-marketplace Social Care Feature Preferences

Functionality	Features	Importance Value	Utility Estimate
Identity	Photo	21.996	**0.355**
	Sign Up		-0.013
	Account Profile		-0.080
	Forgot Password		-0.262
Conversation	Service Assistance Chat	25.026	**0.146**
	Testimonial		-0.058
	Like		-0.089
Information & Service	Special Request	24.402	**0.302**
	List of Services & Detail of Services		-0.068
	Order & Order History		-0.028
	Search		-0.205
Reputation	Worker Performance Rating	16.351	**0.027**
	Application Rating		**0.303**
	Best Dealer		-0.330
Sharing	Tips & Articles	12.226	**0.011**
	Social Media		-0.011
(Constant)			12.906
Correlations[a]		**Value**	**Sig.**
	Pearson's R	.558	.002
	Kendall's tau	.320	.012

Based on the result of conjoint analysis of 521 respondents on the attributes of social care functionality in the table above, the conclusions are as follows:

1. In the Identity section, based on the result of statistical test can be summed that the respondent prefers the form of feature photo. It can be seen from the utility value which shows a positive result with 0.355 for the photo feature. Meanwhile the Sign Up, Account Profile and Forgot Password have a negative utility value. It can be seen from the utility value which shows a negative result with -0.013 for Sign Up, -0.089 for Account Profile feature, -0.262 for Forgot Password feature.
2. In the Conversation section, based on the result of statistical test can be summed that the respondent prefers the form of Services Assistance Chat feature. It can be seen from the utility value which shows a positive result with 0.146 for the Services Assistance Chat feature. Meanwhile the Testimonial and Like features have a negative utility value. It can be seen from the utility value which shows a negative result with -0.058 for testimonial feature and -0.089 for Like feature.
3. In the Information and Service section, based on the result of statistical test can be summed that the respondent prefers the form of Special Request feature. It can be seen from the utility value which shows a positive result with 0.302 for the Special Request feature. Meanwhile the List of Service & Service Detail, Order & Order History and Search feature have a negative utility value. It can be seen from the utility value which shows a negative result with -0.068 for List of Service & Service Detail feature, -0.028 Order & Order History feature and -0.205 for Search feature.
4. In the Reputation section, based on the result of statistical test can be summed that the respondent prefers the form of Worker Performance Rating and Application Rating feature. It can be seen from the utility value which shows a positive result with 0.027 for the Worker Performance Rating feature and 0.303 for Application Rating feature. Meanwhile the Best Dealer feature has a negative utility value -0.330.
5. In the Sharing section, based on the result of statistical test can be summed that the respondent prefers the form of List of Tips & Articles feature. It can be seen from the utility value which shows a positive result with 0.011 for the List of Tips & Articles feature. Meanwhile the social media feature has a negative utility value -0.011.

4.5. Conjoint Analysis Hypothesis Test on Media Social Functionality

By using the predictive accuracy conjoint test can be measured using Pearson's R Correlation. In this study, the Pearson's R value was 0.558 with a significance level 0.002 and Kendall's Tau correlation showed a value 0.320 with a significance level 0.012. This indicates a correlation between the respondent opinion patterns (estimates part-worth) and actual preferences. For each respondent preference towards factors on social care functionality which is divided into five parts, namely Identity, Conversation, Information and Services, Reputation, Sharing.

H_0 = Estimated opinion patterns of respondents differ greatly
H_1 = Estimated opinion patterns of respondents do not differ that much

Based on the result, it can be concluded that H_0 is rejected and H_1 is accepted, because:

- The significance level of Pearson's R = 0.002 < 0.05
- The significance level of Kendall's Tau = 0.012 < 0.05

The result of conjoint analysis shows that the estimates of respondent opinion pattern (*estimates part-worth)* do not much differ from the respondent actual opinion (*actual preferences*) on the stimuli process.

5. Conclusion

The 6 features that have a positive value from the results of conjoint analysis, namely Photo, Chat Assistance Services, Special Requests, Worker Performance Ratings, Application Ratings, and Tips & Articles, are 6 features that are the main feature preferences in the social e-marketplace system. care and at the same time these 6 features answer the first research question.

As well as to answer the second research question, based on the results of the conjoint analysis can prove that the approximate opinion patterns of respondents are not much different from the actual opinion of the respondent which can be seen from the feature rating given by users as a result of user satisfaction in using the features, and it can answer the problem with determining feature preferences can help increase the intensity of the use of the e-marketplace social care system.

6. Limitation and Future Research

This research is limited to the feature preferences of the e-marketplace social care application system which users assess as a feature rating on the random orthoplan results which later on these features will be maintained and used in the other e-marketplace social care system development.

Further research activity needs to be reassessed in terms of the service capability in meeting user needs, from user satisfaction in using this type of service on the e-marketplace social care so that companies can improve the quality of services offered.

References

[1] L. Åberg, T. Pham, J. Tomsfelt, " A Study about E-loyalty using the Bagozzi framework," 2013.

[2] G.L. Polites, C.K. Williams, E. Karahanna, L. Seligman, “A Theoretical Framework for Consumer E-Satisfaction and Site Stickiness: An Evaluation in the Context of Online Hotel Reservations,” Journal of Organizational Computing and Electronic Commerce, 2012, doi:10.1080/10919392.2012.642242.

[3] S.H. Hsu, W.H. Chen, J.T. Hsueh, “Application of customer satisfaction study to derive customer knowledge,” Total Quality Management and Business Excellence, **17**(4), 439–454, 2006, doi:10.1080/14783360500528197.

[4] E. Christiaanse, M.L. Markus, “Participation in collaboration electronic marketplaces,” in Proceedings of the 36th Annual Hawaii International Conference on System Sciences, HICSS-2003, 2003, doi:10.1109/HICSS.2003.1174383.

[5] S. Wang, N.P. Archer, “Electronic marketplace definition and classification: Literature review and clarifications,” Enterprise Information Systems, 2007, doi:10.1080/17517570601088380.

[6] M.M. Cruz-Cunha, I. Miranda, N. Lopes, R. Simoes, “An e-marketplace of healthcare and social care services: The perceived interest,” Learning Organization, 2013, doi:10.1108/TLO-10-2013-0055.

[7] G. Armstrong, P. Kotler, M. Harker, R. Brennan, “Marketing: An introduction,” 2010.

[8] P. Silayoi, M. Speece, “The importance of packaging attributes: A conjoint analysis approach,” European Journal of Marketing, 2007, doi:10.1108/03090560710821279.

[9] J.F. Hair, R.E. Anderson, R.L. Tatham, W.C. Black, Multivariate Data Analysis, Multivariate Data Analysis, 2018.

[10] H.I. Ghozali, Aplikasi Analisis Multivariate dengan Program IBM SPSS 19, 5th ed., Undip, Semarang, 2005.

[11] R.L. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, Multivariate Data Analysis, 2009.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1598-1608 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

High-Performance Computing: A Cost Effective and Energy Efficient Approach

Safae Bourhnane[1,2,*], Mohamed Riduan Abid[2], Khalid Zine-Dine[3], Najib Elkamoun[1], Driss Benhaddou[4]

[1]*Chouaib Doukkali University, El Jadida, 24030, Morocco*

[2]*School of Science and Engineering, Al Akhawayn University, Ifrane, 53000, Morocco*

[3]*Mohammed V University in Rabat, Faculty of Sciences, Rabat, 10170, Morocco*

[4]*School of Engineering and Technology, University of Houston, Texas, TX 77004, USA*

ARTICLE INFO

Article history:
Received: 05 November, 2020
Accepted: 20 December, 2020
Online: 25 December, 2020

Keywords:
Big data
HPC
Energy efficiency
Green computing
Amdahl's law

ABSTRACT

The world is witnessing unprecedented advancements in ICT (Information & Communication Technology) related fields. These advancements are further boosted with the emergence of big data. It goes without saying that big data requires two major operations: storage and processing. The latter is usually provided through High-Performance Computing (HPC) which is delivered through two main venues: supercomputers or clustering. The second venue has been widely opted for as a cost-effective alternative when compared to supercomputers. However, with the widespread increase in deploying ICT-based applications, the parallel increase in energy consumption has become a real issue. Thus, researchers have been exploring approaches to conceive big data analytics platforms that are both cost-effective and energy-efficient. In this paper, we present a cost-effective and energy-efficient HPC clustering that is based on Raspberry PIs. Our approach leverages the concept of Green Computing. We evaluated our cluster performance and its energy consumption and compared it to a commodity server. We leveraged on the Amdahl's law to set the maximum speedup of the proposed approach. Our approach can be easily deployed for usage in different ICT-based applications that consider energy efficiency as a priority.

1. Introduction

Nowadays, ICT is interweaving into the fabrics of everyday life and touches all aspects of our lives: smart homes, smart cities, smart grids, smart agriculture, eHealth, autonomous vehicles, etc. All these aspects been further enhanced by the emergence of big data.

Indeed, big data is now becoming wide data. By looking at the amount of data generated each day from different sources (e.g. social networks, Internet of Things, business transactions, etc.), we can infer that these data will eventually need large and advanced platforms that can accommodate for their constantly increasing amount.

It is very common that Big Data faces two major challenges: storage and processing. The storage of the big data can be challenging as it needs to be correctly performed in order to allow for an appropriate and efficient processing later on.

As the data becomes more massive and the applications more demanding, the processing becomes very heavy. This implies the use of High-Performance Computing (HPC) as the ultimate solution for the increasing demand on the computing power. The term refers to combining processing powers to deliver a higher performance. Throughout the years, HPC has been provided via either supercomputers, or cluster of commodity computers. The first venue is no longer opted for mainly because of its high cost and not-so-easy maintenance which leaves the clustering as the perfect alternative.

Cluster Computing involves combining similar type of computing machines that work, transparently, like a single computing unit. These machines are connected using dedicated network protocols and usually Local-area networks (LAN). The aggregation of computing powers allows for a better efficiency to

*Corresponding Author: Safae Bourhnane, Al Akhawayn University in Ifrane, +212610880040, Email: s.bourhnane@aui.ma

solve complex operations using faster processing speed than a single machine.

Cluster Computing can be divided into three main categories: load balancing cluster, high availability cluster, and high-performance cluster. The latter is what we are selecting for the work presented in this paper. Clusters of computers represent the environment for different distributed platforms to store and process the big data. Hadoop is the most widely used case in point and it can be utilized for a wide range of applications thanks to the number of benefits it presents.

Hadoop is a famous open-source framework that is conceived for storing and processing data. It serves to run applications on clusters of computers which help it make use of the great processing power provided by the cluster. Hadoop has two main components: MapReduce and HDFS. MapReduce is the programming model brought by the framework while HDFS refers to Hadoop Distributed File System and hence represents the underlying file system.

Testing the cluster with a specific number of nodes can give an idea about the performance as the number of nodes increases. This concept is called "Speedup" and can be given using the renowned Amdahl's Law.

Amdahl's law is a mathematical method that allows for having an idea about the maximum improvement that can be achieved through improving a particular part of the system. In our case, we are making use of Amdahl's law to theoretically find the maximum speed up with multiple processors. Also, it will allow us to infer whether the Raspberry Pi cluster will reach the performance of a single commodity hardware and determine the number of nodes needed in the positive case.

Since we are concerned by the energy efficiency and the cost-effectiveness of the entire system, we are opting for an approach that respects both features. Our approach is based on Raspberry Pis as an alternative to commodity computers and that is because it is known for its low cost and low energy consumption.

Currently, the world is trying to implement a solution that copes with the increasing demand on processing power due to the expanding amount of the data produced, without consuming a large amount of energy. Eventually, researchers have been trying to implement HPC clusters based on low-cost and low-energy hardware. There is a significant amount of work that has been carried in this direction. Most of it used Raspberry Pi as a basis of the cluster and tried to prove that the hardware supports the installation of different big data analytics platforms, e.g. Hadoop. However, this does not give an idea about the real performance of the Raspberry Pi cluster.

For the Raspberry Pi cluster to pass the performance test, it needs to be compared to a performant machine/server. This comparison should be accomplished using the same big data analytics platform and by running the same jobs. This does not only imply the use of the same dataset, but also the same dataset size.

In this paper, we are testing our Raspberry Pi approach and comparing it to a commodity server through benchmarking both setups against the Terasort which is one of the mostly used Hadoop benchmarks. In addition to keeping track of the performance, we are measuring the energy consumed by both setups as energy efficiency is one of the main pillars of our study.

Moreover, we are looking at the speedup of the cluster in order to find the maximum value it can reach. This will allow us to have an idea about the size of the cluster corresponding to a higher performance. For this, we are making use of the Amdahl's Law.

The rest of the paper is organized as follows: Section II presents the literature review; we present the background in III and our proposed approach in section IV. Section V contains the results related to the single-node cluster while section VI presents the results of Raspberry Pi multi-node cluster. We compare both setups and present a discussion in section VII. Finally, we present the conclusion and future work in section VIII.

2. Literature Review

During the last few years, implementing Green Computing using Raspberry Pi has been trending among practitioners in the domain.

Raspberry Pi clusters have been the center of different research works that tackled it as a cost-effective and energy-efficient solution to deliver Green Computing. In this section, we are shedding the light on some of the work that has been previously carried out in that sense.

In [1] the authors implemented a Raspberry Pi cluster that consists of 300 nodes. Through their paper, they walked us through the steps of deployment and set up of the hardware and software plus the maintenance and the monitoring of the overall system. Moreover, the work presents some of the limitations and the challenges that would block the deployment of the cluster. These include the overall performance of the card that is considered relatively low due to the design of the flash memory, in addition to the low processor speed.

The next work described another low-cost cluster named Iridis-pi and that was conceived for the sake of demonstration [2]. The cluster contains 64 Raspberry Pis Model B with 700MHz processor and 256MB of RAM. The authors chose to host the cluster on a Lego chassis. The nodes are connected together using Ethernet cables. The entire system has an overall RAM of 16GB and 1TB of storage. In order to assess the numerical computer power, this research used the LINPACK benchmark for the single-node performance, and the High-Performance LINPACK benchmark in order to measure the overall throughput of the cluster. The results revealed a computational performance peak of around 65000kflops. More results showed that the cluster delivers a good scalability with large problems, and a less significant one with small problems where the network overhead becomes more noticeable.

Glasgow Raspberry Pi cluster [3] depicted the use of Raspberry Pi in data processing. The cluster consists of 56 Model B Raspberry Pi that are supposed to emulate the entire stack of the cloud. The devices are interconnected in a tree topology. Each node uses a 16GB SD card that hosts up to three co-located virtualized hosts provided via Linux containers. According to this research, the deployment of this cluster is still not mature. The authors are still investigating an easier and more secure approach

to manage the virtual resources. Also, their future work includes the implementation of live migration through the PiCloud.

The work in [4] introduced a study of the feasibility of implementing Raspberry Pi based data centers for Big Data processing. The paper examined further the advantages using Raspberry Pi for big data applications through a micro data center. The authors conducted a study that tackles the performance, the scalability, energy consumption, ease of management, etc. For testing purposes, they used Hadoop framework. This allowed the authors to discover that the cluster delivers a moderate performance.

Th authors in [5] explored the use of low-power and low-cost devices for pervasive computing (Cloud and Fog computing). The work tests the performance of the cluster using Apache Spark and HDFS. It also uses the same setup for a real and a virtualized environment. The virtualization has proved to be more significant with large amounts of data. Moreover, the virtualization also affects the energy consumed by the Raspberry Pi when the workload becomes higher.

The research in [6] presented the evaluation and comparison of 24 ARM board energy consumption and performance wise. Finally the Raspberry Pi was chosen as the basis for ARM HPC after looking at the tradeoffs. The results of this study have shown that the overall performance of the cluster remains less than that of the x86 server. However, the authors mentioned that this was not the main end goal of the project. The low-cost of the Raspberry Pi allows for affordable and more accessible cluster computing for the public and especially for students. Hence, the work presented in this research encourages the use of low-power parallel programming in education.

In [7], the authors explained the use of Image Processing algorithms to test the performance of the Raspberry Pi cluster. Image processing is known to require a lot of processing power that is why it is being used to assess the performance of the Raspberry Pi cluster. The authors developed an application that counted the wheat grains presented in a given picture. The Raspberry Pi cluster hosted a parallel computing application that was coded in Python in addition to the use of MPI. The application was tested on a single node first and then on the cluster of four nodes. The purpose was to test the speed of the application. The results have shown that the processing time using both single-node and multi-node Raspberry Pi clusters depends on the number of images processed.

The work in [8] presented the design and creation of a high performance Beowulf cluster based on 12 Raspberry Pis. In order to test the cluster, some mathematical benchmarks have been used to measure the performance of the cluster. The authors used a program to calculate the scalar multiplication of a matrix. The runtime and the GFLOPS were logged. The authors have mainly noticed the following: the peak performance of the system went up as the number of nodes increased, the speedup also increased as more nodes were added to the cluster and increasing the problem size increased the performance. The authors also mentioned a number of limitations related to memory. However, according to the study, the cluster can be used to automate some testing operations.

In [9], the research extended some previously done work by comparing a cluster of Raspberry Pi to multi core processors like i5 and i7 processor machines. The results of this work have shown that the Raspberry Pi cluster is ranked third performance wise after the i7 and the i5 processors. The Raspberry Pi is very limited in terms of resources. Even with 14 nodes it could not reach the performance of its competitors. Core i5 architecture has better resources and eventually performs better than the Raspberry Pi cluster. The core i7 architecture remains with the highest performance.

Our paper aims to extend the previously discussed work by conducting a set of experiments in order to compare our cluster to one single performant commodity machine. The work presented here does not provide an ultimate solution, it rather serves as a baseline for researchers to decide on the most suitable configuration with regards to their applications.

3. Background

3.1. High Performance Computing

High Performance Computing (HPC) deals with the implementation of algorithms or code taking into consideration the hardware used for each application [10]. It was conceived by scientists and engineers to cope with the problems that require more resources (CPU, memory, etc.) than what one single machine can provide.

There are two very known venues to implement HPC: either through owning supercomputers, or clustering commodity computers. The second option is what is generally opted for. HPC is usually provided via a set of computers that have almost the same characteristics, and that are connected to each other. Each one of these machines is referred to as a node [11].

HPC-based solutions usually have three main components: Compute, Network, and Storage. Compute servers are connected forming a cluster which is also connected to the storage component. These connections are not visible to the user who deals with the entire system as a single machine in a transparent manner

HPC has several use cases that can either be deployed on premises, at the edge, or on the cloud [12]:

- Research: HPC clusters help researchers find solutions to very complex problems, and hence bring contributions to the community.
- Entertainment: HPC is also used in media in order to edit films and come up with mind-blowing special effects.
- Artificial Intelligence: HPC implements machine learning algorithms that requires huge processing power.
- Finance: HPC clusters can be used to track stocks, design new products, simulate test scenarios, etc.

3.2. Green Computing

Currently, Green Computing (GC) is considered a shift that is supposed to deal with the constantly increasing energy consumption of the existing computing solutions. This is due to the increasing demand on computing power by different applications.

Green Computing consists a tailor-made solution for many applications as it is based on energy-efficiency as one of the pillars. GC as a concept, refers to the deployment of efficient and eco-friendly computing solutions. It is provided through four main technologies: green data centers, virtualization, cloud computing, grid computing [13].

3.3. Hadoop

It goes without saying that data have been generated in a massive manner throughout the years due to the advancements in all the fields and industries. Talking about huge datasets implies bringing up storage. Storage capacities of hardware systems have increased to keep up with the amount of streaming data. However, access speeds did not keep up. Thus, the traditional Database Management Systems (DBMS) are no longer considered an option to go for. Also, hardware failure is a problem that needs to be overcome.

Hadoop was introduced to solve the previously mentioned problems in addition to others. It is an opensource framework that deals with big data through providing a platform for distributed storage and processing. As Hadoop is used over a set of clustered computers, the possibility of a machine failing is relatively high. This may imply serious data loss if not dealt with correctly. Data replication is a way to tackle this issue effectively: having redundant data all over the system.

3.4. Raspberry Pi

The Raspberry Pi is a computer with the size of a credit-card that can be plugged to a monitor, a keyboard, a mouse, and other devices. It provides a low-cost option to explore computing and programming in different languages (e.g. Python). It is also capable of providing all the options of a normal computer: internet browsing, playing games, etc. [14].

Moreover, the Raspberry Pi has the ability to interact with the outer world through the connection with sensors and actuators. This is done through what is called General-Purpose Input Output (GPIO). This allows the computer to be the best option for robotics related projects and prototypes.

Since its first launching, RP has gone through a set of changes that resulted in different versions and models available in the market. Table 1 presents the models of Raspberry Pi and their characteristics.

Table 1: Raspberry Pi Models [15]

Model	Date	Price ($)	Core	Num. Cores	RAM
RP B	15/2/2012	35	ARM1176JZF-S	1	512M
RP 2	1/2/2015	35	Cortex-A7	4	1G
RP Zero	30/11/2015	5	ARM1176JZF-S	1	512M
RP 3	26/2/2016	35	Cortex-A53 64-bit	4	1G
RP A+	10/11/2014	35	ARM1176JZF-S	1	256M
RP Zero W	28/02/2017	10	ARM1176JZF-S	1	512M
RP WH	12/1/2018	14	ARM1176JZF-S	1	512M
RP 3 B+	14/3/2018	39	Cortex-A53 64-bit	4	1G
RP 3 A+	15/11/2018	25	Cortex-A53 64-bit	4	512M
RP 4 B	24/6/2019	37	Cortex-A72 (ARM v8) 64-bit	4	1/2/4

3.5. Project Scope

The work presented in this paper falls under the scope of a USAID sponsored research project named MiGrid. The project aims at developing a holistic approach that couples renewable energy storage and production in smart buildings.

The general architecture of the project is depicted in the Figure 1 below:

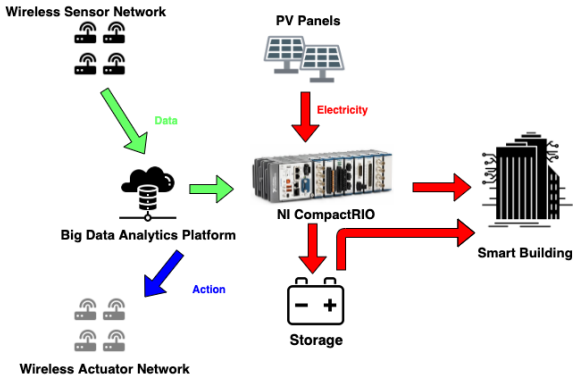


Figure 1: Smart Micro-Grid General Architecture

The architecture contains six main components [16]:

1. Wireless Sensor Network: (WSN) A number of wirelessly connected sensors that are supposed to sense data about the environment.
2. Wireless Actuator Network: Actuators that are connected and interfaced with the WSN. They are responsible for translating the digital signals into electrical ones.
3. Big Data Analytics Platform: A platform that implements a number of algorithms responsible for the processing and analysis of data provided by the WSN.
4. NI CompactRIO: the main controller of the system that decides whether the energy produced is to be stored, injected in the grid, or used to power appliances.
5. Storage: Batteries that store the excess of energy produced and make it available for later use.
6. Solar Panels: the main renewable energy source of the system.

The work tackled in this paper fits in the third (3) component as it presents and tests a potential implementation of the big data analytics platform that is based on Raspberry Pi as the main piece of hardware.

4. Proposed Approach

Our approach to the green cluster relies on using Raspberry Pis instead of commodity computers to deliver HPC. Our work consists of conducting a set of experiments that compare our green Raspberry Pi cluster to one single server. The comparison consists of looking at the performance in terms of CPU time, in addition to the energy cosumption of each solution. Furthermore, we are invistigating the impact of the number of CPUs and memory size on the performance of the single machine, and the impact of the number of nodes and the network performance on the Raspberry Pi cluster.

For the sake of this experiment we made use of two different architectures: a multinode Hadoop architecture, and a single node architecture.

4.1. Multi-node Hadoop Architecture

This setup consists of five Raspberry Pis. The hardware specifications are presented in Table 2:

Table 2 Raspberry Pi Specifications

SoC	Broadcom BCM2837
CPU	4xARM Cortex – A35, 1.2 GHz
GPU	Broadcom VideoCore IV
RAM	2GB
Networking	10/100 Ethernet, 2.4GHz 802.11n Wireless
GPIO	40-pin header, populated
Ports	HDMI, 3.5mm analogue audio-video jack, 4xUSB 2.0, Ethernet, Camera Serial
Price	$25-$195

The hardware architecture consists of the five Raspberry Pis connected together and to a switch through Ethernet connection. Figure 2 below describes the hardware architecture:

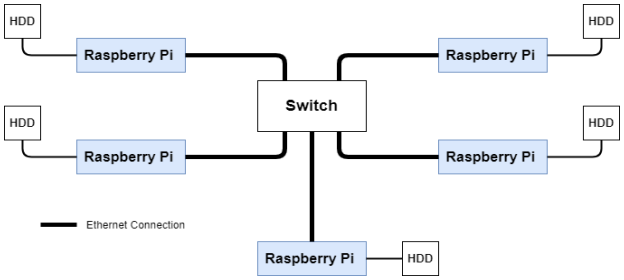


Figure 2: Multi-node Cluster Hardware Architecture

Each one of the Raspberry Pis has the software architecture presented in Figure 3:

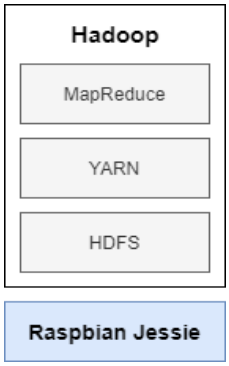


Figure 3: Multi-node Cluster Software Architecture

4.2. Single-node Cluster Architecture

The single node cluster consists of the following server with the specifications mentioned in Table 3 below. The corresponding software architecture is depicted in Figure 4:

Table 3: Single-Node Machine Specifications

Manufacturer	Dell, Inc
CPU	Intel Core i7 (6th Gen) 3.4 GHz
Number of cores	Octa-core
RAM	8GB
Hard Drive	SATA, HDD, 1TB
Networking	Ethernet, Fast Ethernet, Gigabit Ethernet
Graphic Controller	NVIDIA Quadro K620 2 GB
Price	$900

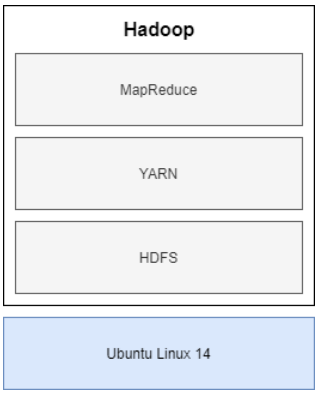


Figure 4: Single Node Machine Software Architecture

4.3. Benchmark

For this experiment, our setups were benchmarked against one of the most famous Hadoop benchmarks: Terasort.

The Terasort is a sorting algorithm implemented using MapReduce. There are two more functions related to Terasort that that are also implemented in Hadoop:

- Teragen: it generates the data to be sorted. We can specify the size of the dataset to be sorted.
- Terasort: the actual sorting jar that takes the result of the Teragen function.
- Teravalidate: it validates the output (the sorted result of the input data).

For the execution, we followed the steps presented in Figure 5:

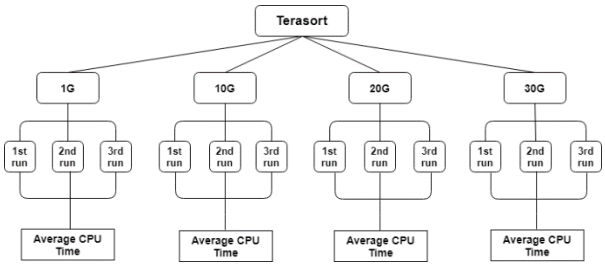


Figure 5: Terasort Execution Steps

We ran the experiment on four different dataset sizes: 1GB, 10GB, 20GB, and 30GB. For each run, and in order to avoid outliers in the results obtained, we ran each experiment many times and we kept the best (statistically) three results, and calculated the mean for both the energy consumption and the execution time.

4.4. Energy Consumption Measurement

In order to measure the energy consumed by each of the setups, we created a sensor node based on the Arduino Uno

microcontroller in addition to the SCT013 current sensor. Arduino is an open-source platform that is easy to use and mainly serves as a prototyping platform. The boards are able to read input form sensors and turn it into an output. We opted for Arduino due to its price affordability and accessibility user-experience wise [17].

The SCT013 sensor is a non-invasive current transformer that measures the intensity of the current in a conductor. These measurements are provided using the electromagnetic induction. They come in the form if clamps that can be wrapped around the equipment. The sensor is very accurate with an error rate of only 1-2% [18].

The Arduino circuit that we used is shown in Figure 6 below:

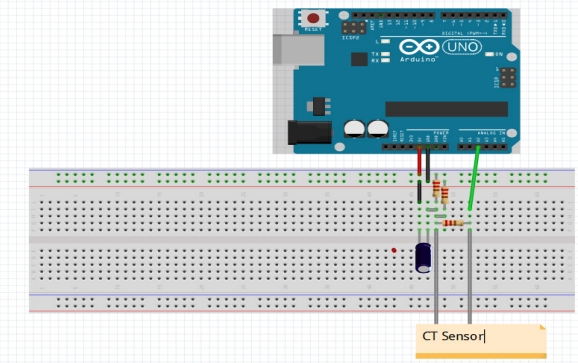


Figure 6: Current Sensor Arduino Circuit

5. Performance Evaluation

5.1. Single Node Cluster

a. Impact of the Number of CPUs

In this section, we are going to study the impact of the number of CPUs on the performance (i.e. the CPU time). The server we are working with is an octa-core machine. We ran the same job 8 times: limiting the number of CPUs from 1 to 8.

The results are given in Figure 7.

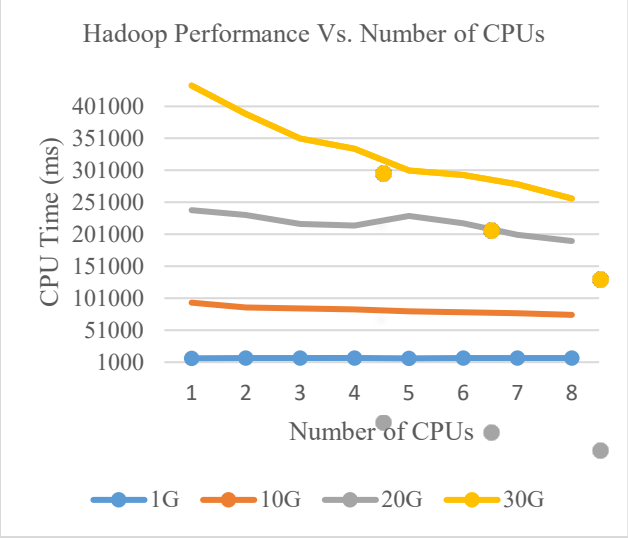


Figure 7: Hadoop Performance Vs. Number of CPUs

The job used in this comparison is the Terasort benchmark as mentioned previously in the paper. The sorting algorithm used as input files of different sizes.

As we can notice from the graph above, the difference in the CPU time when sorting 1Gb of data is not significant: it varies between 7350 ms and 7100 ms. For the CPU time corresponding to sorting 10Gb of data, it showed a slight decrease as the number of CPUs increased: it went from 94100 ms to 75005 ms. The most significant difference was shown in the performance of sorting 20Gb and 30Gb of data.

Concerning the CPU time resultant from sorting 20Gb of data, it scaled down from around 239000 ms to 191000 ms. Sorting 30Gb of data took eventually more time and went from 433590 ms using 1 CPU to about 251000 ms using 8 CPUs.

The number of CPUs used in the Dell Precision Tower machine has a significant impact on the performance of Hadoop jobs. This impact becomes more and more noticeable as the size of the job increases.

In the next section, we are going to have a closer look at the impact of memory size on the performance of Hadoop as a single node cluster.

b. Impact of Memory Size

In order to investigate the impact that the size of the memory has on the performance of Hadoop, we made use of the same benchmark (i.e. Terasort) and using the same input dataset sizes. The machine that we are working with in this experiment has a maximum of 8Gb of RAM. For each job, we had to limit the memory usage starting from 1Gb to 8Gb.

Figure 8 below shows the results of the experiment.

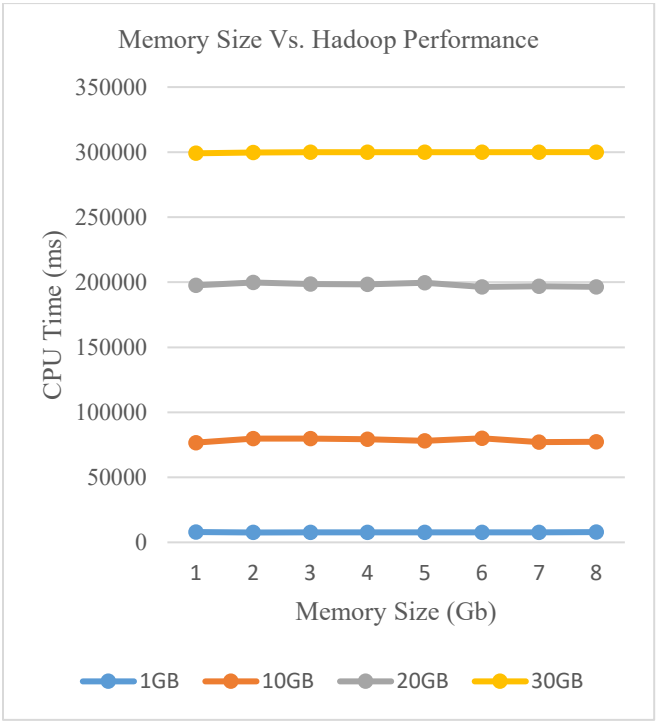


Figure 8: Memory Size Vs. Hadoop Performance

As shown in the figure above, scaling up in the size of the memory did not impact the performance of the cluster. The CPU

time remained the same throughout all the experiments. Sorting 1Gb of data took about 7500 ms with all the sizes of the memory. Concerning CPU time of sorting 10Gb of data, it was about 78000 ms and did not significantly change with all the memory sizes. Similarly, sorting 20Gb of data was stable at around 199000 ms. Also, the CPU time corresponding to sorting 30Gb of data was constant at around 300000 ms and was not impacted by the memory size. This mainly due to the nature of Hadoop and its technologies: they are disk-based and not memory-based. Unlike Spark, Hadoop does not work in memory, which somehow explains the delay that it presents compared to Spark platform.

c. Energy Consumption

Since we are concerned by the energy efficiency as a primary matter, we measured the energy consumed by the singe-node cluster for the same benchmark using 1Gb and 30Gb of input data. For the sake of this experiment, we used all 8 CPUs of the machine and unlimited memory resources.

The energy consumption measured is shown in Figure 9 and 10.

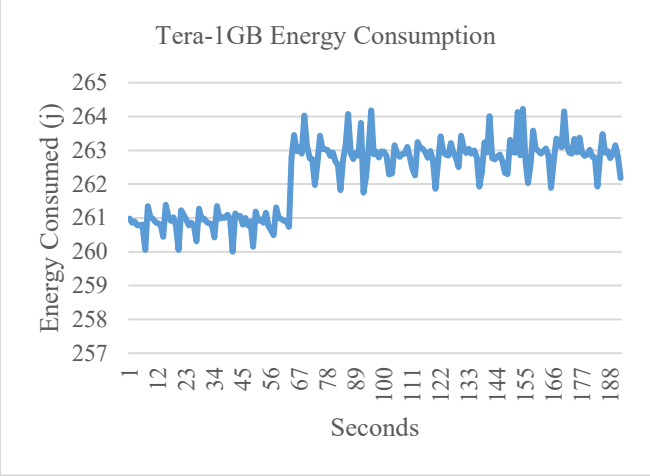


Figure 9: Tera-1GB Energy Consumption

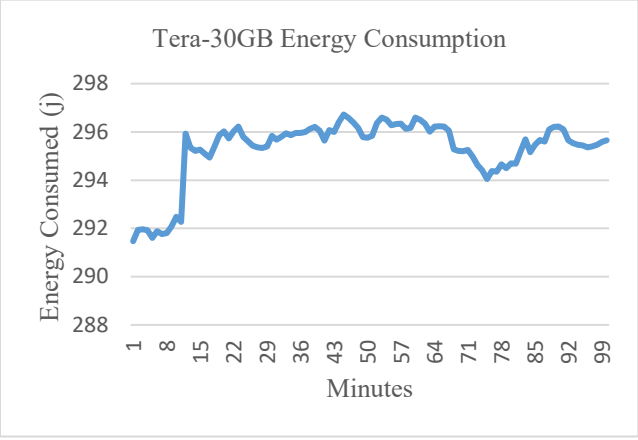


Figure 10: Tera-30Gb Energy Consumption

From the graphs above, we can notice that the energy consumption during the execution time of the Terasort benchmark with 1GB dataset went from 260 J/s to 290 J/s with an average of about 263 J/s. Running the same benchmark on a 30GB dataset size consumed an average of 295 J/s with the max being 296 J/s and the min 291 J/s.

5.2. Multi Node Raspberry Pi Cluster

a. Impact of the Number of Nodes

In this section, we are exploring the impact of the number of nodes in a multi-node cluster on the performance of Hadoop.

As described previously in the paper, we are dealing with a cluster of 5 Raspberry Pis that are connected together through ethernet. Each node has the same software architecture that is based on the same version of Hadoop.

For the sake of comparison, we are using the same benchmark as the previous experiment with the same dataset input sizes.

The result of running Terasort on Raspberry Pi is mentioned in the figure below.

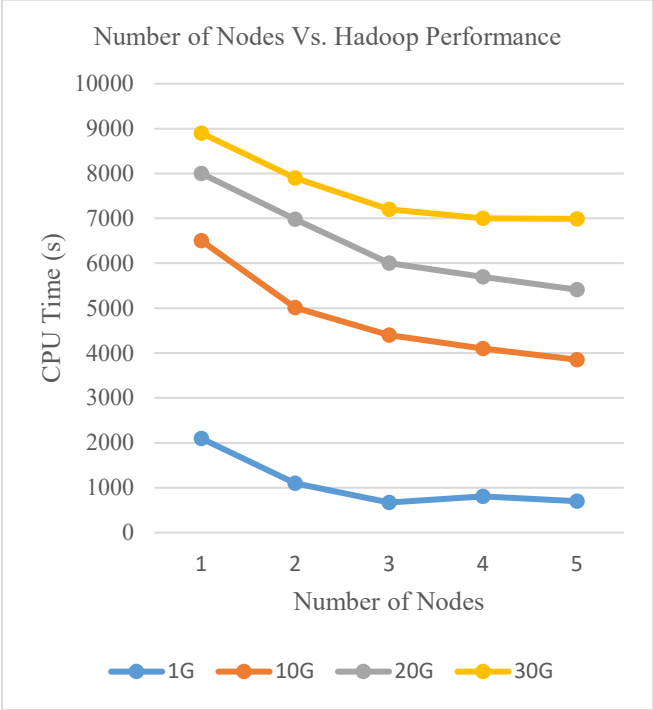


Figure 11: Number of Nodes Vs. Hadoop Performance

As we can infer from Figure 11, sorting all datasets using one node takes always longer than sorting them with more nodes. The time taken by the job using 1 node to sort 1Gb was about 2100s, and to sort 10Gb was 6500s, 8000s to sort 20Gb, and around 9000s to sort 30Gb of data. While using 5 nodes, the CPU time decreased significantly as it took only 700s to sort 1Gb of data, around 3800s to sort 10Gb of data, 5400s to sort 20Gb, and 6900s to sort 30Gb of data.

This means that adding Raspberry Pi nodes to the cluster increases the performance and reduces the CPU time taken by the jobs. Adding nodes means adding resources: memory size and CPUs. Knowing that each Raspberry Pi has 4 CPUs and 1Gb of RAM, working with 5 nodes implies working with 20 CPUs and 5Gb of RAM.

b. Impact of the Network

It goes without saying that any system that is distributed over a number of physical machines requires a network connection

between the nodes. This connection brings an overhead that affects the response time of the cluster.

In order to have a closer look at the impact of the network on the performance of Hadoop, we are connecting the 5 nodes of the cluster using two different types of switches with different throughputs: 100M and 1000M. Table 4 below describes both switches and presents their characteristics.

Table 4: NetGear Switch Vs. Fujitech Switch

	NetGear	**Fujitech**
Dimension	235.5x100.8x27	14.5x8.5x2.6
Network Characteristics	Gigabit Ethernet	Megabit Ethernet
Ports	8x10/100/1000	8x10/100
Energy Consumption	14W	-
Transfer Rate	1Gb/s	1Mbps
Weight	760g	-

Similar to the previous experiments, we are using the same benchmark and the same dataset sizes. The results of running Terasort using the two switches are shown in Figure 12.

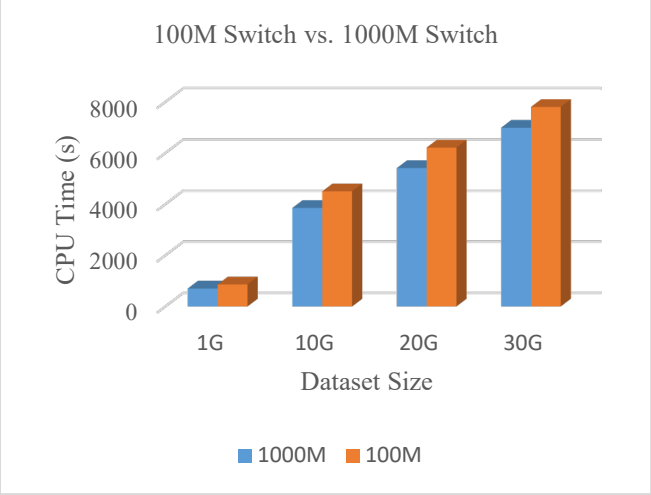


Figure 12: 100M Switch Vs. 1000M Switch

The results show the slight increase in the performance of the cluster as executing the same jobs using the 1000M switch takes less time than the other switch. The network clearly impacts the response time of the cluster. Hence, improving the network quality and equipment would result in a better performance.

c. Energy Consumption

Similar to the previous setup, we also looking at the energy consumption of the entire cluster.

Using the method described previously, we were able to measure the energy consumption of the cluster when sorting 1Gb of data and 30Gb of data. We used the same sensor for each of the slaves and we noticed that the power varies between 4W when jobs are being performed and 11W in an idle state. Regarding the master, the sensor showed a max of 6.7W and a min of 13W and that is because of the different external devices that are connected to it (i.e. mouse, keyboard, monitor, in addition to the external hard disk drive).

The sum of the values given by the sensors is shown in Figure 13 below.

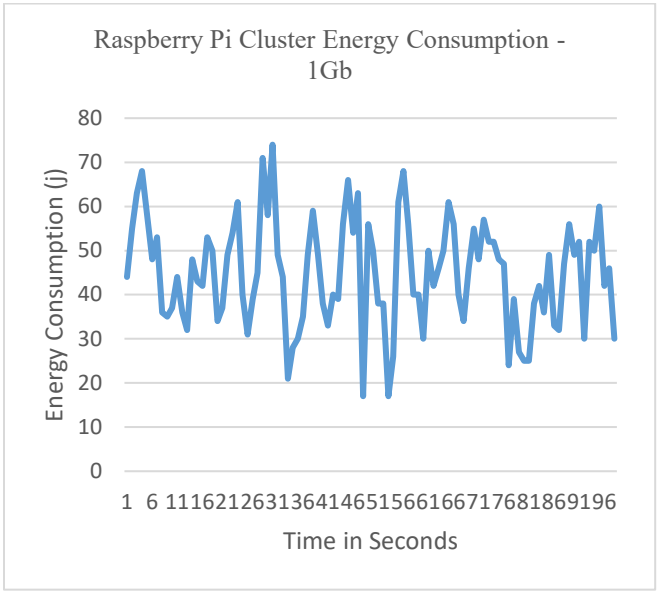


Figure 13: Raspberry Pi Cluster Energy Consumption - 1Gb

As we can notice, the entire 5-node cluster consumes between 17J/s and 70J/s. This is considered relatively low; we are looking into a detailed comparison later in this paper.

The energy consumed when sorting 30Gb of data using the same setup is shown in Figure 14.

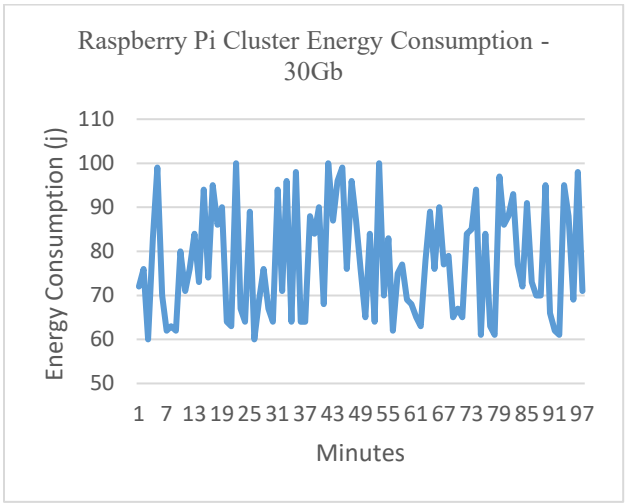


Figure 14: Raspberry Pi Cluster Energy Consumption - 30Gb

The energy consumed during the job is varying between 60J/s and 100J/s.

6. Comparison and Discussion

6.1. Performance Comparison: Single-node Vs. Multi-node Clusters

In order to have a general look over the difference between the two clusters, we are comparing the results given by all CPUs and full memory in the single-node cluster, and all nodes in the multi-node clusters.

The graph below shows the CPU time taken to sort all dataset sizes using both clusters.

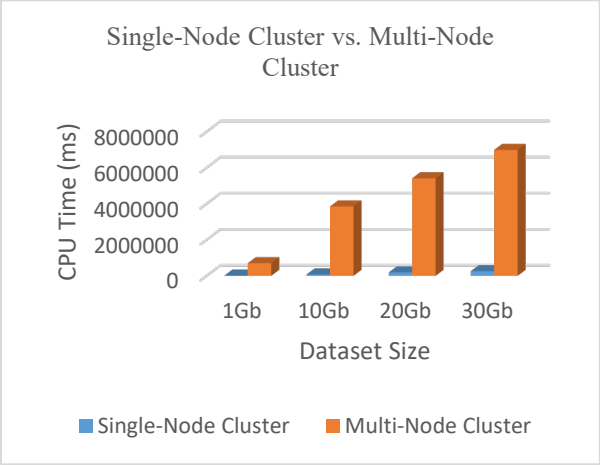


Figure 15: Single-Node vs. Multi-Node Cluster

Figure 15 above shows the clear difference between sorting the same dataset using the two different clusters. As the job becomes heavier (i.e. more data to be processed), the difference between the CPU times becomes more significant.

This implies that the 5-node cluster that we have is not enough to provide the same performance given by the single node cluster.

6.2. Energy Consumption Comparison: Single-node vs. Multi-node Clusters

As mentioned previously in this paper, our approach is supposed to be energy efficient as it tackles that concept of green computing. Thus, we are comparing the energy consumed by both clusters during the execution of the same job. We are comparing the first 100 seconds of sorting 1Gb of data using all CPUs and full memory for the single-node cluster, and all 5 nodes for the multi-node cluster. The result is shown in Figure 16.

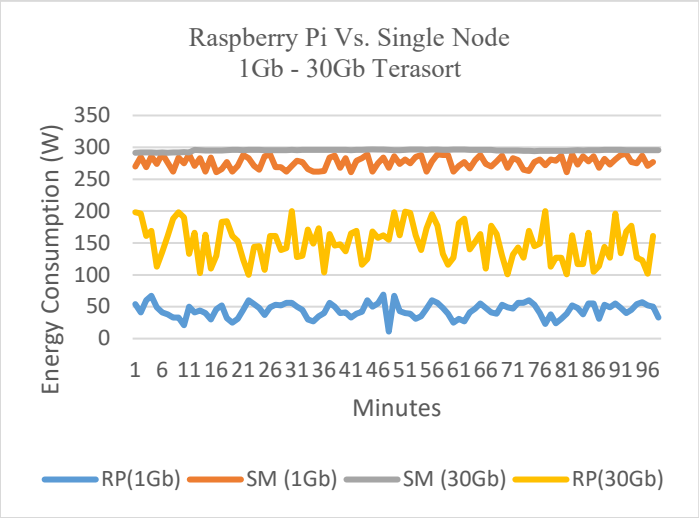


Figure 16: Raspberry Pi Cluster Vs. Single-Node Energy Consumption

The graph above shows the difference in the energy consumed by the two setups when sorting 1Gb and 30Gb of data using Hadoop. The results show that, not only the single machine consumes more energy, but also the RP cluster has a consumption that varies a lot compared to the single machine.

6.3. Discussion

The different experiments conducted in this paper were for the sake of determining how performant the Raspberry Pi green cluster can be.

According to the study conducted in [13] where the authors compared a commodity hardware cluster with a Raspberry Pi cluster using two different Hadoop benchmarks: Terasort and TestDFSIO. The results showed that the commodity hardware cluster outperformed the Raspberry Pi cluster. This was mainly due to the low computing power of the Raspberry Pi. The work performed in this paper sustains these results; In addition to that, since we are only using a single machine vs. a multi-node cluster, the network overhead is added and the performance is eventually decreased.

Another research has been done to test a Raspberry Pi Hadoop cluster against an image analysis in a cloud robotics environment [19]. Their research has proven that the Hadoop cluster lacks in performance compared to a Hadoop cluster that is based on virtual machines running on top of commodity computers.

Authors in [20] investigated the use of Raspberry Pi computers to implement an efficient solution for augmented computing performance. They used Hadoop along with benchmarks and compared the outcome to the one of a single commodity computer. The results showed that the single machine outperforms the clusters in almost all the operations. However, the low-cost and the light weight of the Raspberry Pi based solution makes it more suitable.

For the sake of our experiment, we are using the Amdahl's law to predict the performance of the Raspberry Pi cluster and decide on when the cluster will reach the performance of the single machine we are using.

According to the paralleled and serialized sections of the algorithm used, based on the Amdahl's law formula, the speedup is supposed to be as presented in Figure 17.

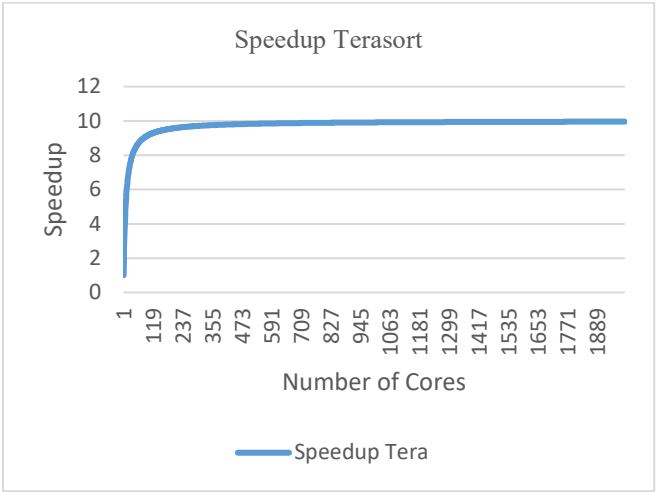


Figure 17: Speedup Terasort for Terasort Benchmark

We applied the formula for the number of cores from 1 to 2000 cores. Based on the speedup found, the performance is expected to increase according to the graph shown in Figure 18.

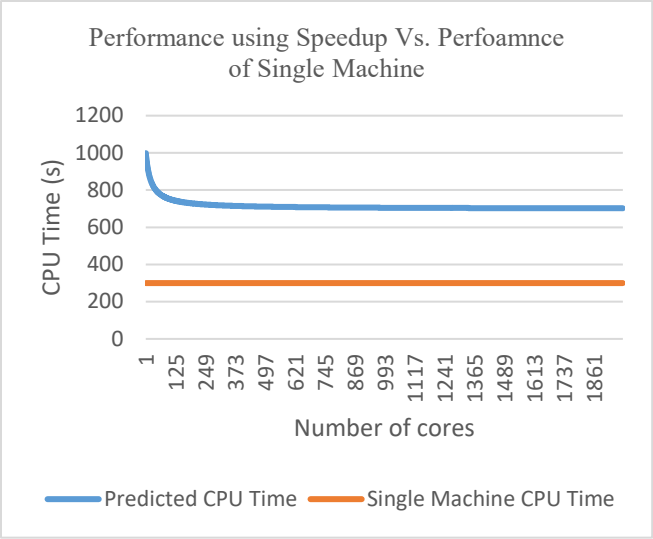


Figure 17: Performance of cluster using Speedup vs Performance of a Single Machine

Although we worked with 2000 cores (knowing that each RP has 4 cores, we are dealing with 500 Raspberry Pi nodes), we could not get the same performance as the single machine performance.

The best CPU time we achieved is 700 seconds using the Raspberry Pi cluster while the single machine performs the exact same job in 300 seconds.

At this level, we need to keep in mind that this study is theoretical, the real-world speedup can be given by the real measurements that we conducted and presented earlier in this paper. Also, the real-world measurements are affected by a number of other parameters and features: the network communication between the nodes that is, itself affected by the quality of the equipment used (i.e. switches as presented earlier).

The difference between the real-world and the theoretical results are shown in Figure 19.

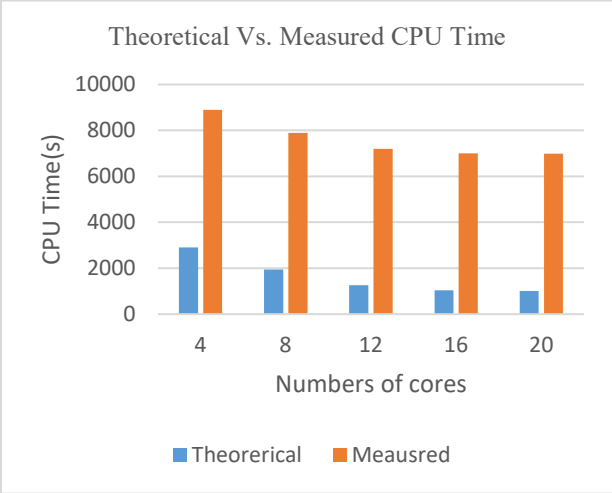


Figure 19: Theoretical vs. Measured CPU Time

Based on the theoretical approach that is built on Amdahl's law, the 500-node cluster will consume a minimum of 2000 W and 5000 W. In addition to that, the cluster will cost a minimum of $10000. Table 5 below summarizes the comparison between the single-node commodity computer and the Raspberry Pi 500-node cluster.

Table 5: Single-Node Vs. 500-Node RP Cluster

	Single-Node	**500-node RP cluster**
Price	$900	$10000
Max Energy Consumption	292W	Around 5000W
Min Energy Consumption	261W	Around 2000W
Scalability	Medium	Easy
Maintenance	Medium	Medium
Ease of use	Medium	Easy
Reliability	High	Medium

Oracle has been working on the world's largest Pi cluster where they used 1060 Raspberry Pi nodes to create a supercomputer.

According to [21], the cluster has 4240 cores for processing and it costs around $37100 without counting the external storage devices, the cables, 3D printed holders, etc.

7. Conclusion and Future Work

In this paper, we investigated the use of a Raspberry Pi clusters to provide High Performance Computing. The cluster performance was benchmarked against the Terasort algorithm and compared to a single legacy server machine. Both experimental setups run Hadoop. The experiments investigated the impact of the number of nodes and the network bandwidth on the performance of the Raspberry Pi cluster. Besides, we also tracked the impact of the number of cores and the RAM memory size on the performance of the single-node setup. We measured the energy consumed by both setups while performing the same operations.

The results of the experiments showed that the single-node cluster outperforms the Raspberry Pi cluster when sorting 1G, 10G, 20G, and 30G of data, but consumes less energy. However, and based on the Amdahl's law formula, the Raspberry Pi delivers a performance that is closer to the one of the single-node cluster when the number of cores reaches 2000. This implies having a cluster of 500 nodes that would cost around $10000 with a max energy consumption of 5000W. These results present a decent ground for researchers to base their choices on. If the performance is not a priority but the cost-effectiveness and the energy efficiency are, then the Raspberry Pi cluster is suitable. However, when the response time is essential, the most adequate solution is to opt for a more stable and performant solution.

As a future work, we are performing more projections on the number of CPUs used in a Raspberry Pi cluster, in order to determine (using prediction methods) the most suitable configuration according to the criteria/need of each application. Furthermore, we are examining more big data analytics solutions and investigating the possibility of putting in place a private cloud

that is based on Raspberry Pi. This will fall under the realm of green private clouds.

Conflict of Interest

The authors declare no conflict of interest

Acknowledgement

This work is sponsored by US-NAS/USAID under the PEER Cycle 5 project grant# 5-398, entitled 'Towards Smart Microgrid: Renewable Energy Integration into Smart Buildings".

References

[1] P. Abrahamsson, S. Helmer, N. Phaphoom, L. Nicolodi, N. Preda, L. Miori, M. Angriman, J. Rikkilä, X. Wang, K. Hamily, S. Bugoloni, "Affordable and energy-efficient cloud computing clusters: The Bolzano Raspberry Pi cloud cluster experiment," Proceedings of the International Conference on Cloud Computing Technology and Science, CloudCom, **2**, 170–175, 2013, doi:10.1109/CloudCom.2013.121.

[2] S.J. Cox, J.T. Cox, R.P. Boardman, S.J. Johnston, M. Scott, N.S. O'Brien, "Iridis-pi: A low-cost, compact demonstration cluster," Cluster Computing, **17**(2), 349–358, 2014, doi:10.1007/s10586-013-0282-7.

[3] F.P. Tso, D.R. White, S. Jouet, J. Singer, D.P. Pezaros, "The Glasgow raspberry Pi cloud: A scale model for cloud computing infrastructures," Proceedings - International Conference on Distributed Computing Systems, 108–112, 2013, doi:10.1109/ICDCSW.2013.25.

[4] N.J. Schot, "Feasibility of Raspberry Pi 2 based Micro Data Centers in Big Data Applications," 2015.

[5] W. Hajji, F.P. Tso, "Understanding the performance of low power raspberry pi cloud for big data," Electronics (Switzerland), **5**(2), 1–14, 2016, doi:10.3390/electronics5020029.

[6] M.F. Cloutier, C. Paradis, V.M. Weaver, "A raspberry Pi cluster instrumented for fine-grained power measurement," Electronics (Switzerland), **5**(4), 2016, doi:10.3390/electronics5040061.

[7] D. Marković, D. Vujičić, D. Mitrović, S. Ranđić, "Image Processing on Raspberry Pi Cluster," Ijeec - International Journal of Electrical Engineering and Computing, **2**(2), 2019, doi:10.7251/ijeec1802083m.

[8] D. Papakyriakou, D. Kottou, I. Kostouros, "Benchmarking Raspberry Pi 2 Beowulf Cluster," International Journal of Computer Applications, **179**(32), 21–27, 2018, doi:10.5120/ijca2018916728.

[9] A. Ashari, M. Riasetiawan, "High performance computing on cluster and multicore architecture," Telkomnika (Telecommunication Computing Electronics and Control), **13**(4), 1408–1413, 2015, doi:10.12928/TELKOMNIKA.v13i4.2156.

[10] M. Alexander, W. Gardner, B. Wilkinson, M.J. Sottile, T.G. Mattson, C.E. Rasmussen, Y. Robert, F. Vivien, Computational Science Series petascale computing: algorithms and applications edited by david a. bader process algebra for parallel and distributed processing.

[11] Introduction to HPC & What is an HPC Cluster? | WekaIO, Dec. 2020.

[12] High performance computing: Do you need it? | Network World, Dec. 2020.

[13] S. Bourhnane, M.R. Abid, R. Lghoul, K. Zine-Dine, N. Elkamoun, D. Benhaddou, "Towards Green Data Centers," in Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, Springer: 291–307, 2020, doi:10.1007/978-3-030-45694-8_23.

[14] What is a Raspberry Pi?, Dec. 2020.

[15] Comparatif des modèles de Raspberry PI | Tableaux comparatifs - SocialCompare, Dec. 2020.

[16] M.R. Abid, R. Lghoul, D. Benhaddou, "ICT for renewable energy integration into smart buildings: IoT and big data approach," 2017 IEEE AFRICON: Science, Technology and Innovation for Africa, AFRICON 2017, 856–861, 2017, doi:10.1109/AFRCON.2017.8095594.

[17] Arduino - Introduction, Dec. 2020.

[18] Non-Invasive Sensor: YHDC SCT013-000 CT used with Arduino. (SCT-013) – PowerUC, Dec. 2020.

[19] B. Qureshi, Y. Javed, A. Koubâa, M.F. Sriti, M. Alajlan, "Performance of a Low Cost Hadoop Cluster for Image Analysis in Cloud Robotics Environment," Procedia Computer Science, **82**(March), 90–98, 2016, doi:10.1016/j.procs.2016.04.013.

[20] K. Srinivasan, C.Y. Chang, C.H. Huang, M.H. Chang, A. Sharma, A. Ankur, "An efficient implementation of mobile Raspberry Pi Hadoop clusters for Robust and Augmented computing performance," Journal of Information Processing Systems, **14**(4), 989–1009, 2018, doi:10.3745/JIPS.01.0031.

[21] Oracle: This 1,060 Raspberry Pi supercomputer is "world's largest Pi cluster" | ZDNet, Dec. 2020.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 1609-1617 (2020)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

Time-to-Event Analysis for Recovery from Coronavirus Disease (COVID-19): A Case Study on Wuhan and Elsewhere in China from Jan 1 to Feb 11, 2020

Murtada Khalafallah Elbashir[*,1,2], Saleh N. Almuayqil[1]

[1]*Department of Information Systems, College of Computer and Information Sciences, Jouf University, Sakaka, 72311, Kingdom of Saudi Arabia*

[2]*Department of Computer Science, Faculty of Mathematical and Computer Sciences, University of Gezira, Madani, 585-0032, Sudan*

ARTICLE INFO

Article history:
Received: 19 September, 2020
Accepted: 21 December, 2020
Online: 25 December, 2020

Keywords:
Kaplan-Meier method
Cox regression model
Proportional Hazard model

ABSTRACT

COVID-19 is a viral disease that became a pandemic representing a very great challenge worldwide. The purpose of this article is to analyze COVID-19 patients' data based on time-to-event analysis and identify the factors that affect the recovery time from COVID-19. The datasets that are used in this study are for cases that are clinically diagnosed and confirmed where the date of onset is recorded in Wuhan and elsewhere in China from Jan 1 to Feb 11, 2020. We used the regression imputation technique to replace the missing dates in the onset-symptoms based on the dates of the report. We fitted the Kaplan-Meier estimator and Cox regression model to our data. The predictor variables (factors) that we used are age, sex, and onset time to hospitalization. The results show that the young age group is better than the old age group in recovering from COVID-19 (the p-value of the log-rank is 0.00012) and at any time 1.9 as many patients in the young age group are having an event (recovery) proportionally to the old age group. Also, the results show that there is a non-significant difference between male and female groups in recovering from COVID-19 (the p-value of the log-rank is 0.63). The results also show that the early time to hospitalization group can recover from COVID-19 better than the late time to hospitalization group (the p-value of the log-rank is 0.0052). This study demonstrates the association of recovery time from COVID-19 with age, sex, and time to hospitalization.

1 Introduction

The novel coronavirus disease also known as COVID-19 is a viral disease that became a pandemic and turn out to be a great challenge that the world faced since world war two. This virus is originated in Wuhan city, which located in the Hubei province of China and it is spreading at a fast rate around the globe. The diseases caused by viral infection continue to emerge and raise a serious issue in public health worldwide. Several viral epidemics appeared in the last twenty years [1]. In 2002 the severe acute respiratory syndrome coronavirus, which is known as SARS-CoV, which is still circulating in China [2]–[4] has appeared followed by H1N1 influenza in 2009. Most recently in 2012, the Middle East respiratory syndrome coronavirus, which is known as MERS-CoV have been recorded. The main component of the viral genome is a positive-stranded RNA and it has a different structure [5]. There are four genera of Coronavirinae family: α, β, γ, and δ. it is believed that there is a viral gene in wild animals since it has been isolated from bats and other animals [6]. The novel COVID-19 causes mild to moderate respiratory illness, but some people and people with health problems can develop serious illness. Worldwide this disease affected more than five billion people and the number of people who died due to the infection with it exceeds five hundred thousand according to the World Health Organization (WHO) report on the time of writing this research. According to WHO the mild or asymptomatic COVID-19 infections represent 80% of the cases while the severe infections, which require oxygen and critical infections, which require ventilation represents 15% and 5% of the cases respectively. So far, the mortality for COVID-19, which is the total number of deaths divided by the total cases is 5% (this percentage is calculated according to figures that are taken from the WHO web site on 09 July 2020, which shows a total infection of (12,196,982) and total deaths of (552,781)). This mortality is higher than that of seasonal influenza, which is below 1% according to WHO.

[*]Corresponding Author: Murtada K. Elbashir, Jouf University, mkelfaki@ju.edu.sa

Normally, data mining and machine learning methods can be used to analyze datasets of biomedical data [7]. When data include survival data, it requires a different analysis approach. This approach or the study of the time from the entry of a study until a subsequent event occurs is known as survival analysis. Survival analysis is applied in different disciplines such as medicine, engineering, social sciences or behavioral sciences and biology [8]–[16]. When it is applied to medicine, survival analysis is used to study people at risk of experiencing a negative event such as death, where the name survival analysis comes from. Survival analysis is also applicable to areas other than mortality such as analyzing the time taken to recover from certain diseases or the time taken to practice certain exercises to maximum tolerance [17]–[20]. Normally we compare two or more groups of patients with respect to the time of event. More than one event can be considered in the same analysis, but we normally take one event at a time as the event of interest in the study and it can be death or recovery [21].

Many methods can be used for survival analysis, these methods include Kaplan-Meier method which is an estimator of survival probabilities [19, 22] and the Cox regression model, which is now known as the Cox Proportional Hazard Model (CPHM) [23]. These two methods are considered among the methods that contribute significantly to the development of the survival analysis field.

Many studies were conducted to model the survival time and to predict the mortality risk for COVID-19. Guillermo Salinas-Escudero et al. applied survival analysis to study the effect of COVID-19 in the Mexican [24]. The factors they used include age, sex, comorbidities, hospitalization, and admission to the intensive care unit. They applied the Kaplan-Meier and Cox regression models to their data. Their results show that men and older people have higher mortality than women and young people respectively. Monira Mollazehi et al. modeled survival time to recover from COVID-19 [25]. They used data from Singapore in the period between January 23 and March 13, 2020. Their purpose is to identify the factors affecting the recovery time from COVID-19. They used patient's age and nationality as predictors and they found that younger patients can recover from COVID-19 faster than old patients and Singaporean patients can recover faster than non- Singaporean. They compared the results of different models and they found that the Weibull model is the best in fitting their data. Using the Weibull model, they obtained a Hazard rate of 1.01 and 0.76 for age and nationality respectively. Qinxia Wang et al. used survival-convolution models to model the duration of the patient remaining infectious to others [26]. Noam Barda et al proposed a hybrid methodology to construct a multivariable prediction model. In their hybrid method, they used a baseline model which they trained on population data to discriminate the risk then they used a multicalibration algorithm for the risk predation [27]

Different factors may have an influence on the mortality or the recovery time from COVID-19. These factors can be used to divide the patients into two or more groups and they include age, gender, and time from acquiring the illness to hospitalization. This study aims to investigate whether these factors affect recovery time. The datasets that are used in this study are downloaded from Github (`https://github.com/mrc-ide/COVID19_CFR_submission`). From these datasets, we used two datasets. The first one is for cases that died from COVID-19 in Hubei and the second dataset is for patients returning to their home, which obtained from six flights that departed between Jan 30 and Feb 1, 2020.

The rest of the paper is organized as follows: The next section describes the materials and methods. The material and methods section starts by showing how we prepared the dataset that we used followed by describing the imputation technique we used to replace the missing data. Kaplan–Meier survival curve, Log-rank test, and Cox proportional hazards (PH) model also are explained in the materials and methods section. we present the results and the discussion in the third section and the conclusion in the last section.

2 Material and Methods

2.1 Dataset

The datasets that we used in this study are for cases that are clinically diagnosed and confirmed where the date of onset is recorded in Wuhan and elsewhere in China from Jan 1 to Feb 11, 2020. From these datasets, we used two datasets. The first one is for cases that died from COVID-19 in Hubei. It contains the features: sex, age, date of symptom onset, date of hospitalization, and date of death or recovery from COVID-19. Some of the data for the date of symptom onset are not available for some cases so used imputation based on regression to replace the missing data. The second dataset is for patients returning to their home, which obtained from six flights that departed between Jan 30 and Feb 1, 2020. Also, the cases with incomplete date of symptom onset were replaced using regression imputation and then we merged the two datasets. We removed the cases where the sex or the age or date of hospitalization are not available and we end up with 693 cases, which represent recovered and died patients. The used datasets were downloaded from Github (`https://github.com/mrc-ide/COVID19_CFR_submission`). They were used by [28], which extracted it from WHO–China Joint Mission report to estimates the severity of COVID-19 based on the model-based analysis.

2.2 Regression imputation

The datasets that we downloaded has missing data on the date onset symptoms, therefore, instead of deleting all the cases that have missing data, we need imputation to replace these missing data with estimated values, because it is important to have the timing of the onset-symptoms to study the recovery time. We used regression imputation to preserves all cases by replacing the missing date onset symptoms with a probable value estimated by the date of the report because it is clear from Figure 1 that there is a strong correlation between these two dates. In Figure 2, the scatter plot shows the relationship between these two dates, and the value of R^2 (0.7266) emphasizes the strength of this relationship. The model that is used to estimate the missing data in the date onset symptoms is $y = 0.8227x - 0.5428$ also shown in Figure 2, where x and y represent the report date and the onset symptoms respectively. Preserving the cases with missing data using regression imputation has several advantages. In addition to avoiding the deletion of the cases with missing data that can alter the variance of the shape distribution, it can also substitute the missing value based on another variable and no novel information will be added therefore we will be having

an increased sample size and therefore a reduced standard error [29]–[31].

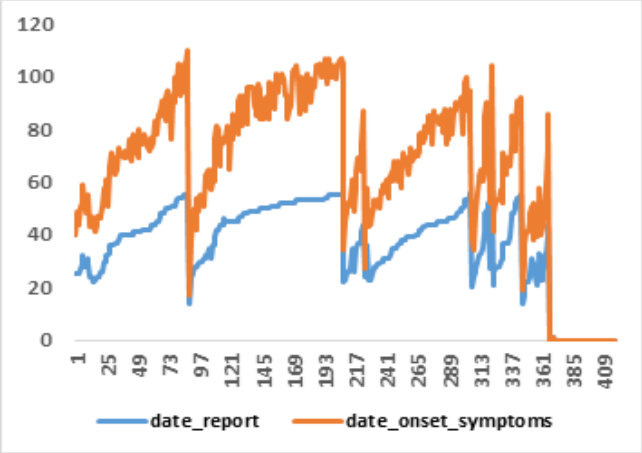


Figure 1: The correlation between date_report and date_onset_symptoms.

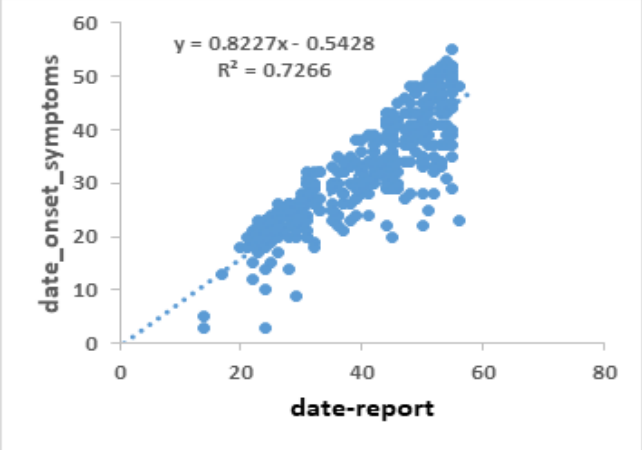


Figure 2: Regression prediction for date_onset_symptoms.

2.3 Kaplan–Meier survival curve

In Kaplan–Meier survival curve, the survival times that include the censored data (the observation that does not get the event) is assumed to be $t_1, t_2, ..., t_n$. These times are entered to the study ordered by increasing duration of a group of n subjects, We can estimate the proportion (survival rate) of subjects $S(t)$ surviving beyond any follow-up time t_p as [17]:

$$S(t) = \frac{(r_1 - d_1)}{r_1} \times \frac{(r_2 - d_2)}{r_2} ... \times ... \frac{(r_p - d_p)}{r_p} \quad (1)$$

Here r_i represents the number of subjects alive just before time t_i given that t_p is the largest survival time and i is any value between 1 and p, d_i represents the number of subjects who died at the time t_i, therefore $d_i = 0$ for censored observations. Before the occurrence of the first event all the patients are alive, therefore, $S(t) = 1$. Considering time t_i, where the number of events(deaths) is d_i and the number of alive is r_i just before t_i then $S(t_i)$ can be calculated as:

$$S(t_i) = \frac{(r_i - d_i)}{r_i} \times S(t_{i-1}) \quad (2)$$

In the censored data we will not have information about the survival time, therefore, $S(t_i)$ will not be calculated for censored observations since the survival curve will not change at the time of a censored observation. At the next event, the number of patients at risk is reduced by the number of censored observations between the two events [32].

2.4 Log-rank test

Normally, we need to compare two survival curves of two groups. For this sake, we use Log-rank test, which is related to a test that uses the logarithms of the ranks of the data and it is used under the assumptions: i) the survival times are continuous or ordinal, ii) one group's risk of an event relative to the other does not change with time. When the death event occurs at time t_i then we will consider the total number alive (r_i) and the total number still alive up to the time t_i in a specific group (say group A) r_{Ai}. Consider that d_i is the total number of deaths i.e event at the time t_i. Then the expected number of deaths in group A at time t_i can be calculated as

$$E_{Ai} = \frac{r_{Ai}}{r_i} \times d_i \quad (3)$$

Then the total number of expected deaths for group A can be calculated as:

$$E_A = \sum E_{Ai} \quad (4)$$

The total number of the expected deaths in group B can be calculated based on the total number of expected deaths for group A given that the total number of events is n as follows:

$$E_B = n - E_A \quad (5)$$

In the Log-rank test, the data for the two groups combined are ordered and then each event, in turn, is considered starting at time $t = 0$. Then the log-rank statistics is calculated for two groups based on the summed observed minus expected score for a given group and its variance estimate and it is given as follows:

$$\chi^2 = \frac{(O_A - E_A)^2}{E_A} + \frac{(O_B - E_B)^2}{E_B} \quad (6)$$

Here O_A and O_B represent the total number of events in groups A and B respectively and E_A and E_B represent the total number of expected events in group A and group B respectively. This statistic is compared with χ^2 statistics to decide whether there is a significant difference between the two groups or not using a specific confidence interval or level of significance.

2.5 Cox regression model

The Cox regression model also known as the Cox proportional hazards model (CPHM) is used to investigate the association between the survival time of patients and one or more predictor variables. CPHM is a regression model that has a dependent variable and independent variables and it is used to know the effect of specific

variables on the event. Its formula is written as shown in the following equation:

$$h(t, X) = h_0(t)e^{(\sum_{i=1}^{p} B_i x_i)} \quad (7)$$

where $h_0(t)$ is the baseline hazard, $X's$ here are time-independent, and B_i are the regression coefficients. It is important to note that Kaplan-Meier curves and log-rank tests work with categorical predictor variable and they can describe the survival according to only one factor under investigation. CPHM model can work for quantitative predictor variables as well as categorical predictor variables and it can assess at the same time the effect of several factors on the survival time of patients.

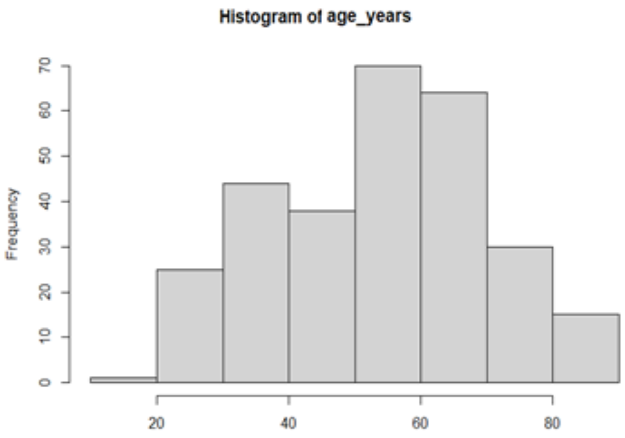


Figure 3: Histogram shows the distribution of age values.

3 Results and discussion

To analyze COVID-19 data, we used survival and survminer functions under R software. In this COVID-19 patients' data, the event of interest is the recovery of the patients and the outcome is time in days until the recovery. We must consider an important analytical problem called censoring, which occurs when we have sick people at risk who have died or the recovery time for them is not known due to losing their follow up, therefore the patients' exact recovery time will not be known at least in the period of the study so the patient survival time is considered censored. In other words, in this study, we will consider censoring if the patient is died or lost follow-up in the determined period given that the period of the study is from Jan 1 to Feb 11, 2020, as shown in the dataset subsection. After extracting the data and preparing it we read it in R software. We consider three predictive variables (patient gender, age, and time to hospitalization). The gender variable is a categorical variable and it can be easily analyzed using Kaplan–Meier survival curve and log-rank since we have two groups male and female. The patient age is a continuous variable, therefore, we need to convert it into categorical to be able to use it as a predictive variable. To do so we need to use a cutoff, where we will consider the age greater than this cutoff as old and the age less than the cutoff as young. To determine the cutoff, we should look at the overall distribution of age values using the histogram shown in Figure 3, where the cutoff of 50 is obviously suggested to be used. Also, we converted time to hospitalization to a categorical variable by considering hospitalization within 6 days as 'Early' and hospitalization in a time greater than 6 days as 'Late'.

Table 1: summary of the Kaplan-Meier estimates for the age groups up to day 16.

time	n. risk	n. event	non-recovery probability	std. err.	Lower. 95% CI	Upper. 95% CI
Old age group						
4	343	1	0.997	0.003	0.991	1.000
5	321	1	0.994	0.004	0.986	1.000
6	302	1	0.991	0.005	0.980	1.000
7	280	4	0.977	0.009	0.959	0.994
9	261	1	0.973	0.010	0.954	0.992
10	250	1	0.969	0.010	0.949	0.989
11	240	5	0.949	0.013	0.923	0.975
13	216	3	0.936	0.015	0.906	0.966
14	205	2	0.926	0.016	0.895	0.959
15	197	4	0.908	0.019	0.872	0.945
16	183	4	0.888	0.021	0.848	0.929
Young age group						
4	270	1	0.996	0.004	0.989	1.000
5	240	1	0.992	0.006	0.981	1.000
6	219	2	0.983	0.008	0.967	1.000
7	207	3	0.969	0.012	0.946	0.992
8	198	2	0.959	0.013	0.933	0.986
9	195	5	0.934	0.017	0.902	0.968
10	185	3	0.919	0.019	0.883	0.957
11	177	4	0.899	0.021	0.858	0.941
12	169	1	0.893	0.022	0.852	0.937
13	167	3	0.877	0.023	0.833	0.924
14	158	1	0.872	0.024	0.826	0.919
15	156	5	0.844	0.026	0.794	0.896
16	148	12	0.775	0.030	0.718	0.837

To analyze the data based on the age group, we created a survival object and we fit the Kaplan-Meier curves by passing the created survival object to survfit function. We obtained the results given in Table 1. Normally, the results obtained from the survfit function are the probability of non-recovery as shown in the 4th column of Table 1 i.e. death or negative event (Table 1 shows the results up to day 16). In this study, we are looking for a positive event (recovery) therefore we can calculate the recovery rate as (1- the probability of non-recovery). The results show that in the old age group over the four days period 1 recovered out of 343, therefore, the probability of non-recovery is (343-1)/343=0.997 (see Table 1 the first row) so the recovery rate is (1- 0.997) =0.003. Over the five days period as shown in the table (see Table 1 the second row), 21 patients of the remaining 342 patients lost follow-up (censored) so the number of remaining patients on the 5th day is 321. One of the remaining patients is recovered in the 5th day therefore, the proportion not recovered is 0.994. We could calculate the survival at a specific time t as the product of the observed survival rates until t i.e $S(t) = p.1 * p.2 * .... * p.t$, where $p.1$ is the rate of the surviving patients who past the first time point and $p.2$ is the rate of the surviving patients who past the second time point, and so forth.

It is very important to take into account that starting from $p.2$ we should consider only those patients who survived past the previous time point to calculate the survival rate for the following time point, in other words, $p.2$, $p.3\ldots$, $p.t$ are survival rates that are conditional on the previous survival rates. Given the assumption of independent and random censoring, we assume that the 21 patients who were censored were similar to the 321 who remain at risk regarding their survival experience. Since 1 of the 321 who remained and survived on the 5th day recovered and we have 1 recovered on the 4th day then the total number of patients who recovered on the course of 5 days is 2. Subtracting 2 from the original number, which is 343 will yield 341. Then the recovery rate in the 5th day will be (1-341/343) = 0.006. The same analysis for the old age group is applied to the young age group (the results up to day 16 out of 30 days are shown in Table 1). The lower 95% confidence interval and upper 95% confidence interval tell us how accurate the estimate of the mean is [33]. In the first row in Table 1 the lower 95%CI and the upper 95%CI show us that we are 95% confident that the interval (0.991, 1.000) contains the true value of the parameter. Also, we can see that this interval is very narrow, which means that the certainty of the results is very high. In other words, we are 95% certain about the results. This narrow interval is associated with a very small standard error (0.003).

The corresponding survival curve can be obtained using the function ggsurvplot on the survival object. The obtained curves (see Figure 4) are step functions that allow us to compare the survival time of two age groups. Typically, the curve starts at 1 representing the fact that all of the patients are not having the event at entry into the study (see Figure 4 A). Over time the curve represents the probability of remaining non-recovered patients. Since we are interested in the probability of the recovered patients, we drew the survival curve starting from 0 to represent the portion of the recovered patients as shown in Figure 4 B. It is clear from Figure 4 B that the survival function of the young age group consistently lies above that for the old age group. This indicates that the young age group is better in recovering from COVID-19 than the old age group. We note that the two functions are somewhat close to each other in the first few days. This indicates that the young age group can survive COVID-19 later after infection than its early one. The estimate of the median recovery time for the young age group can be obtained from Figure 4 by selecting the value in the time axis that corresponds to the survival probability of 0.5. From the figure, it is clear that the median recovery time is greater than 20 days. The p-value of the log-rank is 0.00012, which indicates that the results are significant considering $p < 0.05$ indicates statistical significance, in other words the results show that there is a significant difference between young and old patients regarding the recovery from COVID-19.

To analyze the data based on the gender (Male, Female), we directly created a survival object since we don't need to convert the gender of the patient to a categorical variable (it is already a categorical variable). Then we fit the Kaplan-Meier curves by passing the created survival object to survfit function. We obtained the results given in Table 2 (we showed the results for the first 15 days). The results show that in the Female group over the four days period 1 recovered out of 295 therefore the probability of non-recovery is 294/295=0.997 (see the first row). Then the recovery rate in the 4th day will be (1-0.997) = 0.003. The rest of the Female group results and the male group results can be described as we did with the age group results that are given in Table 1.

Table 2: summary of the Kaplan-Meier estimates for the sex groups up to day 15.

time	n. risk	n. event	non-recovery probability	std. err.	Lower. 95% CI	Upper. 95% CI
Female group						
4	295	1	0.997	0.003	0.990	1.000
5	261	3	0.985	0.007	0.971	1.000
6	234	1	0.981	0.008	0.964	0.998
7	216	4	0.963	0.012	0.939	0.987
8	210	1	0.958	0.013	0.933	0.984
9	202	5	0.934	0.016	0.903	0.967
10	193	3	0.920	0.018	0.885	0.956
11	184	5	0.895	0.021	0.855	0.937
13	169	2	0.884	0.022	0.842	0.928
15	160	1	0.879	0.022	0.836	0.924
Male group						
4	346	2	0.994	0.004	0.986	1.000
5	325	1	0.991	0.005	0.981	1.000
6	310	3	0.982	0.007	0.967	0.996
7	293	4	0.968	0.010	0.949	0.988
8	283	1	0.965	0.010	0.944	0.985
9	275	2	0.958	0.012	0.935	0.981
10	262	1	0.954	0.012	0.931	0.978
11	252	4	0.939	0.014	0.912	0.967
12	240	1	0.935	0.014	0.907	0.964
13	233	4	0.919	0.016	0.888	0.952
14	220	4	0.902	0.018	0.868	0.938
15	210	8	0.868	0.021	0.828	0.910

The corresponding survival curve for the sex is shown in Figure 5, where the step functions allow us to compare the survival time of two sex groups. Figure 5 A is the probability of remaining unrecovered patients based on gender. The survival function of the Female group and that for the male group from the time 0 up to 40 follow similar paths, therefore the p-value (0.63) from the log-rank test is not significant considering $p < 0.05$ indicates statistical significance. Figure 5 B shows the survival curves starting from 0 and they represent the proportion of the recovered patients based on sex.

We used Cox regression model to measure the effect of the different factors on the recovery from COVID-9. in Cox regression the measure of the effect is hazard rate. The hazard is the instantaneous event rate or the probability of a patient at time t has an event at that time. Here the assumption is non-recovery if the event does not occur up to time t [23, 34]. Hazard ratio of 1 means that event rates are the same in the members of the same group. Figure 7 shows that at any time 1.9 as many patients in the young age group are having an event (recovery) proportionally to the old age group, which is taken as a reference, and the value 0.001** shows that this result is statistically significant. The result of the hazard ratios supports the results that we obtained in the step functions that are depicted in Figure 4 B. Regarding the sex group, the results in the figure shows that the hazard ratio is 1 which indicate that three is no difference between male patients and female patients in recovery from

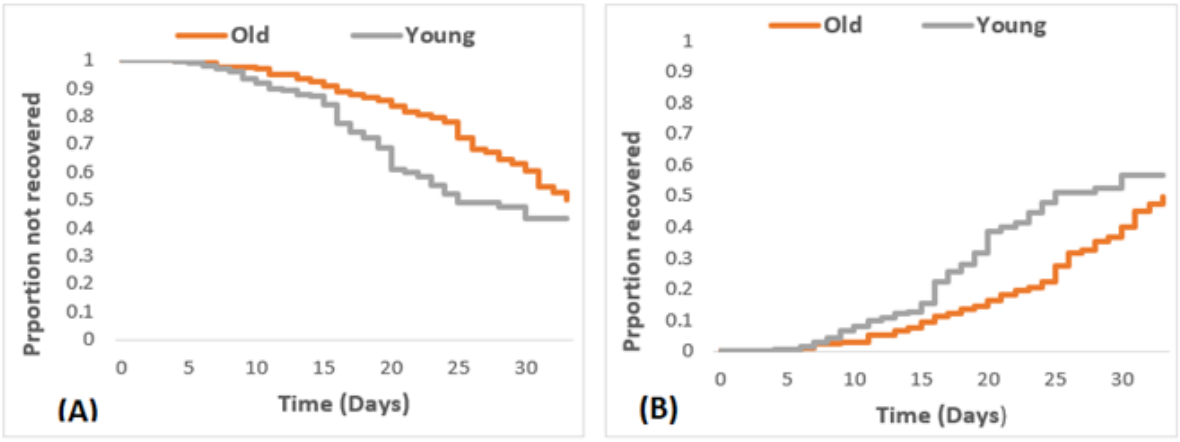


Figure 4: Survival curves for days to recovery from COVID-19 (age groups). A) Proportion not recovered; B) Proportion recovered.

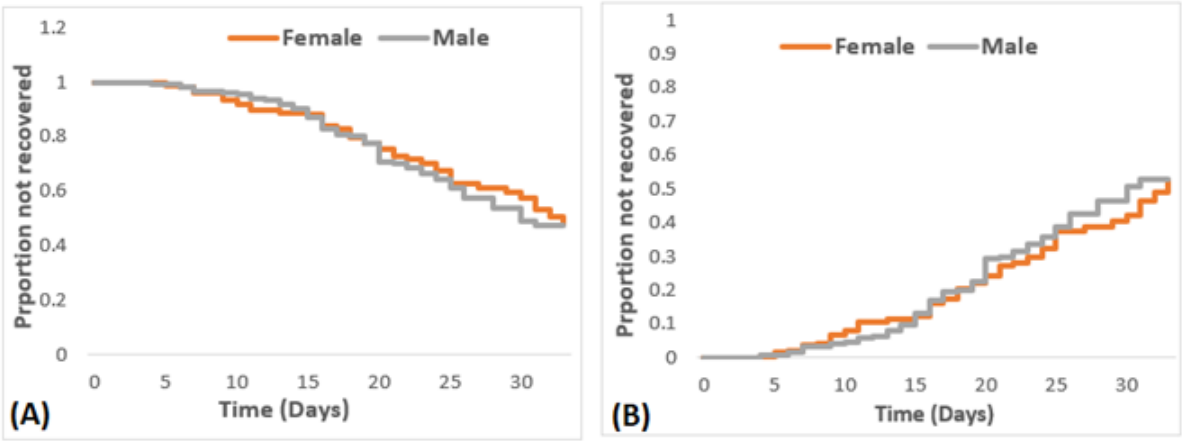


Figure 5: Survival curves for days to recovery from COVID-19 (sex groups). A) Proportion not recovered; B) Proportion recovered.

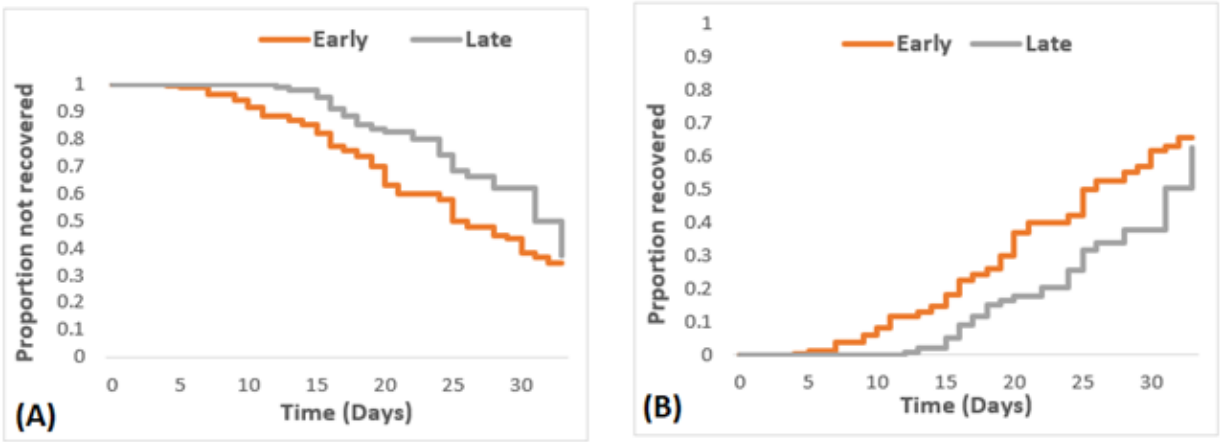


Figure 6: Survival curves for days to recovery from COVID-19 (time to hospitalization groups). A) Proportion not recovered; B) Proportion recovered.

COVID-19. This result supports the results that we obtained in the step functions. We note that the P-value is quite different from what is shown with the Kaplan-Meier estimator and the log-rank test that is because the hazard ratio calculates the hazard ratio and respective risk of death whereas Kaplan-Meier estimator and the log-rank test estimate the survival probability [35]. Therefore, we can see that the results yielded by these different methods are different in terms of significance.

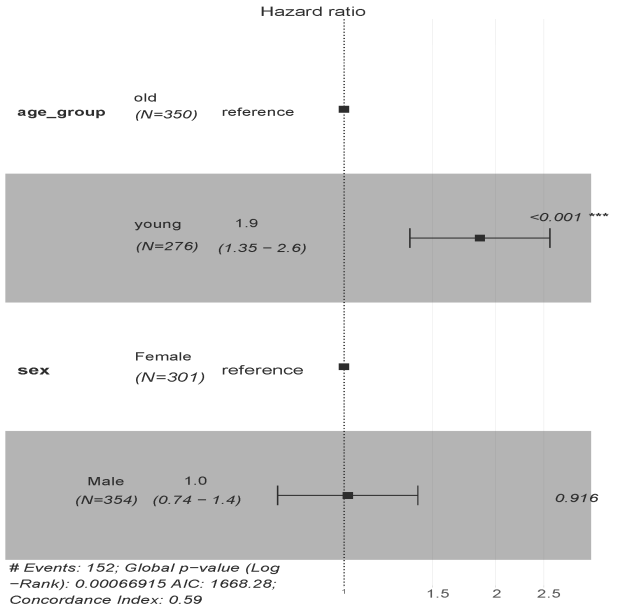


Figure 7: Forest plot shows the hazard ratios of the age and sex groups.

Also, we analyzed the data based on the time to hospitalization (early, late), we considered time to hospitalization as early if the patient is hospitalized within 6 days from catching the disease and as late if is hospitalized in a time greater than 6 days from catching the disease. Since not all the COVID-19 patients are hospitalized, we deleted the cases that have no hospitalization date. We then created a survival object and fit the Kaplan-Meier curves by passing the created survival object to survfit function. We obtained the results given in Table 3, which shows the results for the first 20 days. The results show that in the early time to hospitalization group over the four days period 1 recovered out of 159, therefore, the proportion not recovered rate is 158/159=0.994 (see Table 3 the first row), and therefore the proportion recovered rate is 1-0.994=0.006. In the late hospitalization group, the results show that over 12 days 1 recovered out of 103 so the proportion not recovered is 102/103=0.990 and hence the proportion recovered rate is 1-0.990=0.010. We note that in the early to hospitalization group the recovery starts at day 4, while in the late to hospitalization group the recovery starts at day 12.

The survival curve of the time to hospitalization groups is shown in Figure 6. It is clear from Figure 6 B that the survival function of the early time to hospitalization group consistently lies above that for the late time to hospitalization group. This indicates that the early time to hospitalization group is better recovering from COVID-19 than the late time to hospitalization group. Also, we note that the two functions are somewhat close to each other in the first few days (up to day 4). This indicates that the early time to hospitalization group can survive COVID-19 later after 4 days from infection than its early one. The p-value of the log-rank is 0.0052, which indicates that the results are significant considering $p < 0.05$ indicates statistical significance, in other words, the results show that there is a significant difference between the early time to hospitalization group and late time to hospitalization group.

Table 3: summary of the Kaplan-Meier estimates for the time to hospitalization groups.

time	n. risk	n. event	non-recovery probability	std. err.	Lower. 95% CI	Upper. 95% CI
Early time to hospitalization group						
4	159	1	0.994	0.006	0.981	1.000
5	157	1	0.987	0.009	0.970	1.000
7	147	4	0.961	0.016	0.930	0.992
9	133	3	0.939	0.020	0.901	0.978
10	124	3	0.916	0.023	0.872	0.963
11	118	4	0.885	0.027	0.833	0.940
13	110	2	0.869	0.029	0.814	0.928
14	103	2	0.852	0.031	0.794	0.915
15	100	4	0.818	0.034	0.754	0.887
16	92	5	0.774	0.037	0.704	0.851
17	85	2	0.755	0.039	0.683	0.835
18	81	2	0.737	0.040	0.662	0.819
19	77	4	0.698	0.042	0.620	0.786
20	70	7	0.629	0.046	0.545	0.724
Late time to hospitalization group						
12	103	1	0.990	0.010	0.010	0.972
13	100	1	0.980	0.020	0.014	0.954
15	96	3	0.950	0.050	0.022	0.908
16	89	4	0.907	0.093	0.030	0.851
17	82	2	0.885	0.115	0.033	0.823
18	74	3	0.849	0.151	0.037	0.779
19	68	1	0.837	0.163	0.039	0.764
20	63	1	0.823	0.177	0.040	0.748

Cox regression model for time to hospitalization yielded the hazard ratio, which represents relative that compares the early time to hospitalization group with the late time to hospitalization group as shown in Figure 8. A hazard ratio of 0.54 for the late hospitalization group tells us that patients who sent to hospital late have less opportunity of recovering compared to patients who sent to the hospital early, which served as a reference to calculate the hazard ratio. As shown by the forest plot, the respective 95% confidence interval is (0.35 – 0.84) and this result is significant (p-value=0.006). Using this model, we can see that the time to hospitalization variable significantly influences the patients' recovery from COVID-19. Also, We note that the obtained p-value is quite different from what is shown with the Kaplan-Meier estimator and the log-rank test and that is due to the same justification that we presented when analyzing the sex and age groups.

Salinas-Escudero et al. study [24], which applied Kaplan-Meier and Cox regression models to the Mexican found that the age factor has a significant effect in recovering from COVID-19. This finding agrees with our finding on the data we used. In another hand, their

study found that sex group has significant effects, which disagrees with our finding. Monira Mollazehi et al study [25] applied Weibull model in Singapore. The factors they used are age and nationality. Their finding agrees with ours regarding the age group.

The limitations that need to be declared in this research are: First, the dataset is for a specific region and in a specific period. Second, the dataset is relatively small compared to the total infected cases.

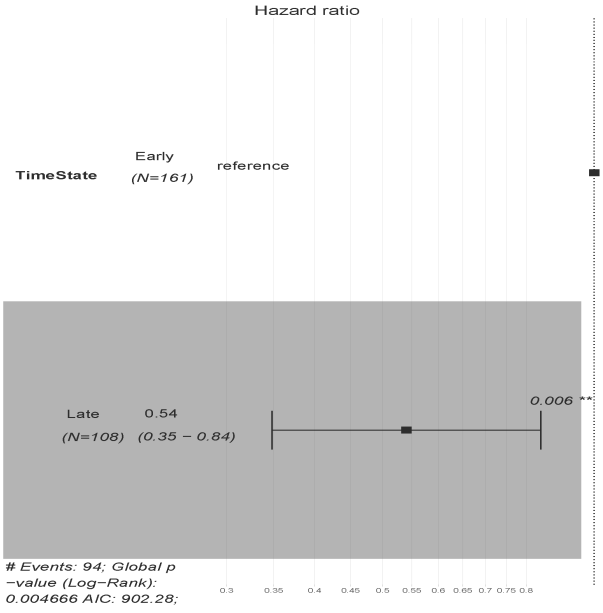


Figure 8: Forest plot shows the hazard ratios of time to hospitalization group.

4 Conclusion

In this work, we used survival analysis to analyze COVID-19 data that we obtained from the clinically diagnosed and confirmed cases where the date of onset is recorded in Wuhan and elsewhere in China from Jan 1 to Feb 11, 2020. We used the Kaplan-Meier method which is an estimator of survival probabilities and the Cox regression model, which is known as the Proportional Hazard Model for the analysis. The event of interest in our analysis is the recovery of the patients from COVID-19 and the outcome is time in days until the recovery. The predictor variables that we used are sex, age, and time to hospitalization. The results show that the young age group is better in recovering from COVID-19 than the old age group with a significant difference (P-value = 0.00012) and at any time 1.9 as many patients in the young age group is having an event (recovery) proportionally to the old age group. The step functions of the sex group show that the female and male groups are somewhat close to each other in recovering from COVID-19 and the p-value =0.63 indicates that there is a non-significant difference in the results between Male and Female considering $p < 0.05$ indicates statistical significance. The results also show that early time to hospitalization group can recover from COVID-19 better than late time to hospitalization group (the p-value of the log-rank is 0.0052)

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors would like to acknowledge the Jouf University for all the support that it provides.

References

[1] M. Cascella, M. Rajnik, A. Cuomo, S. Dulebohn, N. R. Di, "Features, evaluation and treatment coronavirus (COVID-19)," Statpearls [internet]: StatPearls Publishing, 2020.

[2] N. Zhu, D. Zhang, W. Wang, X. Li, Y. et al., "A novel coronavirus from patients with pneumonia in China, 2019," New England Journal of Medicine, 2020, doi:10.1056/NEJMoa2001017.

[3] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. et al., "A new coronavirus associated with human respiratory disease in China," Nature, **579**(7798), 265–269, 2020, doi:10.1038/s41586-020-2008-3.

[4] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, Z. et al., "A pneumonia outbreak associated with a new coronavirus of probable bat origin," nature, **579**(7798), 270–273, 2020, doi:10.1038/s41586-020-2012-7.

[5] V. M. Corman, D. Muth, D. Niemeyer, C. Drosten, "Hosts and sources of endemic human coronaviruses," Advances in virus research, **100**, 163–188, 2018, doi:10.1016/bs.aivir.2018.01.001.

[6] A. E. Gorbalenya, L. Enjuanes, J. Ziebuhr, E. J. Snijder, "Nidovirales: evolving the largest RNA virus genome," Virus research, **117**(1), 17–37, 2006, doi:10.1016/j.virusres.2006.01.017.

[7] M. K. Elbashir, M. Ezz, M. Mohammed, S. S. Saloum, "Lightweight Convolutional Neural Network for Breast Cancer Classification Using RNA-Seq Gene Expression Data," IEEE Access, **7**, 185338–185348, 2019, doi: 10.1109/ACCESS.2019.2960722.

[8] A. Ancarani, C. Di Mauro, L. Fratocchi, G. Orzes, M. Sartor, "Prior to reshoring: A duration analysis of foreign manufacturing ventures," International Journal of Production Economics, **169**, 141–155, 2015, doi:10.1016/j.ijpe.2015.07.031.

[9] E.-Y. Jung, C. Baek, J.-D. Lee, "Product survival analysis for the App Store," Marketing Letters, **23**(4), 929–941, 2012, doi:10.1007/s11002-012-9207-0.

[10] N. E. Buckley, P. Haddock, R. D. M. Simoes, P. et al., "A BRCA1 deficient, NFκB driven immune signal predicts good outcome in triple negative breast cancer," Oncotarget, **7**(15), 19884, 2016, doi:10.18632/oncotarget.7865.

[11] S. R. Gross, B. O'Brien, C. Hu, E. H. Kennedy, "Rate of false conviction of criminal defendants who are sentenced to death," Proceedings of the National Academy of Sciences, **111**(20), 7230–7235, 2014, doi:10.1073/pnas.1306417111.

[12] M. A. Alvi, D. G. McArt, P. Kelly, M.-A. Fuchs, A. et al., "Comprehensive molecular pathology analysis of small bowel adenocarcinoma reveals novel targets with clinical utility," 2015, doi:10.18632/oncotarget.4576.

[13] D. Murray, A. Carr, C. Bulstrode, "Survival analysis of joint replacements," The Journal of bone and joint surgery. British volume, **75**(5), 697–704, 1993, doi:10.1302/0301-620X.75B5.8376423.

[14] T. Sørlie, C. M. Perou, R. Tibshirani, A. et al., "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications," Proceedings of the National Academy of Sciences, **98**(19), 10869–10874, 2001, doi:10.1073/pnas.191367098.

[15] F. Emmert-Streib, M. Dehmer, Medical biostatistics for complex diseases, John Wiley & Sons, 2010.

[16] W. Zhang, T. Ota, V. Shridhar, J. Chien, B. Wu, R. Kuang, "Network-based survival analysis reveals subnetwork signatures for predicting outcomes of ovarian cancer treatment," PLOS Computational Biology, **9**(3), e1002975, 2013, doi:10.1371/journal.pcbi.1002975.

[17] M. J. Campbell, T. D. V. Swinscow, Statistics at square one, John Wiley & Sons, 2011.

[18] M. J. Hancock, C. G. Maher, L. d. C. M. Costa, C. M. Williams, "A guide to survival analysis for manual therapy clinicians and researchers," Manual therapy, **19**(6), 511–516, 2014, doi:10.1016/j.math.2013.08.007.

[19] S. J. Staffa, D. Zurakowski, "Competing risks analysis of time-to-event data for cardiovascular surgeons," The Journal of thoracic and cardiovascular surgery, **159**(6), 2459–2466, 2020, doi:10.1016/j.jtcvs.2019.10.153.

[20] J. T. Rich, J. G. Neely, R. C. Paniello, C. C. Voelker, B. Nussenbaum, E. W. Wang, "A practical guide to understanding Kaplan-Meier curves," Otolaryngology—Head and Neck Surgery, **143**(3), 331–336, 2010, doi: 10.1016/j.otohns.2010.05.007.

[21] G. David, K. Kleinbaum, Survival analysis: a self-learning text, Springer-Verlag New York, 2016.

[22] E. L. Kaplan, P. Meier, "Nonparametric estimation from incomplete observations," Journal of the American statistical association, **53**(282), 457–481, 1958, doi:10.2307/2281868.

[23] D. R. Cox, "Regression models and life-tables," Journal of the Royal Statistical Society: Series B (Methodological), **34**(2), 187–202, 1972.

[24] G. Salinas-Escudero, M. F. Carrillo-Vega, V. Granados-García, S. Martínez-Valverde, F. Toledano-Toledano, J. Garduño-Espinosa, "A survival analysis of COVID-19 in the Mexican population," BMC public health, **20**(1), 1–8, 2020, doi:10.1186/s12889-020-09721-2.

[25] M. Mollazehi, M. Mollazehi, A.-S. G. Abdel-Salam, "Modeling Survival Time to Recovery from COVID-19: A Case Study on Singapore," 2020, doi: 10.21203/rs.3.rs-18600/v1.

[26] Q. Wang, S. Xie, Y. Wang, D. Zeng, "Survival-Convolution Models for Predicting COVID-19 Cases and Assessing Effects of Mitigation Strategies," medRxiv, 2020, doi:10.1101/2020.04.16.20067306.

[27] N. Barda, D. Riesel, A. Akriv, J. Levy, U. Finkel, G. Yona, D. Greenfeld, S. Sheiba, J. Somer, E. Bachmat, et al., "Developing a COVID-19 mortality risk prediction model when individual-level data are not available," Nature communications, **11**(1), 1–9, 2020, doi:10.1038/s41467-020-18297-9.

[28] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, W. et al., "Estimates of the severity of coronavirus disease 2019: a model-based analysis," The Lancet infectious diseases, 2020, doi:10.1016/S1473-3099(20)30243-7.

[29] M. Pampaka, G. Hutcheson, J. Williams, "Handling missing data: analysis of a challenging data set using multiple imputation," International Journal of Research & Method in Education, **39**(1), 19–37, 2016, doi:10.1080/1743727X.2014.979146.

[30] R. J. Little, R. D'Agostino, M. L. Cohen, K. Dickersin, S. S. Emerson, J. T. Farrar, C. Frangakis, J. W. Hogan, G. Molenberghs, S. A. Murphy, et al., "The prevention and treatment of missing data in clinical trials," New England Journal of Medicine, **367**(14), 1355–1360, 2012, doi:10.1056/NEJMsr1203730.

[31] J. W. Graham, P. E. Cumsille, A. E. Shevock, "Methods for handling missing data," Handbook of Psychology, Second Edition, **2**, 2012, doi: 10.1002/9781118133880.hop202004.

[32] D. G. Kleinbaum, M. Klein, Survival analysis, Springer, 2010.

[33] J. H. McDonald, Handbook of biological statistics, volume 2, sparky house publishing Baltimore, MD, 2009.

[34] D. W. Hosmer Jr, S. Lemeshow, S. May, Applied survival analysis: regression modeling of time-to-event data, volume 618, John Wiley & Sons, 2011.

[35] N. H. Ng'andu, "An empirical comparison of statistical tests for assessing the proportional hazards assumption of Cox's model," Statistics in medicine, **16**(6), 611–626, 1997, doi:10.1002/(sici)1097-0258(19970330)16:6⟨611::aid-sim437⟩3.0.co;2-t.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1618-1633 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Elasticity Based Med-Cloud Recommendation System for Diabetic Prediction in Cloud Computing Environment

Karamath Ateeq, Manas Ranjan Pradhan*, Beenu Mago

School of Information Technology, Skyline University College, Sharjah, UAE

ARTICLE INFO

Article history:
Received: 14 August, 2020
Accepted: 05 November, 2020
Online: 25 December, 2020

Keywords:
Diabetes
Data Mining
Cloud Computing
Medical Cloud
Prediction
Fuzzy Clustering
Neural Network
Genetic Algorithm
Recommendation System

ABSTRACT

Day to day huge medical data have been accumulating for diabetic diseases. The complexity of storing, processing ,analyzing and predicting the data related to diabetics is not so easy for healthcare professionals .The prediction of accurate results also has the limitation due to scale of data increasing worldwide for patients, symptoms and test results .In this paper ,it has been tried to considered the diabetic related data storage on cloud and adopt integrated computational algorithms of datamining for better prediction system to various diabetic types(Type 1, Type 2 and Gestational).Though many computational prediction model and recommended system have been proposed by many researchers ,the proposed model has the novelty of considering the elasticity in data analysis due to frequent data changes of patients due to diabetic test time to time. In this work, Elasticity based Med-Cloud Recommendation System (EMCRS) is proposed for predicting the diabetic disease types and providing recommendations for the patients diagnosed with diabetes. Moreover, elastic resource allocation mechanism is proposed to provide cloud resources an on-demand basis to EMCRS.Various computational algorithms have been used for different proposed to make EMCRS to predict results as compared other existing system. The Adaptively Toggle Genetic Algorithm (ATGA) is applied for elastic resource allocation while increase in the number of data sets. ATGA has taken toggle genetic algorithm that shifts between Roulette Wheel Selection Operator. Hybrid Classification and Clustering Algorithm (HC2A) is used for classifying and clustering the diseased patients as Type 1, Type 2 and Gestational Diabetic patients. Fuzzy C Means clustering based attribute weighting (FCMAW) was used for classifying the diabetic data set. The accuracy of the system tested on Pima Indian Diabetic Dataset (PID) and US Diabetic Dataset (USD) from UCI website which is approximately 98% classification accuracy.

1. Introduction

Diabetes is a chronic disease that begins with the failure of pancreas. The pancreas fails to produce sufficient insulin required by the body [1]. The internal changes prompt to an increased concentration of glucose in the blood. It is a condition of high blood glucose level in diabetic patients. It can cause either Type I or Type II diabetes. Type I is known as insulin dependent diabetes, which occurs when there is lack of insulin production. Type II diabetes is non-insulin dependent which is caused by the ineffective use of insulin by human body. This will result in excess body weight and physical inactivity. An earlier prediction or recommendation system is needed to save the patients from the risk of diabetes. In such a condition, Data Mining is suggested and found to be a better diagnostic tool which can be used by the medical practioners too.

Data Mining is the process of selecting, exploring and modeling large amounts of data [2], [3]. This process has become an increasingly pervasive activity in all areas of medical science research. Data mining has resulted in the discovery of useful hidden patterns from massive databases. Consequently, data mining tools are now being used for clinical data. The bottle neck in data analysis is now raising the most appropriate clinical questions and using proper data and analysis techniques to obtain

*Corresponding Author: Manas Ranjan Pradhan, School of IT, Skyline University College, Sharjah, UAE, manas.pradhan@skylineuniversity.ac.ae

clinically relevant answers [4]. But as the accumulated data is increasing abundantly, Data Mining algorithm alone is not enough to retrieve hidden pattern from the data.

In order to improve the existing algorithm performance, the computing and storage resources are insufficient in traditional data mining environment. In order to overcome the resource challenges, now day's resources are utilized on-demand elastic manner with the development of Cloud Computing [5]. National Institute of Standards and Technology (NIST) defines [6] "Cloud Computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (eg., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".

The main features of cloud computing are on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service. Cloud computing has three service models such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). It also has four deployment models such as public cloud, private cloud, hybrid cloud and community cloud. Elasticity is used to utilizing the cloud computing resources (Storage, Virtual machines, Servers, Platforms, network) in an elastic way based on the workload requirement. Elasticity manages the ability to increase or decrease the cloud system resources. Elasticity is defined by NIST [7], "Rapid elasticity: Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time."

Elastic resource allocation is completed with the help of proper workload prediction and scheduling which are used to improve the performance of the cloud system. There are no existing algorithms that focus on workload prediction with elasticity in cloud computing environment [8]-[10].

In this research work, an Elastic Medical Cloud Recommendation System is proposed for diabetic classification, where Adaptively Toggle Genetic Algorithm is applied for the Elastic Resource Allocation. HC2A is involved to classify the diabetes patients and in case of diseased patients, they are clustered into three types and the recommendation system is suggested based on the cluster.

The remaining work of the paper is formed as follows. Related works are presented in section 2. In section 3, the proposed algorithm is explained. Results and discussions are presented in section 4. Section 5 shows the conclusion and future directions of the work.

2. Related Works

One of the important feature in cloud computing is elasticity which provides resources in an on-demand elastic manner for cloud users and providers. In [11]the authors concentrated energy consumption in cloud computing resources with the support of elasticity. In [12] the authors designed a CBIHCS (Cloud Based Intelligent Health Care Service) for supervising user health data for diabetic detection and proposed simple heuristic for dynamic resource elasticity.

Resource scaling methods are explained by in [13] based on the queuing models. It contains database level, application level and storage level. Each levels were analyzed with the parameters such as throughput, resource utilization rate and response time. Based on the results to scale applications in cloud environment, they recommended the elastic cloud resource allocation algorithm. In [14] authors proposed an improvised genetic approach for effective cloud resource allocation by maintaining vertical elasticity in cloud IaaS environment. They concentrated on vertical elasticity which focuses virtual machines allocation on cloud servers based on the workloads to improve resource allocation. The Enhanced Genetic algorithm concentrated only on cloud IaaS environment.

In [15], the scaling techniques of elastic resource management were classified as reactive or predictive imminent. Based on the workload requirements the system react without pre-planning in reactive imminent. In predictive imminent, the workload predicts the requirement of cloud servers that are allocated to handle the workload. In [16] it is proposed a cloud based framework with minimum resource setting for monitoring and diagnosis the PD (Parkinson's Disease). The framework used cloud database, voice data and FBANN (Feedforward Backpropagation based Artificial Neural Network) classifier in cloud platform.

The role played by evolutionary algorithms and Clustering algorithms in Medical Data Mining is incredible. K means algorithm was used in [17] to remove noisy data and Genetic Algorithm was applied to find the optimal set of features. Finally Support Vector Machine (SVM) was used for classifying the diabetic patients. The proposed method has obtained an accuracy of about 96.71%. Fuzzy C Means Algorithm (FCM) [18] was used to determine the small clusters in Diabetic data set and Outlier detection method was applied for classification. The accuracy was 93%. Attribute Weighting method [19] called Fuzzy C Means clustering based attribute weighting (FCMAW) was used for classifying the diabetic data set. This method reduced the variance within the attributes and improves the classification accuracy. It was done by transforming the non-linear separable datasets to linearly separable datasets. FCM was applied to find the center of the attributes and the dataset is weighted according to the ratios of the means of attributes to centers of theirs attributes. Then SVM and KNN classifiers were applied to classify the dataset. The accuracy of the method was 84.38%.

3. Proposed Work

3.1. The system design

In order to get maximum benefit from cloud computing, Cloud providers designed the architecture and deployment model. Various cloud architectures are designed by different cloud service providers. The conceptual diagram of cloud architecture is shown in Figure 1. Cloud service providers provide services to cloud users through internet and datacenter. Cloud servers contains memory utilization vector, processing unit and storage. Virtualization in cloud is to run multiple VMs on a single server and sharing these cloud resources among multiple cloud users. Based on the workloads, the VMs are allocated to process the data as per the command of elastic resource manager.

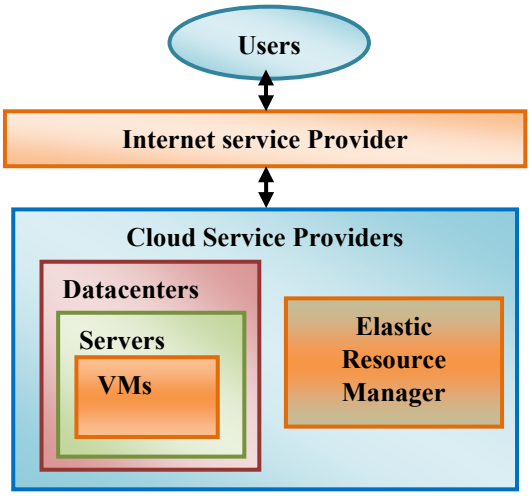


Figure 1: Conceptual diagram of Cloud Computing Architecture

In this work, the task of the ERM (Elastic Resource Manager) has been concentrated with the goal of elastic resource allocation for improving the resource utilization. ERM is responsible for VMs scheduling, Monitoring VMs and Workloads. Figure 2 explains the proposed framework of EMCRS and Figure 3 explains the flows of the proposed EMCRS for diabetic disease prediction.

3.2. The Elastic Resource Manager

The Elastic Load Manager task is to generate the schedule for the allotment of VM to the available cloud servers with the objective to improve the resource utilization. VM handler is responsible for handling VM queue which contains the user requests information. The scheduling algorithm brings the user request from VM queue and assigns it to the scheduler.

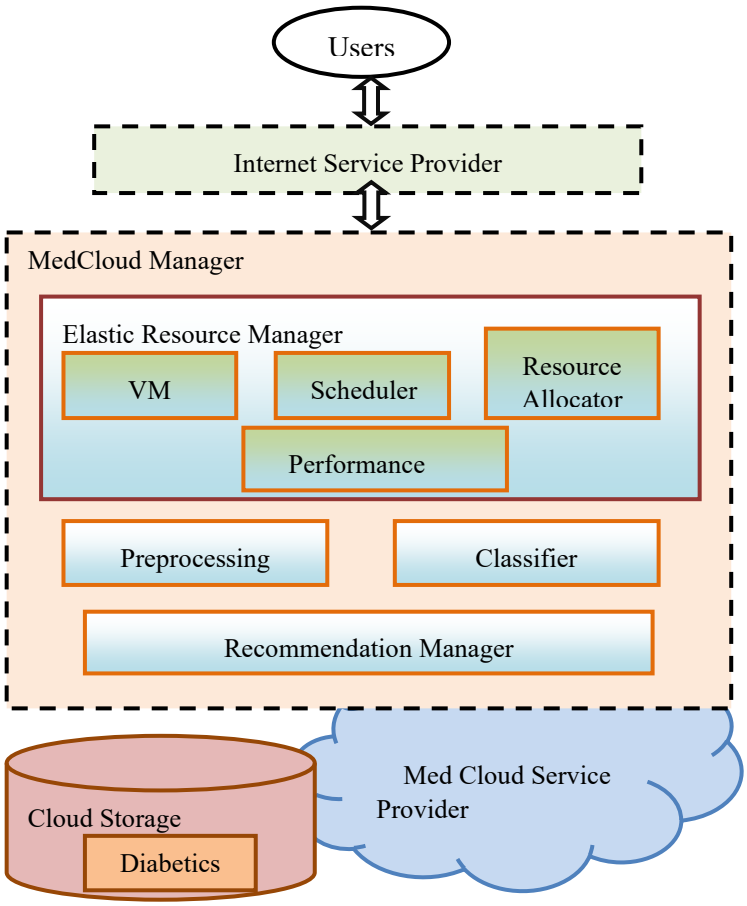


Figure 2: Framework for Elasticity based Med-Cloud Recommendation System

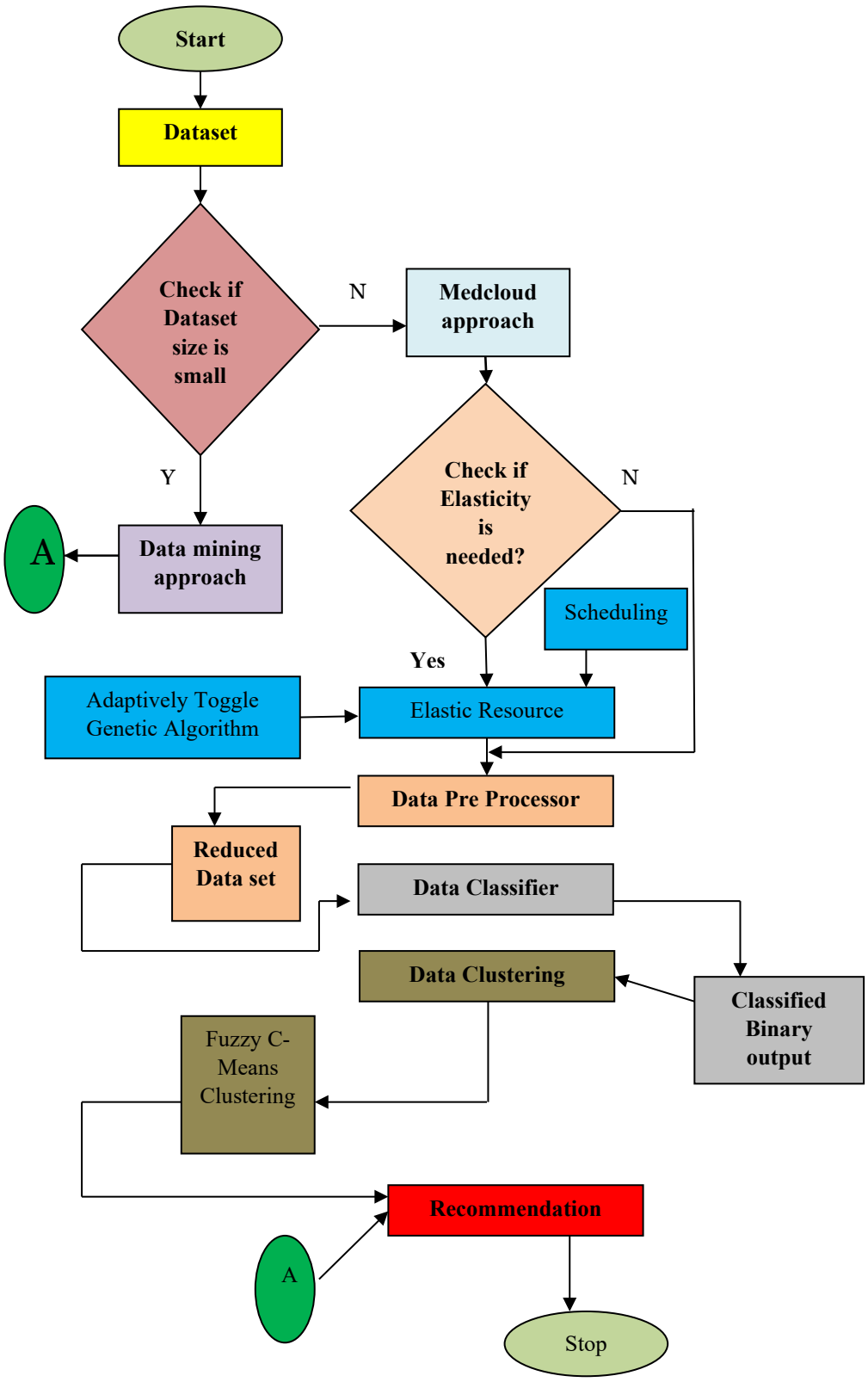


Figure 3: The EMCRS for predicting Diabetes

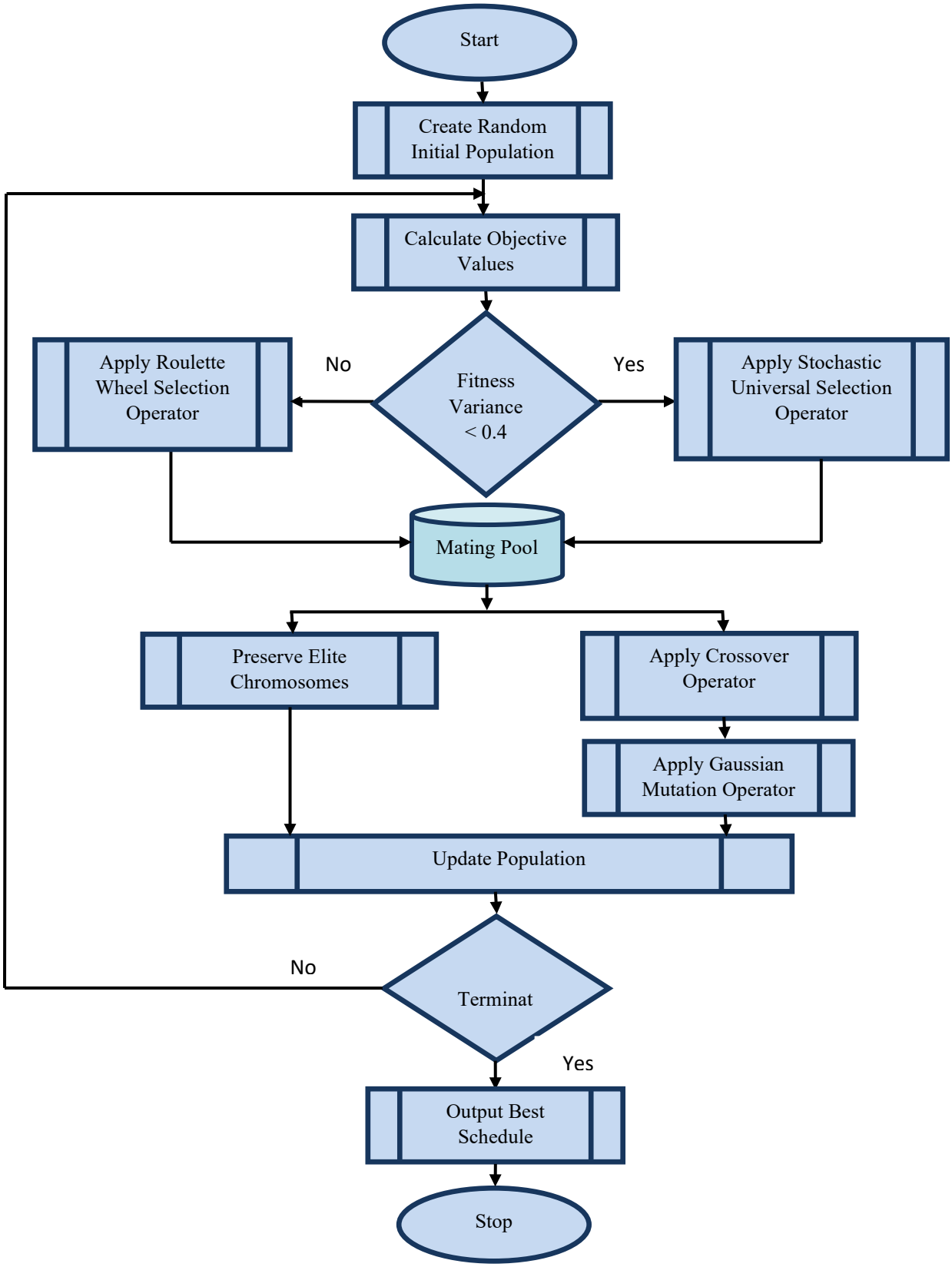


Figure 4: The flow of Adoptively Toggle Genetic Algorithm

3.3. Pseudocode of ATGA (Adaptively Toggle Genetic Algorithm)

Step 1: Begin

Step 2: Generate initial population with randomly generated chromosomes

Step 3: Calculate the objective Values for each chromosome in the current population

Step 4: Observe the fitness distribution of current population based on the fitness variance

Step 5: If the fitness variance is below 0.4

Step 6: Apply the Stochastic Universal Sampling Selection operator

Step 7: Else

Step 8: Apply the Roulette Wheel Selection operator

Step 9: End If

Step 10: Store the parent chromosomes in mating pool

Step 11: Store the best chromosomes from the current population as elite chromosomes

Step 12: Apply the crossover operator for generating offspring chromosomes

Step 13: Apply the Gaussian mutation operator to adaptively mutate genes

Step 14: Update the current population with the offspring chromosomes

Step 15: If termination condition is not reached

Step 16: Go to Step 3

Step 17: Else

Step 18: Output Best Chromosome

Step 19: End If

Step 20: End

3.3.1. Flow chart Adaptively Toggle Genetic Algorithm

The flow chart of adaptively toggle genetic algorithm is presented in figure 4.

3.3.2. Working Principle of Adaptively Toggle Genetic Algorithm

Even though it is a successful tool, many phenomena can affect the performance of Genetic Algorithm and such an important factor is the parameter settings of Genetic Algorithm. Most of the existing research works related to Genetic Algorithm were carried out with static parameter settings. But, inherently the Genetic Algorithm supports self-learning, parallelism and dynamism. Hence, this research work got the insight to improve the Genetic Algorithm by utilizing the adaptive nature of genetic operators. The core idea behind the development of the proposed Adaptively Toggle Genetic (ATG) algorithm is that to intelligently manipulate the selector operator and to adaptively control the mutation operator in order to enhance the performance of the algorithm. So, the two key operations of the proposed ATG algorithm are as follows.

- Toggle the selection operator
- Adaptively change the mutation probability

3.3.2.1. Toggle the selection operator

As selection operator is one of the crucial operators for the convergence of the Genetic Algorithm, the selector operator must be careful chosen and designed in order to make the best individuals to be survived in the successive generations. Various selection operators are already available, but the before applying the selection operators, the characteristics and suitability of them must be analyzed. In this research work, two selection operators are taken namely, roulette wheel selection and stochastic universal sampling selection operators. Though roulette wheel selection operator is the most commonly used selection operator in Genetic Algorithm, it fails when the fitness values of chromosomes in the population are highly scattered. This is due to the state that the high portion of the roulette wheel is assigned to the fittest chromosome where all the other chromosomes only obtain least sectors of the wheel. Therefore, the roulette wheel may provide biased output and it may lead to the premature convergence state. Hence, the proposed ATG algorithm calculates the fitness variance of the population at first and then it determines which selection operator to use at the current iteration of the algorithm. If the fitness variance is high the stochastic universal sampling selection is used, otherwise the roulette wheel selection is used. In this way, the proposed ATG algorithm intelligently toggles between the two selection operators and helps to conquer the premature convergence problem.

3.3.2.2. Adaptively change the mutation probability

The main task of the mutation operator is to provide exploitation in the search space. Further, it can also help to promote the gene diversity in the gene population. But, when the value of mutation probability is set as a constant, it may disturb the solutions in the global convergence state. Thus, the mutation probability value must be changed according to the convergence state of the algorithm. Hence, the mutation probability value in the proposed ATG algorithm is calculated based on the Gaussian curve as a function of the current iteration. In this way, the proposed ATG algorithm can escape from the local optima by adaptively changing the mutation probability value.

Therefore, the two frequent problems associated with Genetic Algorithm namely the local optimum and premature convergence are eliminated by the proposed ATG algorithm through the dynamic switching between selection operators and adaptive control of mutation operator with additional benefits. With help of ATG algorithm, the scheduler optimally allocates the virtual machines for processing the user requests. The user queries are processed for performing the classification task using the proposed HC2A algorithm.

3.4. Proposed Algorithm for Classification

The HC2A is proposed to classify the diabetic patient data set. The Framework works in three phases as Pre-Processing Phase, Classification Phase and Clustering Phase. Figure 5 explains the

flow of the proposed HC2A, In the Pre-Processing Phase, the Genetic-Relative Reduct (G-RR) Algorithm is applied to remove the noisy data and eliminate the irrelevant data. The output at this phase will be the reduced data set. The attributes are classified as Conditional and Decision Attributes. At the initial state the variable R is initialized to null value. And the variable γ_{best} is assigned to 0. The γ_{best} value is stored temporarily in another variable γ_{tmp} and R is stored in T.

The consistency of the data set is checked after removing every attributes. If the decision table is consistent, the attributes is removed and the reduced data set is stored. If the classification accuracy of the Conditional Attributes obtained is greater than the classification accuracy of the Decision attributes ($\gamma_{RU(X)}$ (C)>γ_X(D)), then the first Generation of Offspring is constructed. The obtained attributes are selected and then mutation and crossover operations taken place.

Input: Diabetic Data Set

Step 1 Start

Step 2 Load Diabetic dataset

Step 3 Initialize the parameters (R,))

Step 4 Call Genetic – Relative Reduct Algorithm

Step 5 Check for Reduced Data set

Step 6 **If** *the data set is reduced*

Go to Step 7

else

Go to step 3

Step 7 Initialize the attributes

Step 8 Apply Modified Particle Swarm Optimization – Neural Network Algorithm

Step 9 Check for maximum number of epoch

Step 10 Calculate error and accuracy

Step 11 **If** *the error is minimized*

Go to Step 12

else

Go to step 7

Step 12 Apply Fuzzy C Means Clustering Algorithm

Step 13 Check if the output is clustered and ready to suggest the recommendation system

Step 14 **If** *the clustered output is ready*

Go to step 15

Else

Go to step 12

Step 15 Stop

Output: Clustered output for the recommendation system

The obtained Reduct set is stored in T. The Decision Attributes which are best are stored in the variable γ_{best} . The obtained reduced set is stored in R. The process goes on repeating until the optimal data set is obtained. The obtained output is Reduct Set R. After obtaining the optimal data set, the algorithm terminates.

Step 1: Start

Step 2: R= ϕ

Step 3: γ_{best} = 0

Step 4: **do**

Step 5: $\gamma_{tmp} = \gamma_{best}$

Step 6: T=R

Step 7: **for** x in (C-R)

Step 8: ***If*** $\gamma_{RU(X)}$ (C)>γ_X(D)

Step 9: Construct the First Generation

Step 10: Selection

Step 11: Crossover

Step 12: Mutation

Step 13: T=RU{X}

Step 14: $\gamma_{best} = \gamma_c$(D)

Step 15: R=T

Step 16: *Until* $\gamma_{best} == \gamma_{tmp}$

Step 17: Return R

Step 18: **end**

The next phase is classification phase. In this phase, Modified Particle Swarm Optimization – Neural Network (MPSO-NN) Algorithm is employed to classify the dataset. This phase optimizes the reduced data set and classifies it.

Step 1: Initialize the population

Step 2: Evaluate the fitness of the attribute

Step 3: For each attribute, find the maximum fitness and compare it with the best found so far

Step 4: $Pbest_i$ is equal to the location of maximum fitness function

Step 5: Compare Fitness evaluation with population overall Pbest

Step 6: **If** particle best is greater than gbest, then reset gbest(i) is equal to the current Pbest's array index and value

Step 7: Calculate Convergence factor

Step 8: Calculate Inertia weight

Step 9: Update velocity and position and new population is generated

Step 10: Adjust acceleration

Step 11: **If** the data set is optimized

Go to Step 12

else Go to Step 2.

Step 12: Choose the initial weight

Step 13: **If** the error is minimum

Go to step 21

else

Go to Step 14

Step 14: Apply the optimized dataset to the network

Step 15:Calculate output for every neuron through hidden layer(s) to output layer

Step 16: Calculate Error value at the output layer

Step 17: Update the weight and bias at the output layer

Step 18: Calculate the Error value at the hidden layer

Step 19: Update the weight and bias at the hidden layer

Step 20: Check if the maximum number of epochs reaches

If yes

Go to Step 21

else

Go to Step 14

Step 21 Evaluate the network performance

Step 22: Classified output

At the preliminary stage, the population is initialized and the fitness of the population is calculated. The maximum fitness of each particles is compared with the best found so far as given in (1).

$$p_{best,i} = \begin{cases} P_{best,i} \; if \; f(x_i) > P_{best,i} \\ x_i \; if \; f(x_i) \le P_{best,i} \end{cases} \quad (1)$$

The P_{besti} obtained will be equal to the location of the maximum fitness function. The fitness evaluation is then compared with the population's overall best as in (2).

$$G_{best} = \{\min \{ P_{best,i} \} \quad (2)$$

$$i \in [1, ..., n] \text{ where } n > 1$$

The convergence factor is calculated as

$$\lambda = \frac{2}{\left\| 2 - C - \sqrt{C^2 - 4C} \right\|} \quad (3)$$

where C=C1-C2

C1 is the cognitive learning parameter and C2 is the social collaboration parameter. The C1 and C2 always lie between 0 and 2. Then the inertia value is calculated as given in (4). The inertia value provides the balance between the exploration and exploitation. Generally, the inertia value lies between 0.8 and 1.2. Lower value of inertia component speeds up the convergence of swarm to optima. Higher values encourage the exploration of entire search space.

$$\omega_{id} = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{iter_{\max}} \quad (4)$$

The velocity and position of the particle is updated using the (5).

$$V_{id} = \lambda(\omega_{id} V_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \quad (5)$$
$$x_{id} = x_{id} + V_{id}$$

V_{id} is the momentum of the particle and r1, r2 is random numbers (0, 1). The acceleration of the particle is adjusted as

$$V_{id} = \begin{cases} V_{\max} \; if \; V_i > V_{\max} \\ -V_{\max} \; if \; V_i < V_{\max} \end{cases} \quad (6)$$

After the reduced data set is optimized, then the optimized data set is fed as input to Multi-Layer Perceptron Network. Back propagation learning is used to train the network. The output for very neuron is calculated in the hidden unit and output unit. It is calculated as

$$Z_{inj} = V_{oj} + \sum_{i=1}^{n} x_i v_{ij}$$
$$Z_j = f(Z_{inj}) \quad (7)$$

$$where \; f(Z_{inj}) = \frac{1}{1 + e^{-\lambda Zinj}}$$

The sigmoid activation function is applied in every layer. The number of neurons in the hidden layer is fixed by

$$N_h = \sqrt{N_i N_o} \quad (8)$$

where N_h is the number of hidden neurons in the hidden layer. N_i and N_o represent the number of neurons in the input layer and output layer. The error at the output layer is calculated using (9) and the weight and bias are updated using (10).

$$\delta_k = (t_k - y_k) f'(y_{ink}) where \; f'(Y_{ink}) = \lambda f(Y_{ink})[-f(Y_{ink})] \quad (9)$$

$$\Delta W_{jk} = \alpha \delta_k z_j \; \Delta w_{ok} = \alpha \delta_k \tag{10}$$

If the maximum number of epochs is reached and the error is minimized the training process is stopped, and the classified output and results are taken. The final output from this framework will be classified as binary output 0 or 1.

In Clustering Phase, the data with value 0, that is the non-diabetic patients are ignored and the diabetic patients with binary value 1 is again clustered using FCM to cluster the diabetes patients into three types as Type 1, Type 2 and Gestational Diabetes. Fuzzy C-Means (FCM) is a clustering method that allows each data point to belong to multiple clusters with varying degrees of membership. The objective function to minimize the FCM is:

$$J_m = \sum_{i=1}^{D} \sum_{j=1}^{N} \mu_{ij}^{m} \left\| x_i - c_j \right\|^2 \tag{11}$$

where

- D is the number of data points.
- N is the number of clusters.
- m is fuzzy partition matrix exponent for controlling the degree of fuzzy overlap, with $m > 1$. Fuzzy overlap refers to how fuzzy the boundaries between clusters are, that is the number of data points that have significant membership in more than one cluster.
- x_i is the ith data point.
- c_j is the center of the jth cluster.
- μ_{ij} is the degree of membership of x_i in the jth cluster. For a given data point, x_i, the sum of the membership values for all clusters is one.

FCM performs the following steps during clustering:

Step 1 Randomly initialize the cluster membership values, μij.

Step 2 Calculate the cluster centers:

$$c_j = \frac{\sum_{i=1}^{D} \mu_{ij}^{m} x_i}{\sum_{i=1}^{D} \mu_{ij}^{m}} \tag{12}$$

Step 3 Update μij according to the following:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{N} \left(\frac{\left\| x_j - c_j \right\|}{\left\| x_k - c_k \right\|} \right)^{\frac{2}{m-1}}} \tag{13}$$

Step 4 Calculate the objective function, Jm.

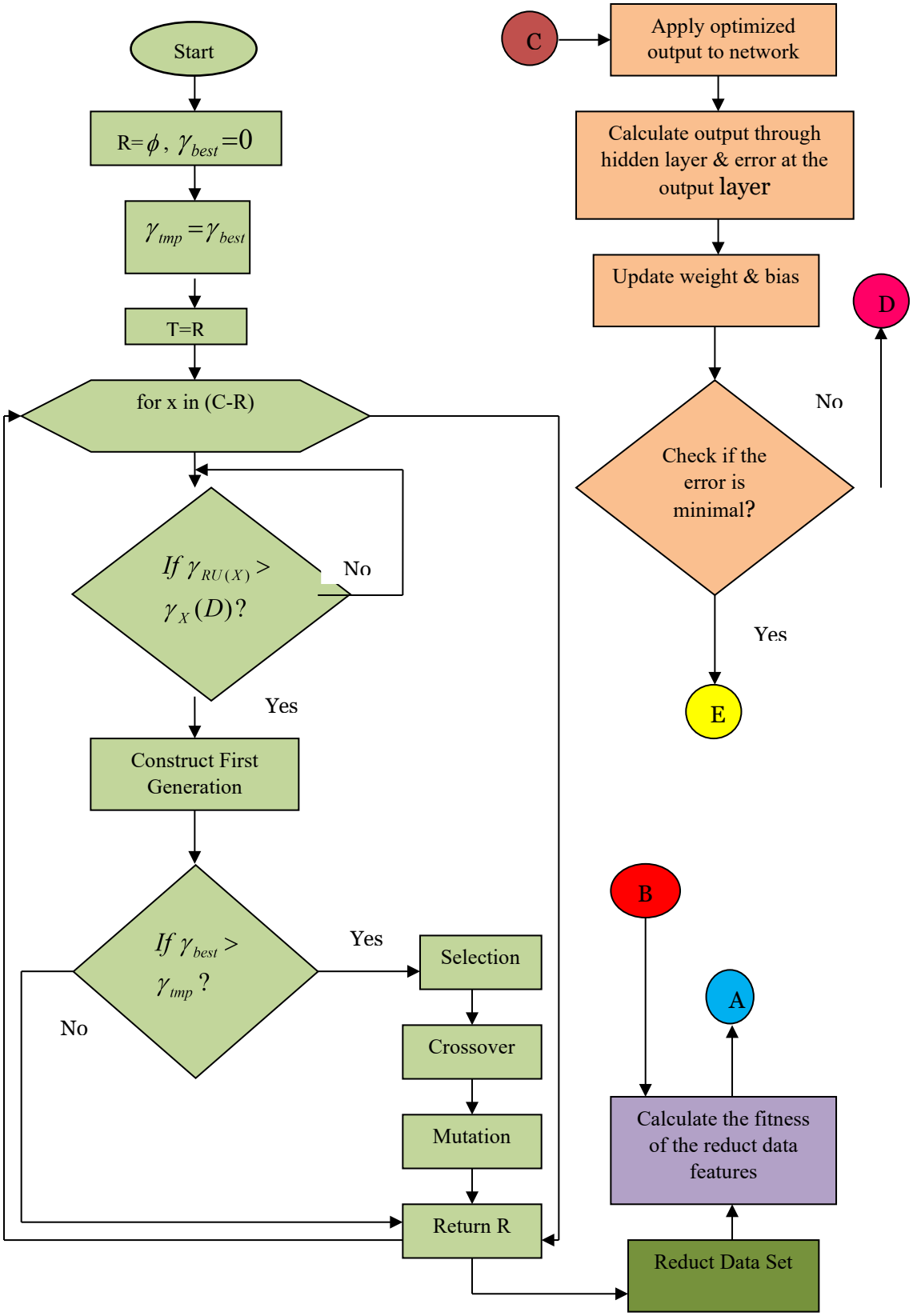
Step 5 Repeat steps 2–4 until Jm improves by less than a specified minimum threshold or until after a specified maximum number of iterations.

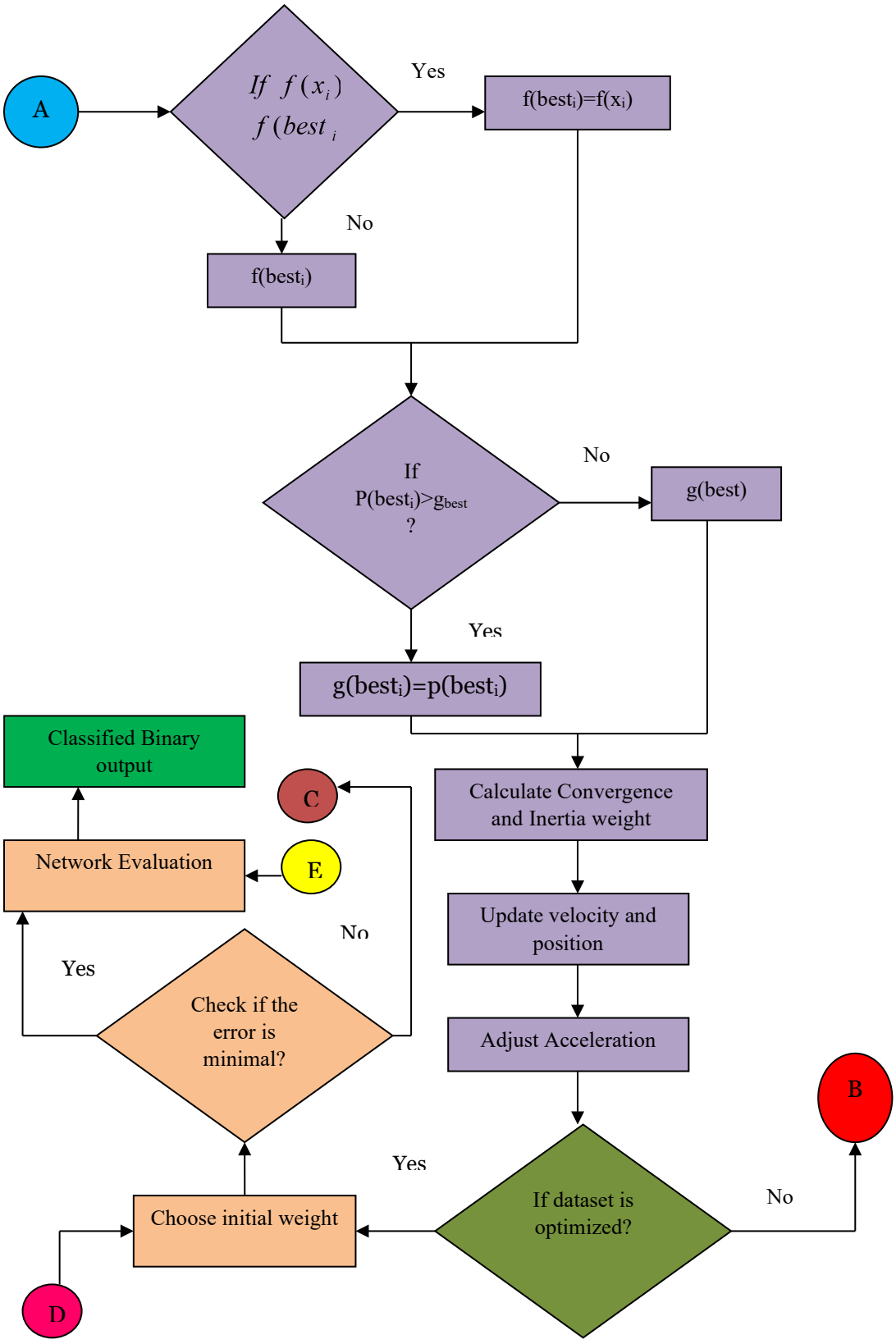
Table 1: Confusion matrix for MPSO-ANN – PID

	0	1
0	558 72.66	80 10.42
1	60 7.81	70 9.11

Table2: Confusion matrix for GA-RR+MPSO-ANN- PID

	0	1
0	612 79.69	40 5.99
1	50 6.51	60 7.81


Start
R=ϕ , γ_{best}=0
γ_{tmp} =γ_{best}
T=R
for x in (C-R)
If $\gamma_{RU(X)}$ > $\gamma_X(D)$?
No
Yes
Construct First Generation
If γ_{best} > γ_{tmp} ?
Yes
No
Selection
Crossover
Mutation
Return R
C
Apply optimized output to network
Calculate output through hidden layer & error at the output layer
Update weight & bias
D
Check if the error is minimal?
No
Yes
E
B
A
Calculate the fitness of the reduct data features
Reduct Data Set


A
If $f(x_i)$ $f(best_i)$
Yes
f(best$_i$)=f(x$_i$)
No
f(best$_i$)
If P(best$_i$)>g$_{best}$?
No
g(best)
Yes
g(best$_i$)=p(best$_i$)
Classified Binary output
C
Calculate Convergence and Inertia weight
Network Evaluation
E
Update velocity and position
No
Yes
Check if the error is minimal?
Adjust Acceleration
B
Yes
If dataset is optimized?
No
Choose initial weight
D

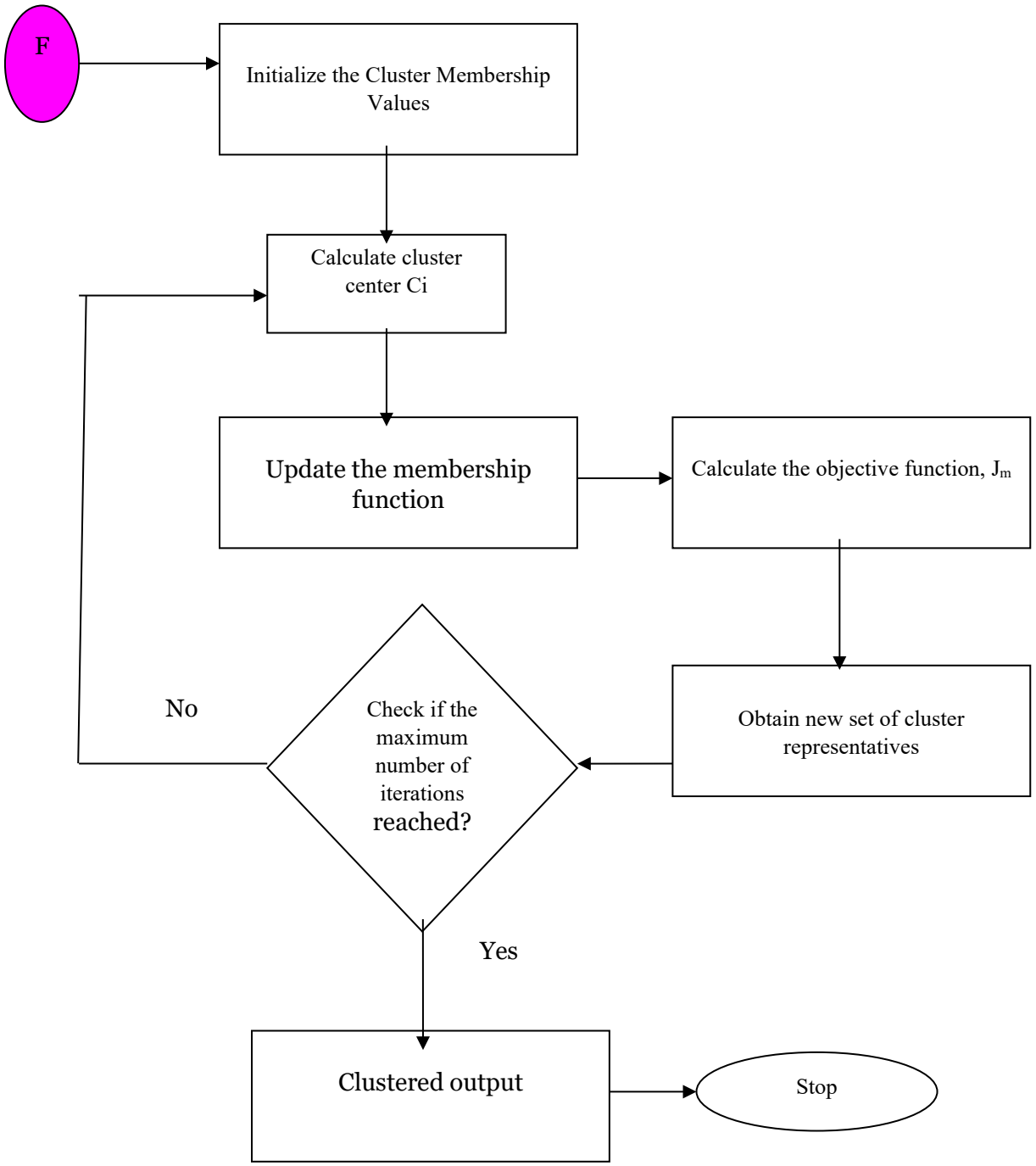


Figure 5: The work flow of HC^{2}A

4. Experimental Setup

The proposed hybrid algorithm is implemented in Matrix Laboratory (MATLAB). The source code developed are run on system with Inter ® i5 with 8 GB RAM. The dataset required for experimentation is taken from the University of California (UCI) Website [20], [21].

5. Results and Discussion

The framework is implemented on Cloudsim and Matlab. The metrics like Sensitivity, Specificity, Precision or Positive Predictive Value (PPV), Negative Predictive Value, False Positive Rate (FPR), False Negative Rate and Accuracy is used to evaluate the performance of the network with different number of neurons. Sensitivity is also called as True Positive Rate which measures the proportions of positivies that are correctly identified. Specificity is defined as the measure of proportions of negatives that are correctly identified. It is also denoted as True Negative Rate.

Table 3: Comparison of results obtained using different algorithms for PID

Techniques	Sensitivity	Specificity	Precision	NPV	FPR	FNR	Accuracy
MPSO-ANN	0.90	0.47	0.87	0.54	0.53	0.09	0.81
GA-RR+MPSO-ANN	0.92	0.60	0.94	0.55	0.40	0.07	0.87
GA-RR+MPSO-ANN+FKM	0.99	0.51	0.93	0.88	0.49	0.01	0.92
HC^{2}A	0.99	0.94	0.99	0.94	0.06	0.005	0.98

$$Sensitivity = \frac{TP}{(TP + FN)}$$

$$Specificity = \frac{TN}{(TN + FP)}$$

$$\Pr ecision = \frac{TP}{(TP + FP)}$$

$$NPV = \frac{TN}{(TN + FN)}$$

$$FPR = \frac{FP}{(FP + TN)}$$

$$FNR = \frac{FN}{(FN + TP)}$$

$$Accuracy = \frac{(TP + TN)}{Total\ no\ of\ ins\tan ces}$$

where,

TP- True Positive

FN- False Negative

FP- False Positive

TN- True Negative

Pima Indian Diabetes (PID) Data Set is taken from UCI to evaluate the performance of the proposed algorithm. It consists of a total of 8 attributes and one output class A total of 768 instances are taken for experimentation. Table 1 depicts the results obtained using different algorithms. The proposed algorithm after scheduling using the ATGA has given more accuracy when compared with other algorithms.

The proposed HC2A is compared with Modified Particle Swarm Optimization – Artificial Neural Network (MPSO-ANN) Algorithm [22], Genetic Algorithm-Relative Reduct Algorithm combined with MPSO-ANN GA-RR+MPSO-ANN) Algorithm [23], GA-RR and MPSO-ANN combined with Fuzzy K-Means Clustering (GA-RR+MPSO-ANN+FKM). Table1, 2, 4 & 5 explains the Confusion Matrix obtained for different algorithms.

Below Table 3 shows the results that compares different algorithms for PID. Table 4 below shows the confusion matrix results.

Table 4: Confusion matrix for GA-RR+MPSO-ANN+FKM – PID

	0	1
0	558 72.66	80 10.42
1	60 7.81	70 9.11

The Table 5 below shows the results of Confusion Matrix for HC^{2}A- PID.

Table 5: Confusion Matrix for HC^{2}A- PID

	0	1
0	558 72.66	80 10.42
1	60 7.81	70 9.11

The sensitivity rate for PID is 0.99 with the proposed algorithm. The sensitivity rate is increased since ATGA is applied for scheduling. The prediction accuracy is increased to 98%. The

other metrics which are used to measure the performane of the proposed framework also supports the algorithm in classifying the diabetes disease. Figure 6 explains the results obtained by using the different algorithms.

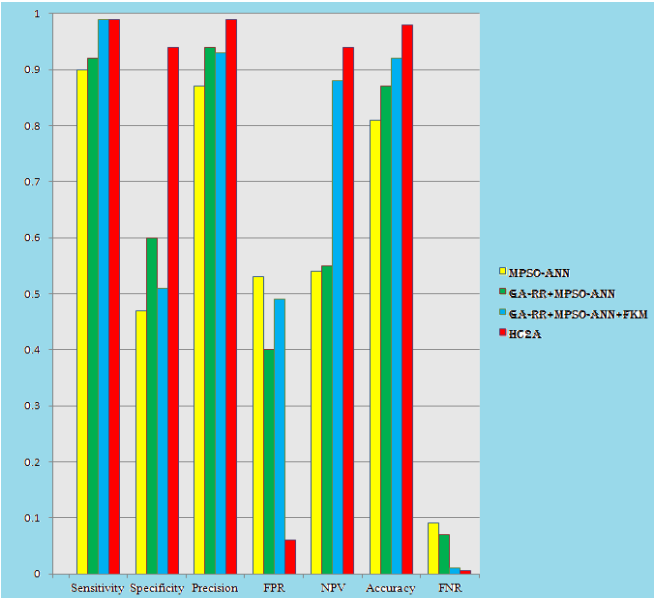


Figure 6: Performance of different algorithms

Table 6: Confusion Matrix for MPSO-ANN – USD

	0	1
0	7000 70	431 4.31
1	1569 15.69	1000 10

Table 7: Confusion Matrix for GA-RR+ MPSO-ANN – USD

	0	1
0	8115 81.15	987 9.87
1	513 5.13	385 3.85

This algorithm is again experimented with US diabetes data set (USD) collected from UCI. This data set consists of 10000 instances, with 54 attributes and one output classes. The promising results obtained supports the proposed framework in classifying the diabetes data set even when the number of data set increases, since the ATGA is applied for scheduling the Cloud Server. The results obtained are tabulated in Table 10. Confusion Matrix for MPSO-ANN, GA-RR+MPSO-ANN, GA-RR+MPSO-ANN+FKM and HC^{2}A is given in tables 6,7,8 and 9. Figure 7 gives the performance of different algorithms on USD. Figure 8 compares the performance of the algorithm on PID and USD based on Accuracy. It was observed and evident that the proposed Framework which comprises the ATGA and GA-RR+MPSO-ANN+FCM is most suitable for diabetes prediction even with increase in the number of data set.

Table 8: Confusion Matrix for GA-RR+ MPSO-ANN+FKM – USD

	0	1
0	8563 85.63	643 6.43
1	357 3.57	437 4.37

Table 9: Confusion Matrix for HC^{2}A – USD

	0	1
0	9015 60.15	115 1.15
1	77 0.77	793 7.93

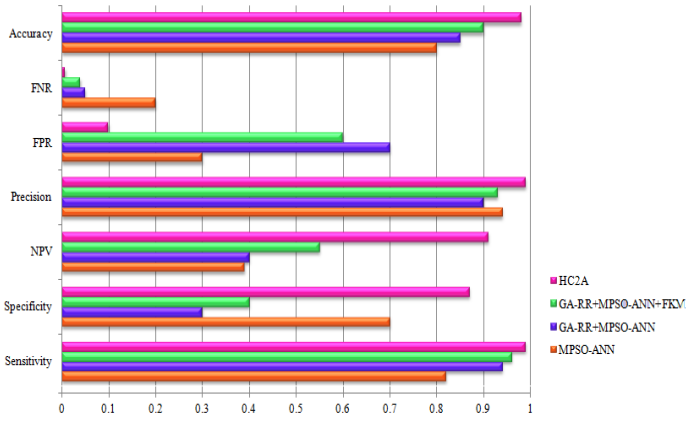


Figure 7: Performance comparison of different algorithms – USD Data set

The below table 10 shows the Comparison of results that obtained using different algorithms for USD.

Table 10: Comparison of results obtained using different algorithms for USD

Techniques	Sensitivity	Specificity	Precision	NPV	FPR	FNR	Accuracy
MPSO-ANN	0.82	0.70	0.94	0.39	0.30	0.20	0.80
GA-RR+MPSO-ANN	0.94	0.30	0.90	0.40	0.70	0.05	0.85
GA-RR+MPSO-ANN+FKM	0.96	0.4	0.93	0.55	0.60	0.04	0.90
HC^{2}A	0.99	0.87	0.99	0.91	0.10	0.008	0.98

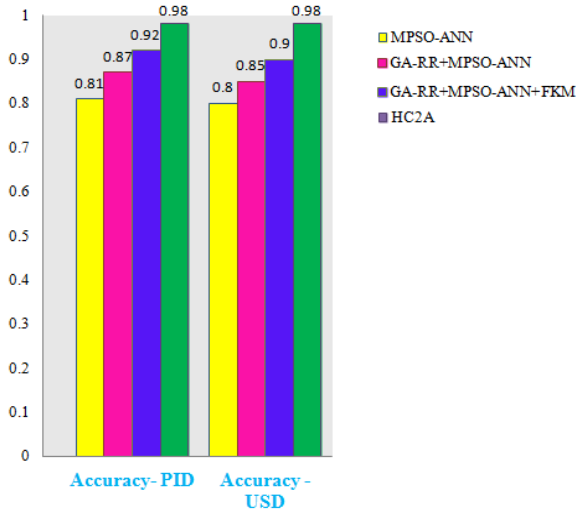


Figure 8: Comparison of performance of different algorithms in PID and USD Data sets based on Accuracy

6. Conclusion

Elasticity-based Med Cloud Recommendation System(EMCRS) is proposed for diagnosing the diabetes disease and providing recommendation for the diabetic patients. This framework is implemented on Cloud and hence the resource sharing is made more comfortable. Adaptively Toggle Genetic Algorithm is applied to allocate the cloud resources in elastic manner. In the proposed algorithm when huge spikes of data occur, the classification process and clustering is handed over to virtual manager to scale the data across virtual machines while maintaining the security features of medical data. The proposed algorithm focuses on workload prediction with elasticity in cloud environment, in case of scaling conditions which is a unique feature of the algorithm. Hybrid Classification and Clustering algorithm is applied to classify and cluster the diabetic patients. The accuracy of about 98% is acquired for both PID and USD datasets. Performance metrics taken for evaluation also explores the effectiveness of EMCRS. Hence, it is evident that the proposed recommendation system is well suited for diabetic prediction. The limitation of the algorithm is the level of security on cloud data. The framework can be enhanced to predict and classify other types of financial and industrial data in future.

References

[1] P.G. Jiménez, J. Martín-Carmona, E.L. Hernández, "Diabetes mellitus," Medicine (Spain), **13**(16),883-890,2020, doi:10.1016/j.med.2020.09.010.

[2] V. Singh, K. Kumar, "Data Mining and Knowledge Management", International Research Journal of Engineering and Technology, **4**(2),200-209, 2017.

[3] V. Hedau, P. Pant, K. Sharma, "Material Selection Using Association Rule Mining," International Journal of Advanced Research in Computer Science, **7**(3), 230-234,2016.

[4] M.S. Janani and M.P. Saravana, "Data Mining On Diabetics", International Journal of Innovative Science, Engineering & Technology, **1**(10), 129-134, 2014.

[5] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, A. Warfield, "Xen and the art of virtualization", in: Proceedings of the 19th ACM Symposium on Operating Systems Principles, SOSP 2003, Bolton Landing, NY, USA, 2003, doi:10.1145/945445.945462.

[6] P. Mell, T. Grance,"The NIST definition of cloud computing", Technical Report, National Institute of Standards and Technology, Information Technology Laboratory, 2011, doi:10.1016/b978-0-12-804018-8.15003-x.

[7] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia,"A view of cloud computing", Communications of the ACM, **53**(4),50–58, 2010, doi:10.1145/1721654.1721672.

[8] P. Mell, T. Grance, "The NIST Definition of Cloud Computing", Tech. rep., U.S. National Institute of Standards and Technology (NIST), Special Publication 800-145, 2011.

[9] M. Durairaj, P. Kannan, "A study on Virtualization Techniques and Challenges in Cloud Computing", International Journal of Scientific &Technology Research, **3**(11), 147-151, 2014.

[10] B. Jennings, R. Stadler, "Resource Management in Clouds: Survey and Research Challenges," Journal of Network and Systems Management, **23**(3),567–619 ,2015, doi:10.1007/s10922-014-9307-7.

[11] R. Buyya, R. Ranjan, R.N. Calheiros, "InterCloud: Utility-oriented federation of cloud computing environments for scaling of application services," In: Hsu CH., Yang L.T., Park J.H., Yeo SS. (eds) Algorithms and Architectures for Parallel Processing. ICA3PP 2010. Lecture Notes in Computer Science,. Springer, Berlin, Heidelberg, 6081, 19-24, 2010, https://doi.org/10.1007/978-3-642-13119-6_2.

[12] P.D. Kaur, I. Chana, "Cloud based intelligent system for delivering health care as a service," Computer Methods and Programs in Biomedicine, **113**, 346–359, 2014, doi:10.1016/j.cmpb.2013.09.013.

[13] D.A. Bacigalupo, J. Van Hemert, X. Chen, A. Usmani, A.P. Chester, L. He, D.N. Dillenberger, G.B. Wills, L. Gilbert, S.A. Jarvis, "Managing dynamic enterprise and urgent workloads on clouds using layered queuing and historical performance models,"Simulation Modelling Practice and Theory, 1-30, 2011, doi:10.1016/j.simpat.2011.01.007.

[14] M. Durairaj, P. Kannan, "Improvised Genetic Approach for an Effective Resource Allocation in Cloud Infrastructure," International Journal of Computer Science and Information Technologies, **6**(4), 4037-4046, 2015, doi:ISSN: 0975-9646.

[15] T. Lorido-Botrán, J. Miguel-Alonso, J.A. Lozano, "Auto-scaling Techniques for Elastic Applications in Cloud Environments," Technical Report: EHU-KAT-IK-09-12, University of the Basque Country, 1-44,2012.

[16] K.A. Al Mamun, M. Alhussein, K. Sailunaz, M.S. Islam, "Cloud based framework for Parkinson's disease diagnosis and monitoring system for remote healthcare applications," Future Generation Computer Systems, **66**, 36–47, 2017, doi:10.1016/j.future.2015.11.010.

[17] T. Santhanam, M.S. Padmavathi, "Application of K-Means and genetic algorithms for dimension reduction by integrating SVM for diabetes diagnosis," Procedia Computer Science, Elsevier, **47**, 76-83, 2015, doi:10.1016/j.procs.2015.03.185.

[18] K. Vembandasamy, T. Karthikeyan, "Novel outlier detection in diabetics classification using data mining techniques," International Journal of Applied Engineering Research, **11**(2),1400-1403, 2016.

[19] K. Polat, "Intelligent Recognition of Diabetes Disease via FCM Based Attribute Weighting", World Academy of Science, Engineering and Technology, International Journal of Computer and Information Engineering, **10**(4), 783-787, 2016. doi.org/10.5281/zenodo.1124023.

[20] D. Sisodia, D.S. Sisodia, "Prediction of Diabetes using Classification Algorithms," in Procedia Computer Science, **132**, 1578-1585, 2018, doi:10.1016/j.procs.2018.05.122.

[21] W. Chen, S. Chen, H. Zhang, T. Wu, "A hybrid prediction model for type 2 diabetes using K-means and decision tree," in Proceedings of 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 386-390,2017 doi: 10.1109/ICSESS.2017.8342938.

[22] K. Ateeq, G. Ganapathy, "The novel hybrid Modified Particle Swarm Optimization-Neural Network (MPSO-NN) Algorithm for classifying the Diabetes", International Journal of Computational Intelligence Research, **13**(4), 595-614, 2017.

[23] K. Ateeq and G. Ganapathy, "A Hybrid Genetic-Relative Reduct Algorithm for Pre-Processing the Diabetic Dataset", International Journal of Advanced Research in Computer Science, **8**(1), 60-64, 2017.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1634-1642 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

On Design of IoT-based Power Quality Oriented Grids for Industrial Sector

Nesma N. Gomaa[1,2], Khaled Y. Youssef[*,3], Mohamed Abouelatta[1]

[1]*Faculty of Engineering, Electronics and Electrical Communication, Ain Shams University (ASU), Cairo, 11757, Egypt*

[2]*Faculty of Engineering, Electronics and Electrical Communications, October University Modern Sciences and Arts (MSA), Cairo, 11757, Egypt*

[3]*Faculty of Navigation Science and Space Technology, Beni-Suef University (BSU), Beni Suef, 62511, Egypt*

ARTICLE INFO

Article history:
Received: 16 September, 2020
Accepted: 16 December, 2020
Online: 25 December, 2020

Keywords:
Energy Management System
Internet of Things (IoT)
Power system
Power Factor
Harmonics
Power quality

ABSTRACT

Manufacturing industry is facing several challenges in power quality and energy consumptions that play a significant role in the cost of goods sold in addition to the operational efficiency of manufacturing plants. The current techniques are not enough to manage the manufacturing processes from power perspectives as it is focusing only on the monitoring of the power grids using smart meters connected to SCADA systems. In this paper, a novel technique is proposed that is based on energy-aware manufacturing process control model using internet of things technology. The model is applied on cabling industry use case in addition to implementation of a MATLAB model for this purpose whereas the IoT physical layer is collecting , analyzing and communicating the electric power parameters correlated with the manufacturing process parameters received from PLC IoT nodes. Accordingly, the effective power quality is enhanced using manufacturing process control rather than additional correction nodes in traditional techniques. On applying the model on cabling use case, the Total Harmonic Distortion (THD-current) is improved 10 times to be 3.1% against 31.4% and the power factor is improved by 33% from 0.7 to 0.93 without additional correction nodes.

1. Introduction

This paper is an extension of work originally presented in 15th International Computer Engineering Conference (ICENCO) [1]. Power Quality (PQ) has recently become a very important topic because of its effect on energy producers, equipment's, manufacturers and consumers. There are so many industries nowadays that produce and use high technology for production thus high quality and high reliability are needed for these supplying strength [2]. The advantages of gathering real time monitoring is to develop the perception of awareness for power quality in factories.

Despite these circumstances, the executions of energy quality steps in improving energy efficiency have not yet satisfied expectations. The reason for the deficits in energy efficiency realization include lack of time and clarity on energy consumption [3]. The widespread use of non-linear loads leads to a huge distortion of voltage and especially current waveforms that affect any power electronic system [4]. In power systems, harmonic distortion and power factor disturbance increase on daily basis due to an increase in the use of nonlinear loads such as system resonances, voltage distortion, interference with communications networks, transformers overloading, a negative impact on energy metering and a decrease in the overall system efficiency. Thus, these two factors are set to be a serious problem in terms of power quality and this may affect other pieces of equipment connected to a particular system. The optimal performance between increasing in energy demand without severe increases in energy costs come to be our challenge now days. Hence, according to IEEE and International Electro-technical Commission (IEC) standards, industries must devote a greater attention to keeping harmonics and Power Factor (PF) within permissible limits [5, 6]. Furthermore, (IoT) could have significant contribution on energy management, hence several (IoT) applications to reduce the energy consumption have been developed [7, 8].

Power quality (PQ) monitoring methods that are recently produced have the capability to calculate and store data

[*]Corresponding Author: Khaled Y. Youssef, khalid_youssif@yahoo.com

https://dx.doi.org/10.25046/aj0506194

automatically with a high sampling rate. Significant quantities of measured data, such as current signals, voltage signals, total harmonic distortion, and power factor, can be obtained by monitoring devices. Yet, they do not identify distortions automatically, and involve off-line analysis of the data recorded. Consequently, in order to identify the type of end-use equipment, harmonic variations, and power factor detection a real time technique study is provided in this paper to develop the perception of awareness for power quality in factories and examine the impact of energy efficiency in industries. In today's power systems connected to nonlinear loads, voltage and current waveform distortion are more prevalent. The prevalence and interaction of loads of this kind with the power supply system increases the need to use power monitoring equipment to assess the quality of electric power. This paper proposes a new architecture which makes use of real-time data to ensure prediction of machine failure, using power quality data and ways to improve the energy efficient process. The algorithm proposed can be used for automated correction of industrial energy profile using Internet of Things (IoT) technology.

2. Literature review

Systematic approaches to eliminate the consumption of energy has been studied. Hence in order to explain the significance of problems with power quality, a detailed representation of any nonlinear system may be needed in the case of weak systems as interactions may occur between harmonics of different orders in such systems that are not predicted by linear time invariant models. Thus to identify and evaluate energy quality indices for various categories of electricity users, a fuzzy logic-based approach is suggested in [9]. Low energy quality is a serious concern for domestic, commercial and industrial users, in [10] issues of power quality in electrical systems is illustrated briefly.

Harmonics in electricity networks have recently gained even more attention from the power industry as a key cause of PQ problems, in [11], the total harmonic and unbalance distortion is well-defined to help factories to reduce harmonic distortion. In [12], an index is determined to represent the thermal effects on three phase induction motors as a result of harmonic distortions. Researchers in [13,14] illustrate other methods of PQ disturbances such as inter harmonics and voltage fluctuations. Furthermore , a Harmonic Trap Filter in conjunction with a line reactor designed and implemented by [15], using Simulink to demonstrate the principle of harmonic analysis in a three-phase system.

Additionally, the dramatic increase of electricity each year take the initiative on the aspect of energy efficiency in order to address the environmental issues. One of these aspects is the power factor correction. In [16], it is studied the importance of power factor which reflects how much the industries or organizations will consume the electricity efficiently. In [17], the authors investigate a system to carry out the effect of power factor on inductive loads using capacitor bank. Moreover, a system using a thyristor switching capacitor was proposed by [18], which helped to improve the quality of power and regulate the reactive power under the regularization of power factor.

Decreasing energy consumption is important for increasing productivity thus the relationship between electricity and economic growth is vital .To conserve energy and enhance its efficiency, smart grids (SG) were developed to do that [19]. John S. In [20], it is highlighted and clarified the different demand response (DR) programs. A number of DR-based programs are offered to organization managers and consumers, in general, where they choose which program they want to follow in order to decrease the overall energy consumption. Likewise, as energy saving has been one of the critical problems in designing electronic devices, Smart home appliance control system was developed in the IoT environment, a smart control system based on the technologies of internet of things has been proposed in [21].

3. Architecture for power quality management using IoT technology

In this paper a new architecture is proposed to increase the levels of power efficiency in industries. A smart energy management technique is illustrated in Figure 1. Using this block diagram will help factories to reduce their energy consumption. This architecture depends on tracking different loads in the downstream zone.

The architecture shown in Figure 1 is focused on monitoring the different loads attached in a Sensor Area Network (SAN), for the purpose of automated real-time monitoring of transient behaviour and accordingly the detection of anomalies related to power quality issues as voltage dwells, low power factor, and high order harmonics, that might happen as a result of grid's load non-linear behaviour. Most of this phenomenon is currently detected by capture of a measurement sample from the sub-system under test and that is usually reactive rather than proactive.

The concept emphasized by the study is about converting the grid system into an autonomous grid that has the ability to auto detect the phenomenon using a transit-time domain analysis and thus paves the way for artificial intelligence techniques on board to automatically detect the abnormal phenomena even on odd occurrence based on the intra-load communication, source-load communication and load on board processing. The concept also states the edge of machine to machine communications that is the core base of the well-known IoT technology.

According to Figure 1, the central processing server (CPS) is the node that is responsible for orchestration of the correction nodes on the correction plane according to the dynamicity of the load, and thus the CPS calculates and signals the configuration parameters to the corresponding correction nodes based on the measurements done by the IoT layer. The key difference between the existing techniques of correction and the proposed technique is mainly the holistic management of correction for the detected problem according to the multi-factor causal analysis.

One example is the power factor transient's detection that will be captured by IoT layer that is then escalated to CPS layer which in turn take the holistic decision to off –load the power factor capacitor bank and signal the synchronous power factor correction node to be in place in addition to active harmonic cancellation node in place of the passive one. The solution needs a level of intelligence as the classical capacitor bank response time to the change in loads will cause the failure of the classical technique. Moreover, the active cancellation will become effective due to the significant presence of several orders of harmonics rather than the

classical passive filter mode. The edge of the IoT layer in this example is in the consolidated view for the potential sources of problem with the decision of act on the optimum solution of correction.

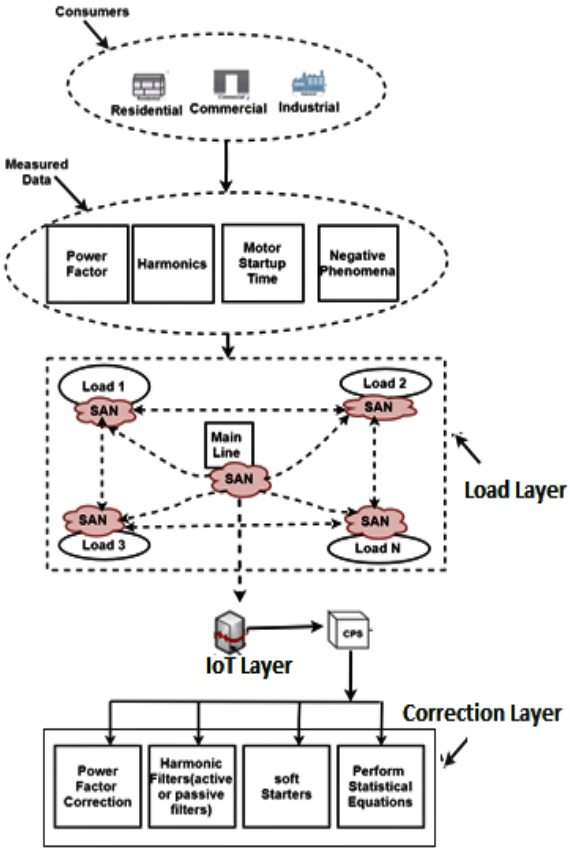


Figure 1: Schematic for the modules combining an intelligent energy management system in factories.

Illustrated in Figure 2 the concept of IoT smart grid which adds intelligence at each step by using online monitoring of transmission lines for disaster prevention and efficient use of power in order to decrease the overall energy consumption inspired by the factory. The three-layer architecture provides a high level framework through which different approaches might be implemented. The first layer consists of Machine drives, which are pumps, electric motors, and fans, account for about half of the manufacturing sector's delivered electricity, layer two is the IoT layer which consists of IoT beacon which are used as linking bridge that helps to provide total tracking and the last layer is named by correction plane layer in which the power factor correction bank and the harmonic filter exists.

The energy efficiency scheme is demonstrated by the flow chart in Figure 3, which begins by monitoring various production lines where each line has different load numbers, X is the order of each load, and N is the total number of loads. Two different parameters are evaluated on each load; power factor and harmonics. The CPS can detect if any of these parameters are not functioning effectively. A signal is sent to the CPS in the event of any errors. Where the problem can be overcome by the CPS,

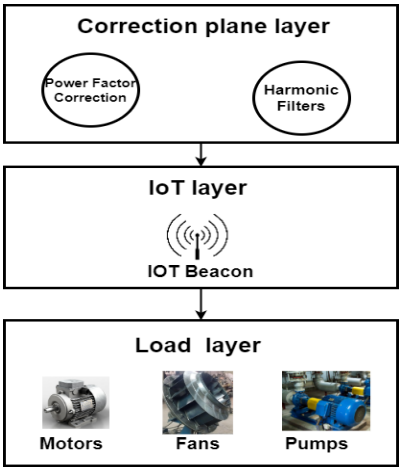


Figure 2: IoT architecture for smart grid.

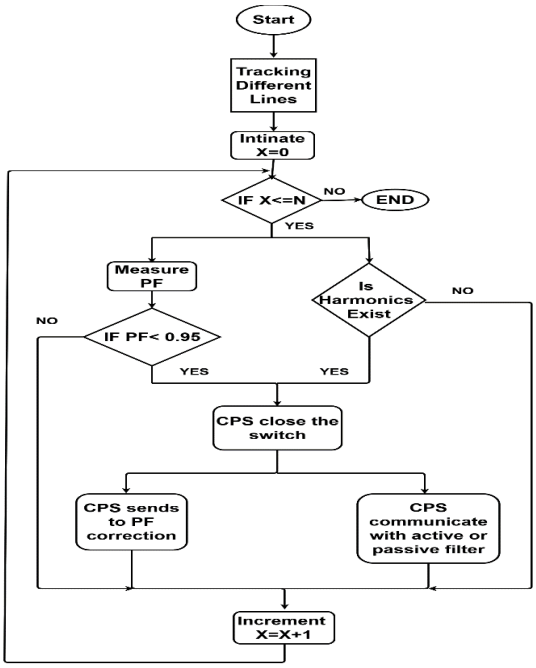


Figure 3: Flow Chart for energy saving analysis.

4. Power Factor correction and Harmonics detection Matlab simulation model results and analysis

A Matlab Model is designed to limit the voltage and current harmonics to be applicable with the IEC standards and to improve the power factor value. A Three-phase harmonic filters with power factor correction model for High-Voltage Direct Current (HVDC) is designed in Figure 4, using Matlab Simulink platform to reduce the voltage and current distortion in the power system. The model is designed using a frequency of 60 HZ and a firing angle α= 19°. One capacitor bank C1 of 150 Mvar designed by a "Thee-Phase Series RLC Load". Three types of different filters combination were also used including one C-type high pass filter tuned to the 3rd F1 of 150 Mvar ,one doubled –tuned filter 11/13th F2 of 150 Mvar and one high pass filter tuned to the 24th F3 of 150 Mvar.

The model used a switch control device for reducing the load imbalance and improving the efficiency of data processing. This

switch control is used to switch on or off when the CPS detect that the data received from the load is not achieving the threshold. The filter can be activated and deactivated using the breaker in the circuit. When the breaker (i.e. CPS) is closed, the filter is activated, the current and voltage will pass through the filter and will successively be filtered. Conversely when the breaker is opened the filter will be deactivated and will have no effect on the output waveforms of the system.

After running the developed Simulink model with different values of α. A various responses are observed. The waveforms in Figure 5, 6 respectively illustrate that when the breaker is opened, the filter is deactivated, a highly distorted waveform was detected because of the nonlinear loads that can be existing in the system. Hence when the breaker is closed, the filter can remove the harmonics given the desired waveform which is free from distortion.

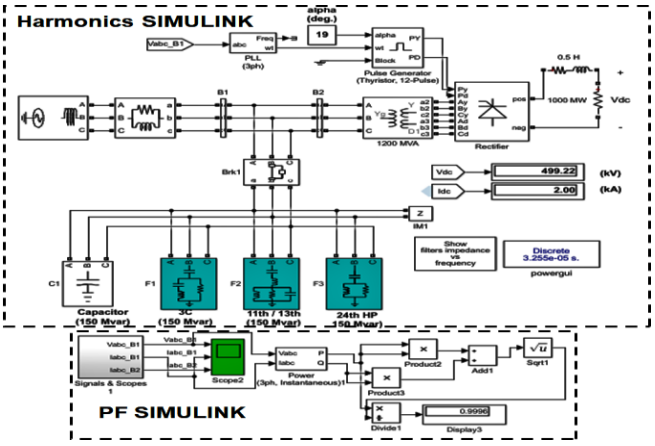


Figure 4: Three Phase Harmonic Filter with aid of PF correction (Simulink Design).

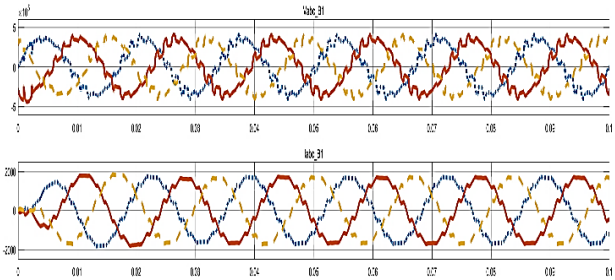


Figure 5: Output Waveforms when filter is disconnected.

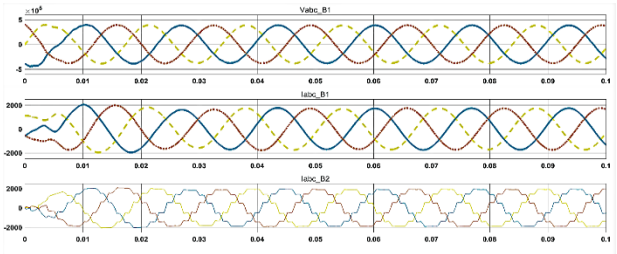


Figure 6: Output Waveforms when filter is connected.

Table 1 and table 2 illustrate that by using a suitable harmonic filter with the aid of a suitable capacitor value, we can control the flow of these fluctuations. From theses measureable results, it can be inferred that the system has worked successfully. As for example, it is clear that with the assistance of the harmonic filters and a capacitor, the THD voltage has fallen dramatically from 17.1 percent to 0.6 percent with an impedance of 407 Likewise, from 7.4 % to 0.3 percent, the new THD has decreased. The PF has risen from 0.3 to almost 0.9, coming close to the ideal 1.0.0. The proposed technique is therefore highly effective. 1.

Table 1: Comparison between voltage THD_v and current THD_i before and after adding the capacitor with the aid of the filter

Z (Ω)	THD_v Without Filter (%)	THD_i Without Filter (%)	Capacitor (F)	THD_v with Filter (%)	THD_i with Filter (%)
407	17.1	7.4	6.2E-06	0.6	0.3
400	16.2	7.1	8.0E-06	0.7	0.4
301	17.0	7.8	9.3E-06	0.71	0.4
261	17.01	7.12	1.0E-05	0.72	0.4
224	17.04	7.11	1.1E-05	0.69	0.5
193	17.08	7.14	1.2E-05	0.73	0.5
169	18	7.51	1.2E-05	0.8	0.7
151	17.53	7.19	1.1E-05	0.98	0.95
338	17.98	7.18	8.4E-06	0.02	0.1
488	16.82	7.00	6.1E-06	0.74	0.4
319	16.80	7.20	9.5E-06	0.7	0.4
335	17	7.31	9.1E-06	0.7	0.4
242	17.01	7.08	1.1E-05	0.7	0.4
187	17.2	7.16	1.3E-05	0.80	0.5
279	17.52	7.3	1.0E-05	0.8	0.4
156	17.69	7.4	1.9E-05	0.8	0.5
138	17.03	7.5	1.39E-05	0.9	0.51
152	17.7	7.22	1.40E-05	0.6	0.49
131	18	7.20	1.41E-05	0.7	0.71

Table 2: Comparison between PF before and after adding the capacitor

Z (Ω)	Distorted P.F	Capacitor (F)	Corrected P.F
406	0.30	6.2E-06	0.8
359	0.34	8.0E-06	0.91
300	0.4	9.4E-06	0.98
260	0.47	1.0E-05	0.99
223	0.55	1.1E-05	0.92
192	0.64	1.2E-05	0.94
168	0.7	1.2E-05	0.95
150	0.82	1.1E-05	0.95
338	0.36	8.5E-06	0.96
487	0.25	6.1E-06	0.97
318	0.15	9.5E-06	0.91
334	0.14	9.1E-06	0.99
241	0.41	1.1E-05	0.99
186	0.53	1.4E-05	0.98
279	0.17	1.0E-05	0.97
156	0.32	1.9E-05	0.95
137	0.72	1.5E-05	0.94
152	0.65	1.5E-05	0.99

5. Existing and future power quality system model for different loads

In the current technological transition scenario, power is very precious. A lot of researches need to improve the power quality system and find out the causes of power loss. In industries the huge increase in the use of inductive loads will leads to deficiency of the power system hence a routinely techniques have been used to detect the failures of their loads and correct it[22]. The classical approach shown in Figure 7 is to connect a direct act device with a current transformer (C.T) across the industrial input power lines. The C.T is directly installed at each load. In the classical scenarios, the technique of switching method (on/off) is used. Each load is supported by a controller device which is responsible for detection or cancellation to any failure on each line.

For example, to correct the power factor, these old techniques make use of a microcontroller to control the switching phase of the linked capacitor bank at the end of the main power supply. Each load has its specific control device connected to the main power supply close to the variable load. The controller's job is to read the power factor, voltage and current of the device. Then the controller determines, according to a proven algorithm, to cut off the power for the specified load and decide how many capacitors should be added to raise the power factor to the value required.

With this classical approach, there are two major problems: (1) the correction is a fixed one, so that if the industrial load differs , the correction becomes either less efficient or on the other hand, potentially overly efficient, resulting in an undesirable low leading power factor , (2) many industrial capacitor banks have inherently unpredictable maintenance problems. It would therefore be highly desirable to provide a device that would automatically correct the power factor of an induction motor in real time, so that as the applied mechanical load is varied, the motor runs at or near optimum efficiency continuously.

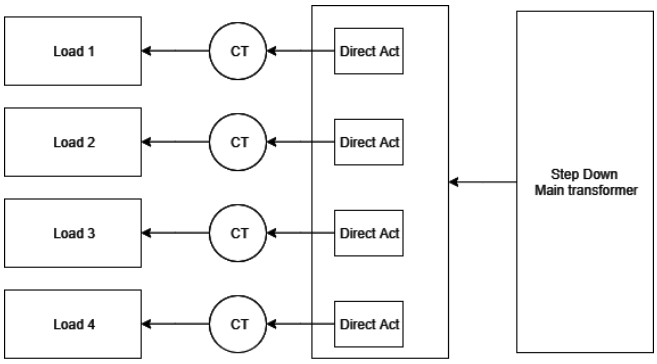


Figure 7: The classical power quality model for different load.

Substantial energy savings could be achieved through the Building Energy Management System (BEMS) within the Smart Electric Grid setting. A new technique based on IoT real time data acquisition using Message Queuing Telemetry Transport (MQTT) protocol which is the most commonly used protocols in IoT projects. The proposed architecture make use of real-time data to ensure prediction of machine failure using power quality data and ways to improve energy efficient process. The development of an energy monitoring system with the ability to acquire, process and store data in real-time is required for improving energy efficiency operations of the manufacturing industries and to enable proactive maintenance .This architecture prove the concept of sense, infer and act. As shown in Figure 8, the IoT beacons can contact with each other through the MQTT protocol. Thus, they can agree together that for example load number (2) will be compensated only. Consequently, tune tone generated by centralized harmonic cancellation to connect central node to a specific load.

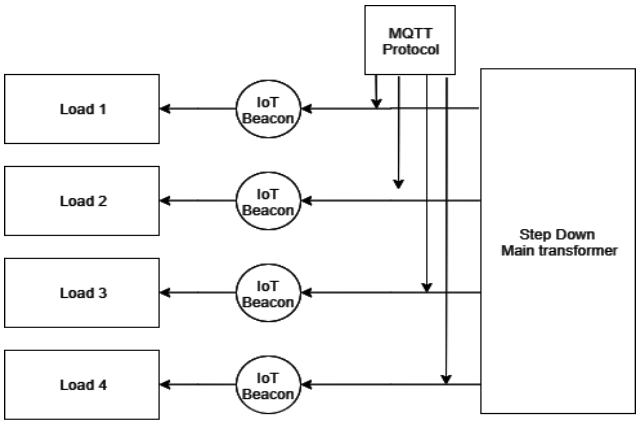


Figure 8: Future power quality system model for different loads.

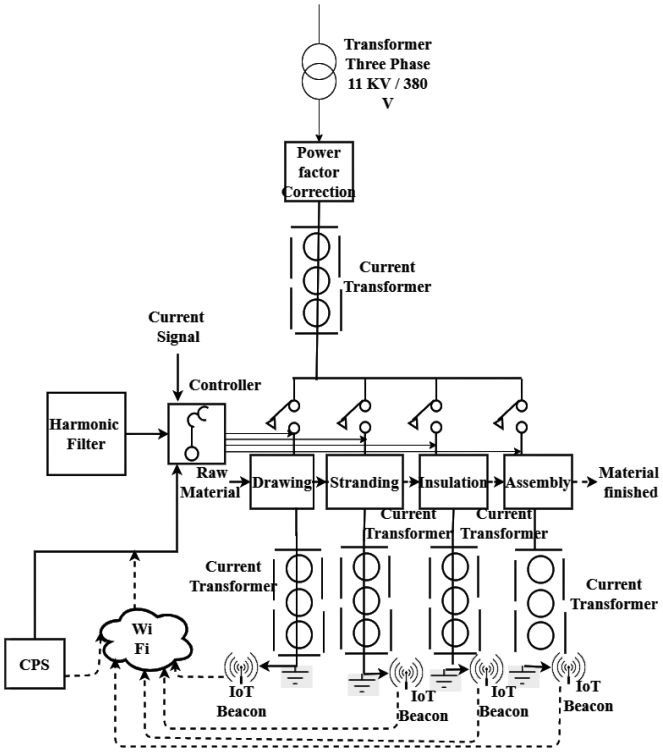


Figure 9: The proposed model for energy management technique.

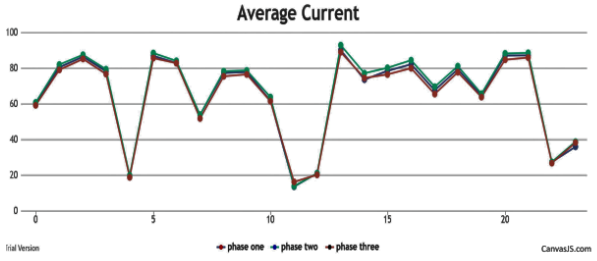
6. Power factor and harmonics improvement: A Case study and Analysis of the processes of the production of a metals in Factories

This work is a case study and review of the technological and energy aspects of the recommended design of the metal factory production lines in Egypt. The study provide the factory with some technologies which can be integrated into the factory production

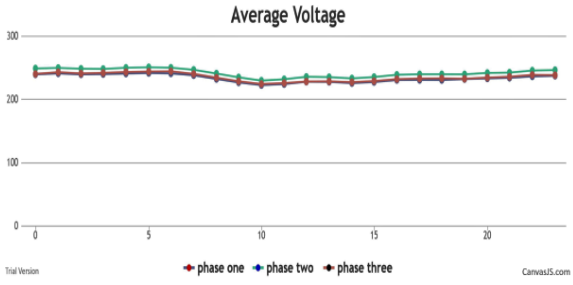
processes in order to improve the product quality, minimize the need for electrical energy, reduce maintenance time and cost, and increase reliability and real-time data accuracy.

This study reviewed the production line of metals in factories. The metal production line consists of four main process illustrated in Figure 9 which are drawing, stranding, insulation and assembly. In this power quality assessment, a power quality analysis of metal production process has been shown where power is given to different loads through transformer. Readings of power quality parameters such as voltage unbalance, current unbalance, power factor, harmonics and voltage fluctuation in each phase of the metal production have been studied.

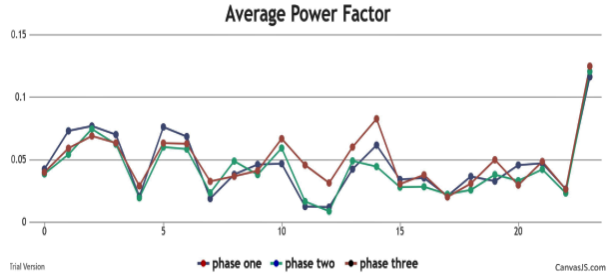
The purpose of the electrical distribution system is to support proper operation of the loads. When any one of the loads fails to meet its purpose, it is time to investigate the problem, find the cause and initiate corrective action. When a load does not operate properly, the other loads agree together through the proposed architecture shown in Fig.9 that this load should be compensated only. Figure 10-13 show the average voltage distortion, average current distortion, average power factor and average harmonic in the drawing phase, stranding phase, insulation phase and assembly phase respectively from the metal factory. Finally all these parameters are measured at the main line, so that the power factor and the harmonics at the main line can be fixed directly. Thus, Figure 14 and Figure 15 illustrated the reading at the main line before and after correction correspondingly.



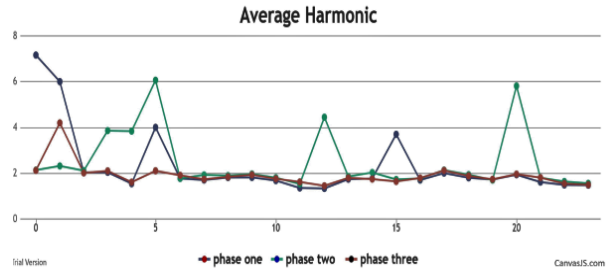
(a)



(b)

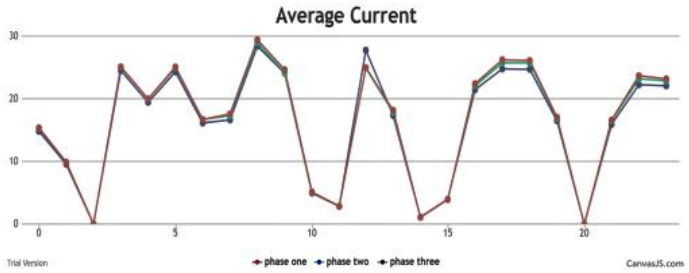


(c)

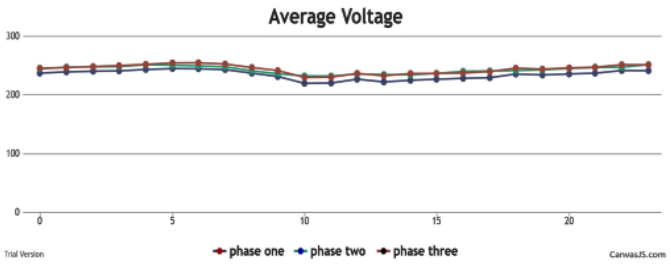


(d)

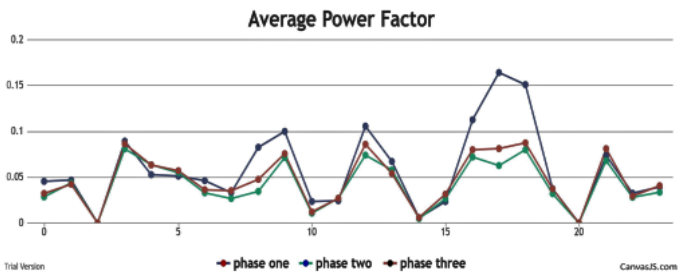
Figure 10: (a) The average current distortion, (b) average voltage distortion, (c) average power factor and (d) average harmonic in the drawing phase respectively.



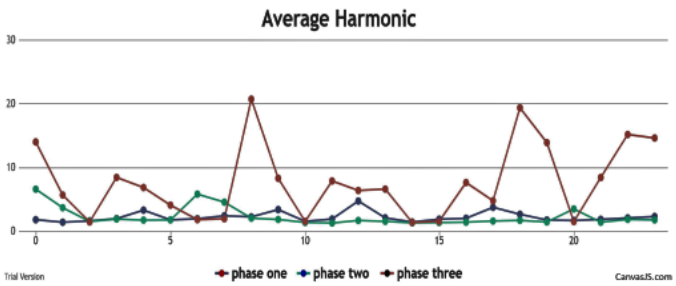
(a)



(b)

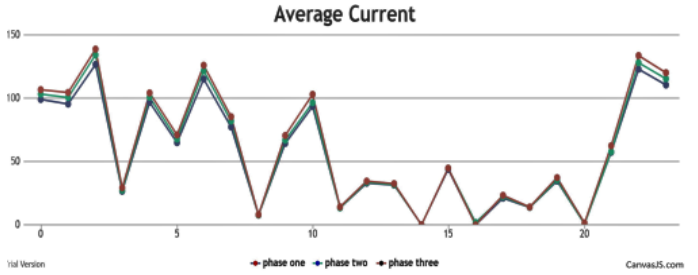


(c)

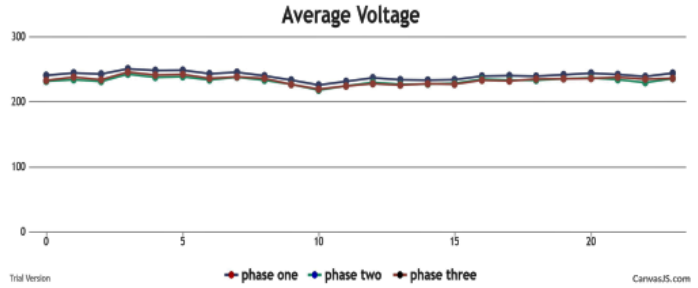


(d)

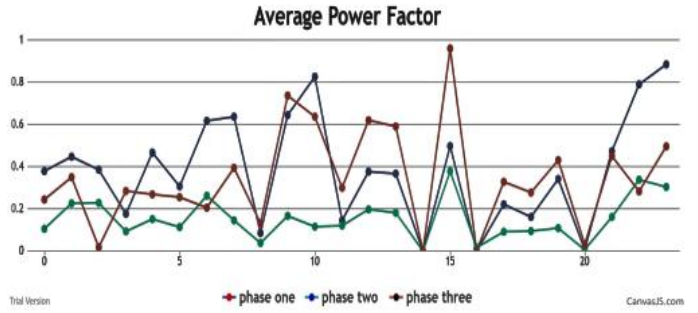
Figure 11: The average current distortion, average voltage distortion, average power factor and average harmonic in the Stranding phase respectively.



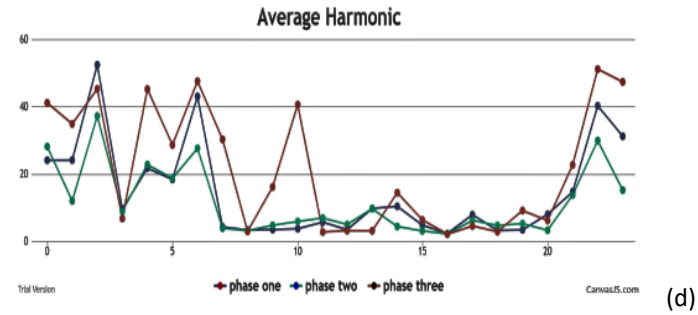
(a)



(b)

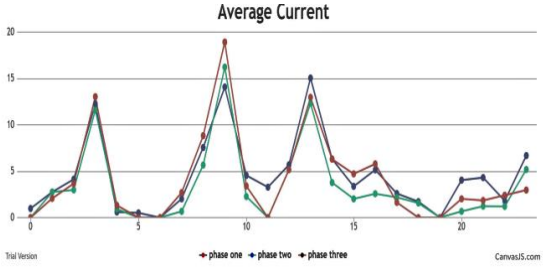


(c)

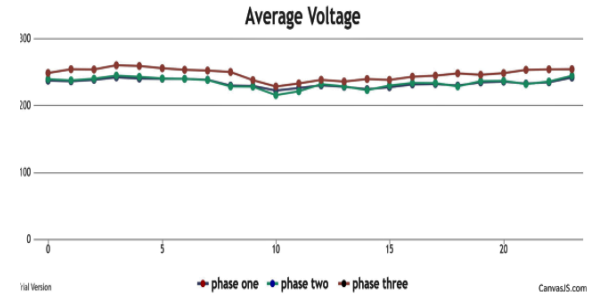


(d)

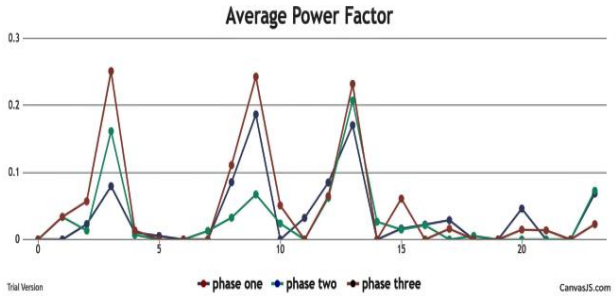
Figure 12: (a) The average current distortion, (b) average voltage distortion, (c) average power factor and (d) average harmonic in the insulation phase respectively.



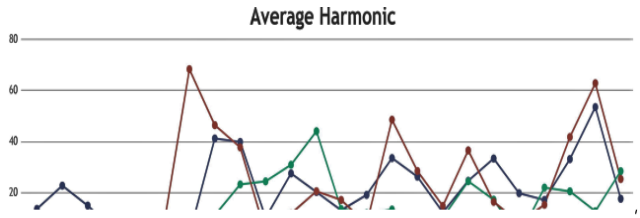
(a)



(b)

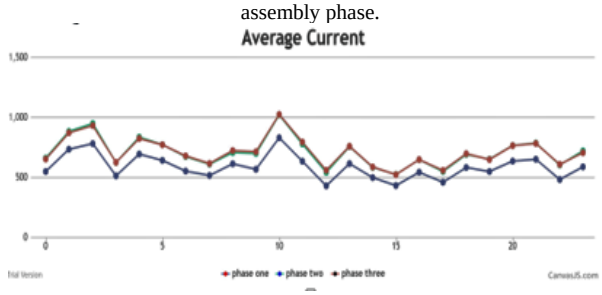


(c)

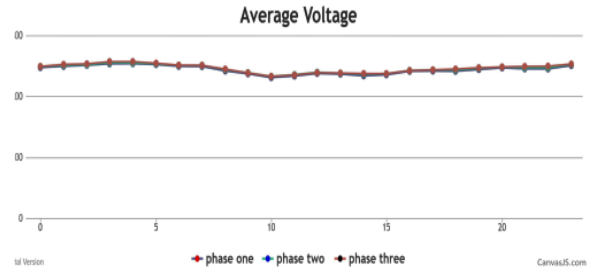


(d)

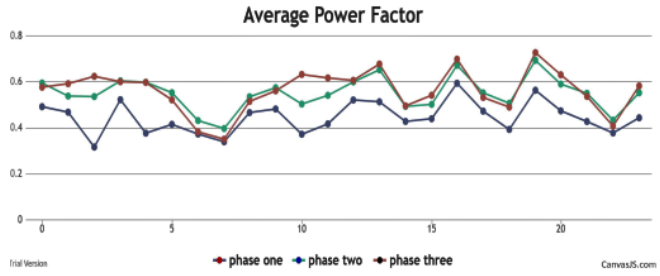
Figure 13: The Assembly phase (a) The average current distortion, (b) average voltage distortion, (c) average power factor and (d) average harmonic in the assembly phase.



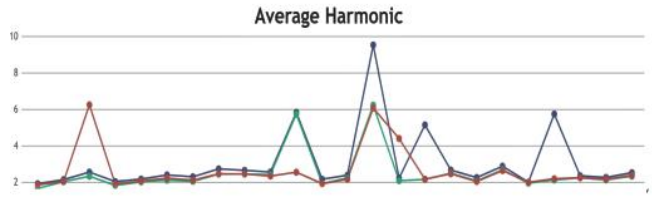
(a)



(b)

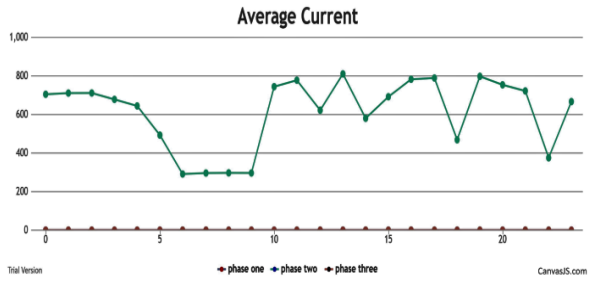


(c)

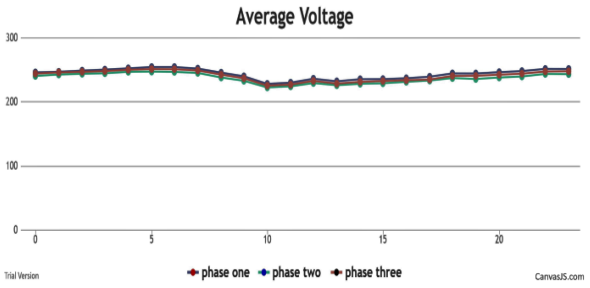


(d)

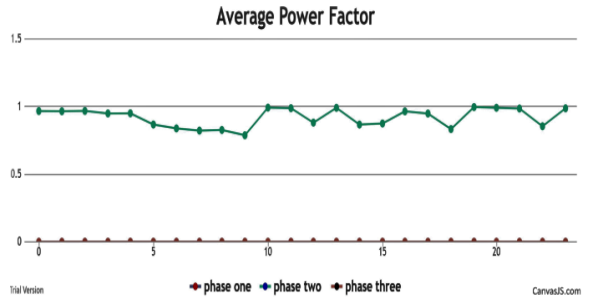
Figure 14: The average current distortion, average voltage distortion, average power factor and average harmonic readings respectively for the main line before correction.



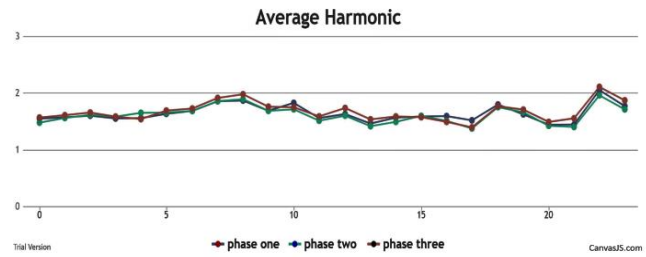
(a)



(b)



(c)



(d)

Figure 15: The average current distortion, average voltage distortion, average power factor and average harmonic readings respectively for the main line after correction.

The proposed system shown in Figure 9 will monitor the load current to detect the harmonics wave up to 49 th order and performs the computing process to analyse and calculate the power factor, then to transfer them to the high current capacity rectifier module along with the module processor in order to generate the compensation current in reverse direction, with overall response time less than 10ms and perform exceeding 97%.

Furthermore, Figure 16 illustrated the power consumption for one of the loads in four days. The measuring is taken by calculating the consumption of power for the three phases of one of the loads over time. Thus, it can be concluded that when using the proposed architecture a daily real time monitoring can be conducted which offers an understanding of energy usage and efficiency across different loads.

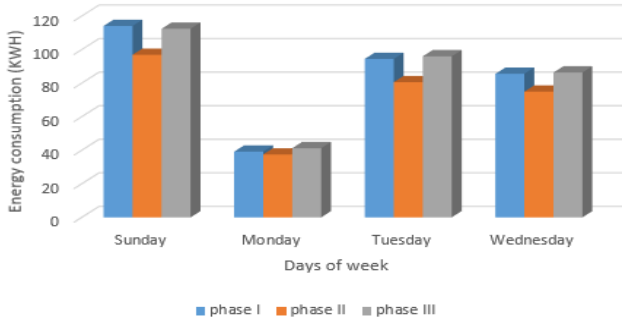


Figure 16: Energy consumption on one of the loads through four days

7. Discussion and conclusion

The manufacturing industry has been challenged with a lot of power quality problems in the overall including power factor, harmonics, unbalance ratios, power swells and dwells. Traditional techniques are moving towards correction steps as power factor correction using capacitor banks, synchronous power factor correction on individual loads, passive harmonic cancellations, active harmonic cancellations, soft switches to absorb spikes, voltage regulators, inverting technologies and others. However, those techniques are more focusing on correction of power rather than correction of manufacturing process that leads to the impact.

In the presented work , energy-aware process control model is proposed that is depended on IoT technology that enables the communication between PLC on machines and the energy management analytic platform on board to pinpoint the process that causes the phenomena to happen rather than the machine. For example, if two different machines working in parallel and doing the same operation type, machine x and machine y, that are installed in a factory. If the two machines are operating during three time shifts daily .Once the proposed model explore that machine x is doing wide spectrum of harmonic compared to machine y. The model will operate machine y for only two time shifts while machine x will operate for only one time shift. Hence the proposed model is able to stop the noisy machine and improve the overall power quality.

The model is simulated using MATLAB and is applied on a use case in factory that is specialized in cable industry. The results shows that the model pinpoints key areas of improvement of reducing the operation hours of a specific malfunctioning machine by one shift (eight hours) and increasing on the other side of operation hours of other two healthy machines doing the same operation type by 3 hours each has led to significant improvement of the power factor by 33% to reaches 0.93 versus 0.7 without impacting average OEE of the operation. On the other hand, the current total harmonic distortion is improved 10 times from 31.4% to 3.1% as shown in Figure 17 and 18 respectively.

The proposed model enables the cost-effective process management concept to achieve the same production results with better power quality and less energy consumption with minimum addition of correction nodes.

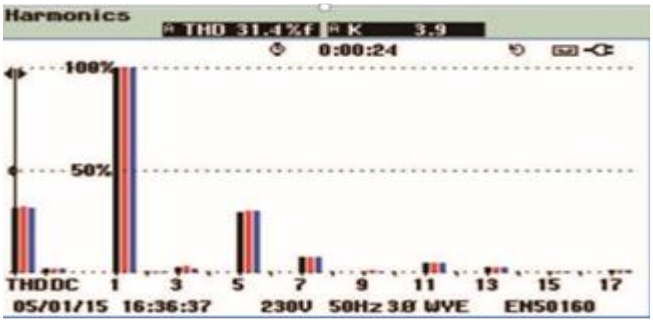


Figure 17: The Total Harmonic distortion (THD) before applying the proposed technique.

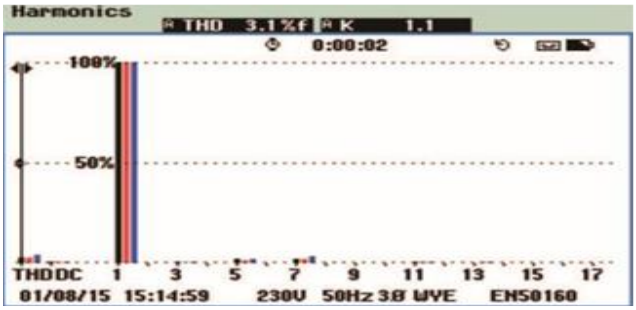


Figure 18: The Total Harmonic distortion (THD) after applying the proposed technique.

8. References

[1] N.N. Gomaa, K.Y. Youssef, M. Abouelatta, "An IoT-based energy efficient system for industrial sector," ICENCO 2019 - 2019 15th International Computer Engineering Conference: Utilizing Machine Intelligence for a Better World, 132–137, 2019, doi:10.1109/ICENCO48310.2019.9027440.

[2] V. Muni, R. Sangepu, "Effect of Power Quality Issues in Power System and Its Mitigation by Power Electronics Devices," Discovery, **28**(105), 72–79, 2015.

[3] M. Krones, E. Müller, "An approach for reducing energy consumption in factories by providing suitable energy efficiency measures," Procedia CIRP, **17**, 505–510, 2014, doi:10.1016/j.procir.2014.01.045.

[4] M.Z. Fortes, L.B. De Sousa, "Analysis of the Effects of Voltage Unbalance on Three-Phase Induction Motors," **3**(2), 29–37, 2018, doi:10.1007/s12053-015-.

[5] S. Rüstemli, E. Okuducu, M.N. Almalı, S.B. Efe, "Reducing the effects of harmonics on the electrical power systems with passive filters," Bitlis Eren University Journal of Science and Technology, **5**(1), 1–10, 2015, doi:10.17678/beujst.57339.

[6] S. A, "Power Quality and Harmonic Analysis in Three Phase Systems," Journal of Electrical & Electronic Systems, **7**(4), 2018, doi:10.4172/2332-0796.1000291.

[7] Z. Cao, Y.W. Chung, Y. Xiong, C.C. Chu, R. Gadh, "IoT based manufacturing system with a focus on energy efficiency," IEEE PES Innovative Smart Grid Technologies Conference Europe, 545–552, 2016, doi:10.1109/ISGT-Asia.2016.7796443.

[8] R.O. Aburukba, A.R. Al-Ali, T. Landolsi, M. Rashid, R. Hassan, "IoT based energy management for residential area," 2016 IEEE International Conference on Consumer Electronics-Taiwan, ICCE-TW 2016, 2016, doi:10.1109/ICCE-TW.2016.7521035.

[9] J. Huang, Z. Jiang, "Power Quality Assessment of Different Load Categories," Energy Proceda, **141**, 345–351, 2017, doi:10.1016/j.egypro.2017.11.041.

[10] D. Ogheneovo Johnson, "Issues of Power Quality in Electrical Systems," International Journal of Energy and Power Engineering, **5**(4), 148, 2016, doi:10.11648/j.ijepe.20160504.12.

[11] M.J. Ghorbani, H. Mokhtari, "Impact of harmonics on power quality and losses in power distribution systems," International Journal of Electrical and Computer Engineering, **5**(1), 166–174, 2015, doi:10.11591/ijece.v5i1.pp166-174.

[12] S.X. Duarte, N. Kagan, "A power-quality index to assess the impact of voltage harmonic distortions and unbalance to three-phase induction motors," IEEE Transactions on Power Delivery, **25**(3), 1846–1854, 2010, doi:10.1109/TPWRD.2010.2044665.

[13] M. Mazadi, S.H. Hosseinian, W. Rosehart, D.T. Westwick, "Instantaneous voltage estimation for assessment and monitoring of flicker indices in power systems," IEEE Transactions on Power Delivery, **22**(3), 1841–1846, 2007, doi:10.1109/TPWRD.2007.893188.

[14] Ö. Salor, B. Gültekin, S. Buhan, B. Boyrazoğlu, T. Inan, T. Atalik, A. Açik, A. Terciyanli, Ö. Ünsar, E. Altintaş, Y. Akkaya, E. Özdemirci, I. Çadirci, M. Ermiş, "Electrical power quality of iron and steel industry in Turkey," IEEE Transactions on Industry Applications, **46**(1), 60–80, 2010, doi:10.1109/TIA.2009.2036547.

[15] R. M Patil, M.S. Nagaraj, P.S. Venkataramu, "Design and Implementation of Harmonic Trap Filter in Conjunction with a Line Reactor using Simulink," International Journal of Electrical and Electronics Engineering, **3**(9), 7–16, 2016, doi:10.14445/23488379/ijeee-v3i9p102.

[16] A. Sharma, "Dependence of Power Factor on Inductive Loads for Microcontroller based Power Systems," IOSR Journal of Electrical and Electronics Engineering, **7**(2), 30–35, 2013, doi:10.9790/1676-0723035.

[17] P. Papechen, S. Sebastian, "Automatic Power Factor Correction Using Capacitor Banks," International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, **4**(7), 2321–2004, 2016, doi:10.17148/IJIREEICE.

[18] J.D.M. Mohamed Ibrahim A1 , Karthikeyan A2, Aravind K3, Gokulakrishna S4, "Minimizing Industrial Power Consumption Penalty By Using Thyristor Switching Capacitor," International Journal of Recent Trends in Engineering and Research, **4**(3), 741–749, 2018, doi:10.23883/ijrter.2018.4192.qidou.

[19] A.H. Bagdadee, L. Zhang, "A Review of the Smart Grid Concept for Electrical Power System," International Journal of Energy Optimization and Engineering, **8**(4), 105–126, 2019, doi:10.4018/ijeoe.2019100105.

[20] J.S. Vardakas, N. Zorba, C. V. Verikoukis, "A Survey on Demand Response Programs in Smart Grids: Pricing Methods and Optimization Algorithms," IEEE Communications Surveys and Tutorials, **17**(1), 152–178, 2015, doi:10.1109/COMST.2014.2341586.

[21] M.C.L. Padilla syamala, "An IOT Based Appliances Control for Smart Homes," in 2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP), 170–174, 2017, doi:10.1109/ICICIP.2013.6568171.

[22] M.I. Muhamad, N. Mariun, M.A.M. Radzi, "The effects of power quality to the industries," 2007 5th Student Conference on Research and Development, SCORED, (January), 2007, doi:10.1109/SCORED.2007.4451410.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1643-1650 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

FPGA-Based Homogeneous and Heterogeneous Digital Quantum Coprocessors

Valeriy Hlukhov*

Computers Department, Lviv Polytechnic National University, Lviv, 79013, Ukraine

ARTICLE INFO

Article history:
Received: 15 September, 2020
Accepted: 19 December, 2020
Online: 25 December, 2020

Keywords:
Digital quantum computer
Digital quantum coprocessor
Digital qubit
Quantum Fourier transform
FPGA

ABSTRACT

Quantum computers are heterogeneous device. It consists of a main CPU and a quantum accelerator. True quantum accelerator (or coprocessor) is analog and probabilistic device. Qubits are the basic building blocks of quantum computers. But qubits can be digital. A digital qubit is similar to RISC processor pipeline and is an unique chain of digital gates.

In this work, it is proposed to execute quantum routines in quantum computer not on the quantum chip but on the chip of a digital FPGA. This paper presents the architecture of such FPGA – an architecture of digital quantum coprocessor. The paper presents two types of digital quantum coprocessors - heterogeneous and homogeneous. The advantage of a homogeneous coprocessor is shown.

The IP Core generator was developed to create VHDL descriptions of digital quantum elements and digital quantum coprocessors in general.

In this paper heterogeneous quantum computer which consists of a main CPU and a FPGA-based quantum accelerator (coprocessor) has been proposed. And these FPGA-based digital quantum coprocessors can have a homogenous or heterogeneous structure. Quantum coprocessors have up to 1024 qubits in one FPGA. A homogeneous quantum coprocessor performs better than a heterogeneous one. Also, its implementation is easier.

The measured ratio of correct results for a 1024-qubit homogenous coprocessors is more then 50 %.

1. Introduction

This paper is an extension of work originally presented in IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT) [1].

Quantum algorithms are a mixture of classical logic and quantum routines which can be executed on the quantum chip. In [2] a heterogeneous quantum computer architecture was presented. According to [2] a quantum computer consists of both a classical and quantum computing part.

In this work, it is proposed to execute quantum routines in quantum computer not on the quantum chip but on the chip of a digital field programmable gate array, on a FPGA. This paper presents the architecture of such FPGA – an architecture of digital quantum coprocessor.

As an extension of the results described in [1] heterogeneous and homogeneous digital quantum coprocessors are considered. It is important to note that a heterogeneous quantum computer and proposed heterogeneous quantum coprocessor are completely different concepts.

Also, the number of digital qubits in the coprocessors under study was increased to 1024.

Von Neumann architecture had for a long time a single processor. Then homogeneous multi-core processor dominated the processor development. In the era of microprocessors, the understanding came that heterogeneity is the best way forward to improve the compute power. System architecture with heterogeneous accelerators includes the main CPU and heterogeneous coprocessors such as floating-point math coprocessors, graphics and neural accelerators, FPGAs. In [2] and [3] heterogeneous quantum computer which consists of a main CPU and a quantum accelerator has been proposed. In this paper heterogeneous quantum computer which consists of a main CPU and a FPGA-based quantum accelerator (coprocessor) has been proposed. And these FPGA-based digital quantum coprocessors can have a homogenous or heterogeneous structure and can be used at the lowest level of full-stack quantum accelerators along with quantum chips and quantum simulators [3].

A quantum computer contains N qubits and a digital quantum coprocessor contains N digital qubits. In both cases as a result of any calculations they can produce any of 2^N results. But the same

*Corresponding Author: Valeriy Hlukhov, Lviv Polytechnic National University, Computers Dept., Lviv, 79013, Ukraine, Glukhov@polynet.lviv.ua

calculations can lead to different results. And every i-th result will be produced with probability p_i.

A qubit can be thought of as a device that has a group of inputs for data and instructions that control its behavior. Unlike a real qubit, a digital qubit can have an additional group of outputs. The exact qubit state code is generated on this group of outputs. An important element of the digital qubit is a pseudo random number generator (PRNG). Together with the qubit state code, the pseudo-random code is used to generate a probabilistic result at the one-bit output of the qubit (Figure 1) [4].

Quantum volume VQ is a metric that measures the performance of a quantum computer's capabilities and error rates [5]. In the simplest case, the quantum volume is d·N. Now the quantum volume of a real quantum computers is very small (at the level of milliseconds), but for digital quantum coprocessors it is practically unlimited.

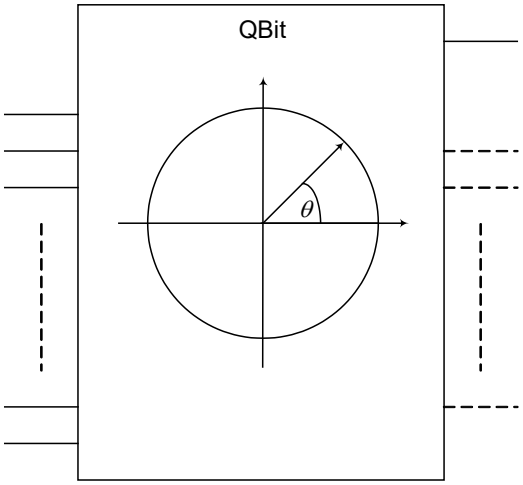


Figure 1: Digital qubit symbol

The DiVincenzo [6] criteria are conditions necessary for constructing quantum computer. Not a single word about the physical nature of a quantum computer is included in this criterion. Therefore, you can try to create digital quantum computer. It can be created either as a software model or as a hardware device.

"You can run a simulator on your phone that can run a 20-qubit system. But once you get up close to 50, the amount of memory you need to run a simulator gets into the petabytes [of memory]. Over 50 and there is no way you can pretend to be a quantum computer. At that point you really need a true quantum computer" [7].

The number of qubits required for solving practical problems is now estimated at several thousand [8]. Creating a true quantum computer with so many qubits is a very difficult task, and a digital quantum coprocessor can already be created on one FPGA.

The purpose of the article is to demonstrate the advantages of digital quantum computers over true ones. For this, it is necessary to develop a circuit for a digital qubit and a digital quantum coprocessor as part of a digital quantum computer. To compare different coprocessors, it is necessary to develop methods for their comparison. In this works digital quantum computers with 2^n qubits (n = 5, 6, …, 10) have been used for research.

Also, the purpose of the work is to determine the timing and hardware characteristics of the developed digital quantum coprocessors.

An important task is the problem of the reliability of the results of modeling the operation of digital quantum coprocessors. In this work, it was successfully solved, it was shown that simulation results coincide with the results obtained during prototyping.

2. Theoretical Foundations of Quantum Computing

$$|0\rangle, |1\rangle, ..., |N-1\rangle$$

$$|j\rangle$$

Along with measurement, the quantum superposition can be changed under external influence.

$$|\psi\rangle = \sum_{j=0}^{N-1} \lambda_i |j\rangle \quad p_j = \lambda_j^2 \quad P = \sum_{j=0}^{N-1} \lambda_i^2 = 1$$

$$p_0 = \cos^2\theta \qquad p_1 = \sin^2\theta$$

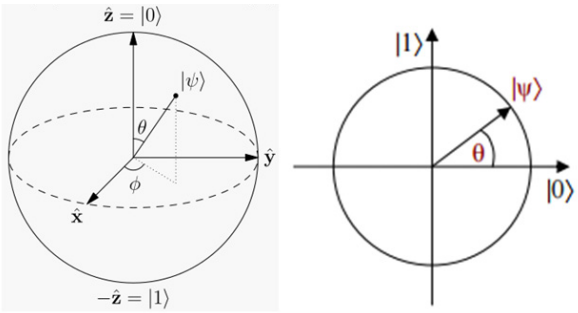


Figure 2: A Bloch sphere (left) and a unit circle (right)

A true quantum computer is an analog device and it has no memory. It has only gates. Therefore, there are no quantum programs. In the drawing of a true analog quantum computer, a sequence of quantum gates shows only the time sequence of qubit state changes. Software tool [14] allows to create circuits from quantum gates, simulate their work and manage it with C-like language instructions. This is very similar to the FPGA design.

In the schema of a digital quantum computer, a sequence of digital quantum gates shows both the time sequence of qubit state changes and the relative physical position of the gates themselves in the space.

A good illustration of a qubit can be an electron spin. An electron can change its spin in time from 1 to 20 nanoseconds [15]. This time can serve as a base operation time for compare the performance of true and digital quantum coprocessors.

To compare the capabilities of true and digital quantum computers, one can calculate the quantum volume as a metric that measures the performance of a quantum computer's capabilities and probability that the qubit will work for some time t without failure:

$$QV = N*p(t),$$

$p(t) = e^{-t/MTBF}$ is probability that the qubit will work for some time t without failure where (MTBF - Mean Time Between Failures);

N is qubits number.

For true qubits MTBF is microseconds and millisecond and p(t) run to 0 when t > 1 ms, QV << N.

For digital qubits MTBF is practically unlimited (MTBF of modern FPGA is at level 50 years) and p(t) run to 1. So, for digital quantum coprocessor QV = N.

The results of comparison of analog data processing methods, which are used in true quantum computers, and digital methods are summarized in the Table 1. All this speaks about the prospects of creating hardware digital quantum computers.

3. The Structure of FPGA-based Digital Quantum Coprocessor

A classical computer controls the operation of a quantum coprocessor (Figure 3) provides it with an input data, instructions and checks the result of its work [4]. This interaction is well described as a full-stack of the layers of an accelerator at [3].

A generalized functional diagram of full-stack hardware resources of quantum computer with realized in FPGA quantum coprocessor is given in Figure 4 [4].

Table 1: Comparison of analog and digital data processing methods

Characteristic	Analog processing methods	Digital processing methods
Speed	+	-
Versatility	-	+
Microminiaturization	-	+
Accuracy	-	+
Zoom	-	+
Transmission in space	-	+
Transmission in time (memory)	-	+
Immunity	-	+
Reliability	-	+
Testing, debugging, diagnostics	-	+
Quantum volume	-	+

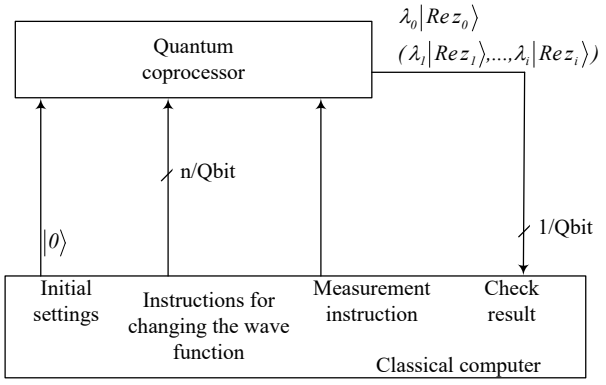


Figure 3: A classical computer with quantum coprocessor.

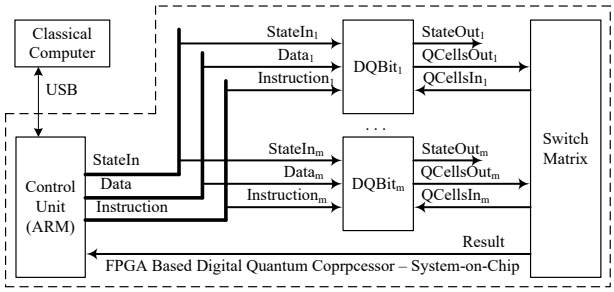


Figure 4: A digital quantum coprocessor for classical computer.

The top-level functionality of this stack is provided by the classic host computer. Functioning at the microarchitecture level is provided by embedded in FPGA microprocessor (AWP, control unit).

And directly quantum computing is provided by a set of digital qubits and a switch matrix which connects the qubits to each other and transmits the final state code of all of them or only those required at the moment to the control unit.

The connections between qubits can be static or dynamic. In this work static connections have been used. They do not change while the computer is running.

A digital qubit consists of j series-connected digital quantum cells (Figure 5).

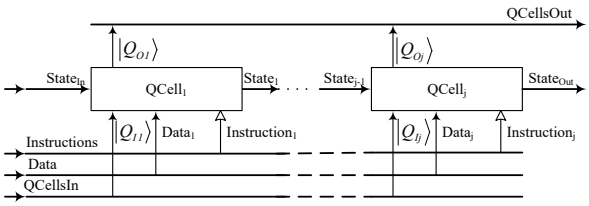


Figure 5: RISC processor-like structure of a digital qubit DQBit.

$|Q_{Oj}\rangle$

$|Q_{Ij}\rangle$

$D = arcsin\sqrt{A}$

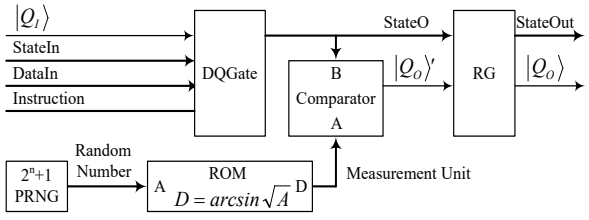
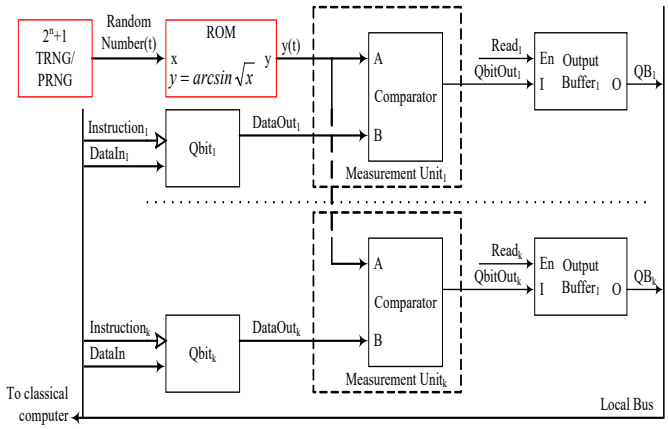


Figure 6: A digital quantum cell QCell.



A digital quantum gate transforms the input qubit state code into the output code under the influence of the instructions and their data, as well as under the influence of measured states of another qubits.

Inverse functional transformation allows the creation and research of two types of digital quantum coprocessors:

each quantum cell of the heterogeneous coprocessor has its own pseudo-random number generator and its own functional transformer;

a homogeneous coprocessor contains only one pseudo-random number generator and only one functional transformer for all quantum cells, for all quantum qubits (Figure 7).

A simplified topology of a part of a FPGA chip with resources available to the user for design is shown in the Figure 8. These resources are sets of configurable logic blocks CLB, programmable switching machines PSW and I/O blocks IOB. The PSM and the wires laid on the chip between them and CLB make it possible to connect the CLB to each other. CLB can be programmed to act as a qubit (QB) or quantum cell (QC). This allows many digital qubits to be placed on one chip. The interaction of qubits and the control of qubits is carried out by digital methods using PSM. The use of COTS FPGAs makes it possible to create digital quantum coprocessors with a large number of qubits, which can operate for a long time under normal conditions.

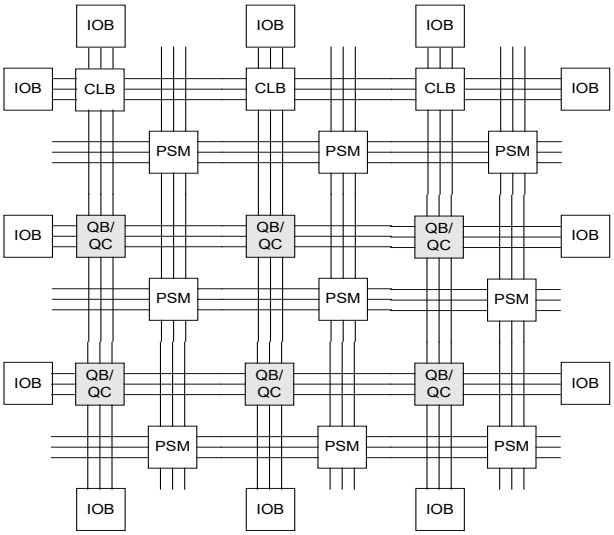


Figure 8: Simplified topology of digital quantum coprocessor FPGA chip

Recently, other developers have also tried to connect digital technologies with true quantum computers:

place true qubits on the crystal (similar to the Figure 8 shown) and organize interaction between them using digital methods [16]. It uses a custom crystal that can operate at ultra-low temperatures (near 4° K);

carry out optical control of true qubits embedded in the chip [17];

use a pseudo-random number generator in true quantum computers [18].

In this study digital quantum coprocessors were implemented for the case of wave function real amplitudes (Figure 2) and for polar coordinate system to represent the movement of the vector (Figure 13) [12].

Algorithm design steps for digital quantum computer design on FPGA are standard for FPGA design:

to create or to find an algorithm for solving the problem;

to find or to create a mathematical description of the solution to the problem;

for FPGA-based circuits to create graphic symbols of library elements and their descriptions in hardware description language;

- to create a schema from library elements;
- to simulate the created schema;
- to implement the project;
- to verify the project;
- to make a prototype project.

The IP Core generator was developed to create VHDL descriptions of digital quantum elements and schema of digital quantum coprocessors in general.

4. Testing Digital Quantum Coprocessors

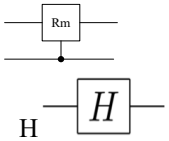
The quantum Fourier transform (QFT) as a part of Shor's algorithm [19] (Figure 9) for factorization [11] was chosen to compare different variants of quantum coprocessors.

The QFT is defined as

$$\sum_j \alpha_j |j\rangle \rightarrow \sum_k \tilde{\alpha}_k |k\rangle$$

$$\tilde{\alpha}_k \equiv \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{2\pi i j k / N_{\alpha_j}}$$

Only 2 types of digital quantum gates are required during QFT (Figure 10, Figure 11):



- Hadamard transform H

$$R_m = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{2\pi i / 2^m} \end{pmatrix}, \quad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

When implementing a digital qubit in the form of a chain of digital quantum gates and implementing only a QFT circuit on FPGA, there is no need to change the functions of the quantum gates. Therefore, the functions of each gate are determined only by their circuit. And there is no need for instructions that change the

function of the gates. Therefore, they are not shown in the Figure 12, but are shown in the general diagram Figure 4. And the data path in the Figure 12 is a chain of links named Q*, along which the changing qubit state code goes from one gate to another. The measured states of a qubit, which are called q*, are transferred from one gate to another as required by the algorithm for solving the problem. QFT determines spectrum of qubits states.

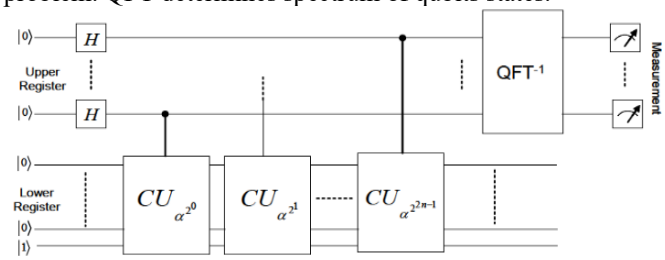


Figure 9: Quantum factorization by Shor's algorithm

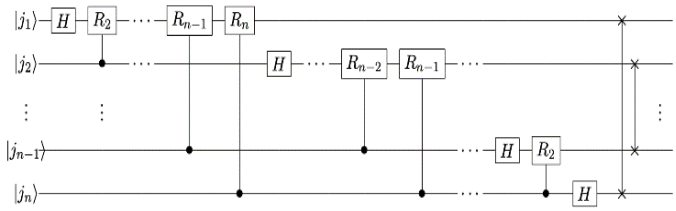


Figure 10: QFT drawing

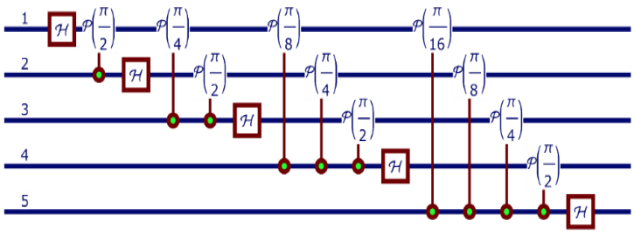


Figure 11: Simplified drawing of QFT (5-qubits)

The diagram for FPGAs corresponding to the Figure 11 is shown in the Figure 12.

The QFT results were obtained as described in [12] and [13].

5. Simulation and Verification of FPGA-based Digital Quantum Coprocessor

For analysis the input state of qubits, which can conditionally be described as |XXX..X0>, where X corresponds to the neutral position of the vector in the unit circle - at angle of π/4, was selected. The probability of measuring the input state with odd code is p_{odd}=0, and with even codes is $p_{even} = 100/2^{n-1}$ %, where n is qubits quantity and the spectrum of qubit states at the QFT input will look like in Figure 14 (for 4 qubits).

QFT simulation results for this case are shown in Figure 15. The 4 qubit digital quantum coprocessor determines the period of the input states spectrum with probability 38,5 %, so it correctly executes a QFT.

Also digital quantum coprocessors with up to 1024 qubits, have been investigated [13], [20]. Similar studies were performed for other periods in input states spectrum (for 0, 1, 2, 4, …, 2^{n-1} periods) [13].

Implementation results are summarized in Table 2.

Table 2: Implementation Results of Heterogonous Digital Quantum Co-Processors (2^{n-1} periods)

Qubit number, n	CLK period (ns)	LUT number	RAM number	FPGA resoirces, % R	FPGA 7z type [21].
128	2.353	32352		14	045
64	4.289	16043		15	010
32	3.874	7983			010
16	3.606	4000			010
10	3.976	2522			010
10	4.180	4307		24	010
10	5.137	3754			010
10	5.281	3650	20	16	010

32-qubit digital quantum coprocessor (Figure 18) was implemented on ZedBoard Zynq-7000 ARM_FPGA SoC Development Board (Digilent, Figure 17) for verification.

Verification and simulation results are almost the same (Table 3). This indicates the creation of both a high-quality model of a digital quantum coprocessor and its prototype, which can be used in further research. Each experiment was repeated 4000 times to determine the probability of getting the correct result (Figure 19). The average level of correct results during the verification of a heterogeneous coprocessor is 9%, during the simulation of a homogeneous processor is 61%, during the verification of a homogeneous processor is 58%.

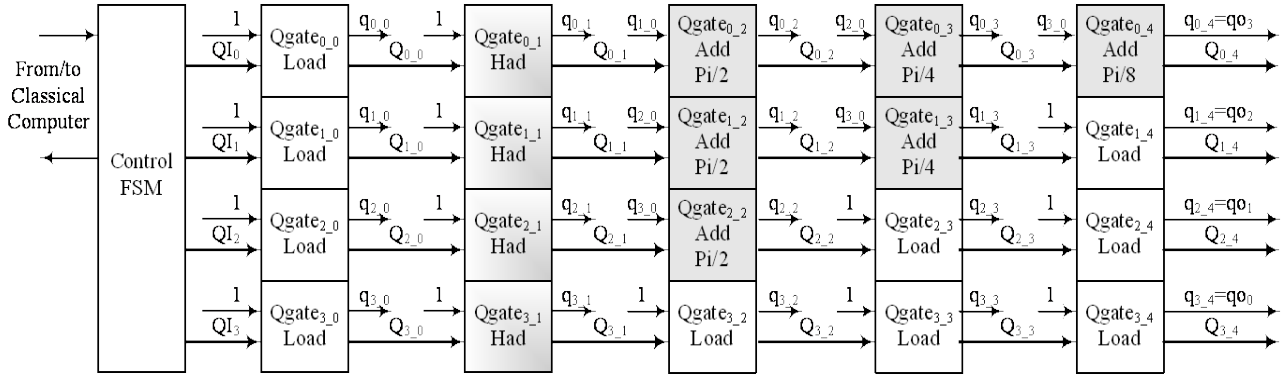


Figure 12: FPGA schema of QFT (4-qubits)

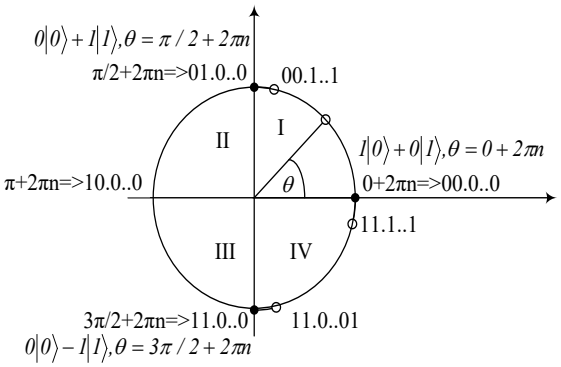


Figure 13: A polar coordinate system (angle θ codes)

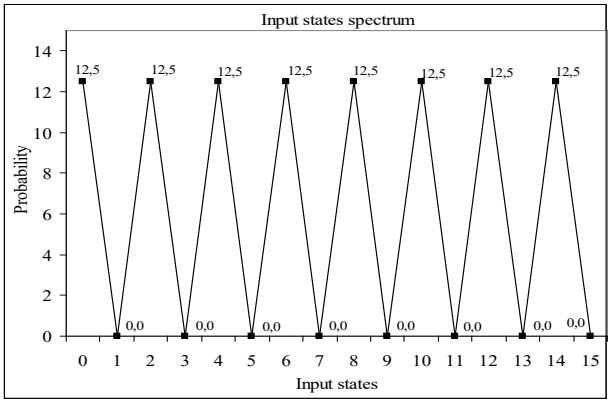
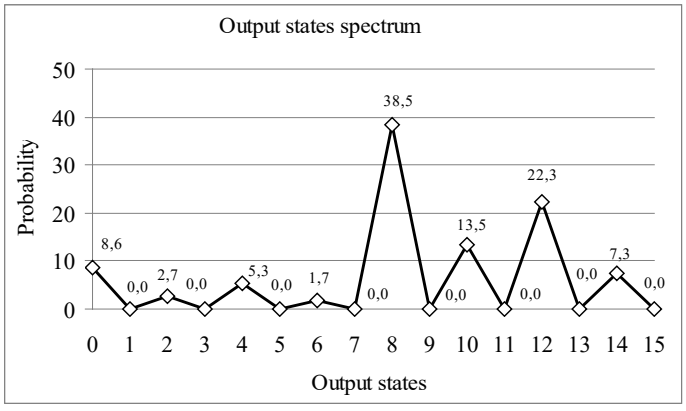


Figure 14: Input state spectrum

$$\theta_3 = \theta_2 = \theta_1 = 45°, \theta_0 = 0°$$

A homogeneous digital quantum coprocessor has shown better results than a heterogeneous one. Therefore, further research was carried out on a homogeneous digital quantum coprocessor.

The probabilities of obtaining correct results during the verification for the operation of multi-qubit homogeneous digital quantum coprocessors are shown in the Figure 21. The work of coprocessors with 32, 64, 128, 256, 512 and 1024 qubits were investigated.



$$\theta_3 = \theta_2 = \theta_1 = 45°, \theta_0 = 0°$$

Figure 16: Digital quantum computer

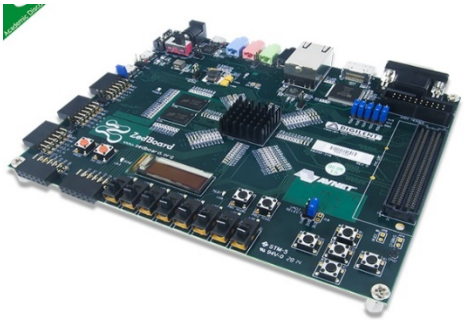


Figure 17: FPGA Development Board

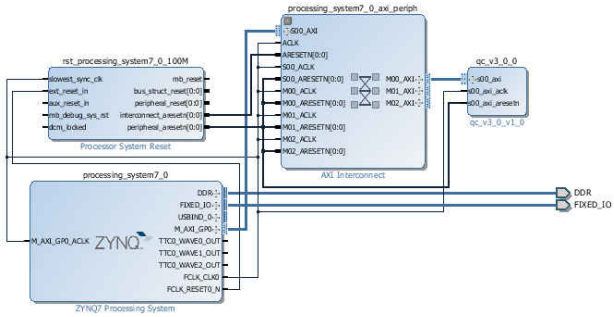


Figure 18: Schema of digital quantum coprocessor for FPGA

Table 3: Simulation vs verification

Qubits number N	Qubits width W, bits	True results probability % H	Cycle T, ns	Process
32	3	61	3.874	Simulation
32	3	58	8.0	Verification

FPGA resources for 32-qubit heterogeneous and homogeneous coprocessors are presented in the Table 4.

FPGA topology of 32-qubit digital quantum coprocessor is presented in Figure 20.

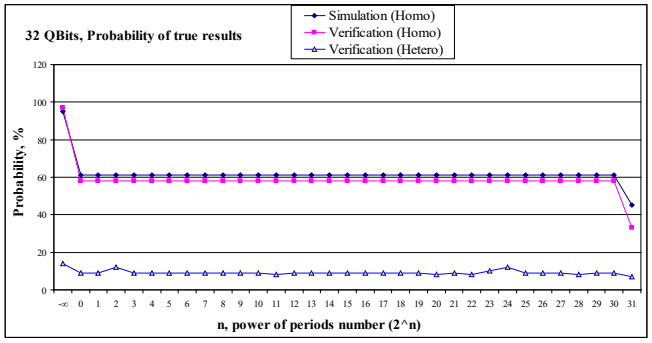


Figure 19: Probability of true results in digital quantum coprocessors

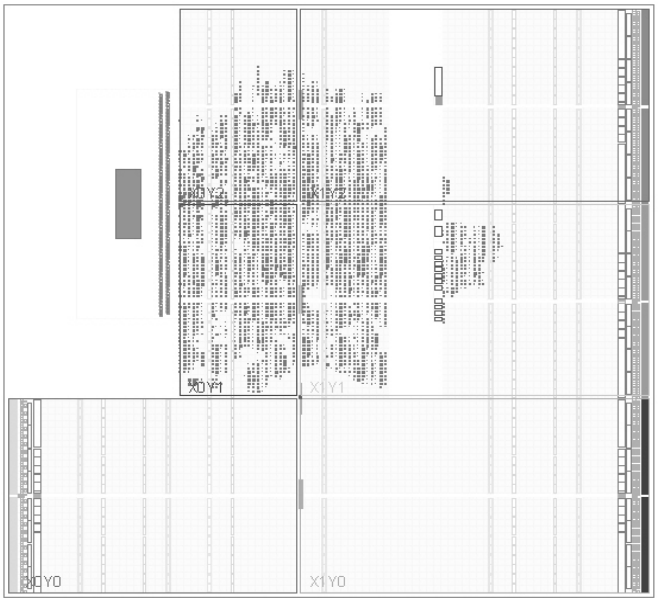


Figure 20: 32-qubit digital quantum coprocessor topology

This result indicates the creation of both a high-quality model of a digital quantum coprocessor and its prototype, which both can be used in further research.

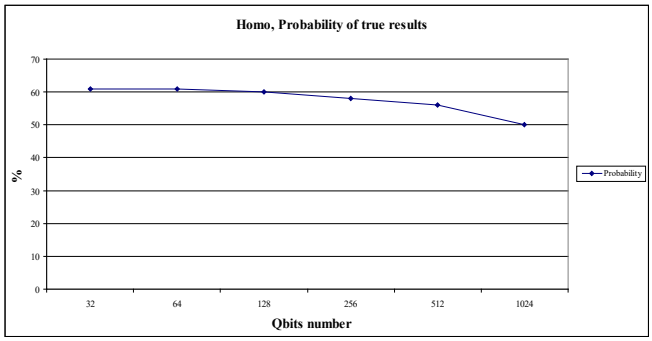


Figure 21: Probability of true results in homogenous digital quantum coprocessors

Table 4: FPGA resources for 32 qubit coprocessors

	Homo	Homo	Hetero	Hetero
	Used	Util %	Used	Util %
Slice LUTs	5831	10.96	7669	14.42
# of logical nets	21575		28147	

6. Conclusion

The paper shows the advantage of digital quantum computers over true computers in terms of quantum volume. The paper presents heterogeneous quantum computer which consists of a main CPU and a FPGA-based quantum accelerator (coprocessor). This FPGA-based digital quantum coprocessors can have a homogenous or heterogeneous structure. Structures of digital quantum qubits and cells are also presented.

The IP Core generator was developed to create VHDL descriptions of digital quantum elements and digital quantum coprocessors in general.

32 qubits homogeneous digital quantum coprocessor generates correct quantum Fourier transform results with a probability of 61% during simulation and 58% during prototyping.

1024 qubits homogeneous digital quantum coprocessor generates correct quantum Fourier transform results with a probability more than 50 % during simulation.

Homogeneous digital quantum computer generates the correct results of the quantum Fourier transform 7 times more often than heterogeneous one.

The hardware cost of homogeneous digital quantum coprocessors is 20% less than that of heterogeneous coprocessors.

The number of logical nets in homogeneous digital quantum coprocessors is 20% less than in heterogeneous coprocessors.

Digital quantum coprocessors verification results are practically the same as simulation results. This indicates the creation of both a high-quality model of a digital quantum coprocessor and its prototype, which both can be used in further research.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] V. Hlukhov, "FPGA Based Digital Quantum Computer Verification," in 2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DeSSerT), 178-182, 2020, doi: 10.1109/DESSERT50317.2020.9125077.

[2] X. Fu et al., "A heterogeneous quantum computer architecture," in 2016 ACM International Conference on Computing Frontiers (CF'16), 323–330, 2016, doi: 10.1145/2903150.2906827.

[3] K. Bertels, A. Sarkar, T. Hubregtsen, M. Serrao, A. Mouedenne, A. Yadav, A. Krol, I. Ashraf, "Quantum Computer Architecture: Towards Full-Stack Quantum Accelerators," in 2020 EDAA Design, Automation & Test in Europe Conference & Exhibition (DATE), 1–6, 2020, doi: 10.23919/DATE48585.2020.9116502.

[4] V. Hlukhov, B. Havano, "FPGA-based Digital Quantum Coprocessor," Advances in Cyber-Physical Systems, **3**(2), 12-31, 2018, doi: 10.23939/acps2018.02.067.

[5] N. Moll, P.Kl. Barkoutsos, L.S. Bishop and others, "Quantum optimization using variational algorithms on near-term quantum devices," Quantum Science and Technology, **3**(3), 030503, 2017, doi:10.1088/2058-9565/aab822.

[6] D. DiVincenzo, "The physical implementation of quantum computation," Fortschritte der Physik: Progress of Physics, **48**(9–11), 771–783, 2000.

[7] E. Scannell, "IBM quantum computers' usefulness in sight - using binoculars," in TechTarget News. available at: https://searchdatacenter.techtarget.com/feature/IBM-quantum-computers-usefulness-in-sight-using-binoculars, accessed 25 Nov. 2020.

[8] "Applying Moore's Law to Quantum Qubits," Quantum Computing Report, avaliable at: https://quantumcomputingreport.com/our-take/applying-moores-law-to-quantum-qubits, 2019, accessed 25 Nov. 2020.

[9] E. Grumbling, M. Horowitz (eds.), Quantum computing: progress and prospects, The National Academies of Sciences, Engineering, and Medicine, Washington, DC: National Academies Press, 2019.

[10] M. Khalil-Hani, Y. Lee, M. Marsono, "An accurate FPGA-based hardware emulation on quantum Fourier transformm," in 13th Australasian Symposium on Parallel and Distributed Computing (AusPDC 2015), 23–30, 2015.
[11] S. Gushanskiy, V. Pereverzev, "Simulation of quantum computing using hardware cores," (in Russian), Scientific Journal of KubSAU, 123(09), 545–557, 2016, avaliable at: http://ej.kubagro.ru/2016/09/pdf/37.pdf, accessed 02 Dec. 2020.
[12] V. Hlukhov, "Implementing quantum Fourier transform in a digital quantum coprocessor," Advances in Cyber-Physical Systems, **4**(1), 6–13, 2019, doi: 10.23939/acps2019.01.006.
[13] V. Hlukhov, "FPGA-based K-qubit digital quantum coprocessor," Electrotechnic and Computer Systems, **31**(107), 104–117, 2019, doi: 10.15276/eltecs.31.107.2019.10.
[14] "Microsoft Quantum Documentation," 2020, available at: https://docs.microsoft.com, accessed 02 Dec. 2020.
[15] J. Smith, P. Dalgarno, R. Warburton, A. Govorov, K. Karrai, B. Gerardot, P. Petroff, "Voltage control of the spin dynamics of an exciton in a semiconductor quantum dot," Physics Rev. Lett., **94**, 197402, 2005.
[16] S. Moore, "What Intel Is Planning for The Future of Quantum Computing: Hot Qubits, Cold Control Chips, and Rapid Testing", IEEE Spectrum, 2020, available at: https://spectrum.ieee.org/tech-talk/computing/hardware/intels-quantum-computing-plans-hot-qubits-cold-control-chips-and-rapid-testing, accessed 24 Aug 2020.
[17] Ch.P. Anderson, A. Bourassa, K.C. Miao, G. Wolfowicz, P.J. Mintun, A.L. Crook, H. Abe, J.U. Hassan, N.T. Son, T. Ohshima, D.D. Awschalom, "Electrical and optical control of single spins integrated in scalable semiconductor devices", Science, **366**(6470), 1225–1230. 2019. doi: 10.1126/science. aax9406.
[18] K. Miyamoto, K. Shiohara, "Reduction of Qubits in Quantum Algorithm for Monte Carlo Simulation by Pseudo-random Number Generator," Physical Review A., 022424, 2020, doi: 10.1103/physreva.102.022424.
[19] P. Shor, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer," in 35th Annual Symposium on Foundations of Computer Science, 124–134, 1994, available at: https://www.jstor.org/stable/2653075?seq=1, accessed 25 Nov. 2020.
[20] V. Hlukhov, "Comparison of Homogeneous and Heterogeneous Digital Quantum Coprocessors," in International Workshop on Computational Methods and Information Transformation Systems, a satellite of IEEE 2020 XV International Scientific and Technical Conference on Computer Science and Information Technologies, 70–73, 2020.
[21] "DS190 (v1.11.1) Zynq-7000 SoC Data Sheet: Overview," 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1651-1656 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Type 2 Diabetes Risk and Physical Activity in outpatients treated in Health Centers in a District of North Lima, 2020

Deisy Chipana-Collahua[1], Mariluz Chipana-Collahua[1], Rosa Villegas-Ortiz[1], Brian Meneses-Claudio[2,*], Hernan Matta-Solis[1]

[1]*Faculty of Health Sciences, Universidad de Ciencias y Humanidades, 15314, Lima-Perú*

[2]*Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, 15314, Lima-Perú*

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 09 December, 2020
Online: 25 December, 2020

Keywords:
Diabetes risk
Lifestyle
Outpatients
Physical activity
Risk factors

ABSTRACT

Diabetes Mellitus Type 2 is a chronic disorder that affects the way the body metabolizes sugar (glucose) in the blood and depends on a combination of risk factors, such as genes and lifestyle. Although certain risk factors such as family history, age, or ethnicity cannot be changed, those related to diet, physical activity, and weight can be changed. These lifestyle changes can affect the probability of developing type 2 diabetes and its complications. In this research work, the prevention of the patient to develop type 2 diabetes with the option of leading a healthy lifestyle is proposed and reinforcing the population of each establishment with health education. It is a quantitative approach, a non-experimental, descriptive, and correlational study, with a population of 300 outpatients from a district of North Lima, who answered a questionnaire with sociodemographic data and the instruments of the level of physical activity IPAQ and risk of FINDRISK type 2 diabetes mellitus. In the results with respect to the level of physical activity, low 125 (41.7%) predominated and at risk of type 2 diabetes mellitus, 113 (27.7%) predominated. In conclusion, the population should be educated to have an adequate lifestyle by improving physical activity to prevent the risk of contracting diabetes.

1. Introduction

Type 2 diabetes is currently considered one of the non-communicable diseases with the greatest negative impact and projection on the quality of life of the world population and represents a world health problem [1]; it is diagnosed in 85-95% of all those who suffer from it, with a higher percentage in less developed regions (one in 14 adults will have diabetes mellitus). It is estimated that nearly four million deaths a year are directly related to this condition, based on changes in the lifestyle of the population, associated with sedentary lifestyle, obesity, high blood pressure and other cardiovascular risk factors [2]. For these reasons, it is the fourth cause of death in the entire universe [3].

Globally, the number of patients diagnosed with type 2 diabetes mellitus has exceeded the references by more than 100% increase in the last 20 years [4]. In its development, it is related to lifestyle, age, sex, ethnicity/race and family history of Diabetes Mellitus Type 2 (DMT2) [5]. The Body Mass Index (BMI) increased means the probability of developing DMT2, to a greater extent in individuals with a Normal BMI compared to the obese, identified physical activity as a fundamental strategy in its prevention [6].

In Peru, the number of people with diabetes mellitus is increasing very rapidly and the main cause of its rapid increase is the important change in the lifestyle of the Peruvian population, whose current characteristics represent an excessive intake of high-calorie foods (junk food and sugary drinks), and what is equally worrisome: a reduction in physical activity that directly leads to a high morbidity of overweight and obesity [7].

Regular physical activity is essential for disease prevention, particularly chronic non-communicable diseases; Health promotion; quality of life; and reduction of mortality [8].

In [9], the study prevalence and incidence of type 2 diabetes mellitus in the general adult population of Peru, they define the prevalence of type 2 diabetes mellitus nationwide according to the National Survey of Nutritional, Biochemical, Socioeconomic and Cultural Indicators study was 5.1% in 2004-2005, and it increased to 7.0% in 2010-12, Perú Diabetes (PERUDIAB, national study).

* Corresponding Author: Brian Meneses-Claudio, Sr., +51 1 950159924 & bmeneses@uch.edu.pe

The Risk Factors of Non-communicable Diseases study in Lima and Callao reported a prevalence of 3.9% in 2006, and in 2010-2012 the Centre of Excellence in Chronic Diseases (CRONICAS) study also in Lima found this indicator at 5.5%. Regarding the incidence, three of the nine studies analyzed PERUDIAB estimated the incidence of diabetes in 19.5 new cases per 1,000 person-years. The CRONICAS study, in selected populations of Lima, Puno and Tumbes, found the same result. The Perú's Rural to Urban Migrants (PERU MIGRANT) study estimated a 50% lower incidence than the previous two studies, although more than half of the population in this study lived in rural areas or were migrants from rural to urban areas.

In [10], the author studied physical activity in outpatients with diabetes mellitus type 2 concluded that 88% of diabetic patients had inactivity or minimal physical activity, estimating that patients with diabetes have a higher level of inactivity than the general population.

In [11], the author indicates that leading a healthy lifestyle pattern could lead to a lower risk of diabetes mellitus type 2, which identifies 142 associations related to the risk of developing DMT2. In which, they pointed out that Eleven associations (8%) presented evidence for the risk of DMT2 (low consumption of whole grains, metabolically healthy obesity, increased sedentary time). Thirty-four associations (24%) were associated with an increased risk of DMT2 and presented highly suggestive evidence, they were the following: high BMI, low educational level, gestational diabetes, increased consumption of processed meat, high level of total physical activity. Twenty-nine associations had suggestive evidence (20%), and 42 associations had weak evidence (30%) of DMT2 risk. All but six associations with convincing or highly suggestive evidence were based exclusively on prospective cohort studies.

The main objective of the study is to determine the risk of developing diabetes mellitus type 2 in relation to its level of physical activity in adults treated in health centers in a district of North Lima. It is important the need to modify the lifestyles of the population, associated with sedentary lifestyle, obesity, high blood pressure and other cardiovascular risk factors and thus identify physical activity as a fundamental strategy in its prevention.

Diabetes mellitus risk is understood as the probability of presenting or suffering from type 2 diabetes in adults attended in District-Rímac health centers, which can be evaluated by the presence of indicators such as age, body mass index, anthropometry waist, physical activity, consumption of vegetables or fruits, blood pressure medication, diagnosis of high glucose values and a history of diabetes mellitus [12].

The technique to be used in the research work will be the survey, which a significant amount of data will be obtained optimally and effectively. To measure the risk of type 2 diabetes mellitus, the data collection instrument will be used, the Finnish Diabetes Risk Score (FINDRISK) test that considers eight items grouped into eight dimensions and the level of physical activity will also be evaluated using the International Physical Activity Questionnaire (IPAQ), which considers 27 items grouped into three dimensions.

The data collection process will be carried out during the month of June through surveys, after the data will initially be entered into a data matrix designed in the Microsoft Excel 2013 program, then will be exported to the IBM SPSS Statistics version 24.0 program. It will apply descriptive statistical tests such as absolute frequencies, relative frequencies, and measures of central tendency, determining the risk of developing diabetes and the level of physical activity.

The following research work is structured as follows: In section II, the methodology that will determine the research work will be presented, with emphasis on the instrument, techniques to be used, scope of the study, valid inclusion, and exclusion criteria for the estimated population, among others. In section III, the results will show the tables and graphs obtained after data processing with an emphasis on determining the risk of developing type 2 diabetes mellitus and the level of physical activity of the studied population. In section IV, it will present the discussions of the research work and finally in section V, the conclusions obtained through the present research work, as well as the future work that leads to the timely detection of type 2 diabetes mellitus and the adoption of healthy styles.

2. Methods

The steps to carry out the survey as the population and sample were obtained, type of study, inclusion and exclusion criteria, analysis of the variables, data collection, location and application of the IPAQ and Finnish Diabetes Risk Score (FINDRISK) instruments. The paper before its application went through an ethics committee in which it was approved.

2.1. Population and Sample

It was applied to 300 outpatients at risk of developing diabetes, in the nine Health Centers of the Rimac district, considering the total distribution of patients who participated in the study, the sample corresponds to 35 participants from each center: H.C. Rímac, H.C. Caquetá and H.C. Flor de Amancaes. Likewise, 34 participants from Leoncio Prado, San Juan de Amancaes, Mariscal Castilla, Ciudad y Campo and Villa Los Ángeles.

2.2. Inclusion and Exclusion Criteria

Inclusion criterion is that all patients who receive regular care and wish to collaborate voluntarily in being evaluated to obtain a risk diagnosis participate and exclusion criteria; Patients who did not wish to participate in the study are excluded.

2.3. Type of study

The present work, due to its data collection characteristics and the measurement of the variables, has a quantitative approach, a non-experimental, descriptive and correlational study.

2.4. Analysis of the variables

The present study presents a main variable that is the risk of diabetes, the data collection instrument was used, the FINDRISK test that comprises eight items grouped into eight dimensions and also can be evaluated through the presence of indicators such as age, body mass index, anthropometry of the waist, physical activity, consumption of vegetables or fruits, medication for blood

pressure, diagnosis of high glucose values and a history of diabetes mellitus [13].

The level of physical activity will be evaluated using the IPAQ questionnaire. The short version evaluates three specific characteristics of activity: intensity (mild, moderate or vigorous), frequency (measured in days per week) and duration (time per day). Activity of moderate intensity is considered as which produces a moderate increase in respiration, heart rate and sweating for at least 10 continuous minutes, and vigorous activity, such as which produces a greater increase in the same variables, for 10 minutes or more [14].

2.5. Instruments

The data collection technique was the survey and the selected instruments were the FINDRISK test and the short version IPAQ questionnaire to measure the level of physical activity.

The data collection process will be carried out during the month of June through the surveys, later the data will be initially entered into a data matrix designed in the Microsoft Excel 2013 program, and then they will be exported to the IBM SPSS Statistics version 24.0 program. And the data analysis will apply descriptive statistical tests such as absolute frequencies, relative frequencies and measures of central tendency, determining the risk of developing diabetes and the level of physical activity.

The instrument was applied in the nine establishments of the Rímac district. The respective permission was requested from Directorate of Integrated Health Networks (DIRIS) Lima Norte, to have access to health centers, for this purpose an enrollment was made in the ethics committee to approve the instrument to apply; After approval, it was presented to the head of each establishment and permission was given to the triage health personnel in charge of the area to develop the questionnaire.

Figure 1: The patient is triage taking anthropometric measurements, before going to the external office to be evaluated

As shown in Figure 1, the instructions on how to carry out the respective filling in of the questionnaires are explained to the patients, they were informed to mark a single alternative in the questions avoiding smearing, and the patients were also consulted to ask before any doubt that they would not understand, to obtain real data.

Figure 2 considers the distribution according to sex of the total number of people who participated in the study, segmented by each health center in the Rímac district, an analysis that considers 35 HC respondents: Rímac, Caquetá and Flor de Amancaes; the second block considers the HC with 34 respondents: Leoncio Prado, San Juan de Amancaes, Mariscal Castilla, Ciudad y Campo and Villa Los Angeles; and finally the block of 25 respondents the HC Piedra Liza, an activity that was carried out according to schedule and considered the minimum established quota.

Figure 2: Filling out the survey of outpatients treated in the health center of a district of North Lima - 2019 (N = 300)

2.6. Developing the surveys

The patients were interested in reading the questionnaire; some patients were hesitant to answer because of their advanced age they did not understand the questions, they did not interpret it in the health field.

Based on good data collection in most of the patients with good attitude and with consent, they collaborated in the survey. In some patients, they did not give their consent because of the time or they had another meeting, the discomfort of the patient and/or relatives when waiting for care was also observed, so they arrived before the mentioned time and were not treated, since that this has an impact on answering the survey.

It is important in the data collection, the technical staff of the triage service since their function is to take the anthropometric measurements (weight, height) and the abdominal perimeter of all the patients before entering the medical office and thus be diagnosed and evaluated.

3. Result

After collecting the data, graphic tables were generated where it will synthesize the information and graphics so that the results obtained are easier to understand.

Table 1: Sociodemographic data on the risk of type 2 diabetes and physical activity in outpatients treated at health centers in a district of North Lima, 2019 (n=300)

Sex of hospitalized patient	**Frequency (f)**	**Percentage (%)**
Female	214	71.3
Male	86	28.7
Health Center	**Frequency (f)**	**Percentage (%)**
CMI Rímac	35	11.7
Caquetá	35	11.7
Flor de Amancaes	35	11.7
Leoncio Prado	34	11.3

San Juan de Amancaes	34	11.3
Mariscal Castilla	34	11.3
Ciudad y Campo	34	11.3
Villa los Ángeles	34	11.3
Piedra Liza	25	8.3
Degree of Study	**Frequency (f)**	**Percentage (%)**
No studies	2	0.7
Primary Incomplete	2	0.7
Primary Complete	4	1.3
High School Incomplete	18	6.0
High School Complete	91	30.3
University study Incomplete	41	13.7
University study Complete	142	47.3
Marital Status	**Frequency (f)**	**Percentage (%)**
Single	121	40.3
Cohabitant	87	29.0
Married	65	21.7
Separated	7	2.3
Divorced	6	2.0
Widowed	14	4.7
Type of Family	**Frequency (f)**	**Percentage (%)**
Nuclear	111	37.0
Single parent	36	12.0
Extended	71	23.7
Expanded	40	13.3
Reconstituted	7	2.3
Equivalent	9	3.0
Single Person	26	8.7
Mental health disorder	**Frequency (f)**	**Percentage (%)**
Depression	3	1.0
Family violence	3	1.0
Not have	294	98.0

Table 1 considers the socio-demographic data regarding the sex of the participants, most of the surveyed sample, equivalent to 71.3%, is female. There are nine health centers in North Lima that were part of the study, which between 34 and 35 patients (11.7% and 11.3% respectively) for each health center made up the study sample, except for Piedra Liza Center, which had 25 patients (8.3%). Likewise, the data of the surveyed sample (47.3%) has completed higher education, followed as a percentage of those with complete secondary education (30.3%). Regarding marital status, most respondents (40.3%) were "single". In a similar way, 37% of the surveyed patients denied having a mental health disorder.

Figure 3 considers the history of diabetes, which the risk is considered considering direct affinity, in which 108 patients are observed, representing 36% have no history, that is, relatives with a diagnosis of diabetes, followed by 104 patients representing 34.7% consider that there is a history of diabetes (grandparents, aunt, uncle, cousin, brother) and lastly 88 patients representing 29.3% consider that there is a history of diabetes (parents, siblings, children), they consider the line Transmissible with higher prevalence.

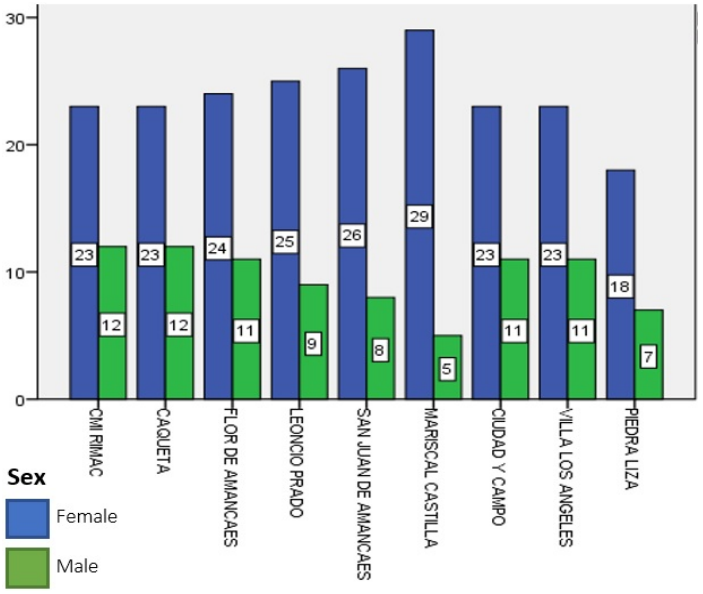


Figure 3: History of diabetes mellitus in outpatients attended in different health centers of a district of North Lima - 2019 (N = 300)

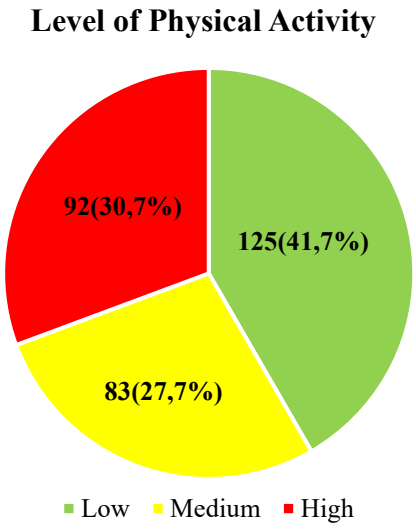


Figure 4: Patients surveyed who perform physical activity, cared for in the establishments of a District of North Lima - 2019 (n=300)

In Figure 4 regarding the level of physical activity, it is observed that 92 (30,7%) of the patients present a high level of physical activity, 83 (27,7%) a medium level of physical activity and 125 (41,7%) low level of physical activity.

In Figure 5 regarding the risk of type 2 diabetes mellitus, it is observed that 79 (26,3%) of the patients present a low risk, 113 (27,7%) present a slightly high risk, 46 (15,3%) present a moderate risk, 52 (17,3%) present a high risk and 10 (3,3%) of the patients present an extremely high risk.

Table 2: Relationship between the Risk of type 2 diabetes mellitus and Level of physical activity in outpatients attended in health centers of a district of North Lima, 2019 (n = 300) based on the Pearson's Chi-square statistical test (X^2)

			Level of physical activity			Total
			Low	Moderate	High	
Risk of type 2 diabetes mellitus	Low	Count	30	27	22	79
		% of total	10,0%	9,0%	7,3%	26,3%
	Slightly high	Count	55	28	30	113
		% of total	18,3%	9,3%	10,0%	37,7%
	Moderate	Count	17	10	19	46
		% of total	5,7%	3,3%	6,3%	15,3%
	High	Count	19	17	16	52
		% of total	6,3%	5,7%	5,3%	17,3%
	Very High	Count	4	1	5	10
		% of total	1,3%	0,3%	1,7%	3,3%
Total		Count	125	83	92	300
		% of total	41,7%	27,7%	30,7%	100,0%

Chi-square tests	Value	gl	Sig. (p)
	9,653	8	0,290

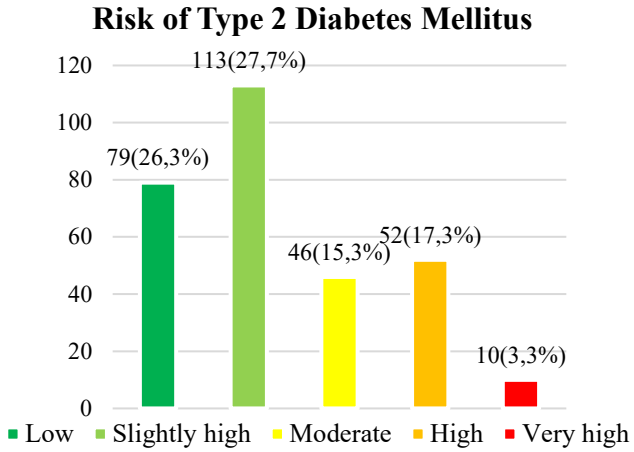


Figure 5: Risk of type 2 diabetes mellitus in patients treated in establishments in a District of North Lima - 2019 (n = 300)

Table 2 shows the relationship between risk of type 2 diabetes mellitus and level of physical activity in outpatients treated in health centers in a district of North Lima, where the significance value of the chi-square test is of 0,290 (p> 0.05) (X^2 = 9,653; g.l. = 8), therefore, there is insufficient evidence to deny the dissociation hypothesis. Given the above, it is concluded that there is no significant relationship between the risk of type 2 diabetes and physical activity in outpatients in health facilities in a district of northern Lima, 2019 with a confidence level of 0,05 (95%).

The importance of obtaining data allows us to identify the relationship between physical activity and the risk of T2DM and which factors are risk factors, which we can carry out prevention strategies that allow maintaining the well-being of the person, also serves for future research with studies that allow evaluating the relationship between physical activity and risk of T2DM in which they can look for the effectiveness of the strategies applied in this research work.

4. Discussion

For the present research work, the focus was given to the patient on their physical health, although we can affirm that the risk of having type 2 diabetes mellitus today is one of the problems in our country due to the sedentary lifestyle that people present due to physical inactivity.

In the results on the level of physical activity, the patients present a low level of physical activity 125 (41,7%), this is since the patients have an inappropriate lifestyle, where sedentary lifestyle is the first factor present when physical activities are not performed, therefore patients tend to have an increased risk of non-communicable diseases that can compromise their health throughout their lives. In [15], the author argues that sedentary lifestyle is a predisposing factor for metabolic diseases where long hours spent at a computer, being on the phone and the eating of fast-food foods that provides a high caloric index, that is not waste by not performing any physical activity, compromises the health of the person generating health problems.

In the risk results for T2DM, patients present a slight risk of having this disease 113 (27,7%), this is since the prevalence of morbidity for contracting T2DM is increasingly considered higher, because people nowadays they manage an inadequate lifestyle, a poor diet, diminished family activities, also family history, where the person who has a relative with this disease if it does not have

an adequate lifestyle can develop the disease. Therefore, adopting prevention and control strategies will reduce the risk of contracting T2DM. In [16], the author argue that family history of diabetes risk tends to increase the degree of susceptibility to the disease and that it is not possible to modify it, but also changes in the lifestyle of the person as a preventive measure by modifying food consumption, increasing physical activity and reducing body weight, are actions that allow the person to increase the benefits in their own health, ensuring well-being and trying to maintain a healthy life. Likewise, in [17], they argue that the self-care that patients with diabetes mellitus should exercise in themselves is important because it is related to the actions that each person take for their own benefit of their health including physical activity, proper nutrition, and self-control.

5. Conclusion

It is concluded that the population should be educated about the importance of increasing physical activity and of a healthy lifestyle where the well-being of health in the person is balanced.

It is concluded that nutritional counseling should be provided when making home visits to patients with diabetes mellitus.

It is concluded that the strengthening and control of prevention programs in primary care related to chronic diseases, including type 2 diabetes mellitus, should be considered.

It is concluded that preventive actions should be taken in the patient with diabetes mellitus such as a healthy lifestyle, a diet program that suits the patient and physical activity.

The main limitations in this research is the understanding that patients had about filling out the surveys, also, the time they had to fill out the surveys was also limited because the patients were waiting to be attended in the health centers and some did not want to fill out the survey; most of the patients voluntarily agreed to participate in the research work to find out if they had any risk of contracting T2DM.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] A. Olokoba, O. Obateru, L. Olokoba, "Type 2 diabetes mellitus: A review of current trends.," Oman Medical Journal, **27**(4), 269–273, 2012, doi:10.5001/omj.2012.68.

[2] M. Mazidi, P. Toth, M. Banach, "C-reactive Protein Is Associated With Prevalence of the Metabolic Syndrome, Hypertension, and Diabetes Mellitus in US Adults.," Angiology, **69**(5), 438–442, 2018, doi:10.1177/0003319717729288.

[3] L. Slater, S. Stuckey, R. Chandra, "Tratamiento actual de la diabetes mellitus tipo 2.," Neuroimaging Pharmacopoeia, **20**(1), 289–292, 2016, doi:10.1007/978-3-319-12715-6_40.

[4] A. Koopman, J. Beulens, E. Voerman, P. Simone, A. Heijden, T. McDonald, M. Langendoen, F. Rutters, "The association between GAD65 antibody levels and incident Type 2 Diabetes Mellitus in an adult population: A meta-analysis.," Metabolism: Clinical and Experimental, **95**, 1–7, 2019, doi:10.1016/j.metabol.2019.03.001.

[5] A. Sharma, L. Sikora, S. Bush, "Management of Diabetes Mellitus in Adults at the End of Life: A Review of Recent Literature and Guidelines.," Journal of Palliative Medicine, **22**(9), 1133–1138, 2019, doi:10.1089/jpm.2018.0614.

[6] F. Petermann, X. Díaz, Á. Garrido, A. Leiva, M. Martínez, C. Salas, F. Poblete, C. Celis, "Association between type 2 diabetes and physical activity in individuals with family history of diabetes.," Gaceta Sanitaria, **32**(3), 230–235, 2018, doi:10.1016/j.gaceta.2017.09.008.

[7] A. Pinilla, M. Barrera, "Prevención en diabetes mellitus y riesgo cardiovascular: enfoque médico y nutricional.," Revista Facultad de Medicina, **66**(3), 459–468, 2018, doi:10.15446/revfacmed.v66n3.60060.

[8] S. Seclén, "Diabetes Mellitus en el Perú: hacia dónde vamos.," Revista Medica Herediana, **26**, 3–4, 2015, doi:10.20453/rmh.2015.2340.

[9] R. Carrillo, A. Bernabé, "DIABETES MELLITUS TIPO 2 EN PERÚ: UNA REVISIÓN SISTEMÁTICA SOBRE LA PREVALENCIA E INCIDENCIA EN POBLACIÓN GENERAL.," Rev. Perú Med. Exp. Salud Publica, **36**(1), 26–36, 2019, doi:10.17843/rpmesp.2019.361.4027.26.

[10] A. Manzaneda, M. Lazo, G. Málaga, "Physical activity in outpatients with type 2 diabetes in a national hospital of Perú.," Rev. Perú Med. Exp. Salud Pública, **32**(2), 311–315, 2015.

[11] V. Bellou, L. Belbais, L. Tzoulaki, E. Evangelou, "Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses.," PLoS ONE, **13**(3), 1–27, 2018.

[12] Å. Bjørn, K. Midthjell, S. Krokstad, V. Rangul, A. Bauman, "Prolonged sitting may increase diabetes risk in physically inactive individuals: an 11 year follow-up of the HUNT Study, Norway.," Diabetologia, **60**(5), 830–835, 2017, doi:10.1007/s00125-016-4193-z.

[13] I. Mendiola, I. Urbina, A. Muñoz, G. Juanico, G. López, "Evaluación del desempeño del Finnish Diabetes Risk Score (findrisc) como prueba de tamizaje para diabetes mellitus tipo 2.," Atención Familiar, **25**(1), 22–26, 2017, doi:10.22201/facmed.14058871p.2018.1.62925.

[14] R. Barrera, "Cuestionario Internacional de actividad física (IPAQ)," Revista Enfermería Del Trabajo, **7**(2), 49–54, 2017.

[15] B. Pérez, "Salud: entre la actividad física y el sedentarismo.," An Venez Nutr, **27**(1), 119–128, 2014.

[16] Y. Llorente, P. Soca, D. Rivas, Y. Borrego, "Factores de riesgo asociados con la aparición de diabetes mellitus tipo 2 en personas adultas.," Revista Cubana de Endocrinología, **27**(2), 123–133, 2016.

[17] M. Soler, E. Pérez, M. López, D. Quezada, "Conocimientos y autocuidado en pacientes con diabetes mellitus tipo 2.," Archivo Médico de Camagüey, **20**(3), 244–252, 2016.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1657-1665 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

A Novel Way to Design ADS-B using UML and TLA+ with Security as a Focus

Pranay Bhardwaj*, Carla Purdy, Nawar Obeidat

College of Engineering and Applied Science, University of Cincinnati, Cincinnati, OH 45220, USA

ARTICLE INFO

Article history:
Received: 01 September, 2020
Accepted: 04 December, 2020
Online: 28 December, 2020

Keywords:
ADS-B
TLA+
Security

ABSTRACT

Automatic Dependent Surveillance-Broadcast (ADS-B) is the future of aviation. It is a vast system that provides situational awareness for the aviator and regulator at a very low cost and does so with the aid of multiple disparate systems working closely together and communicating with one another. ADS-B uses the Global Navigation Satellite System (GNSS/ GPS) to locate elements. Weather information and ground-based information is also transmitted wirelessly. The system is designed to be open, unencrypted, and accessible to actors throughout the world. However, this leaves it open to attacks. The use of GNSS and other wireless technologies also carries over their security vulnerabilities into ADS-B. Certain issues have arisen due to both component-system failures and malicious attacks. Most obvious solutions impinge on the openness and transparency of the system. Past research has indicated that security must be built into a system design itself and cannot be retrofitted. We want to showcase such a design process for ADS-B. Our pathway to do so is to first create Universal Modeling Language (UML) diagrams to showcase security and safety issues and responses. These UML diagrams will then help us to model state and sequence diagrams. These will then be used to create a TLA+ model of one selected security methodology. We then run the TLC model checking on it to find loopholes and plug gaps in our scheme. We managed to create such models and prove deadlock-free running using only software tools. Our eventual goal is to develop a comprehensive formal specification for ADS-B model-creation and checking.

1. Introduction

This paper is an extension of work originally presented at the 2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS), Dallas, TX, USA [1].

Under the 2020 ADS-B mandate [2, 3], the FAA has designated that all commercial traffic and most private traffic must make itself compatible with ADS-B by hardware and software upgrades by the date of January 1. 2020.

ADS-B will replace 'hard' sites like primary/secondary surveillance radar (PSR/SSR) stations with satellite-based GNSS. The onus to report position is now on the aircraft. Frequencies of 1090 MHz and 978 MHz will be used for the actual information transmission. Aircraft will also beam down velocity, altitude, and a host of other data. In exchange, weather data will be beamed up to them. Aircraft will share position information with each other as well, to prevent conflicts.

ADS-B is a lot more economical than the current system. In 2007 the cost to monitor 200 nautical miles of air space was estimated to be $10-14 million using PSR, $6 million using SSR, and $380,000 for ADS-B [4]. Cost savings occur due to not having to set up and maintain expensive PSR/SSR RADAR stations plus increased coverage of formerly 'RADAR-dark' areas, resulting in more safety dividends, among other advantages.

The cost factor makes the choice look easy, but the security and safety weaknesses of ADS-B present a conundrum. Security issues are noted in [5-7], and an example of a deliberate system shutdown is given in [8].

Any attempts to gatekeep ADS-B go against its ethos. Security thus becomes a challenge. Attempts are now being made to introduce security measures into an already deployed ADS-B. But in [9], the authors have postulated that security must be considered at the design phase itself and any attempts to fit it in later will result in underestimation of the system capacity required to run security.

The scope of ADS-B is not restricted to 1090/978 MHz. In [10], the authors talk about a future where the Internet Protocol is the communications backbone for this system, even for voice

* Corresponding Author: Pranay Bhardwaj, Email: bhardwpy@mail.uc.edu

communication for aircraft flying. The radio has not been the sole point of communication for an aircraft for a long time. Now component subsystems of an aircraft independently 'talk' to maintenance and route planning departments of the operator via Aircraft Communications Addressing and Reporting System [11]. IP is also slated to take this communication load.

Following [9], the goal of this research is to rethink the ADS-B system design from a security viewpoint. We want to show what models such a hypothetical design process could have made, and what tools they could have used. A similar methodology is adopted by the authors in [12]. Here we are not recommending specific solutions for the numerous safety and security problems, nor are we devising a way to 'fix' ADS-B as it exists. We want to model how the different subsystems that makeup ADS-B interact with each other when under attack.

Our methodology is to pick a particular attack, pick specific solutions to that attack, and then first model the attack and response using use case diagrams. These will be used to create state and sequence diagrams to give a step-by-step description of the security response process to that attack. Once the sequence of events is clear, we can develop a TLA+ model of our system and run it through the TLC model checker to see if there are any deadlocks or failure points.

The specific attack scenarios and response tools are not ours, nor do we want to create any. We just want to observe how these response tools would look if they were all designed into the system, rather than retrofitted and to see if their in-built design into the system will cause any problems with any other system.

To put it simply, we do not want a situation where, for example, a cyberattack is successfully stopped by the security system but with an unintended result that airplanes are left without GNSS coverage. This could happen if the inter-relationships of competing cybersecurity tools are not well- understood at design time. We want to showcase that our methodology can be followed to model all of ADS-B, and perhaps other large systems in this manner at a low cost to spot deadlocks and bugs early in the design process.

So why do we want to use UML?

In [13], the author has used UML, the Unified Modeling Language [14], to model the Controller Area Network Bus (CANBUS). It has also been used to create security parameters [15,16]. UML focuses on whole 'objects' and each component may be a whole system unto itself. Thus hardware, software, and cyber-physical systems can all be modeled in UML.

Looking from a very high level, we can visualize the ADS-B as similar to the CANBUS. Both are unencrypted, both have various 'nodes' in the system, both rely on messages, both use broadcast protocols and both have no 'central brain.' Finally, security was not a consideration in the original design of either.

In ADS-B, the 'nodes' are the aircraft, the GNSS is a 'sensor' and the bus itself could be the pathways of the 978/1080 MHz frequencies. This is quite an interesting comparison and can be made between ADS-B and many other systems of various sizes and scales.

If ADS-B analogies can be found with the CANBUS, then methodologies to secure the CANBUS can theoretically also be proposed for ADS-B. As in [13], we will begin by modeling selected security techniques in UML and create use case diagrams.

Aviation is a heavily regulated sector and the basic requirement in all these systems↓ operational and security↓ is that they all work well together. Thus, the early introduction of security into the design is necessary because a late addition may introduce unintended security flaws. That is exactly the model we suggest in this paper- a redesigned system that does not stray too much from the current design and does not require major changes in protocol or infrastructure.

Our inspiration to use TLA+ is to get a workable model of how a potential 'built-in security' ADS-B system will look and behave. If successful, we will also be able to spot any failure points or illegal states in our models.

We envision that our paper can be used as a guidebook to model any discrete distributed system and test it for security loopholes at a very low cost. This will lead to the introduction of formal methods in civil aviation cybersecurity.

Formal methods have been used for many years in modeling localized systems [17], and we believe large systems consisting of geographically discrete subsystems can also be modeled this way. ADS-B uses independent systems for position, weather, and collision avoidance↓and each one of these is important to ensuring security and safety.

Our end goal is to create an overall super-design of the ADS-B system incorporating all the different security measures we talked about. In the future, we will attempt to create state and sequence charts for all of ADS-B. We will seek to validate the TLA+ based model checking and formal methods for complete ADS-B design.

In section 2, we discuss some weaknesses of ADS-B security and some well-known attacks. We then do the same for its largest sub-system, GNSS in section 3. In section 4 we create UML use-case diagrams to show how some selected methods can protect us from a cyberattack, and we also look at many other methods specified in the literature and select some for the next stage. In Section 5, we create a state diagram and sequence diagram of how ADS-B could respond to an attack using tools described in section 4. This is not a general view↓ just a specific attack scenario↓ and it feeds into the next section. In section 6, we will finally convert the state diagram to an actual TLA+ text specification/model. We will run it through the TLC model checker in section 6.1 to look for deadlocks and failures. We will then suggest how to apply our method more broadly to deal with the safety and security issues we have identified for the ADS-B system in the concluding remarks and the future work described in section 7.

2. Major Weaknesses of ADS-B

Because of its accessible design, ADS-B is missing even minimal security features [5]. This makes attacking easy and defense complicated. Some overarching vulnerabilities are:

- Physical layer: The 1090 MHz channel used for commercial aircraft ADS-B is severely congested, with up to 50% message loss rate measured at 174 miles [5]. This will only get worse with time, and we are in a very early phase yet.
- Unencrypted: Because ADS-B is unencrypted, the positions of aircraft can be tracked by anyone with homemade equipment and a free open-source application like flightradar24.com (FR24). Eavesdropping is the first step for many attacks [5]. FR24 details a step-by-step guide to making an ADS-B 'snooper' and getting real-time aircraft identification, position, and performance data.
- Unauthenticated: ADS-B has no authentication on messages↓no signatures, no handshake protocol. Aircraft identifiers, like the 24-bit ICAO code, are publicly available in a platform like FR24. Thus, there is no way to identify the source of a message [3]. If the constraint of powerful transmission equipment is fulfilled, a well-funded actor can theoretically beam fake data about an existing contact, or a fake contact itself, to air traffic control. Even worse, it can remove data as well.

Some common attacks that can happen in the above environment are:

- Ghost contacts: Two works on cybersecurity [18,19] mention the possibility of fake, or 'ghost' targets in the system. An attacker could use the openness of the system to download authentic aircraft data and then broadcast it over the frequency with sufficient power. This would appear as a legitimate contact on the ATCS screen. Without confirmation of GPS position, it would be impossible to catch such an attack. An example that combined eavesdropping with ghost injection is described. The appearance of multiple ghost aircraft can cause system and sensory overload and can cause pilots and ATC to make dangerous decisions [6].
- Hybrid attacks: Hybrids of various attacks can be made. In [6], the authors talk about an attack that 'deletes' legitimate data from the system and injects false data in its place.
- Legacy systems: It is not uncommon to find aircraft flying today that are 20-30 years old or more. This is especially true in developing countries, cargo airlines, and various armed forces. Legacy aircraft systems were not designed with an interconnected super system like the ADS-B in mind. In the past, each aircraft was an independent entity. Legacy systems have always been a security challenge when hooked up to large networks. A determined adversary with ample support↓like a state-backed actor↓can overcome all compatibility and air gap barriers to infect an 'old' system with a 'new' bug, as was seen in [20].
- Physical access: Physical isolation or 'air- gapping' that ensured the protection of many systems before will cease to exist. In the past, the output of a PSR/SSR was inaccessible to the public. Now, the detection system, position system, and the results therein are all readable on a home computer in real-time. Physical isolation of infrastructure has given way to an open, unencrypted, unauthenticated setup.

Risks of GNSS Use in Aviation

- Survivability: As of today, four nations possess the capability to destroy satellites in orbit the USA, Russia, the PRC, and India [21]. These 4 nations have also all developed their own global or regional GNSS (GPS, GLONASS, BEIDOU, NAVIC respectively). All these nations have also militarized their GNSS systems↓to guide bombs, missiles, warplanes, naval vessels [22] and even to plan large-formation land maneuvers↓as was seen in the 1990 gulf war. Law enforcement and coast guarding also use GNSS. However, this also makes the GNSS a ripe target in the event of great power warfare. Knocking out GNSS satellites using anti-satellite (ASAT) missiles could lead to a mass blackout of ADS-B position reporting for all countries. This could have disastrous unintended consequences for civil aviation within seconds. In a future PSR/SSR-less world, this could mean crippling the global flight infrastructure. Although the probability of this is low, the authors find this to be one of the most disturbing eventualities.
- Deliberate shutdown: Like any other large system, GNSS may be on reduced performance from time to time, as noted, e.g., in [23]. Besides, deliberate shutdowns may be conducted for various purposes, like military training. The degraded signal quality sent an aircraft into high-speed oscillations in 2016 [24]. As the dependence of aircraft on GNSS grows, any shutdown or breakdown will become a safety issue for civilian aircraft. Dependence of autopilot systems on GNSS to navigate means aircraft could go into abnormal flight regimes without even alerting the crew if a sudden 'quiet death' of GNSS occurred locally during normal flight.
- Jamming and spoofing: GNSS is a satellite-based wireless system, and like every wireless system ever made, GNSS is susceptible to jamming of signals and spoofing↓incorrect location reporting. The Russian Federation has already demonstrated a GNSS spoofing/ jamming capability precise enough to make a single human difficult to locate [25]. It can be devastating for unsuspecting civilian aircraft in an ADS-B based system and could very well be the model of a terror attack.

3. Use Case Diagrams for ADS-B Security

For such a large and distributed system as ADS-B, it must be understood that no single solution exists. In [7], the authors tested the viability of many solutions and found none without a loophole. This is a glaring after-effect of not keeping security paramount in the design process. Of course, we cannot talk about a cybersecurity model without talking about the actual 'weapons' that will make it possible. We will now make UML diagrams of how the various cybersecurity tools we are familiar with could be deployed, in isolation or in combination, to ensure protection. Figure 1 provides the convention we use in the use case diagrams that follow.

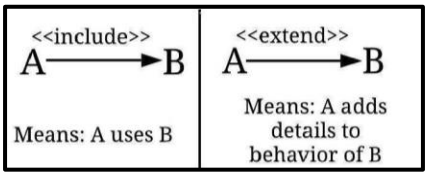


Figure 1: Convention for use case diagrams

In Figure 2, multilateration uses the time delay of signal arrival at multiple (>2) points to estimate the location of the transmitter [7]. Time Delay of Arrival (TDOA) of the signal is

calculated with the use of time stamps stamped on any message. Comparison of timestamps of transmission and reception gives a rough idea of the travel time, from which a rough distance can be calculated. If used at 3 receivers, a close X-Y-Z coordinate can be back-calculated. This is not very different from the SONAR equation.

If this back-calculated position is reasonably close to the reported GNSS coordinate of the transmitter, the allegiance of the message sender can be confirmed. Multilateration can be used to combat ghost aircraft, message modification, and man-in-the-middle attacks, and can provide a backup to GNSS breakdown in the absence of PSR.

Figure 3 is the most important here as it was the most complex one in our original paper. We have thus chosen to extend this diagram to the state diagram, sequence diagram, and later, to the TLA+ specification.

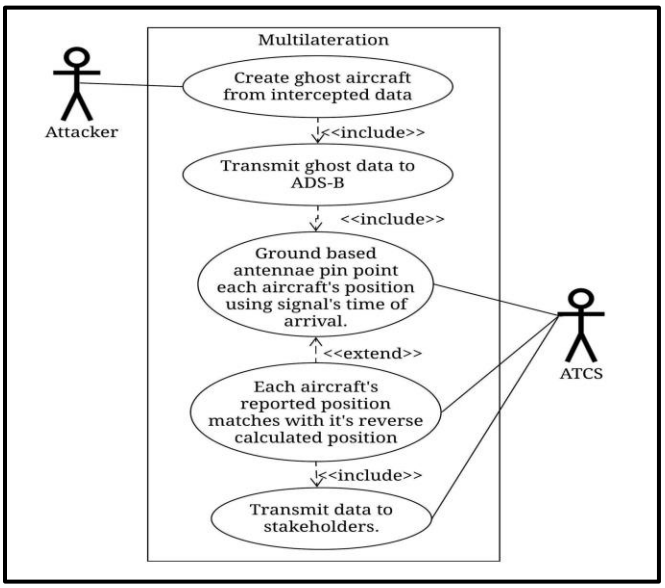


Figure 2: Multilateration and group verification can protect against ghost aircraft injections by matching the physical source of the signal to its reported one.

One of the weaknesses of using static ground posts as receivers for multilateration is that the system may be fooled by the attacker. If the location of these ground antennas is known, the attacker can tailor his transmission time stamps to present a 'false but accurate' picture to the multilateration-systems. A solution to this is to use dynamic receiver posts↓make the other, trusted aircraft in the airspace as multilateration, or triangulation points themselves!

Figure 3 shows multi-point multilateration. Group verification [7] is a kind of multilateration performed in the air. Aircraft already multilaterate each other using TDOA↓it is exactly how the Traffic Collision Avoidance System (TCAS) works. The equipment and algorithms are thus available on board already, and this is in line with our philosophy to cause minimal modification.

Each aircraft locates each member using TDOA of received ADS-B IN signals and estimates the others' positions. If the calculated position differs from the reported one, a 'suspect' airframe is identified, and group members move away from it. This can be successful only when 100% implementation of ADS-B is achieved. Location fixes from different sources are independent of GNSS. The fusion of this data can be an alternate method to locate aircraft. This will counter GNSS spoofing and GNSS jamming attacks.

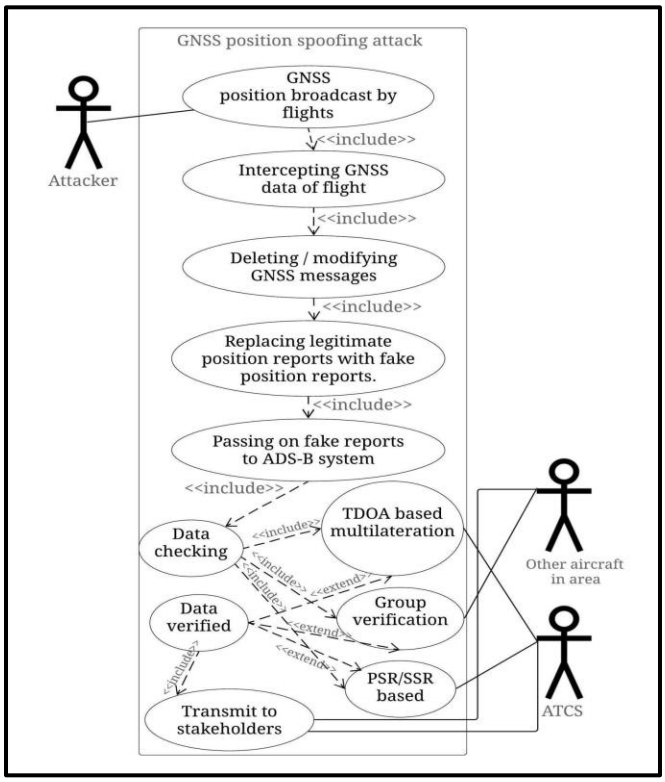


Figure 3: Multilateration may help with GNSS spoofing and unreliable position reports.

Other issues and solutions we will use in our overall model:

- Authentication of messages can be used to identify which messages are from what source. Authentication just confirms that the message is from who you think it to be. This directly allows us to trust or not trust it. Authentication of messages is the closest we come to a silver bullet solution for ADS-B as a single tool protects us from multiple issues↓Denial-of-Service (DOS) attacks and ghost aircraft attacks, just for starters. DOS attacks on airport or airline infrastructure are of concern as these are cheap to wage.
- Kalman filtering has many applications↓the ADS-B system can be 'taught' the typically correct values of airplane parameters [7]. According to their capabilities, scheduling, and operators, most aircraft will fly within a certain envelope of speed, heading, and locations usually. The Kalman filter can be rewarded whenever correct values are noticed, and not when unusual values are noticed. This will protect against ghost injection attacks and false data messages. The system will sense the departure from usual data trends and raise an alarm. If data trends change, it can be fed the new data trend and 'taught' any updates to the system this way. The only disadvantage is that the system is equally likely to be taught the 'wrong' way and thus become a 'bad child.'
- Hardware and software fingerprinting [26] have great potential to introduce economical authentication capability

without making major hardware or software changes. The authors in [26] provide a simple solution to 'fingerprint' the transmitters themselves. Every transmitter in the world is unique, just as every crystal is unique. Individual aircraft-borne ADS-B transmitters can be 'fingerprinted' either via unique hardware or through software properties. This can positively identify the transmitter and provide authentication.

- ADS-B as RADAR? A unique solution of using the ADS-B signal itself like a radar signal, by adding a random bi-phase modulation is proposed in [27]. This system can track other aircraft even with no GNSS. It can provide the much-needed reliability of radar at the low cost of ADS-B. This can counter issues of the survivability of GNSS. This agrees with [4], which talks of a fusion of GNSS and non- GNSS sources for locating aircraft. However, this will require changes to ADS-B transmitting recruitment and is thus low on our list of options.
- Weather RADAR for aircraft tracking: Weather RADAR is ubiquitous and big and small weather radar stations exist all over a developed country like the US. Since this RADAR can track small cloud formation, moisture, and bird/ locust flocks, can it be used to detect hard-skinned objects like airplane skin? Some research has been conducted in the use of weather RADAR to track aircraft and we believe that if it is properly developed, it can act as an emergency backup in the case of a major GNSS failure. For small general aviation (GA) aircraft at least, a method is suggested in [28]. This method can distinguish between small aircraft and large birds based on pressure waves created by aircraft propellers. In [29], the authors managed to detect aircraft with weather RADAR using some signal processing. More than 90% of aircraft in the USA are GA and most of these are propeller equipped [30]. Thus, a solution for GA aircraft as presented in [28] solves a huge chunk of aircraft tracking issues in case of GNSS blackout.
- Identification can be achieved by building a small ADS-B receiver like the ones FR24 uses, to receive aircraft data. Thus, in the event of a GNSS failure, weather RADAR can track the physical location of GA aircraft while an ADS-B receiver can identify them. One only must ignore the ADS-B location data in such a case. The use of a non-ADS-B source as a redundant backup is in the spirit of [7].
- TESLA protocol for encryption: In [31], the authors suggest a method that encrypts only the 24-bit ICAO aircraft identifier. It uses Timed Efficient Stream Loss-tolerant Authentication (TESLA) for authentication. It has performed well under testing without reducing security, performance, or affecting any major changes to ADS-B philosophy.

4. TLA+ conversion

TLA+ stands for 'Temporal Logic of Actions' [32]. It is a logical system or toolbox, developed by Leslie Lamport. which allows one to write formal specifications and to include scalable security protocols. The TLA+ specification can then serve as a formal 'model- checker' for any future realization of the ADS-B system and its components. It is not a classical programming 'language.' TLA+ is based on mathematics and uses set theory to model systems. Each system has states, and each state can be defined by a set of certain values. When these values change, so does the state. Any transition (step) is the change of one state to another or a change of one set of values to another set of values.

An entire system can be modeled as a superset of all possible state sets. We do not go too deep into the specifics of each state↓we are more concerned with the what rather than the how. Thus, abstraction and conciseness are a very important skill to master before modeling the system in TLA+. While economics and computer overhead would be major factors in practical systems design, we drop them when dealing with TLA+.

As a testimonial of its effectiveness, the application of TLA+ by Amazon revealed it to be excellent for bug detection [33,34]. Microsoft also used it to verify the design of a subsystem of the Alpha 21364 Microprocessor [17].

Our work here can be viewed as a guide on how to design large systems in the future↓by comparative analysis and extensive model-driven design.

ADS-B has many subsystems that make up the whole. GNSS is responsible for position reporting to aircraft. Aircraft then report their position to air traffic control.

Aircraft get weather and aerodrome information, and share position reposts with each other for collision avoidance. All this communication happens over the 978/1090 MHz spectrum. So not only are there several independent operations but also there are independent failure points. All these systems must be a part of any formal method examination of the ADS-B system. The specific solutions for responding to attack are not our own↓they are specified in section 4. What we are doing is modeling a security solution for the ADS-B system that will have safety and security included by design, and suggesting a procedure to design such a system and simulate it.

Several works on smart systems in TLA+ have been researched. In [35], the authors design a smart school system in TLA+. In [36], the authors model a sewerage system in UML and TLA+.

The authors in [35] start by making a UML specification, a sequence diagram, and a state diagram. These diagrams aid in making the sequence of events clear. The system's pathways, legal and illegal actions, and outcomes are modeled.

4.1. Why do we make sequence and state diagrams?

TLA+ defines every system as a sum of all the states it can exist in. A system can have multiple states, and there will be a defined 'step', or a sequence of events, to go from one state to another. Thus, we first develop a sequence diagram, which shows all the steps between states, and a state diagram which shows the states themselves. Only then can we draw a true TLA+ model.

Let us take the below example of applying methodologies of [35] and [36] to the ghost aircraft and GNSS spoofing issues as noted in Figure 3.

4.2. ADS-B response to Ghost aircraft and GNSS spoofing

In sections 2 and 4, we talked about security issues caused if an aircraft is either a fake contact or is reporting untruthful position information about itself to ADS-B, via GNSS spoofing. Since ADS-B envisions doing away with PSR/SSR, we had

examined some other backup systems that can physically locate an aircraft if GNSS reports are unreliable. The UML diagram for this was Figure 3. Let us now look at a sequence diagram for the situation, in Figure 4.

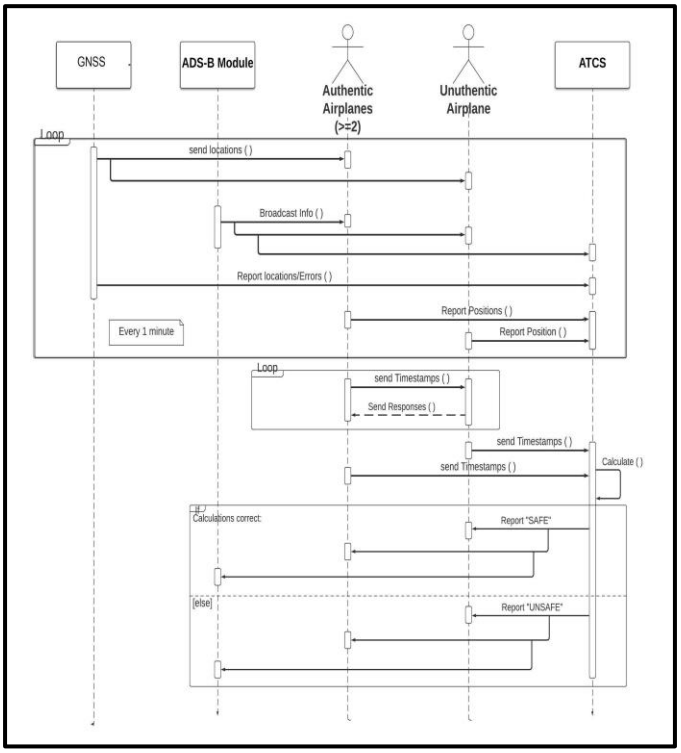


Figure 4: Sequence diagram: Reverse calculation of the aircraft's true position can be done using time-delay analysis of transmitted signals

GNSS reports aircraft position to airplanes, which includes both the 'good' and 'bad' airplane. All aircraft will then report their position to ATCS. A time delay analysis of the signals from all aircraft will be performed, and this will be used to get a location fix of the aircraft in physical space. Aircraft will also be able to perform multilateration and triangulation on each other in airspace↓within a '4-ship-cell' for example↓and get location fixes for each other's locations.

A comparison of the back-calculated position data with the GNSS reported position data will point out which contacts are reporting true data and which are not. Thus, aircraft can be labeled 'SAFE' or 'UNSAFE.' This information can then be provided to other aircraft in the area, and ATCS. Note here that 'ADS-B' stands for the system, but practically it may mean a supervisory body, like the FAA Technical Operations Services, or any other service which needs to be notified of security incidents. Let us now look at a state diagram of the same function in Figure 5.

This specifies an interval of 1 minute after which the whole cycle is repeated. This is arbitrary and will practically be a decision governed by radio traffic and economic factors. Since aircraft move fast and move in all 3 dimensions, this interval cannot be too long. Also, when we say 'ATCS', we mean a local or regional airspace manager. We do not envision a global central ATCS system in this paper.

Let us now use the 3 inputs, Figures 3,4, and 5 to start making the TLA+ formal specification for the above response.

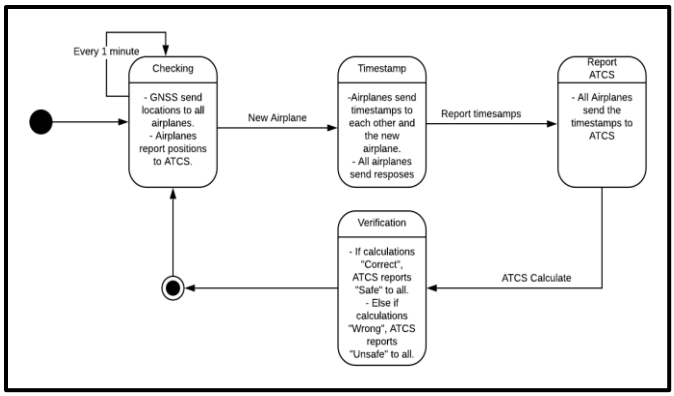
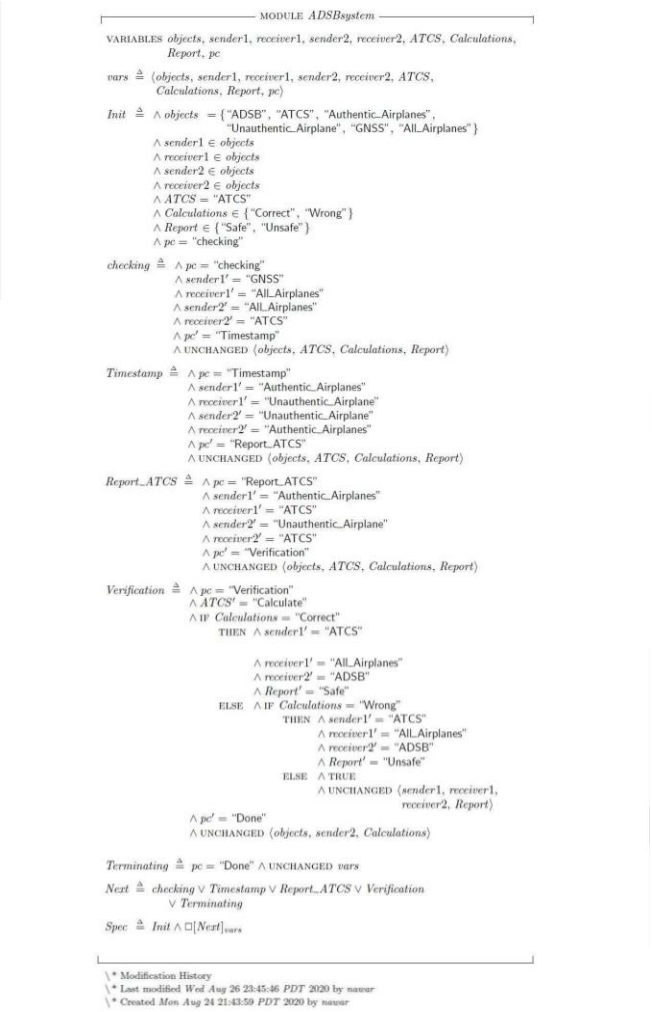


Figure 5: State Diagram of the system under Figure 4 conditions.

5. TLA+ Specification

We present now, the TLA+ specification for the attack response model sequenced in Figure 4 and stated in Figure 5. A block by block explanation follows.



MODULE *ADSBsystem*

VARIABLES *objects*, *sender1*, *receiver1*, *sender2*, *receiver2*, *ATCS*, *Calculations*, *Report*, *pc*

$vars \triangleq \langle objects, sender1, receiver1, sender2, receiver2, ATCS, Calculations, Report, pc \rangle$

$Init \triangleq \land objects = \{\text{"ADSB"}, \text{"ATCS"}, \text{"Authentic_Airplanes"}, \text{"Unauthentic_Airplane"}, \text{"GNSS"}, \text{"All_Airplanes"}\}$
$\land sender1 \in objects$
$\land receiver1 \in objects$
$\land sender2 \in objects$
$\land receiver2 \in objects$
$\land ATCS = \text{"ATCS"}$
$\land Calculations \in \{\text{"Correct"}, \text{"Wrong"}\}$
$\land Report \in \{\text{"Safe"}, \text{"Unsafe"}\}$
$\land pc = \text{"checking"}$

$checking \triangleq \land pc = \text{"checking"}$
$\land sender1' = \text{"GNSS"}$
$\land receiver1' = \text{"All_Airplanes"}$
$\land sender2' = \text{"All_Airplanes"}$
$\land receiver2' = \text{"ATCS"}$
$\land pc' = \text{"Timestamp"}$
$\land \text{UNCHANGED } \langle objects, ATCS, Calculations, Report \rangle$

$Timestamp \triangleq \land pc = \text{"Timestamp"}$
$\land sender1' = \text{"Authentic_Airplanes"}$
$\land receiver1' = \text{"Unauthentic_Airplane"}$
$\land sender2' = \text{"Unauthentic_Airplane"}$
$\land receiver2' = \text{"Authentic_Airplanes"}$
$\land pc' = \text{"Report_ATCS"}$
$\land \text{UNCHANGED } \langle objects, ATCS, Calculations, Report \rangle$

$Report\_ATCS \triangleq \land pc = \text{"Report_ATCS"}$
$\land sender1' = \text{"Authentic_Airplanes"}$
$\land receiver1' = \text{"ATCS"}$
$\land sender2' = \text{"Unauthentic_Airplane"}$
$\land receiver2' = \text{"ATCS"}$
$\land pc' = \text{"Verification"}$
$\land \text{UNCHANGED } \langle objects, ATCS, Calculations, Report \rangle$

$Verification \triangleq \land pc = \text{"Verification"}$
$\land ATCS' = \text{"Calculate"}$
$\land \text{IF } Calculations = \text{"Correct"}$
$\text{THEN } \land sender1' = \text{"ATCS"}$
$\land receiver1' = \text{"All_Airplanes"}$
$\land receiver2' = \text{"ADSB"}$
$\land Report' = \text{"Safe"}$
$\text{ELSE } \land \text{IF } Calculations = \text{"Wrong"}$
$\text{THEN } \land sender1' = \text{"ATCS"}$
$\land receiver1' = \text{"All_Airplanes"}$
$\land receiver2' = \text{"ADSB"}$
$\land Report' = \text{"Unsafe"}$
$\text{ELSE } \land \text{TRUE}$
$\land \text{UNCHANGED } \langle sender1, receiver1, receiver2, Report \rangle$
$\land pc' = \text{"Done"}$
$\land \text{UNCHANGED } \langle objects, sender2, Calculations \rangle$

$Terminating \triangleq pc = \text{"Done"} \land \text{UNCHANGED } vars$

$Next \triangleq checking \lor Timestamp \lor Report\_ATCS \lor Verification \lor Terminating$

$Spec \triangleq Init \land \Box[Next]_{vars}$

* Modification History
* Last modified *Wed Aug 26 23:45:46 PDT 2020* by *nawar*
* Created *Mon Aug 24 21:43:59 PDT 2020* by *nawar*

Figure 6: TLA+ model specification in literal form.

In the module, the keyword "VARIABLES" represents the variables in this module where "vars" is the tuple of variables:

<objects, sender1, receiver1, sender2, receiver2, ATCS, Calculations, Report, pc >

In every TLA+ module, there is an "Init" and a "Next" state. In the Init state, the variables get their initial values, while the next state represents all possible states that come next. In this module, the Next state represents:

Next == checking \/ Timestamp \/ Report_ATCS \/ Verification\/ Terminating

This includes 5 different states. The 'Checking' state occurs when the GNSS sends positioning data to all the airplanes, and then all the airplanes report their locations to ATCS.

In the group verification strategy noted in Figure 3, airplanes were using timestamps on exchanged signals to triangulate each other using time delay of transmission. The 'Timestamp' state represents this communication between airplanes. When a new airplane that possibly might be an attacker appears, it must communicate with other authentic airplanes (even if just for TCAS) and exchange timestamps.

In the 'Report_ATCS' state, all airplanes also report to ATCS with timestamps for routine operations↓and these can be used for position verification from a ground point of view.

The system must then enter the 'verification' state when ATCS does the time delay and radio directional calculations on these signals and compares them to the first GNSS location reports received from all airplanes. Based on these calculations the ATCS will identify which airplanes are reporting their position faithfully and which are not. It will then broadcast the result with a safe/unsafe report to everyone in the airspace.

The termination state allows infinite stuttering to prevent deadlock on the termination (when pc = "Done").

The statement:

Spec == Init /\ [][Next]_vars

represents the specification of the whole system, where Init means the starting state and []Next_vars means Next state must be true for the entire behavior with the option of keeping all variables unchanged.

5.1. TLC Result

The TLC model checker is a tool to validate the TLA+ module and check it for any errors. We run the TLC model checker for our TLA+ ADS-B module and it shows that the module is valid with no errors as well as no deadlock. Also, as appears in the figure below, the TLC model checker shows how many times each state was visited in the module, and what values changed to arrive at the next state.

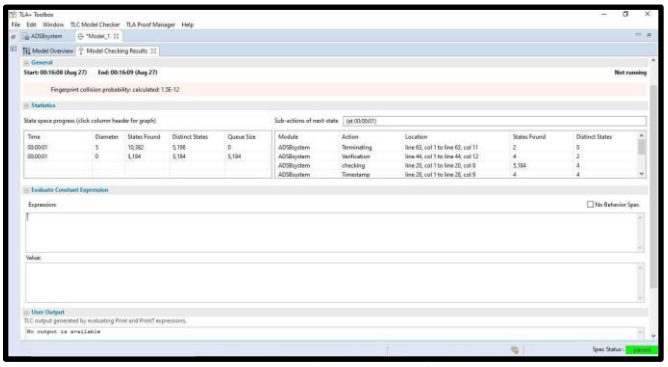


Figure 7: TLC Model check for validity of TLA+ spec in Figure 6.

6. Conclusion and Future Work

At this stage, our model for Figure 3 for verifying physical aircraft locations to combat ghost aircraft and incorrect position reporting issues shows no deadlock. We see that a security system using multilateration and time delay of arrival can identify which aircraft are ghosts and which are not, without getting into a deadlock for resources with the ADS-B system itself.

What this leads us to conclude is that this multi-level modeling of use case-> state diagram-> sequence diagram-> TLA+ specification-> TLC model check can be used to model a large and discrete system like ADS-B to see how it would function and what it would fail at. This is what TLC does it goes through iterations and finds failure points. It throws different numbers at the problem and sees what 'sticks to the wall,' what causes the system to break. We can conclude, at least on this level, that our procedure to model the ADS-B this way and the ability to test the model is valid.

It is always advisable to design security into a system from the beginning of the design process. We hope our method of using UML + state diagrams + sequence diagrams + TLA+ can not only provide the reader with a stable, practical ADS-B model that is secure and realistic but also act as a model of how to design security into other discrete systems in the future.

At the beginning of this paper, we compared the vast ADS-B to a small CANBUS. The roadmap we used here has already been used to formally model smart schools and sewerage systems. When a certain method is successfully used to design various systems of different sizes and scope successfully, it has the potential to develop into a standard. We hope other authors can expand on our work. Our diagrams can be used as an example of how to do this. They also provide a means to examine multiple solutions for strengths, weaknesses, and costs, before any detailed design or implementation takes place.

*6.1. Future work- **An Overall ADS-B** Model in TLA+*

The only weakness of our model is that it is not detailed enough. A true ADS-B system model with all subsystems and security systems and multiple attackers will be much larger in scope. While we model only one attack above, a real-world model will have all attacks and all security systems integrated and will run them in different ways and see when and where two security systems 'collide.'

And that is our future goal.

Following our methodology, the first step towards envisioning the entirety of ADS-B in TLA+ would start with a diagrammatic representation of what we want, as in Figure 8. We believe this will be an extensive task requiring more hands and time than we have right now. We can, however, think about what such a system would look like.

We make some basic assumptions.

No true communication can take place as long as the lines of communication remain unverified. We must find a way to secure the ADS-B physical layer without compromising on openness.

A lot of solutions impinge on trusting information systems like GNSS. Before we trust any data, we must know it came from the right source. Thus, a periodic 'wellness check' on vital infrastructures like transmission systems, radio frequencies, and GNSS should be the second step.

The third step will utilize the data gleaned from steps 1 and 2 to create a model of the system. We would get an idea of how safe or unsafe it is and would be able to deploy mitigation measures accordingly.

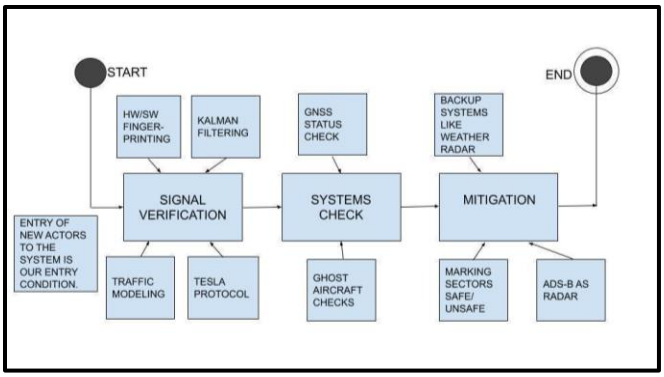


Figure 8: What an overall security solution for ADS-B might look like.

Once again, this would most probably happen on a local level. However, in the spirit of TLA+, we concern ourselves more with the 'what' than the 'how' and do not get too deep into details.

On a higher level, an overall ADS-B specification will consist of multiple figures like Figures 3,4,5, and 6, created for each security challenge, coming together and coalescing into one super-state diagram. This is why we see this paper as a stepping stone to achieving a formally designed ADS-B specification and showing the proper steps to carry out this process.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

All the use case diagrams in section 4 were made using chart-making tools at www.lucidchart.com and the authors would like to thank the website.
The authors would also like to thank Mr. Leslie Lamport for the TLA+ tool which is a fundamental part of this paper.
The authors would also like to thank the University of Cincinnati for its support.

7. References

[1] P. Bhardwaj, C. Purdy, "System design methodologies for safety and security of future wireless technologies in aviation," in 2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS), 235-238, 2019, doi: 10.1109/MWSCAS.2019.8885347.

[2] FEDERAL AVIATION ADMINISTRATION, May 2010, Federal regulation 14 CFR 91.225S, [Online], Available: http://www.ecfr.gov/cgi-bin/text-idx?node=14:2.0.1.3.10#_top.

[3] FEDERAL AVIATION ADMINISTRATION, May 2010, Federal regulation 14 CFR 91.227, [Online], available: http://www.ecfr.gov/cgi-bin/text-idx?node=14:2.0.1.3.10#se14.2.91_1225.

[4] International Civil Aviation Organization, Guidance material on comparison of surveillance technologies (Ed. 1), [Online], Available: http://www.icao.int/APAC/Documents/edocs/cns/gmst_technology.pdf.

[5] M. Strohmeier, V. Lenders, I. Martinovic, "On the security of the automatic dependent surveillance-broadcast protocol," IEEE Communications Surveys and Tutorials, 2015, doi:10.1109/COMST.2014.2365951.

[6] M. Schäfer, V. Lenders, I. Martinovic, "Experimental analysis of attacks on next generation air traffic communication," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2013, doi:10.1007/978-3-642-38980-1_16.

[7] M. Riahi Manesh, N. Kaabouch, "Analysis of vulnerabilities, attacks, countermeasures and overall risk of the Automatic Dependent Surveillance-Broadcast (ADS-B) system," International Journal of Critical Infrastructure Protection, **19**, 16-31, 2017, doi:10.1016/j.ijcip.2017.10.002.

[8] FAA Safety Team, 2019, Team notice NOTC8274, [Online], Available: http://www.faasafety.gov/SPANS/noticeView.aspx?nid=8274.

[9] P. Kocher, R. Lee, G. McGraw, A. Raghunathan, S. Ravi, "Security as a new dimension in embedded system design," in Proceedings - Design Automation Conference, 2004, doi:10.1145/996566.996771.

[10] K. Sampigethaya, R. Poovendran, S. Shetty, T. Davis, C. Royalty, "Future E-enabled aircraft communications and security: The next 20 years and beyond," Proceedings of the IEEE, 2011, doi:10.1109/JPROC.2011.2162209.

[11] ICAO International Communications Group, April 2006, Introduction to ACARS messaging services as implemented via Iridium satellite link, http://www.icao.int/safety/acp/inactive%20working%20groups%20library/acp-wg-m-iridium-7/ird-swg07-wp08%20-%20acars%20app%20note.pdf.

[12] C. W. Lin, A. G. Vincentelli, Security-aware design for cyber-physical systems, Book, Springer, 2017, doi: 10.1007/978-3-319-51328-7.

[13] G. Kalakota, Hierarchical partition based design approach for security of CAN bus based automobile embedded system, Electronic Thesis or Dissertation, University of Cincinnati, 2018.

[14] Object Management Group, 2005, Introduction to OMG's Unified Modeling Language (UML®), http://www.uml.org/what-is-uml.htm.

[15] J. Vidal, F. De Lamotte, G. Gogniat, P. Soulard, J.P. Diguet, "A co-design approach for embedded system modeling and code generation with UML and MARTE," in Proceedings -Design, Automation and Test in Europe, DATE, 2009, doi:10.1109/date.2009.5090662.

[16] J. Jürjens, P. Shabalin, "Tools for secure systems development with UML," in International Journal on Software Tools for Technology Transfer, 2007, doi:10.1007/s10009-007-0048-8.

[17] S. Tasiran, Y. Yu, B.Batson, "Using a formal specification and a model checker to monitor and direct simulation," Design Automation Conference, 2003, doi:10.1109/dac.2003.1219024.

[18] Darlene Storm, August 2012, "Curious hackers inject ghost airplanes into radar, track celebrities' flights," Computerworld, [Online], Available: http://www.computerworld.com/article/2472455/curious-hackers-inject-ghost-airplanes-into-radar--track-celebrities--flights.html.

[19] A. Costin and A. Francillon, July 2012, "Ghost in the air (traffic): on insecurity of ADS-B protocol and practical attacks on ADS-B devices," [Online], Available: http://www.researchgate.net/publication/267557712.

[20] S. Collins, S. McCombie, "Stuxnet: the emergence of a new cyber weapon and its implications," Journal of Policing, Intelligence and Counter Terrorism, 2012, doi:10.1080/18335330.2012.653198.

[21] Niall Firth, MIT Technology Review, June 2019, "How to fight a war in space (and get away with it),"http://www.technologyreview.com/2019/06/26/725/satellite-space-wars/.

[22] Defense Intelligence Agency, January 2019, "Challenges to Security in Space,"http://www.dia.mil/Portals/27/Documents/News/Military%20Power%20Publications/Space_Threat_V14_020119_sm.pdf.

[23] European GNSS Service Center, European Global Navigation Satellite Systems Agency, 23 November 2017, Notice advisory to Galileo users (NAGU) 2017045," [Online], Available: http://www.gsc-europa.eu/notice-advisory-to-galileo-users-nagu-2017045.

[24] Federal Aviation Administration, 2016, "EMB-300 Phenom yaw damper failure due to unreliable or unavailable GPS signal," [Online], available: http://www.faa.gov/documentLibrary/media/Notice/GENOT_7110_711_EMB-300.pdf.

[25] C4ADS innovation for peace, 2019, "Above us Only Stars- Exposing GPS Spoofing in Russia and Syria," [Online], Available: http://www.c4reports.org/aboveusonlystars.

[26] M. Strohmeier, I. Martinovic, "On passive data link layer fingerprinting of aircraft transponders," in CPS-SPC 2015 - Proceedings of the 1st ACM Workshop on Cyber-Physical Systems-Security and/or Privacy, co-located with CCS 2015, 2015, doi:10.1145/2808705.2808712.

[27] M.S. Huang, R.M. Narayanan, Y. Zhang, A. Feinberg, "Tracking of noncooperative airborne targets using ADS-B signal and radar sensing," International Journal of Aerospace Engineering, 2013, doi:10.1155/2013/521630.
[28] S. Bachmann, V. DeBrunner, D. Zrnic, "Detection of small aircraft with doppler weather radar," in 2007 IEEE/SP 14th Workshop on Statistical Signal Processing, 443-447, 2007, doi: 10.1109/SSP.2007.4301297.
[29] S. Rzewuski, K. Kulpa, A. Gromek, "Airborne targets detection using weather radar," in 2015 Signal Processing Symposium, SPSympo 2015, 2015, doi:10.1109/SPS.2015.7168305.
[30] Aircraft Owners and Pilots Association, 2019, "2019 State of general aviation,"http://download.aopa.org/hr/Report_on_General_Aviation_Trends.pdf.
[31] H. Yang, Q. Zhou, M. Yao, R. Lu, H. Li, X. Zhang, "A practical and compatible cryptographic solution to ADS-B security," IEEE Internet of Things Journal, 2019, doi:10.1109/JIOT.2018.2882633.
[32] Leslie Lamport, 4 September 2018, "A High-Level View of TLA+," [Online], Available: http://lamport.azurewebsites.net/tla/high-level-view.html.
[33] C. Newcombe, T. Rath, F. Zhang, B. Munteanu, M. Brooker, M. Deardeuff, "How amazon web services uses formal methods," Communications of the ACM, 2015, doi:10.1145/2699417.
[34] C. Newcombe, "Why Amazon chose TLA+," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2014, doi:10.1007/978-3-662-43652-3_3.
[35] N.H. Obeidat, C. Purdy, "Modeling a smart school building system using UML and TLA+," in Proceedings - 3rd International Conference on Information and Computer Technologies, ICICT 2020, 2020, doi:10.1109/ICICT50521.2020.00028.
[36] S. Latif, A. Rehman, N.A. Zafar, "Modeling of sewerage system linking UML, automata and TLA+," in 2018 International Conference on Computing, Electronic and Electrical Engineering, ICE Cube 2018, 2019, doi:10.1109/ICECUBE.2018.8610971.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1666-1673 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Real-time Gradient-Aware Indigenous AQI Estimation IoT Platform

Hasan Tariq[*,1], Abderrazak Abdaoui[1], Farid Touati[1], Mohammad Abdullah Al Hitmi[1], Damiano Crescini[2], Adel Ben Mnaouer[3]

[1]*Department of Electrical Engineering, College of Engineering, Qatar University, 2713, Qatar*

[2]*Dipartimento di Ingegneria dell'Informazione, Brescia University, 25121, Italy*

[3]*Department of Computer Engineering and Computational Sciences, Faculty of Engineering, Applied Sciences and Technology, Canadian University Dubai, 117781, UAE*

ARTICLE INFO

Article history:
Received: 16 July, 2020
Accepted: 08 October, 2020
Online: 28 December, 2020

Keywords:
Air quality
IoT
Gas sensors node
Multi-variable environmental mapping

ABSTRACT

Environmental monitoring has gained significant importance in outdoor air quality measurement and assessment for fundamental survival as well as ambient assisted living. In real-time outdoor urban scale, instantaneous air quality index estimation, the electrochemical sensors warm-up time, cross-sensitivity computation-error, geo-location typography, instantaneous capacity or back up time; and energy efficiency are the six major challenges. These challenges lead to real-time gradient anomalies that effect the accuracy and pro-longed lags in air quality index mapping campaigns for state and environmental/meteorological agencies. In this work, a gradient-aware, multi-variable air quality sensing node is proposed with event-triggered sensing based on position, gas magnitudes, and cross-sensitivity interpolation. In this approach, temperature, humidity, pressure, geo-position, photovoltaic power, volatile organic compounds, particulate matter (2.5), ozone, Carbon mono-oxide, Nitrogen dioxide, and Sulphur dioxide are the principle variables. Results have shown that the proposed system optimized the real-time air quality monitoring for the chosen geo-spatial cluster (Qatar University).

1. Introduction

Living conditions or probability of life forms are dependent on environmental conditions, i.e. combination of gases, temperature, humidity, pressure, and light [1]. The quality of respiratory life processes is directly related to air quality at a particular geo-location [2]. Gradual deterioration in urban air quality is ambient each year due to the increasing population, chemical emissions from machinery, and depreciation in green ecology [3]. Air quality sustainability is a major concern in the age of imperialism [4] due to exponential development and construction density in the urban landscape.

An increase of 41% in the design and development of gas and pollutant sensing systems was observed in the environmental market since 2001 [5]. The top contributions AirNut, PA-I and PA-II, Egg, PATS+, and S-500, CairClip, Portable ASLUNG, AirSensEUR, Met One, AQY v0.5, Vaisala AQT410, 2B Tech, and AQMesh V3.0 systems had measurement capabilities in specific pollutants and gases presented in [6–8]. FIS SP-61 by FIS, O3-3E1F by CityTechnology, AirSensEUR v.2 by LiberaIntentio, and S-500 by Aeroqual, and AirSensEUR used a built-in AlphaSense OX-A431 limited to O3 [9–11]. The PMS1003 and PMS3003 by Plantower; DC1100 PRO and DC1700 by Dylos; OPC-N2 by AlphaSense had only sensing support for particulate matter (PM) [11–13]. The CO-3E300 by City Technology; CO-B4 by Alphasense, MICS-4515 by and SGX Sensortech, Smart Citizen Kit by Acrobotic and the RAMP had only abilities to measure carbon mono-oxide [4]. It illuminated the requirement of structure real-time sensor data processing framework for multiple gases sensing nodes with warm-up times below 5mins at ambient outdoor environmental conditions.

The top 10 environmental protection agencies (EPAs) unanimously agreed on the standard of four core gases for outdoor air quality [14] i.e. Ozone, Nitrogen Dioxide, Sulphur Dioxide, Carbon Monoxide, and two prime pollutants particulate matter PM2.5 and PM10. The concentration or ratio of four critical gases and air particles part per million (ppm) constitutes air quality index (AQI) is the fundamental design using Smart Rig [15]. The ratio of

[*] Corresponding Author: Hasan Tariq, Lab106A, B09, Qatar University, P.O. Box 2713, Doha, Qatar +97450419852, hasan.tariq@qu.edu.qa

Ozone (O3), Sulphur Dioxide (SO2), Carbon Monoxide (CO), and Nitrogen Dioxide (NO2) are the standard gases discussed in major environmental standards for AQI [4,14,15]. A standard outdoor air quality node (O-AQN) was needed to measure discrete and cumulative AQI with minimum data processing and power consumption.

Related Works

The existing O-AQNs, Urban AirQ, Smart Citizen Kit, SeReNo V1, AirQ Mesh needed improvement in energy efficiency, GPS, time-series vector structuring, and remote data transmission to ensure dynamic air quality mapping [16–20].

1.1. Urban AirQ

The AirQ platform [17] by Vishal et al was a smart and cost-effective solution for air quality monitoring. It supported wireless data communication technologies including WiFi, 4G, Bluetooth, LoRa, etc. It was supported by non-embedded data management and the implementation of smart data analytics algorithms exhibited at an AirQ Dashboard and Smartphone App. It had a dependency on external hardware and software with more processing needs with only air quality sensors.

1.2. Smart Citizen Kit

The Smart Citizen Kit [18] by Gullien et al was dependent on an additional resource named "Smart Citizen Station". Needed improvement in multi-variable time-series measurements and location accuracy, energy efficiency, and reduced data processing. Multi-variable sensing with simultaneous measurement of gases and environmental variables.

1.3. SeReNo V1

The Sensor Node (SeReNo) V1 [19] by Farid et al was the first node developed with indoor air quality monitoring and energy harvesting capabilities. SeReNo V2 presented in this work [20] is the improvement of [19] that needed to be upgraded with GSM and GPS, more Flash to perform AQI data processing needed for optimized energy efficiency, electrochemical gas sensors warm-up times, and outdoor air quality.

1.4. AirQ Mesh

The networked AirQ Mesh [21] by Nuria et al focused on the review and performance of existing AirQ platforms and their maximum utilization leading to credible results by EPA and Air Quality Directive (AQD). Data payloads were very high due to response time estimation errors leading to high storage requirements with energy challenges in the field were the gaps in the Nuria et al contribution.

Furthermore, the improvement potential in context-aware data processing model and thresholding for detection was discussed by Hasan et al in [22,23] for multi-sensor single-board node. The energy efficiency schemes were one major gap highlighted by [24]. The works Lebilium Waspmote [21] was Zigbee but needed GSM/GPRS support for outdoor sensing and have been more contributory if equipped with geo-spatial traces or cadastral layouts support for all the measured variables like AirBox [22]. An infrastructure embedded system was needed to accommodate the maximum sensing, minimum data processing, AQI estimation, and long-haul deployability [19–23].

2. Materials and Methods

This work proposed a smart AQM IoT system with three major components:

- The Real-time Gradient Aware Multi-Variable Sensing Model (GAM)
- The SeReNo V2 AQM Sensing Platform
- Real-time AQI Estimation Engine (AQM-EE)

First, we proposed the GAM and then the pillars of GAM led to the design and fabrication SeReNo V2 node.

3. The Real-time Gradient Aware Multi-Variable Sensing Model (GAM)

A real-time structured multi-variable structured data time-series vector was needed to proceed with AM. Let us consider an EPA standard outdoor air quality index (O-AQI) real-time variables as temperature T in centigrade, pressure P in pascals, humidity H in %, volatile organic compounds as VoC (ppm), particulate matter as PM (ug/m3), Ozone as $O3$ (ppm), Nitrogen Dioxide as $NO2$ (ppm), Carbon Monoxide as $CO2$ (ppm), and Sulphur Dioxide as $SO2$ (ppm). The real-time O-AQI data was proposed as a commutative time series multi-variable vector $V_{O\text{-}AQI}$ of two non-linear time-series with t_1 and t_2 of environmental E and gas G sensors data at a particular geo-location L given as:

$$V_{O\text{-}AQI}(t) = [E(t_1), G(t_2)] : L(t) \quad \text{where } t = (0, 1, 2, 3, \ldots\} \tag{1}$$

The practicality of response time of heterogeneous sensors was taken in account for non-linear time-series decomposition t, gas sensor response time t_2 is greater than the response time of environmental sensors t_1 with relationship $t_2 > t_1$ given as:

$$t_2 = 3t_1 \qquad where\ [t_1, t_2] \in t \tag{2}$$

The environmental sensor variables function E for sensor array $A_E(T, P, H, VoC, PM)$ as $E(A_E, t_1)$; and for gas sensors array $A_G(O3, NO2, SO2, CO)$ as $G(A_G, t_2)$ and position vector L as reference function GPS using GSM network cell locations (using $AT+CIPGSMLOC=1,1$) for L_{GPRS} and GPS module as L_{GPS} (using AT+CGPSINF). For precise AQM the L_{GPS} must belong to the slope of L_{GPRS1} and L_{GPRS2} in a particular slope format by NEMEA specifier for consecutive cells is given as:

$$L_{GPS}(X, Y) \in [L_{GPRS1}(X2, Y2), L_{GPRS2}(X1, Y1)] \tag{3}$$

The agreed L_{GPS} was termed as $L(t)$ where condition (3) was satisfied. From equations (1), (2), and (3) the finalized AQM vector of $V_{O\text{-}AQI}$ was derived as:

$$V_{O\text{-}AQI}(t) = [E(A_E(T, P, H, VoC, PM), t_1), G(A_G(O3, NO2, SO2, CO), t_2)] : L(t) \tag{4}$$

The transmission of this structured AQM vector can be power leeching. Here we introduce a gradient impact vector as a function of warm-up time t_W, cross-sensitivity classifier C_S, geo-location $L(t)$, and energy efficiency E_E. Three bounded value conditions we applied on GAM programmed in the SeReNo V2 firmware:

a) The mandatory system unit to stay on is the power is the micro-controller with power consumption P_M and at sleep mode, it has $P_{M\text{-}SLEEP}$ power with all the critical sensors active only.

b) For E_E the system has to have instantaneous power P_i varying between $P_{M\text{-}SLEEP} < P_i < P_M$.

c) The t_W had to be in the lower section of the curve of P_i. Likewise, at the constant magnitude of vector A_E at t_n and t_{n+1}, the difference cross-sensitivities had to stay constant at a particular $L(t)$.

4. The SeReNo V2 AQM Sensing Platform

A generic multi-variable AQM node consists of an EPA recommended sensors for $V_{O\text{-}AQI}$ based on equation (4), application-specific signal conditioning, power, and communication sections. The block diagram of a generic AQM node is given below in figure 1.

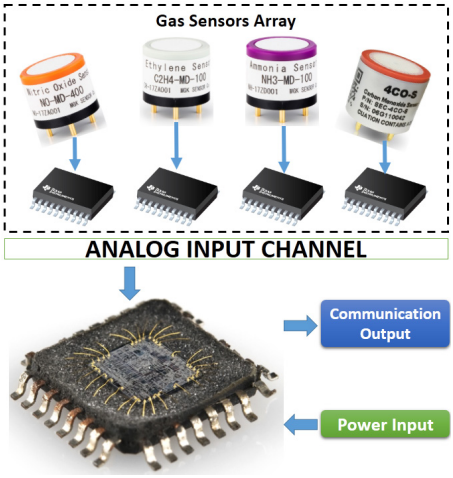
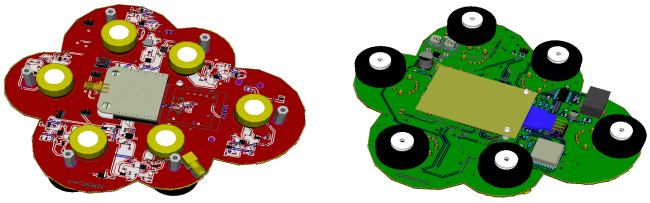


Figure 1: A Generic Multi-Variable AQM Sensing

In our previous works [14,15,19,20], we had filled the design gaps in indoor air quality monitoring and developed a multi-sensor AQM node called Sensors V1 revision 1 and revision 2 presented in figure 1. In this work, work proposed a novel O-AQM node by improving the system architecture of SeReNo V1 as SeReNo V2. In our recent work [24] we had developed SeReNo V2 nodes tested for long-haul autonomous operations of 1.5 years that were improved in this work. The existing SeReNoV2 node was improved with GAM and AQI-EE to meet the state of the art needs.



a. Top View b. Bottom View

Figure 2: Two 3D Layouts of SeReNo V1 AQM Node

For outdoor AQM SeReNo V1 was improved to V2 as per global EPAs standards recommended sensors, energy harvesting long-range communication support. The comparison of major capabilities of SeReNo V1 and V2 is exhibited in the table below.

Table 1: SeReNo V1 and V2 Characteristics

Blocks	SeReNo V1	SeReNo V2
Energy Block	PV and Manual Charging + LiPo Battery 3.7V / 1200mAh + 6 Thermo-Electric Generators	Dual PV + Manual Charging
Communication Block	RF 433Mhz	GSM/GPRS
Sensors Block	6 Indoor Sensors Only interfaced with Analog Input	10 Outdoor Sensors Networked with I2C (4 Gas Sensors + 2 Pollutant Sensors + 4 Environmental Sensors)
Environmental Sensors Block	Absent	4 Sensors
Controller Block	RFC200	ATMega328P
Location Block	NIL	GPS
Local Display	NIL	Present
Energy Optimization	Partial	Full
Sensors Control	NIL	Present
Outdoor Cross Sensitivity	NIL	Present

GAM was a core engine driving the multi-sensor O-AQM IoT system. Mathematically modeled design and equations (1~4) parameters are addressed in the hardware blocks. A 1-1 correspondence electronics and instrumentation system was designed SeReNo V2 presented in figure 3 presented below.

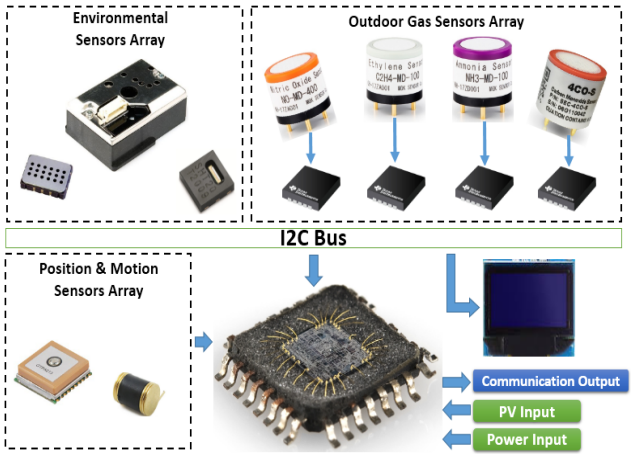
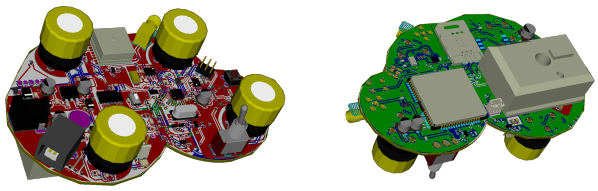


Figure 3: Block Diagram of SeReNo V2 AQM Node

Fig 3 contains three sets of sensors The PCB design and completely assemble functional O-AQM node is presented in figure 4 given below.



a. Top View b. Bottom View

Figure 4: Two 3D Layouts of SeReNo V2 AQM Node

Three SeReNo V2 nodes with GAM and AQI-EE as improved firmware were deployed in QU for outdoor testing. The deployment was focused on the resolution of six key improvements mentioned in the abstract and literature review through proposed GAM and AQI-EE for six months, i.e. QU Greenhouse exhibited in figure 5 below.

Figure 5: SeReNo V2 Deployment in QU to utilize the GAM

This deployment was for six months and all the data was collected from Ubidots IoT platform as CSV files for further processing. The first month was stability assessment of GAM and AQI-EE firmware and the remaining 5 months were the measurement campaign.

5. Real-time AQI Estimation Engine (AQM-EE)

An AQI magnitude is estimated from 5 critical air variables or termed as pollutants, i.e. PM2.5, O3, NO2, SO2, and CO. The critical variable with the maximum AQI magnitude is termed as the "principle pollutant" and entire computation revolves around its significant figures among all the critical air variables. Let I_p be the index of the principal pollutant; C_p as the rounded-off concentration of pollutant p; $B_{P\text{-}high}$ as the breakpoint greater or equal to C_p; $B_{P\text{-}low}$ as the breakpoint less than or equal to C_p; I_{high} is the AQI corresponding to $B_{P\text{-}high}$; I_{low} is the AQI corresponding to $B_{P\text{-}low}$. The AQI is generically estimated as:

$$I_P = [(I_{high} - I_{low})/(B_{P-high} - B_{P-low})] \times (C_P - B_{P-low}) + I_{low} \quad (5)$$

Every pollutant was formulated in using equation (5) and given by in equations (6 to 10).

$$I_{O3} = [(I_{high} - I_{low})/(B_{O3-high} - B_{O3-low})] \times (C_{O3} - B_{O3-low}) + I_{low} \quad (6)$$

$$I_{SO2} = [(I_{high} - I_{low})/(B_{SO2-high} - B_{SO2-low})] \times (C_{SO2} - B_{SO2-low}) + I_{low} \quad (7)$$

$$I_{CO} = [(I_{high} - I_{low})/(B_{CO-high} - B_{CO-low})] \times (C_{CO} - B_{O3-low}) + I_{low} \quad (8)$$

$$I_{NO2} = [(I_{high} - I_{low})/(B_{NO2-high} - B_{NO2-low})] \times (C_{NO2} - B_{NO2-low}) + I_{low} \quad (9)$$

$$I_{PM} = [(I_{high} - I_{low})/(B_{PM-high} - B_{PM-low})] \times (C_{PM} - B_{PM-low}) + I_{low} \quad (10)$$

Furthermore, the standard AQI evaluation template for a specific region provided by EEA as Common Air Quality Index (CAQI) and US EPA as AQI were adapted in this AQM-EE and given as a cumulative state chart given below in figure 6 and 7.

Pollutant	Index level (based on pollutant concentrations in µg/m3)				
	Good	Fair	Moderate	Poor	Very poor
Particles less than 2.5 µm ($PM_{2.5}$)	0-10	10-20	20-25	25-50	50-800
Particles less than 10 µm (PM_{10})	0-20	20-35	35-50	50-100	100-1200
Nitrogen dioxide (NO_2)	0-40	40-100	100-200	200-400	400-1000
Ozone (O_3)	0-80	80-120	120-180	180-240	240-600
Sulphur dioxide (SO_2)	0-100	100-200	200-350	350-500	500-1250

Figure 6: CAQI EEA AQI Cumulative State Chart

The SeReNo V2 was successfully able to give real-time AQI-EE estimations that have plenty of environmental, metrological, and health applications and significance.

Breakpoints							AQI	Category
O_3 (ppm) 8-hour	O_3 (ppm) 8-hour[1]	PM_{10} (µg/m³)	$PM_{2.5}$ (µg/m³)	CO (ppm)	SO_2 (ppm)	NO_2 (ppm)		
0-0.064	–	0-54	0-15.4	0-4.4	0-0.034	(2)	0-50	Good
0.065 0.084	–	55 154	15.5 40.4	4.5 9.4	0.035 0.144	(2)	51 100	Moderate
0.085-0.104	0.125-0.164	155-254	40.5-65.4	9.5-12.4	0.145-0.224	(2)	101-150	Unhealthy for sensitive groups
0.105-0.124	0.165-204	255-354	65.5-150.4	12.5-15.4	0.225-0.304	(2)	151-200	Unhealthy
0.125-0.374 (0.155-0.404)[4]	0.205-0.404	355-424	150.5-250.4	15.5-30.4	0.305-0.604	0.65-1.24	201-300	Very unhealthy
(3)	0.405-504	425-504	250.5-350.4	30.5-40.4	0.605-0.804	1.25-1.64	301-400	Hazardous
(3)	0.505-0.604	505-604	350.5-500.4	40.5-50.4	0.805-1.004	1.65-2.04	401-500	Hazardous

Figure 7: US EPA AQI Cumulative State Chart

The EPA rated sensors produced the results in thresholds of the chart in figure 6.

The AQM-EE uses 5 KPIs to evaluate the performance of GAM given as:

5.1. Energy Efficiency:

The instantaneous energy consumed by node E_N for a minute is the sum of 60 secs for the product of supply voltage V_{DC}, instantaneous current I_i, and time T_i given in equation 11.

$$E_N = \sum_{i=0}^{n} V_{DC} * I_i * T_i \quad (11)$$

The difference between the maximum rated power consumption of node and minimum instantaneous power consumption in terms of watt-hour is termed as energy efficiency or E-Efficiency symbolized as E_e and given by:

$$E_e\ (Wh) \approx \left(\frac{N-M}{N}\right)(E_N) \quad (12)$$

5.2. Cross-Sensitivity Computation Error:

The cross-sensitivity computation error is the average of the total difference in a sensor's reaction to other gases being measured.

$$C_s\ (\%) \approx \left(\sum_{i=0}^{n} \Delta G_i\right)/n \quad (13)$$

It is used as the correction factor and must be in the range defined by the vendor.

5.3. Typo-graphic Error:

The difference in the GPS coordinates by GPS and AGPS was termed as typo-graphic or GPS error, presented in equation below:

$$T_E = |L_{GPS}(X, Y) - L_{AGPS}(X, Y)| \quad (14)$$

The acceptable value for this error was marked as 20 meters for precise AQM.

5.4. The ratio of Magnitude E_{BAT}/E_N or Instantaneous Capacity:

The ratio of instantaneous system battery energy E_{BAT} in watt-hour (Wh) and instantaneous node energy E_N in watt-hour (Wh) was termed the instantaneous capacity magnitude C_{MAG} of the node with current operations given as:

$$C_{MAG} = E_{BAT} / E_N \quad (15)$$

For long haul AQM, the C_{MAG} must be greater than 1 to ensure at least 1-hour autonomous operation.

5.5. Improved Warm-Up Time of Gas Sensors:

The start-up warm-up time is mandatory for the gas sensors to acquire chemical equilibrium with the atmosphere in which they are used. The warm-up time was improved by putting all the micro-controller, GPS, and GSM modules to sleep and then get active based on the gradient as mentioned in section 3.

6. Results and Discussion

After the long haul deployment of six months, the data results obtained were displayed on the Ubidots IoT platform as exhibited in figure 7.

The eleven real-time variables were exhibited in Figure 8 sending data through GSM model QuecTel M10. The application of GAM enabled only meaningful data to be sent to the cloud that made time-series more non-linear as only gradients impacted values were being transmitted. The following plots of individual variables give more insights of GAM in SeReNo V2.

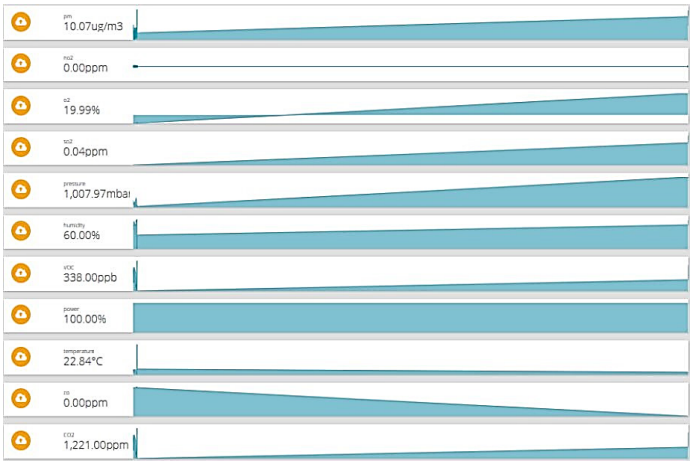


Figure 8: The SeReNo2 IoT Dashboard at Ubidots IoT

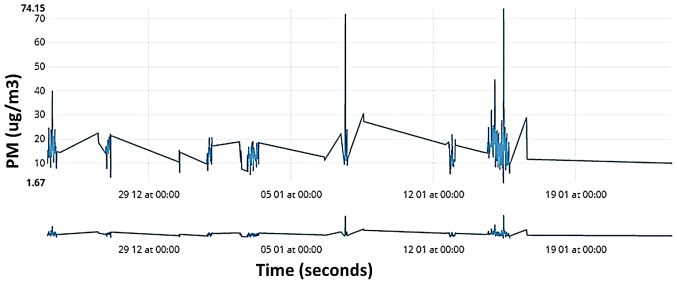


Figure 9: Illustration of Particulate Matter data due to GAM

Figure 9 explained the real-time measurement of PM from December 22, 2019, to January 22, 2020. The max of 74.15 µg/m^3 was observed on January 17, 2020, and a minimum of 1.67 µg/m^3 on December 27, 2019.

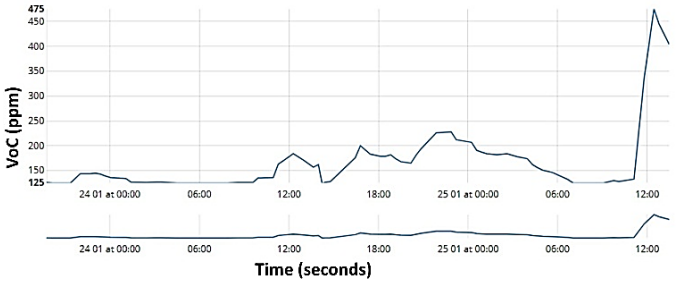


Figure 10: Illustration of VoC and Impact of GAM at 12:00 PM

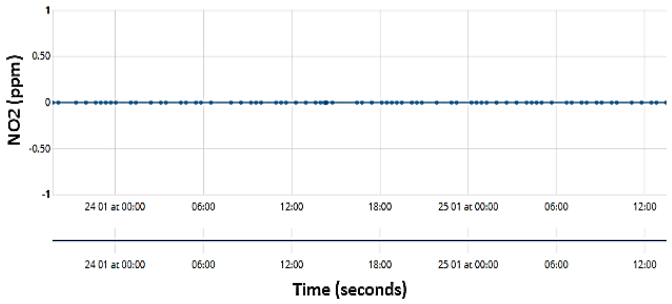


Figure 11: Illustration of Nitrogen Dioxide data due to GAM

In figure 10, it is ambient that amount of VoCs increased significantly due to day time traffic and air pollution in the atmosphere. This the special behavior common in all the days for 6 months.

In figure 11, it is eloquent that NO2 being extremely hazardous stayed zero and was transmitted to cloud whenever there was a signature of cross-sensitivity and time-series for NO2 was also non-linear.

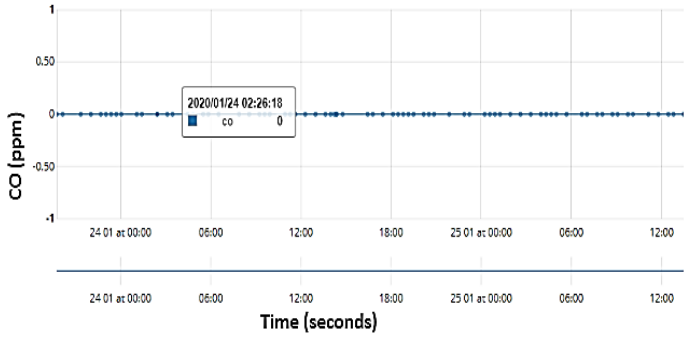


Figure 12: Illustration of Carbon Monoxide data due to GAM

In figure 12, it was observed that CO being hazardous was also observed as a non-linear time-series and the same as NO2 had cross-sensitivity signatures.

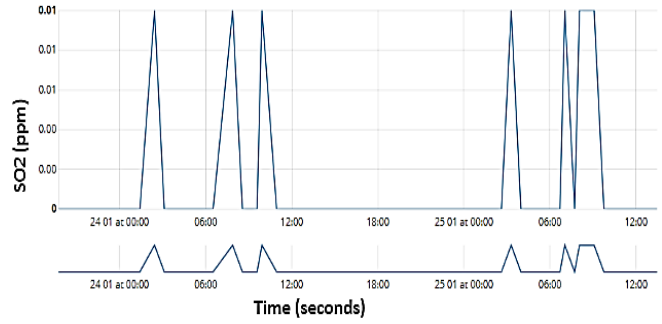


Figure 13: Sulphur Dioxide Signatures in Atmosphere

In figure 13, nominal magnitudes of SO2 were observed in the atmosphere due to plenty of gas processing plants that flare H2S and rest is preserved by Sulphur recovery units (SRUs).

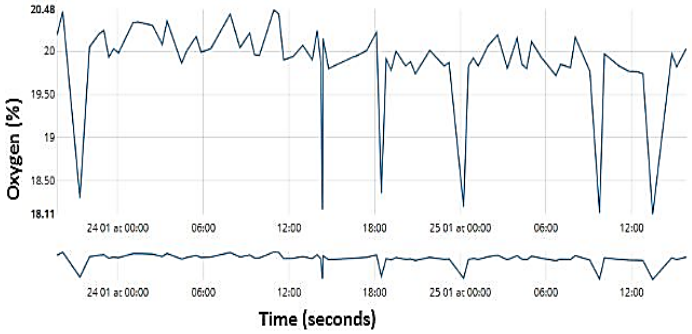


Figure 14: Presence of Oxygen Traces in Air

The SeReNo V2 node was initially equipped with an O2 sensor that was later on replaced by the O3 sensor for 2 years. The breathing conditions were favorable due to the measured average of 19.3 O2 in the air as obvious from figure 14.

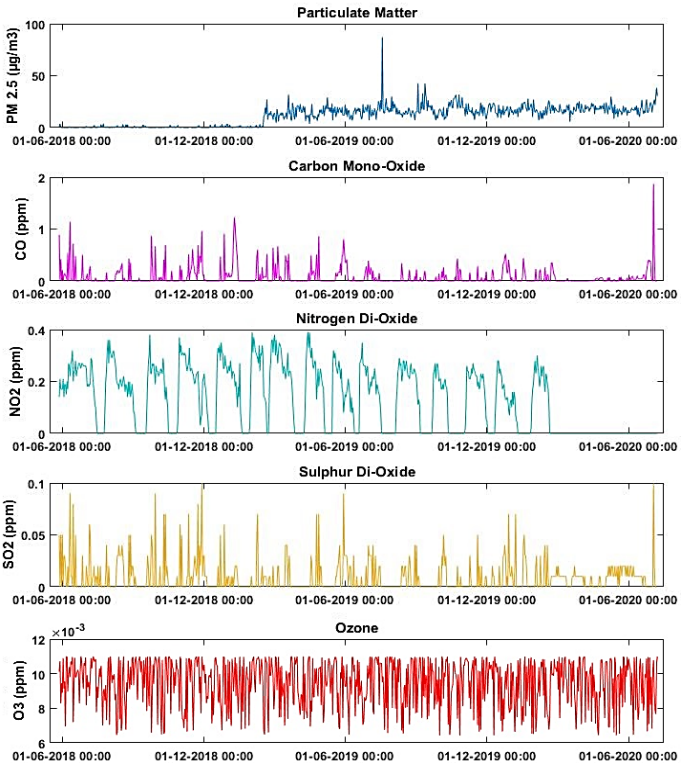


Figure 15: SeReNoV2 Total Deployment Data for EPAs recommended measurement

The first deployment of the SeReNoV2 node was 1.5 years to validate the autonomous operational capabilities as presented in figure 15 at the maximum sampling rate for pollutants and outdoor gases. This deployment was from June 1, 2018, to January 1, 2020.

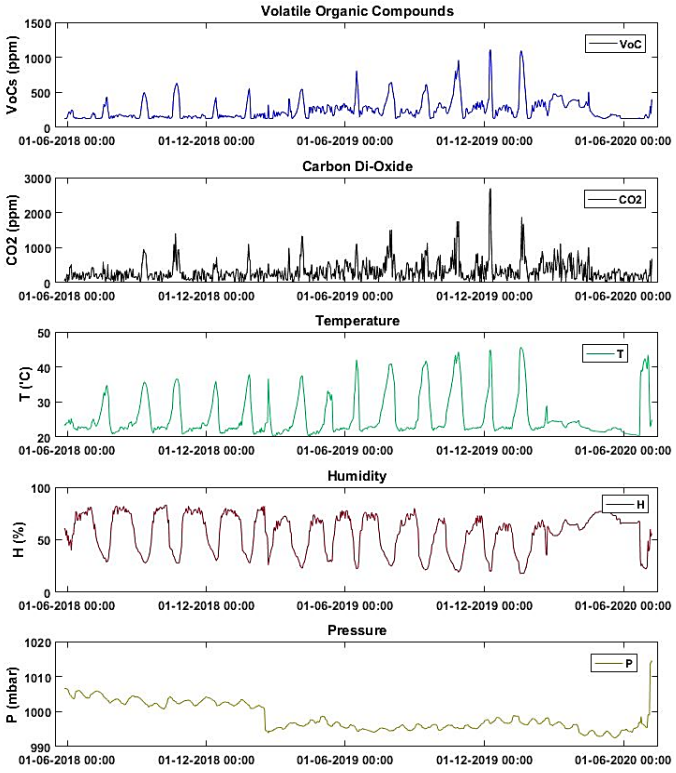


Figure 16: SeReNoV2 Total Deployment Data for EPAs Environmental Variables

In figure 16, the core 5 environmental variables, VoC, CO2, temperature, humidity, and pressure can be observed during the entire deployment tenure.

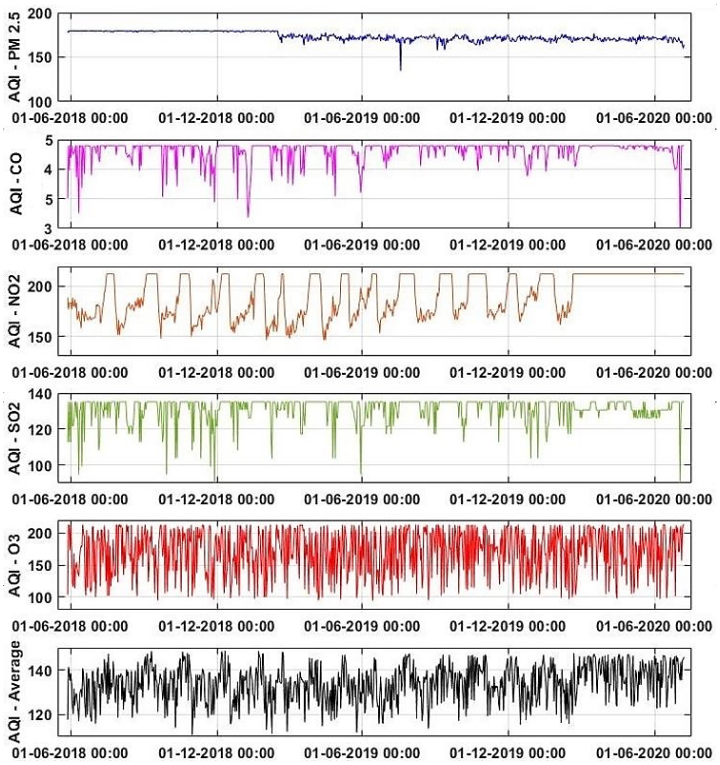


Figure 17: SeReNoV2 Estimation of real-time Discrete and Cumulative AQI

The AQI-EE key performance indicators (KPIs) were estimated for the programmed GAM on SeReNo V2 were the major contributions that enabled all the outcomes presented in figures 17 to 19. 80 meters typographic error can be observed in figure 18.

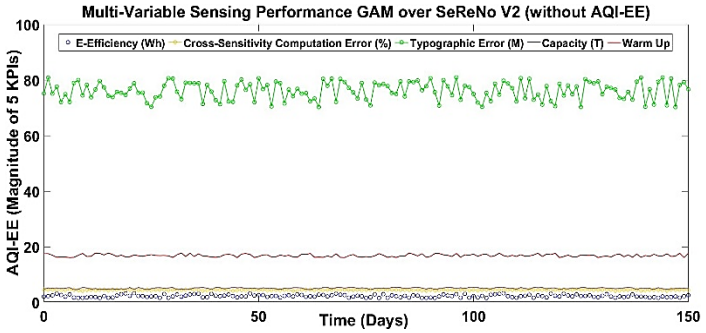


Figure 18: The GAM without AQI-EE KPIs for 5 months

The warm-up time of 18 minutes of electrochemical gas sensors was due to the average humidity of 63%. The warm-up time of electrochemical gas sensors is ideally 2 to 7. The cross-sensitivity error of 5% is normal. The E_e is near 0.8~1% is due to the extremely low power design on SeReNoV2.

The AQI-EE KPIs for GAM efficiency on SeReNo V2 can be observed in these recent 6 months deployment with 5 months of KPI assessment. The correlation of GPS and AGPS minimized typographic error to 3.5 meters. The cross-sensitivities error was also reduced to 2.7~4.8%. The E_e is aligned with the electrochemical gas sensors' operational instances. The capacity has improved to 1.4%. A significant reduction in warm-up times was observed due to the $P_{M\text{-}SLEEP}$ variable. Lastly, the measure data size with respect to power energy efficiency was mapped and significant amounts of data were observed even at very low power consumption as a landmark for future research.

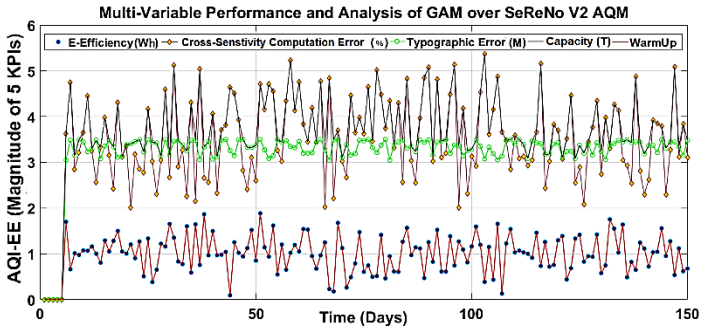


Figure 19: The GAM KPIs for SeReNo V2

7. Conclusion

A gradient-ware urban air quality index estimation multi-variable sensing IoT system was developed. The existing SeReNoV2 nodes were improved with the proposed system. The gradient-aware model and air quality estimation engine contributed to a significant and noticeable reduction in data processing for real-time AQI computation, electrochemical sensors warm-up times, cross-sensitivity errors, geo-location typographic-errors, and improved energy efficiency. The results can be summarized in four key milestones for accurate estimation of urban AQI estimation: 1) implemented gradient-aware sensing model as non-linear time series vector compensating the commutative anomalies in gases and environmental variables measurements; 2) the energy efficiency mechanism was able to keep the node up for 1.5 years constantly without any external charging source; 3) reduced warm times due to the proposed instantaneous power condition, and 4) precise estimation of AQI recommended air quality chart principal parameters. These novel air quality system results demonstrated the potential to handle redundancy in sensing time-triggers and non-linear timers as well.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This publication was made possible by NPRP grant # 10-0102-170094 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "A survey on sensor networks," IEEE Communications Magazine, **40**(8), 102–114, 2002, doi:10.1109/MCOM.2002.1024422.

[2] M. Abderrahim, H. Hakim, H. Boujemaa, F. Touati, "A Clustering Routing based on Dijkstra Algorithm for WSNs," in 2019 19th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), 605–610, 2019, doi:10.1109/STA.2019.8717279.

[3] M. Vinyals, J.A. Rodriguez-Aguilar, J. Cerquides, "A Survey on Sensor Networks from a Multiagent Perspective," The Computer Journal, **54**(3), 455–470, 2011, doi:10.1093/comjnl/bxq018.

[4] F. Touati, H. Tariq, M.A. Al-Hitmi, A. Ben Mnaouer, A. Tahir, D. Crescini, "IoT and IoE prototype for scalable infrastructures, architectures and platforms," International Robotics & Automation Journal, **4**(5), 319–327, 2018, doi:10.15406/iratj.2018.04.00144.

[5] M. Abderrahim, H. Hakim, H. Boujemaa, F. Touati, "Energy-Efficient Transmission Technique based on Dijkstra Algorithm for decreasing energy consumption in WSNs," in 2019 19th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), 599–604, 2019, doi:10.1109/STA.2019.8717210.

[6] P. Temdee, R. Prasad, Communications for Context-Aware Applications, 65–96, 2018, doi:10.1007/978-3-319-59035-6_4.

[7] F. Touati, H. Tariq, D. Crescini, A.B. Manouer, "Design and Simulation of a Green Bi-Variable Mono-Parametric SHM Node and Early Seismic Warning Algorithm for Wave Identification and Scattering," in 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC), 1459–1464, 2018, doi:10.1109/IWCMC.2018.8450277.

[8] N. Zaghdoud, W.H. Alouane, H. Boujemaa, F. Touati, "Secure Performance of AF and DF Relaying in Cooperative Noma Systems," in 2019 19th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), 614–619, 2019, doi:10.1109/STA.2019.8717242.

[9] H. Taheri, P. Neamatollahi, O.M. Younis, S. Naghibzadeh, M.H. Yaghmaee, "An energy-aware distributed clustering protocol in wireless sensor networks using fuzzy logic," Ad Hoc Networks, **10**(7), 1469–1481, 2012, doi:https://doi.org/10.1016/j.adhoc.2012.04.004.

[10] H. Lin, D. Bai, D. Gao, Y. Liu, "Maximum data collection rate routing protocol based on topology control for rechargeable wireless sensor networks," Sensors, **16**(8), 2016, doi:10.3390/s16081201.

[11] M. Abderrahim, H. Hakim, H. Boujemaa, F. Touati, "Multihop transmission strategy using dijkstra algorithm to improve energy efficiency in WSNs," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer International Publishing: 97–107, 2018, doi:10.1007/978-3-030-02849-7_9.

[12] T. Khan, M.A. Shah, K. M., "Communications, Applications, Technology and Beyond: A Surveyon Smart Grid," Sindh University Research Journal-SURJ, **48**(3), 651–656, 2016.

[13] H. Tariq, A. Tahir, F. Touati, M.A.E. Al-Hitmi, D. Crescini, A. Ben Manouer, "Geographical area network-structural health monitoring utility computing model," ISPRS International Journal of Geo-Information, **8**(3), 2019, doi:10.3390/ijgi8030154.

[14] U. Asghar, F. Touati, D. Crescini, A. Galli, A. Ben Mnaouer, "Development of Highly Efficient Multi-invariable Wireless Sensor System Design for Energy Harvesting," 2018.

[15] A. Abdaoui, F. Touati, H. Tariq, A. Ben Mnouar, "A smart rig for calibration of gas sensor nodes: Test and deployment," DTS 2020 - IEEE International Conference on Design and Test of Integrated Micro and Nano-Systems, (i), 18–20, 2020, doi:10.1109/DTS48731.2020.9196182.

[16] H. Luo, W. Li, X. Wu, "Design of indoor air quality monitoring system based on wireless sensor network," IOP Conference Series: Earth and Environmental Science, **208**, 12070, 2018, doi:10.1088/1755-1315/208/1/012070.

[17] V. Choudhary, J.H. Teh, V. Beltran, H.B. Lim, "AirQ: A Smart IoT Platform for Air Quality Monitoring," in 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC), 1–2, 2020, doi:10.1109/CCNC46108.2020.9045550.

[18] G. Camprodon, Ó. González, V. Barberán, M. Pérez, V. Smári, M.Á. de Heras, A. Bizzotto, "Smart Citizen Kit and Station: An open environmental monitoring system for citizen participation and scientific experimentation," HardwareX, **6**, e00070, 2019, doi:https://doi.org/10.1016/j.ohx.2019.e00070.

[19] F. Touati, C. Legena, A. Galli, D. Crescini, P. Crescini, A. Ben Mnaouer, "Environmentally powered multiparametric wireless sensor node for air quality diagnostic," Sensors and Materials, **27**(2), 177–189, 2015, doi:10.18494/sam.2015.1056.

[20] H. Tariq, A. Abdaoui, F. Touati, M.A.E. Al-Hitmi, D. Crescini, A.B. Mnaouer, "An Autonomous Multi-Variable Outdoor Air Quality Mapping Wireless Sensors IoT Node for Qatar," in 2020 International Wireless Communications and Mobile Computing (IWCMC), 2164–2169, 2020, doi:10.1109/IWCMC48107.2020.9148392.

[21] N. Castell, F.R. Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, D. Broday, A. Bartonova, "Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?," Environment International, **99**, 293–302, 2017, doi:https://doi.org/10.1016/j.envint.2016.12.007.

[22] H. Tariq, F. Touati, M.A.E. Al-Hitmi, D. Crescini, A.B. Manouer, "Design and Implementation of Information Centered Protocol for Long Haul SHM Monitoring," in 2019 IEEE International Conference on Design & Test of Integrated Micro & Nano-Systems (DTS), 1–7, 2019, doi:10.1109/DTSS.2019.8915286.

[23] H. Tariq, F. Touati, M.A.E. Al-Hitmi, D. Crescini, A. Ben Mnaouer, "A real-time early warning seismic event detection algorithm using smart geo-spatial bi-axial inclinometer nodes for Industry 4.0 applications," Applied Sciences, **9**(18), 1–25, 2019, doi:10.3390/app9183650.

[24] A. Abdaoui, S.H.M. Ahmad, H. Tariq, F. Touati, A.B. Mnaouer, M. Al-Hitmi, "Energy Efficient Real time Outdoor Air Quality Monitoring System," in 2020 International Wireless Communications and Mobile Computing (IWCMC), 2170–2176, 2020, doi:10.1109/IWCMC48107.2020.9148229.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 6, 1674-1682 (2020)
www.astesj.com
Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Japanese Abstractive Text Summarization using BERT

Yuuki Iwasaki [*1], Akihiro Yamashita[1], Yoko Konno[2], Katsushi Matsubayashi[1]

[1]*National Institute of Technology, Tokyo College, 193-0997, Japan*

[2]*CHOWA GIKEN Corporation, 001-0021, Japan*

ARTICLE INFO

Article history:
Received: 27 August, 2020
Accepted: 12 December, 2020
Online: 28 December, 2020

Keywords:
Abstractive text summarization
BERT
Japanese text summarization

ABSTRACT

In this study, we developed and evaluated an automatic abstractive summarization algorithm in Japanese using a neural network. We used a sequence-to-sequence encoder-decoder model for practical purposes. The encoder obtained a feature-based input vector of sentences using the bidirectional encoder representations from transformers (BERT) technique. A transformer-based decoder returned the summary sentence from the output as generated by the encoder. This experiment was conducted using the Livedoor news corpus with the above model. However, two problems were revealed. One is the repetition of a specific phrase while the model is generating text. The other is that the model can not handle out-of-vocabulary words. As solutions, we use repeat block in n-gram words and WordPiece. In addition, to evaluate the performance of the model, we compared the summarization accuracy between our model and a long short term memory based pointer-generator network. As revealed by the results, our model comprehends the meanings of sentences better than a pointer-generator network but makes more word-based mistakes.

1 Introduction

This paper is an extension of work originally presented in 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI) [1].

Societies with a highly developed information system make it easy for citizens to obtain the information they need. However, for this reason, the information obtained often has redundancy. Therefore, skills are needed to extract only the necessary information. In other words, a summarization skill is important. Thus, we need a system that automatically takes the necessary points of a text and applies deep learning with a neural network. At the present time, models using bidirectional encoder representations from transformers (BERT) [2] have achieved the highest scores in abstractive summarization tasks [3]. Although there are already many models using BERT in English and some other languages, a model in Japanese has yet to be found. We therefore developed models using BERT in Japanese and evaluated them through a comparison with another model. The code is available at `https://github.com/IwasakiYuuki/Bert-abstractive-text-summarization`.

Text summarization is one of the natural language processing that effectively summarizing long sentences. The Algorithms of text summarization used in machine learning are mainly divided into two types: extractive and abstractive summaries. In the extractive text summarization, a summary sentence is generated by combining important sentences in the source text. On the other hand, the latter, abstract type, understands the input sentences and generates the corresponding summary sentences by itself. Although both types are similar in that they summarize the main points of input sentences, they are more flexible in this respect because abstract abstract summarization generates the corresponding summary sentence by itself, while extractive summarization can only process sentences collected from input sentences. In this study, we focused on abstract summarization.

In recent years, various models have been proposed for abstractive summarization. An author proposed an abstractive summarization model using Bidirectional Encoder Representations from Transformers (BERT) [3]. Experimentation results as reported in [3] revealed that the developed model achieved a new state-of-the-art performance on both CNN/Daily Mail and New York Times datasets. Viswani et al. proposed an abstractive summarization model called a pointer-generator network based CopyNet [4]. This pointer-generator network model has advantages in terms of both abstractive and extractive summarizations.

The model developed in this study was built with reference to a text summarization model using BERT and has two stages. In the first stage, the input text was encoded into a contextual representation using BERT, and after processing the input text with BERT, a

[*]Corresponding Author: Yuuki Iwasaki, Tokyo College, s20607@tokyo.kosen-ac.jp

summary draft was generated using a transformer-based decoder. In the second stage, the draft summary text was re-validated using BERT to generate a crisper summary text, which was then processed with BERT. In this experiment, only the first stage was employed.

In section 2, we introduce the several well-know neural network models and layers used in this study, i.e., BERT and multi-head attention (the layers in BERT), and pointer-generator network used for comparison with our model. In section 3, we describe an overview of the proposed model along with its structure. In section 4, we detail the settings and dataset applied while training our model. In section 5, the results of the experiments described in section 4, are evaluated, along with those of another model, i.e., a pointer-generator network. Based on the evaluations, we compare our model with a pointer-generator network based on scores from the Extracted places and ROUGE-N [5]. In section 6, we provide some concluding remarks regarding our experiments and discuss areas of future study.

2 Related Studies

2.1 Pointer-Generator Networks

A pointer-generator network [6] is an LSTM-based model for an abstractive summarization. Long short term memory(LSTM) [7] is a neural network model that is often used for natural language processing. Figure 1 shows a model overview of a pointer-generator network. The most significant differences between this model and our approach are the structures, i.e., the “copy mechanism” and “coverage mechanism”. A pointer-generator network can-not compute in parallel owing to the structure of the model. In addition, it requires a long time and a large amount of data for training to obtain the meanings of both words and sentences at the same time.

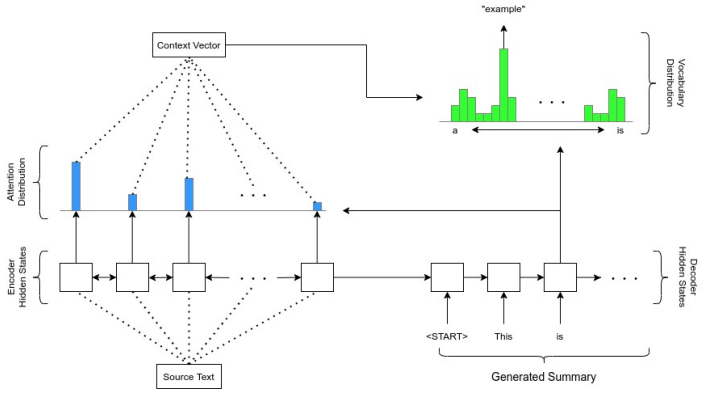


Figure 1: Overview of pointer-generator network. As shown on the left side of the figure, an ’Attention Distribution’ helps handle out-of-vocabulary words. In addition, as shown on the right side, the ’Vocabulary Distribution’ resolves the repeating problem. Referred to [6]

Copy mechanism A neural network outputs the probability of a predetermined group of words when generating them. Therefore, it cannot output proper nouns. However in sentence summaries, proper nouns are often important. Thus, a pointer-generator network dynamically allocates the words in the input text to the IDs using a copy mechanism and thus the model can handle out-of-vocabulary words.

Coverage mechanism Sentence summarization tasks occasionally have a problem of generating the same words repeatedly. A pointer-generator network incorporates a mechanism called a coverage mechanism to solve this problem. The coverage mechanism keeps the distribution of the words generated up to the t-step, and then penalizes them for producing the same word. The penalty is achieved by adding a new term to the loss function.

2.2 BERT

In recent years, pre-training models, such as BERT, have been widely incorporated into neural network models. In particular, models trained with BERT have achieved state-of-the-art performance in natural language processing tasks; BERT has been pre-trained with a large unlabeled corpus and can be fine-tuned with another corpus to achieve better performance. Figure 2 shows the process of pre-training and fine-tuning.

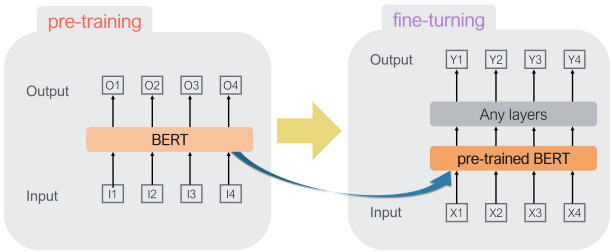


Figure 2: BERT pre-training and fine-tuning processes. A pre-training of the model is shown on the left side of this figure, and a fine-tuning is shown on the right side.

We briefly describe the structure of the BERT model with reference to [2]. BERT has several layers; each layer has a Multi-Head Attention and a linear affine with a residual connection. In our experiment, we utilize the BERT-based model which has 12-layers and 768-hidden sizes.

Pre-training In a general neural network model, the model is trained at once for the tasks. Then, the model is trained for each task at the same time as the meanings of the words and sentences are obtained. For this reason, a large amount of data is needed to train the model. However, there are two processes applied during BERT training, pre-training and fine-tuning. Pre-training tasks of BERT are called masked language model (MLM) and next sentence prediction (NSP). An MLM is a task for predicting masked words in an input text, and an NSP is a task for predicting the next sentence of an input sentence. Through a pre-training, the model obtains the meanings of the words and sentences, as well as the coherence of the texts.

Fine-turning After pre-training of the model, we re-train the model for a specified task, in this case, a summarization. A pre-trained model already has the obtained meanings of words and sentences, and thus we simply tune the model toward our task. This means it can be trained with a low data quantity. At the fine-tuning

stage, we use the pre-trained BERT model and some layers, as shown in Figure 2.

2.3 Multi-Head Attention

BERT and a transformer primarily comprise multi-head attention layers. The multi-head attention divides the attention input into multiple parts and concatenates them with multiple outputs. Multi-head attention is more accurate than attention and can be calculated through (1), (2), and (3) as described in reference [4].

$$Attention\,(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{1}$$

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{2}$$

$$MultiHead\,(Q,K,V) = Concat\,(head_i,\ldots,head_n)\cdot W^O \tag{3}$$

Equation (1) represents a formula for a single input query in multi-head attention, in which the input is split into units of a head as shown in (2), and each output is concatenated as shown in (3) to generate the overall output. In practice, we compute the attention function on a set of queries simultaneously, packed together into a matrix Q. The keys and values are also packed together into matrices K and V, and W_i^Q, and W_i^K, and W_i^V are the parameter matrices. Here, d_k used in (1) is the number of dimensions of key K.

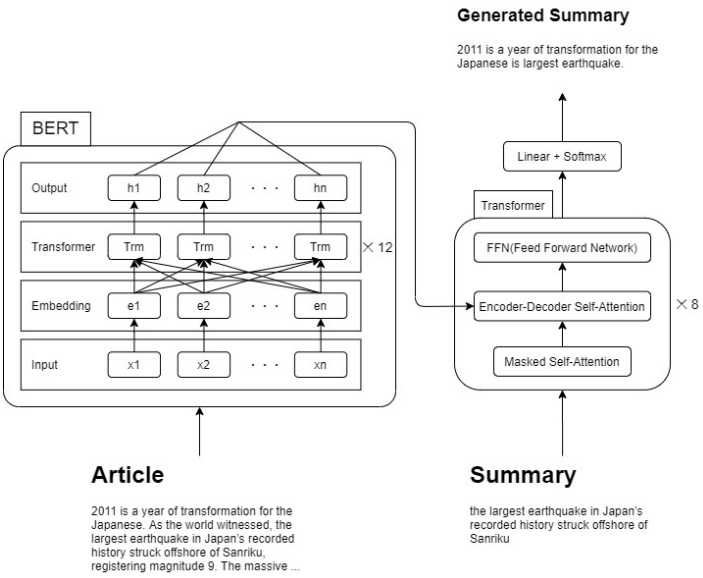


Figure 3: Overview of our text summarization model. Our model includes BERT in the encoder and a transformer in the decoder. During the training phase, we set the article body and summary into the encoder and decoder inputs, respectively.

3 Model

Figure 3 shows an overview of the model used in our experiment.

3.1 Encoder

Several pre-learning models, such as BERT, have been widely utilized in encoder-decoder models. Because BERT is efficient for fast training with high precision, it ensures a higher accuracy in existing models. In a study conducted by Zhang et al. [3], BERT is used as the encoder to achieve state-of-the-art performance for an abstract text summarization task. In this experiment we applied a pre-learning model BERT as the encoder.

3.2 Decoder

In our model, we configured a transformer-based decoder as an encoder. It was not the proposed decoder choice for BERT to send the output generated by the encoder to the input of the decoder at the same time.

Furthermore, we opted for a transformer-based decoder as opposed to a recurrent neural network (RNN) such as an LSTM [7] and GRU [8] for the following reasons: (1) Time for training - Transformer-based decoders are built using multi-head attention, which can perform parallel computations, and are, hence, faster. (2) Accuracy - Transformers are far more accurate than RNNs on machine translation tasks. (3) Long-range dependency - The attention used in the transformers makes it easier to learn long-range dependencies compared to RNNs such as an LSTM.

3.3 Abstractive summarization model

The input is denoted as $X = \{x_1, x_2, \ldots, x_n\}$, sequence representing sentence breaks is denoted as $S = \{s_1, s_2, \ldots, s_n\}$, and the corresponding summary is denoted as $A = \{a_1, a_2, \ldots, a_n\}$. We started by entering X and S into BERT.

If $f_{sen}(x)$ is assumed to be the number of sentence of x, sequence S is computed as $S = f_{sen}(X)\ mod\ 2$. The resulting BERT encoder output is denoted as H. Next, we input H and the output of the decoder at the t-th time step.

The probability of the vocabulary at the t-th time step can be obtained as shown in (4). This probability was conditioned on the decoder output until the t-th time step and the output of the encoder H.

$$P_t(w) = f_{decoder}(w \mid H, Y_{<t}) \tag{4}$$

The loss of training L, is calculated as shown in (4) using the probability of vocabulary $P_t(w)$.

$$L = -\sum_{i=0}^{n} \log P(y_t \mid H, a_{t-1}) \tag{5}$$

4 Experiments

4.1 Setting

During this experiment, we used a pre-trained model with BERT that was developed at the Kurohara and Kawahara laboratories of Kyoto College [9]. Most of the BERT hyperparameters were same as those of BERT-Base (i.e., 12-layers, 768-hidden, and 12heads) in [2]. The model was trained for 30 epochs with 1.8 billion Japanese Wikipedia corpuses. The input text was divided into sub-words with byte pair encoding (BPE [10]) using the morphological analysis system Juman++ [11].

The vocabulary size stood at 32,000 words. The decoder in our model comprises eight multi-head attention layers; the division of hidden size as 3,072; and the embedding vector as 768, similar to the encoder.

We used Adam as the model optimizer, and set the parameters as $\beta_1 = 0.9, \beta_2 = 0.999, and \epsilon = 1.0 \times 10^{-9}$. The maximum learning rate was set at 1.0×10^{-9}. In addition, the dynamic learning rate was adopted as the model learning rate in [4].

The learning rate is computed as shown in (6),

$$lr = max\_learning\_rate * \frac{min\left(cs^{-0.5}, ws^{-1.5} * cs\right)}{ws^{-0.5}} \quad (6)$$

In principle, the learning rate increases linearly up to the warmup step (ws). If the current step (cs) exceeds the warmup step, the learning rate gradually decreases. The learning rate peaks when the current step and warmup step are equal. At this point, the learning rate is denoted as $lr = max\_learning\_rate$. For our experiment, we set $warmup\_step\,(ws) = 4000$ and $max\_learning\_rate = 0.0001$ for training. The model includes BERT with 12 multi-head attention layers. The batch size is set to 4 for GPU memory because the maximum input sequence was set to 512.

Figure 4: An example of Livedoor news corpus. Every article has three sections, a title, a summary and an article body. Each section is labeled in the above figure.

4.2 Dataset

The Livedoor News corpus contains 130,000 Japanese news articles from Livedoor News, each news article is accompanied by a three-line summary. The article text and summary are set as input and output of the experiment, respectively. From the dataset, 100,000 data points were used for training and 30,000 data points were used for validation. The maximum length of the input sequence was set to 512 tokens, but some data values exceeded this limit. In such cases, only the first 512 tokens were entered into the model.

Every article on the Livedoor news website has text and a three-line summary. Figure 4 shows an example of an article body and summary. In this example, the article body at the bottom of the figure is input into the model, and the summary at the middle of the figure is output to the model.

4.3 Training

If 100,000 live news corpus data points in 1 epoch were trained in 15 epochs, the loss of each lexicon in the validation data was about 6.5 words, and the accuracy of identifying the correct lexicon was about 67%. The graphics card used for training was a Titan X (Pascal) with 12 GB of memory. The training took about three days.

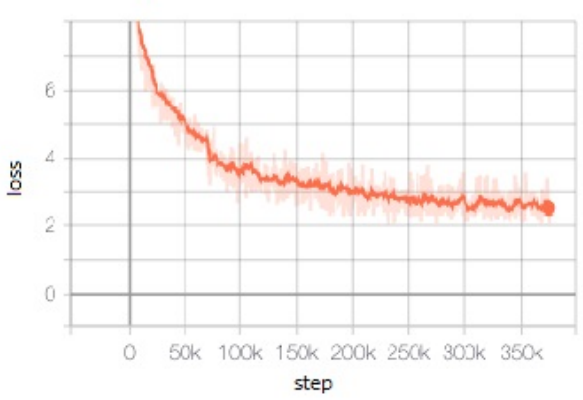


Figure 5: Outputs of loss function for every training step. A gradual decrease in the numbers of loss function outputs is implied in this figure.

5 Results and Discussions

In this section, we can see the evaluations of the generated summaries with two models, i.e., our model (BERT+Transformer) and a pointer-generator network. From the generated summaries, we evaluated the models on the following criteria:

- Grammatical Accuracy
- Vocabulary mistake
- importance level

To check the validity of the generated sentences, we focused on grammatical and vocabulary errors. We also extracted and checked which parts of the original text were important for the generation of the summary text. The above three points are evaluated subjectively. In addition, we calculated and compared their respective ROUGE-N scores.

5.1 Generated summaries

The texts in the appendix are the article bodies and their summaries that generated by the two trained models, our model and the pointer-generator network (translated from Japanese into English). The texts under "Input" are article bodies in the Livedoor news corpus, and the texts under "Output" are summaries generated by our model and the

pointer-generator network from each article body. In the appendix, we provide three article bodies and generated their summaries as examples. When generating the text, a beam search with a width of 4 was used. Prior to the training process, WordPiece [12] was used to further divide out-of-vocabulary words into multiple words. For example, the word "smoke" which is not in the vocabulary of the model, is able to represent the combination of "smo" and "ke" in the vocabulary. For the quantitative effect of WordPiece, refer to an ablation study [13].

One problem that arises during the summary sentence generation is the "repetition problem". This means that an arbitrary phrase will repeatedly appear in the summary sentence. We used a REPEAT block to get around this. The repeat block detects the repetition of a sentence in n-grams increments and returns to the (n-1)-steps before the repetition and reapplies the word generation. In this experiment, we used the repeat block of the tri-gram that had the highest ROUGE-N score.

For example, you can see the *t*-th step generated summaries by the model as follows:

summary : It is going to rain in **the(1) the(2) the(3)**

In the above example, the same three words are generated from the end of the word, and thus the approach returns to the two steps prior to "the(1)" and regenerates words other than the those previously generated.

summary : It is going to rain in **the(1) evening**

Input 1 in the appendix demonstrates that both models were able to learn correctly as, it can be seen, to some extent, that the summary text retained the key points of the input text. In addition, there are no errors in grammar or vocabulary. However, if we look at the summary of a pointer-generator network, the sentence "The Berlin International Film Festival Silver Bear Award is one of the three biggest film festivals in the world for "smoke". " is not extremely important. Therefore, in the above example, we thought the summary generated by our model was better than that of the pointer-generator network owing to the importance.

Input 2 in the appendix is regarding the risks to the human body from sitting for long periods of time. At first glance, both summaries appear to be correct, but the word "skillfully" is included in the summary of our model. This is not the word that should be used. Therefore, we felt that the pointer-generator network was able to produce a more accurate summary in this example than our approach.

Input 3 in the appendix deals with the response of a real estate company when a house is an accidental property. No erroneous words appear in either summary. However, the sentence "It is said that the building should be disliked because it is a place to live continuously." in the summary of our model, is not grammatically correct. In terms of content, however, we believe that our model is able to extract the more important parts. For example, the content regarding payment is fairly important but is not included in the summary of the pointer-generator network.

In the present example, our model extracts the more important parts, but the grammar was correct for the pointer-generator network. From the three examples provided in the appendix, it appears that the pointer-generator network has fewer vocabulary and grammatical errors, although our model is able to extract more of the important content. In addition, you can see that the locations that the pointer-generator network extract is mostly in the first half of the input text. By contrast, our model extracts from all throughout the input text. In other words, our model is more expressive, whereas the pointer-generator network is less prone to grammatical and vocabulary errors. In addition, because the pointer-generator network has a copy mechanism, we thought it might be related to the fact that the model is somewhere between abstract and extractive text.

5.2 *Extracted places*

We will now see which parts of the input text are extracted by the summary statement generated in the previous section. Figure 6, 7 and 8 shows the places that the model extracts from the input texts. Each horizontal axis represents the number of words, and the bar graph shows the word positions referenced by the models.

Figures 6, 7 and 8, indicate that the locations extracted by the pointer-generator network are mostly in the first half of the input texts. By contrast, those extracted by our model(BERT+Trasnformer) are scattered throughout the input text. This leads us to believe that our model may be able to understand the entire content of the input text better than the pointer-generator network.

Figure 6: Input 1 highlighting the extracted locations

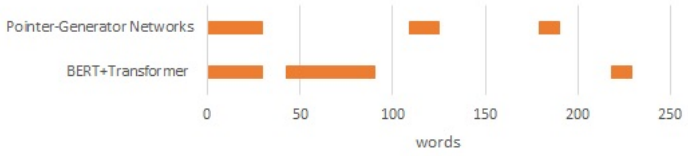


Figure 7: Input 2 highlighting the extracted locations

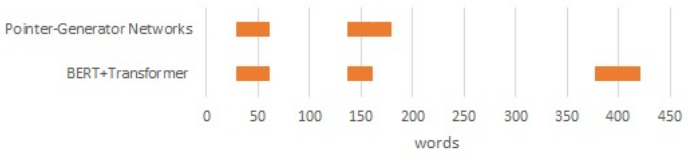


Figure 8: Input 3 highlighting the extracted locations

5.3 ROUGE-N socre

Next, we compare the pointer-generator network with our model using the ROUGE-N [5], which is an index frequently used in the evaluation of summary sentences, which represents the percentage of agreement between words in N-gram units.

As an example, let us try to calculate the ROUGE-N score of the following sentence.

summary : It is going to rain in the evening today
references : It is going to rain today

Referring to [5], the equation for obtaining the ROUGE-N score is shown below.

$$ROUGE_N = \frac{\sum_{C \in Reference} \sum_{Gram_n \in C} Count_{match}(Gram_n)}{\sum_{C \in Reference} \sum_{Gram_n \in C} Count(Gram_n)} \quad (7)$$

Considering the above example, we have the same six pairs on uni-gram, and four pairs on a bi-gram. From this, the ROUGE-1,2 score of the example could be derived as follows:

$$\begin{aligned} ROUGE_1 &= \frac{6}{9} \\ &= 0.667 \\ ROUGE_2 &= \frac{4}{8} \\ &\fallingdotseq 0.500 \end{aligned}$$

In our experiment, we calculated the ROUGE-1,2 scores with 1,000 Japanese news articles (sample size $n = 1,000$) chosen at random from the verification datasets. The standard error, *SE*, calculated using the unbiased standard deviation, u, and the sample size, n, is expressed as follows:

$$SE = \frac{u}{\sqrt{n}} \quad (8)$$

Using the above equation, the margin of error, e, for the 95% confidence interval (95% CI) is expressed as in Equation 9.

$$e = \pm 1.96 \times SE \quad (9)$$

The ROUGE-1,2 scores, the standard errors, *SE*, and the unbiased standard deviations, u, for both the models are presented in Table 1.

Table 1: ROUGE-1, 2 scores for pointer-generator networks and our model (BERT + transformer)

Model		Score	u	e
Pointer-Generator Networks	ROUGE-1	0.463	0.123	±0.008
	ROUGE-2	0.221	0.142	±0.009
BERT + Transformer	ROUGE-1	0.470	0.107	±0.007
	ROUGE-2	0.215	0.126	±0.008

Table 1 indicates that the ROUGE-1,2 scores of our model and the pointer-generator network were similar. However, our model's structure was designed for multiple tasks, whereas the pointer-generator network was a model designed for text summarization tasks because of the copy mechanism. In addition, our model generated all the words in the output sentences. Thus, we hypothesized that our model achieved a better comprehension of the text.

5.4 Word Mover's Distance

Next, we compared the pointer-generator network with our model using word mover's distance (WMD) [14] to evaluate their semantic similarity. The WMD is a novel distance function between text documents. The ROUGE scores are based on literal word overlap; hence, it is necessary to evaluate a score that is not based on literal word overlap. Among a number of recently proposed semantic similarity metrics, WMD is shown to be the most reasonable solution to measure semantic similarity in reformulated texts [15]. We evaluated the WMD for both models with 1,000 Japanese news articles chosen at random from the verification datasets and used the Japanese word embedding vectors [16] to calculate the WMD. Table 2 presents the WMD scores, unbiased standard deviation, and margin of error for the pointer-generator networks and our model.

Table 2: WMD of pointer-generator networks and our model (BERT + transformer)

Model	WMD	u	e
Pointer-Generator Networks	0.533	0.122	±0.008
BERT + Transformer	**0.521**	0.106	±0.007

As shown in Table 2, our model has a slightly lower WMD score than the pointer-generator network. This implies that the summaries generated by our model were more similar to the article's summaries than those generated by the pointer-generator network. Thus, the hypothecate that our model achieved a better comprehension of the text than the pointer-generator network was strengthened.

6 Conclusion and Future works

We conducted an experiment to demonstrate an abstractive summarization of Japanese text with a neural network model using BERT. In addition, we conducted a comparison between the qualitative and quantitative aspects of our model and a model frequently used for text summarization, i.e., a pointer-generator network. Our model is composed of a BERT encoder and a transformer-based decoder. The dataset used in this paper was the Livedoor news corpus consisting of 130,000 datapoints, of which 100,000 were used for training.

The results of the experiment revealed that the model was able to learn correctly as the summary sentence captured the key points of the text to a certain extent. However, the contents of the summary sentence were repeated, and the model could not handle unknown words. As a solution, we applied two mechanisms, a repeat block and WordPiece. The repeat block detects repeated n-grams in the generated summaries and regenerates summaries at the (t-n-1)-th step excluding the repeat word. To handle this, WordPiece further divides the out-of-vocabulary words into sub-words.

The results of evaluations of our model and the pointer-generator network revealed that their ROUGE-1,2 scores were similar. However, our model's structure was designed for multiple tasks, whereas the pointer-generator network was a model designed for text summarization tasks because of the copy mechanism. Therefore, we hypothesized that our model achieved a better comprehension of the text. In addition, the comparison of the two models to evaluate their semantic similarity, that is, the WMD, strengthened our hypothecate. While extracting phrases, the pointer-generator network extracted phrases from the first half of the input text, whereas our model extracted phrases from throughout the text. However, a qualitative evaluation of our model revealed that it made numerous grammatical and vocabulary mistakes. We believe that this problem can be solved by improving the model.

In the future, we will explore these recommendations through additional experiments and compare the results.

7 Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 19K12906.

References

[1] Y. Iwasaki, A. Yamashita, Y. Konno, K. Matsubayashi, "Japanese abstractive text summarization using BERT," in 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), 1–5, 2019, doi:10.1109/TAAI48200.2019.8959920.

[2] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018, doi:10.18653/v1/N19-1423.

[3] H. Zhang, J. Cai, J. Xu, J. Wang, "Pretraining-Based Natural Language Generation for Text Summarization," 789–797, 2019, doi:10.18653/v1/K19-1074.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, u. Kaiser, I. Polosukhin, "Attention is All You Need," in Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, 6000–6010, Curran Associates Inc., Red Hook, NY, USA, 2017, doi:10.5555/3295222.3295349.

[5] C.-Y. Lin, E. Hovy, "Automatic Evaluation of Summaries Using N-Gram Co-Occurrence Statistics," in Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1, NAACL '03, 71–78, Association for Computational Linguistics, USA, 2003, doi:10.3115/1073445.1073465.

[6] A. See, P. J. Liu, C. D. Manning, "Get To The Point: Summarization with Pointer-Generator Networks," in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 1073–1083, Association for Computational Linguistics, Vancouver, Canada, 2017, doi:10.18653/v1/P17-1099.

[7] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory," Neural Comput., **9**(8), 1735–1780, 1997, doi:10.1162/neco.1997.9.8.1735.

[8] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014, doi:10.3115/v1/D14-1179.

[9] S. Kurohashi, Y. Murawaki, "BERT pretrained model with Japanese corpus," http://nlp.ist.i.kyoto-u.ac.jp/index.php?BERT%E6%97%A5%E6%9C%AC%E8%AA%9EPretrained%E3%83%A2%E3%83%87%E3%83%AB (accessed 2019-07-16).

[10] R. Sennrich, B. Haddow, A. Birch, "Neural Machine Translation of Rare Words with Subword Units," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 1715–1725, Association for Computational Linguistics, Berlin, Germany, 2016, doi:10.18653/v1/P16-1162.

[11] T. Kudo, K. Yamamoto, Y. Matsumoto, "Applying Conditional Random Fields to Japanese Morphologiaical Analysis," IPSJ-NL, **161**, 89–96, 2004.

[12] M. Schuster, K. Nakajima, "Japanese and korean voice search," in 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5149–5152, IEEE, 2012, doi:10.1109/ICASSP.2012.6289079.

[13] K. Bostrom, G. Durrett, "Byte Pair Encoding is Suboptimal for Language Model Pretraining," ArXiv, 2020, doi:abs/2004.03720.

[14] M. Kusner, Y. Sun, N. Kolkin, K. Weinberger, "From word embeddings to document distances," in International conference on machine learning, 957–966, 2015, doi:10.5555/3045118.3045221.

[15] I. Yamshchikov, V. Shibaev, N. Khlebnikov, A. Tikhonov, "Style-transfer and Paraphrase: Looking for a Sensible Semantic Similarity Metric," arXiv, 2020, doi:abs/2004.05001.

[16] M. Suzuki, "Japanese Wikipedia entity vector," http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/ (accessed 2020-10-30).

Appendix

The following are three articles from Livedoor news and a summary of our model and the pointer-generator network generated as examples. The examples below were used for the discussions provided in section 5.

Input 1 Director Wayne Wang from Hong Kong, who has won the Silver Bear Award at the Berlin International Film Festival, one of the three biggest film festivals in the world in the movie "Smoke" (1995), and Japan's world-famous Beat Takeshi and Nishijima Hidetoshi. The movie "When a Woman Sleeps" (published on February 27). It was the first time in 12 years since "Blood and Bone" starring in a film other than his own work in 2004, and said "I was so worried about my acting that I held my head," he said unexpectedly. Is it because of the mysterious style that the boundaries between dreams and delusions and reality are vague? Three people talked to each other. This film is a movie based on a short story by Javier Marias, a Spanish writer, and is set in a quiet resort hotel, and a mysterious couple (Takeshi and Shiori Nana Shiori), two years older than their parents and children. It is a mystery of love and hate that stares at a writer (Nishijima) who is obsessed with a relationship that has become obsessed with and has run out of curiosity. "I thought that the main character was Nishijima," said the smiling Takeshi, and read the script, "I thought I was foolish. A normal script would tell the story and the ending." He said that he had a hard time understanding a complicated story. Nishijima, who praises himself for being more acquainted with movie expressions, relieved, "If I say "good", I'm sure there will be no mistake." I watched all the acting all the time. When I finished watching, I didn't even remember what kind of movie it was," said an unknown side of Takeshi called "Kitano of the World." Nishijima, who immediately decided to appear, "Wayne Wang is a work that shoots Beat Takeshi, so it will come out (laughs)," said "This time especially Beat Takeshi is obsessed with love, even at the shooting site. I was shocked a lot and was really moved. It was wonderful." Director Wang, who worked on Japanese movies for the first time, also said, "I wanted to work with Takeshi rather than want to shoot in Japan." Takeshi played an unusual role in keeping a video recording of a young beautiful woman (Kana) sleeping. When asked, "What do

you want to leave your dear ones?", "I've been in the entertainment world for a long time, so I thought, "I thought that the gag of Takeshi at that time, that laugh was amazing." It's okay if you can leave one," he said. (Interview and text: Megumi Shibata)

Output 1 (Pointer-generator network) The Berlin International Film Festival Silver Bear Award is one of the three biggest film festivals in the world for "smoke." For the first time in 12 years, Takeshi said, "I was so worried about my performance that I held my head." "The boundaries between dreams and delusions and reality are vague and mysterious.

Output 1 (BERT+Transformer) A movie "When a Woman Sleeps", in which Takeshi Beat and Hidetoshi Nishijima formed a tag. Beat Takeshi was said to have been attached to love. "I'm worried about my performance, and I'm worried about the finished product, too," he said.

Input 2 "Mentai Wide-Fukuoka Broadcasting" At the corner of "Special News THE Slide Show" broadcasted on February 9, 2016, we covered the risks to your body from sitting for a long time. The image is an image (taken by Andrea Arbogast, from Flickr). A study of 220,000 men and women over the age of 45 in Australia found that the risk of death was 1.4 times higher for people sitting 11 hours or more a day than for those sitting less than 4 hours a day I got the result. It seems that the risk of obesity, diabetes, cancer, etc. increased, causing death. The detailed cause is unknown, but it is also said that "myokine" that suppresses heart disease is not secreted unless the muscles are contracted. This excessive sitting puts a heavy burden on the lower back, with 1.4 times being normally seated and 1.85 times being leaned forward. Even in Fukuoka, more and more offices are incorporating "sometimes standing work". More than 90% of the desks sold in Sweden and Denmark are of the type that can be moved up and down, and by using this you can do "occasionally standing work". Even in Japan, Rakuten has changed the desks of all employees so that they can be moved up and down, to avoid sitting down. Even in Fukuoka, some companies are making their desks movable up and down and changing the height of the desks themselves. When I asked when to change the height of the desk, this was the most common opinion. Because I may feel drowsy after lunch, switch to standing work at such times. When you are tired from working on a computer or lose your concentration, start a standing job. Stand up when your shoulders, waist, or neck hurts. It seemed that it was good to stand up, and there were interesting opinions. I have less stiff shoulders, lower back pain, and neck pain. Increased communication from the surroundings. When I sat down, I was staring at my computer, but by standing up, I was able to see my surroundings better and start talking. Since I also work as a deskwork worker, I sometimes feel pain in my lower back and shoulders after sitting for a long time. For this reason, I stand up and relax once every hour, but I found it useful to have a desk like this. If you keep sitting down, you may not be able to see the surroundings because you are staring at a desk or a personal computer, so I would like you to incorporate such desks in various offices. (Writer: Piyoko).

Output 2 (Pointer-generator network) At the "Special News THE Slide Show" on the 9th, I heard about the risks to the body from sitting for a long time. It is said that "myokine" that suppresses heart disease is not secreted unless the muscles are contracted. It is said that more and more offices in Fukuoka have adopted "sometimes standing work".

Output 2 (BERT+Transformer) In the program on the 9th, we discussed the risks to the body from sitting for a long time. The risk of obesity, diabetes, cancer, etc. is high, and the risk of death is 14 times higher. In addition, he changed the man's desk to something that can be skillfully moved up and down to avoid sitting.

Input 3 Last year, an incident that involved the illegal construction of a large-scale condominium construction project in Yokohama City reminded me of the difficulty of choosing a property, but I finally got the feeling that I could get a house. If it's a so-called "accident property", can you ask the real estate agent for a refund? Lawyer Masami Takeshita answers. (Consultation). At the end of last year, I bought a second-hand house, but when I looked up the reason because the neighborhood was strange, the previous owner was at home It seems that he tried to hang himself. The real estate company that introduced this property did not tell me that it was an accident property. In this case, is it possible to request the real estate agent to refund the purchase price in full? (Answer). The fact that a building purchased for residential use commits suicide and lacks "comfortable living" means that "a building is a place to live continuously, so there are also psychological flaws due to the historical background of the building that should be disliked. (Defect)" is understood. If the buyer makes a contract without knowing that there is a defect, the seller will be liable for the defect guarantee, and if the defect does not meet the contract purpose, the contract will be canceled, and if it does not reach that level, You are obliged to compensate the buyer for damages. In addition, there are court cases that the contract can be canceled if there is suicide and it is judged that it is reasonable to feel that it is not suitable for living because it lacks comfort to live in. .. It is the case that the cancellation of the sale of condominiums, which was found to have committed suicide six years ago when the deposit was paid, was approved. On the other hand, there are some cases in which the suicide was canceled 7 years ago, and there are also cases in the sale and purchase of a plan to demolish an old house in which the degree of dislike of suicide two years ago due to the removal of a building was not a defect. In your case, depending on the time of suicide and the degree of topicality, you may be able to cancel the contract and request a full refund of the price, and even if you do not go that far, the objective evaluation of the suicide building and the purchase price You can consider the difference claim for damages. The important things that the Homebuilding Act requires an intermediary to explain are not the history of the building, such as suicide, but if the trader knows suicide, which is an important factor in determining the buyer's contract decision, There is a statutory obligation to explain and if you remain silent, you will be liable. However, it is questionable to impose accountability forever, and there seems to be a guidance that sales are 5 years and rent is 3 years. In any case, when you buy a used building, you should ask about its provenance, etc., and you can pursue the default liability

of breach of accountability to the lying seller. (Attorney Profile). Masami Takeshita: Born in Osaka in 1946. Graduated from the Faculty of Law, the University of Tokyo. Registered as a lawyer in 1971. * Weekly post February 26, 2016 issue.

Output 3 (Pointer-generator network) Attorney Masami Takeshita explained why a real estate company can request a refund if the property is an ”accident.” It seems that the building purchased for residential use may lack ”comfortable living” due to suicide. ”There are also psychological flaws due to the disgusting historical background of buildings.

Output 3 (BERT+Transformer) A lawyer explains whether you can request a refund when you purchase the accident property in 2017. It is said that the building should be disliked because it is a place to live continuously. Regarding loans, it seems that there is a possibility that you can cancel the contract and request full refund.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1683-1689 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Sentiment Analysis in English Texts

Arwa Alshamsi[1], Reem Bayari[1], Said Salloum[2,3,*]

[1]*Faculty of Engineering & IT, The British University, Dubai, 345015, UAE*

[2]*Research Institute of Sciences & Engineering, University of Sharjah, Sharjah, 27272, UAE*

[3]*School of Science, Engineering, and Environment, University of Salford, Manchester, M5 4WT, UK*

ARTICLE INFO

Article history:
Received: 23 September, 2020
Accepted: 24 December, 2020
Online: 28 December, 2020

Keywords:
Sentiment Analysis
Balanced Dataset
Unbalanced Dataset
Classification

ABSTRACT

The growing popularity of social media sites has generated a massive amount of data that attracted researchers, decision-makers, and companies to investigate people's opinions and thoughts in various fields. Sentiment analysis is considered an emerging topic recently. Decision-makers, companies, and service providers as well-considered sentiment analysis as a valuable tool for improvement. This research paper aims to obtain a dataset of tweets and apply different machine learning algorithms to analyze and classify texts. This research paper explored text classification accuracy while using different classifiers for classifying balanced and unbalanced datasets. It was found that the performance of different classifiers varied depending on the size of the dataset. The results also revealed that the Naive Byes and ID3 gave a better accuracy level than other classifiers, and the performance was better with the balanced datasets. The different classifiers (K-NN, Decision Tree, Random Forest, and Random Tree) gave a better performance with the unbalanced datasets.

1. Introduction

The recent widening expansion of social media has changed communication, sharing, and obtaining information [1–4]. In addition to this, many companies use social media to evaluate their business performance by analysing the conversations' contents [5]. This includes collecting customers' opinions about services, facilities, and products. Exploring this data plays a vital role in consumer retention by improving the quality of services [6,7]. Social media sites such as Instagram, Facebook, and Twitter offer valuable data that can be used by business owners not only to track and analyse customers' opinions about their businesses but also that of their competitors [8–11]. Moreover, these valuable data attracted decision-makers who seek to improve the services provided [8,9,12].

In this research paper, several research papers that studied Twitter's data classification and analysis for different purposes were surveyed to investigate the methodologies and approaches utilized for text classification. The authors of this research paper aim to obtain open-source datasets then conduct text classification experiments using machine learning approaches by applying different classification algorithms, i.e., classifiers. The authors utilized several classifiers to classify texts of two versions of datasets. The first version is unbalanced datasets, and the second is balanced datasets. The authors then compared the classification accuracy for each used classifier on classifying texts of both datasets.

2. Literature Review

As social media websites have attracted millions of users, these websites store a massive number of texts generated by users of these websites [13–20]. Researchers were interested in investigating these metadata for search purposes [16,17,21–24]. In this section, a number of research papers that explored the analysis and classification of Twitter metadata were surveyed to investigate different text classification approaches [25] and the text classification results.

Researchers of [26] investigated the user's gender of Twitter. Authors noticed that many Twitter users use the URL section of the profile to point to their blogs, and the blogs provided valuable demographic information about the users. Using this method, the authors created a corpus of about 184000 Twitter users labeled with their gender. Then authors arranged the dataset for

*Corresponding Author: Said Salloum, University of Sharjah, UAE. Tel: +971507679647 Email: ssalloum@sharjah.ac.ae

experiments as following: for each user; they specify four fields; the first field contains the text of the tweets and the remaining three fields from the user's profile on Twitter, i.e., full name, screen name, and description. After that, the authors conducted the experiments and found that using all of the dataset fields while classifying Twitter user's gender provides the best accuracy of 92%. Using tweets text only for classifying Twitter user's gender provides an accuracy of 76%. In [27], the authors used Machine Learning approaches for Sentiment Analysis. Authors constructed a dataset consisting of more than 151000 Arabic tweets labeled as "75,774 positive tweets and 75,774 negative tweets". Several Machine Learning Algorithms were applied, such as Naive Bayes (NB), AdaBoost, Support vector machine (SVM), ME, and Round Robin (RR). The authors found that RR provided the most accurate results on classifying texts, while AdaBoost classifier results were the least accurate results. A study by [28] interested as well in Sentiment Analysis of Arabic texts. The authors constructed the Arabic Sentiment Tweets Dataset ASTD, which consists of 84,000 Arabic tweets. The number of tweets remaining after annotation was around 10,000 tweets. The authors applied machine learning approaches using classifiers on the collected dataset. They reported the following: (1) The best classifier applied on the dataset is SVM, (2) Classifying a balanced set is challenging compared to the unbalanced set. The balanced set has fewer tweets than the unbalanced set, which may negatively affect the classification's reliability. In [29], the author investigated the effects of applying preprocessing methods before the sentiment classification of the text. The authors used classifiers and five datasets to evaluate the preprocessing method's effects on the classification. Experiments were conducted, and researchers reported the following findings: Removing URL has no much effect, Removing stop words have a slight effect, Removing Numbers have no effect, Expanding Acronym improved the classification performance, and the same preprocessing methods have the same effects on the classifier's performance, NB and RF classifiers showed more sensitivity than LR and SVM classifiers. In conclusion, the classifier's performance for sentiment analysis was improved after applying preprocessing methods. A study by [30] investigated Twitter geotagged data to construct a national database of people's health behavior. The authors compared indicators generated by machine learning algorithms to indicators generated by a human. The authors collected around 80 million geotagged tweets. Then Spatial Join procedures were applied, and 99.8% of tweets were successfully linked. Then tweets were processed. After that, machine learning approaches were used and successfully applied in classifying tweets into happy and not happy with high accuracy. In [31] explored classifying sentiments in movie reviews. The authors constructed a dataset of 21,000 tweets of movie reviews. Dataset split into train set and test set. Preprocessing methods applied, then two classifiers, i.e., NB and SVM, were used to classify tweets text into positive or negative sentiment. The authors found that better accuracy achieved using SVM of 75% while NB has 65% accuracy. Researchers of [32] used Machine Learning methods and Semantic Analysis for analyzing tweet's sentiments. Authors labeled tweets in a dataset that consists of 19340 sentences into positive or negative. They applied preprocessing methods after that features were extracted; authors applied Machine Learning approaches, i.e., Naïve Bayes, Maximum Entropy, and Support Vector Machine (SVM) classifiers after that Semantic Analysis were applied. The authors found that Naïve Bayes provided the best accuracy of 88.2, the next SVM of 85.5, and the last is Maximum entropy of 83.8. The authors reported as well that after applying Semantic Analysis, the accuracy increased to reach 89.9. In [33], the authors analyzed sentiments by utilizing games. Authors introduced TSentiment, which is a web-based game. TSentiment used for emotion identification in Italian tweets. TSentiment is an online game in which the users compete to classify tweets in the dataset consists of 59,446 tweets. Users first must evaluate the tweet's polarity, i.e., positive, negative, and neutral, then users have to select the tweet's sentiment from a pre-defined list of 9 sentiments in which 3 sentiments identified for the positive polarity, 3 sentiments identified for negative polarity. Neutral polarity is used for tweets that have no sentiment expressions. This approach for classifying tweets was effective.

A study by [34] examined the possibility of enhancing the accuracy of predictions of stock market indicators using Twitter data sentiment analysis. The authors used a lexicon-based approach to determine eight specific emotions in over 755 million tweets. The authors applied algorithms to predict DJIA and S&P500 indicators using Support Vectors Machine (SVM) and Neural Networks (NN). Using the SVM algorithm in DJIA indication, the best average precision rate of 64.10 percent was achieved. The authors indicated that the accuracy could be increased by increasing the straining period and by improving the algorithms for sentiment analysis. authors conclude that adding Twitter details does not improve accuracy significantly. In [35], the authors applied sentiment analysis on around 4,432 tweets to collect opinions on Oman tourism, they build a domain-specific ontology for Oman tourism using Concept Net. Researchers constructed a sentiment lexicon based on three existing lexicons, SentiStrength, SentWordNet, and Opinion lexicon. The authors randomly divide 80% of the data for the training set and 20% for testing. The researcher used two types of semantic sentiment, Contextual Semantic Sentiment Analysis, and Conceptual Semantic Sentiment Analysis. Authors applied Naïve Base supervised machine learning classifier and found that using conceptual semantic sentiment analysis expressively improves the sentiment analysis's performance. A study by [36] used sentiment analysis and subjectivity analysis methods to analyze French tweets and predict the French CAC40 stock market. The author used a French dataset that consists of 1000 positive and negative book reviews. The author trained the neural network by using three input features on 3/4 of the data, and he tested on the remaining quarter. The achieved accuracy 80% and a mean absolute percentage error (MAPE) of 2.97%, which is less than the work reported by Johan Bollen. The author suggested adding more features as input to improve the performance. In [37], the authors examined the relationship between Twitter's social emotion and the stock market. Researchers collected millions of tweets by Twitter API. Researchers retrieved the NASDAQ market closing price in the same period. The authors applied the correlation coefficient. Authors conclude that emotion-related terms have some degree of influence on the stock market overall trend, but it did not meet standards that can be used as a guide to stock-market prediction. While at the same time, there was a fairly close association between positive, negative, and angry mood-words. Particularly sad language tends to have a far greater influence on the stock market than other groups. In [38], the authors investigated telecommunications companies' conversation on

social media Twitter ('indihome,' in Indonesia). The authors collected 10,839 raw data for segmentation. The authors collected data: over 5 periods of time in the same year. Authors found that most of the tweets (7,253) do not contain customers' perception toward Indihome. Only 3,586 tweets are containing the perception of customers toward Indihome. Most of the data contained perception reveal that the customers have the negative perception (3,119) on Indihome and only 467 tweets contain positive perceptions; the biggest number of negative perceptions relate to the first product, the second relates to a process, third relate to people, and fourth relate to pricing. Researchers of [39] examined prevalence and geographic variations for opinion polarities about e-cigarettes on Twitter. Researchers collected data from Twitter by pre-defined seven keywords. They classified the tweets into four categories: Irrelevant to e-cigarettes, Commercial tweets, organic tweets with attitudes (supporting or against or neutral) the use of e-cigarettes, and the geographic locations information city and state. Researchers selected six socio-economic variables from Census data 2014 that are associated with smoking and health disparities. Researchers classified the tweets based on a combination of human judgment and machine-learning algorithms, and two coders classified a random sample of 2000 tweets into five categories. The researcher applied a multilabel Naïve Bayes classifier algorithm; the model achieved an accuracy of 93.6% on the training data. Then the researcher applied the machine learning algorithm to a full set of collected tweets and found the accuracy of the validation data was 83.4%. To evaluate the socio-economic impact related to public perception regarding e-cigarette use in the USA, researchers calculated the Pearson correlation between prevalence and percentage of opinion polarities and selected ACS variants for 50 states and the District of Columbia. In [40], the authors Investigated the link between any updates on certain brands and their reaction. Researchers gathered geographic locations based on the data to see consumer distribution. Researchers collected Twitter data by using the REST API. In total, 3,200, from ten different profiles, then used sentiment analysis to differentiate between clustered data expressed positively or negatively then resampled the result in an object model and cluster. For every answer, the researcher has been evaluated for the textual sentiment analysis from the object model. Researchers used AFINN based word list and Sentiments of Emojis to run comprehensive sentiment analysis; for the data that not existed in the word list, researcher added a separated layer to an analysis by using emoji analysis on top of sentiment analysis, and authors did not see any difference in the level of accuracy when applying this extra layer. The researcher found some Sentiment Analysis weaknesses related to the misuse of emoji, the use of abbreviated words or terms of slang, and the use of sarcasm. In [41], the authors proposed an application that can classify a Twitter content into spam or legitimate. Auhtors used an integrated approach, from URL analysis, Natural Language Processing, and Machine Learning techniques. Auhtors analyzed the URL that derived from the tweets, then convert URLs to their long-form, then compare URLs with Blacklisted URLs, then compare them with a set pre-defined expressions list as spam; the presence of any of these expressions can conclude that the URL is spam. After cleaning data, the stemmed keywords are compared with the per set of identified spam words and, if a pre-defined expressions list are found in the tweet, then the user is classified as spam. Six features were used for classification. The training set has 100 instances with six features and a label. The author used Naïve-Bayes algorithm. Authors manually examined 100 users and found (60 were legitimate and 40 were spam) then the sampled checked by the application and the result presented that 98 were classified correctly.

3. Proposed Approach

In this work, the authors implemented and evaluated different classifiers in classifying the sentiment of the tweets. It's by utilizing RapidMiner software. Classifiers were applied on both balanced and unbalanced datasets. Classifiers used are Decision Tree, Naïve Bayes, Random Forest, K-NN, ID3, and Random Tree.

4. Experiment Setup

In this section, the dataset is described as well as the settings and evaluation techniques are used in the experiments have been discussed. The prediction for the tweet category is tested twice—the first time on an unbalanced data set and the second time on a balanced dataset as below.

- **Experiments on the unbalanced dataset:** Decision Tree, Naïve Bayes, Random Forest, K-NN, ID3, and Random Tree classifiers were applied on six unbalanced datasets.
- **Experiments on the balanced dataset:** In this experiment, the challenges related to unbalanced datasets were tackled by manual procedures to avoid biased predictions and misleading accuracy. The majority class in each dataset almost equalized with the minority classes, i.e., many positive, negative, and neutral, practically the same in the balanced dataset as represented in Table 3.

4.1. Dataset Description

We obtained a dataset from Kaggle, one of the largest online data science communities in this work. It consists of more than 14000 tweets, labeled either (positive, negative, or neutral). The dataset was also split into six datasets; each dataset includes tweets about one of six American airline companies (United, Delta, Southwest, Virgin America, US Airways, and American). Firstly, we summarized the details about the obtained datasets, as illustrated in Table 1 below.

Table 1: Summary of obtained Dataset

	American Airline Companies					
	Virgin America	United	Delta	Southwest	US Airways	American
Number of Tweets	504	3822	2222	2420	2913	2759
Positive Tweets	152	492	544	570	269	336
Negative Tweets	181	2633	955	1186	2263	1960
Neutral Tweets	171	697	723	664	381	463

4.2. Dataset Cleansing

In this section, the authors described the followed procedure in the dataset preparation. The authors utilized RapidMinor software for tweet classification. Authors followed the methods described below:

1) Splitting the dataset into a training set and test set.

2) Loading the dataset, i.e., excel file into RapidMinor software using Read Excel operator.

3) Applying preprocessing by utilizing the below operators.

- Transform Cases operator to transform text to lowercase.
- Tokenize operator to split the text into a sequence of tokens.
- Filter Stop words operator to remove stop words such as: is, the, at, etc.
- Filter Tokens (by length) operator: to remove token based on the length, in this model, minimum characters are 3, and maximum characters are 20 any other tokens that don't match the rule will be removed.
- Stem operator: to convert words into base form.

4.3. Dataset Training

Each of the datasets was divided into two-part. The first part contains 66% of the total number of tweets of the data set, and it is used to train the machine to classify the data under one attribute, which is used to classify the tweets to either (positive or Negative or Neutral). The remaining 34% of tweets were used to classify tweets' attribute to (positive or Negative or Neutral), i.e., test set.

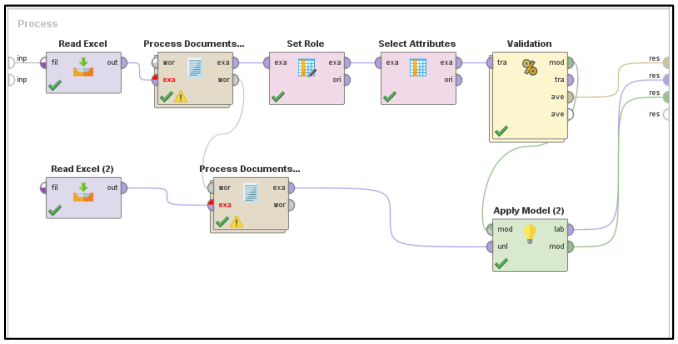


Figure 1: Summarization of the Process Model

4.4. Dataset Classifying

In this section, the authors described the steps in the tweet's classification techniques.

- Set Role operator is used to allow the system to identify sentiment as the target variable,
- Select Attributes operator is used to removing any attribute which has any missing values.
- Then in the validation operator, the dataset is divided into two parts (training and test). We used Two-thirds of the dataset to train the dataset and the last one-third to evaluate the model.
- Different machine learning algorithms are used for training the dataset (Decision Tree, Naïve Bayes, Random Forest, K-NN, ID3, and Random Tree).
- For testing the model, the Performance operator utilized to measure the performance of the model.

5. Experiment Results and Discussion

This section presented the experiment results in terms of accuracy level of prediction for each classifier on both types of datasets (balanced, unbalanced) and a comparison between the two experiments.

5.1. Experiment results for an unbalanced dataset

Figure 2 and Table 2 present the accuracy results of the utilized classifiers on the datasets.

Table 2: Accuracy results on unbalanced dataset

	Accuracy					
	Virgin America	United	Delta	Southwest	US Airways	American
Dataset	504	3822	2222	2420	2913	2759
Training set	333	2523	1467	1597	1923	1821
Test set	171	1299	755	823	990	938
Decision Tree	31.86%	72.03%	42.08%	50.46%	82.72%	68.98%
Naïve Bayes	32.74%	72.38%	42.28%	51.01%	82.72%	72.21%
Random Forest	31.86%	72.03%	42.08%	50.46%	82.72%	68.98%
K-NN	39.82%	11.66%	35.27%	50.46%	82.72%	69.43%
ID3	32.74%	72.38%	42.28%	51.01%	82.72%	72.21%
Random Tree	31.86%	72.03%	42.08%	50.46%	82.72%	68.98%

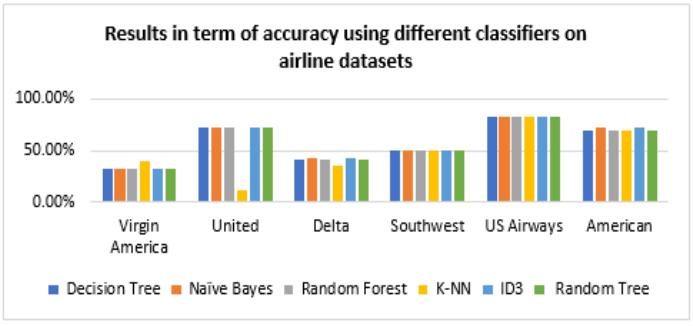


Figure 2: Accuracy results on unbalanced airline datasets using different classifiers

In some datasets, the classifier's accuracy results were very high, while it was low in others. All classifier's performance on the US airways dataset and United dataset provided the best accuracy due to the dataset's size, which was the largest. Naïve Bayes classifier, Decision Tree, and ID3 were mostly better than other classifiers and were given almost the same accuracy level. The classifiers with Virgin America dataset reported the lowest accuracy level due to the dataset's size, which is very small.

5.2. Experiment results for a balanced dataset

Decision Tree, Naïve Bayes, Random Forest, K-NN, ID3, and Random Tree classifiers were applied on the five obtained balanced datasets. (United, Delta, Southwest, and US Airways). The dataset for each was divided into two parts. The first part contains 66% of the total number of tweets of the data set, and it is used to train the machine to classify the data under one attribute, which is used to classify the tweets as either positive, Negative, or Neutral. The remaining 34% of tweets were used to classify tweets' attributes into (positive, Negative, or Neutral), i.e., test set.

Table 3: Number of tweets before and after balancing.

	Number of instances		Percentage		
	Total tweets before balancing	Total tweets after balancing	Positive	Negative	Neutral
United	3822	8276	33%	33%	34%
Delta	2222	2635	33%	33%	34%
Southwest	2420	5518	33%	33%	33%
US Airways	2913	6608	33%	33%	33%
American	2759	5924	34%	34%	33%

After applying different algorithms on the five balanced datasets, the performance, i.e., accuracy results, were reported in Table 4 and Figure 3 below:

Table 4: Accuracy results on the balanced dataset

	Accuracy					
	Virgin America	United	Delta	Southwest	US Airways	American
Dataset	8276	2635	5518	6608	5924	8276
Training set	5464	1743	3642	4363	3911	5464
Test set	2812	892	1876	2245	2013	2812
Decision Tree	35.06%	34.63%	34.35%	35.06%	33.98%	35.06%
Naïve Bayes	97.65%	36.99%	65.48%	97.65%	61.20%	97.65%
Random Forest	35.06%	34.63%	34.35%	35.06%	33.98%	35.06%
K-NN	38.79%	32.77%	35.32%	38.79%	39.47%	38.79%
ID3	97.65%	36.99%	65.48%	97.65%	61.20%	97.65%
Random Tree	35.06%	34.63%	34.35%	35.06%	33.98%	35.06%

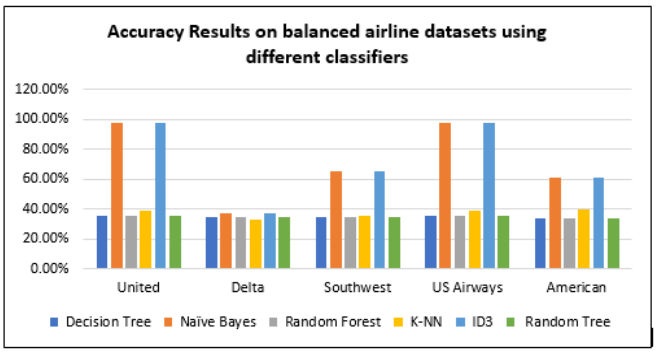


Figure 3: Accuracy results on balanced airline datasets using different classifiers

5.3. Comparison between two experiments results for each classifier

While comparing results between the performance of the classifiers on balanced and unbalanced datasets, it was found the following as seen in Figure 4 below:

5.3.1. Naive Byes and ID3

Gave the best accuracy than other classifiers in the two experiments. The accuracy level with the balanced datasets higher than unbalanced ones. In the unbalanced datasets, the maximum accuracy for both classifiers was 82.7%. In the balanced dataset, the accuracy reached 97.6%; these results confirm that these two classifiers are the best compared to the other selected classifiers in the two experiments:

5.3.2. K-NN and Decision Tree

Show better performance with the unbalanced datasets, and the difference is so apparent. The maximum accuracy with the balanced datasets is 39.4%, while it reached 82.7 % with the unbalanced datasets.

5.3.3. Random forest and Random Tree

It shows better performance with the unbalanced datasets, and the difference is so apparent. The maximum accuracy with the balanced datasets around 35%, while it reached 82.7% with the unbalanced datasets.

In conclusion, Naive Bayes and ID3 gave a better accuracy level than other classifiers, and the performance was better with the balanced datasets. The different classifiers (K-NN, Decision Tree, Random Forest, and Random Tree) gave a better understanding of the unbalanced datasets.

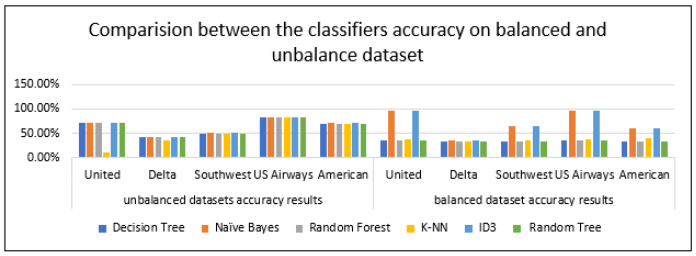


Figure 4: Accuracy results of classifiers on balanced and unbalanced datasets

6. Conclusions

Social media websites are gaining very big popularity among people of different ages. Platforms such as Twitter, Facebook, Instagram, and Snapchat allowed people to express their ideas, opinions, comments, and thoughts. Therefore, a huge amount of data is generated daily, and the written text is one of the most common forms of the generated data. Business owners, decision-makers, and researchers are increasingly attracted by the valuable and massive amounts of data generated and stored on social media websites. Sentiment Analysis is a Natural Language Processing field that increasingly attracted researchers, government authorities, business owners, services providers, and companies to improve products, services, and research. In this research paper, the authors aimed to survey sentiment analysis approaches. Therefore, 16 research papers that studied Twitter's text

classification and analysis were surveyed. The authors also aimed to evaluate different machine learning algorithms used to classify sentiment to either positive or negative, or neutral. This experiment aims to compare the efficiency and performance of different classifiers that have been used in the sixteen papers that are surveyed. These classifiers are (Decision Tree, Naïve Bayes, Random Forest, K-NN, ID3, and Random Tree). Besides, the authors investigated the balanced dataset factor by applying the same classifiers twice on the dataset, one on the unbalanced and the other, after balancing the dataset. The targeted dataset included six datasets about six American airline companies (United, Delta, Southwest, Virgin America, US Airways, and American); it consists of about 14000 tweets. The authors reported that the classifier's accuracy results were very high in some datasets while low in others. The authors indicated that the dataset size was the reason for that. On the balanced dataset, the Naïve Bayes classifier, Decision Tree, and ID3 were mostly better than other classifiers and have given the almost same level of accuracy. The classifiers with Virgin America dataset reported the lowest level of accuracy due to its small size. On the unbalanced dataset, results show that the Naive Byes and ID3 gave a better level of accuracy than other classifiers when it's applied on the balanced datasets. While (K-NN, Decision Tree, Random Forest, and Random Tree) gave a better understanding of the unbalanced datasets.

Conflict of Interest

The authors declare no conflict of interest

Acknowledgment

This is a part of project done in British University in Dubai.

References

[1] S.A. Salloum, C. Mhamdi, B. Al Kurdi, K. Shaalan, "Factors affecting the Adoption and Meaningful Use of Social Media: A Structural Equation Modeling Approach," International Journal of Information Technology and Language Studies, **2**(3), 2018.

[2] M. Alghizzawi, S.A. Salloum, M. Habes, "The role of social media in tourism marketing in Jordan," International Journal of Information Technology and Language Studies, **2**(3), 2018.

[3] S.A. Salloum, W. Maqableh, C. Mhamdi, B. Al Kurdi, K. Shaalan, "Studying the Social Media Adoption by university students in the United Arab Emirates," International Journal of Information Technology and Language Studies, **2**(3), 2018.

[4] S.A. Salloum, M. Al-Emran, S. Abdallah, K. Shaalan, Analyzing the arab gulf newspapers using text mining techniques, 2018, doi:10.1007/978-3-319-64861-3_37.

[5] F.A. Almazrouei, M. Alshurideh, B. Al Kurdi, S.A. Salloum, Social Media Impact on Business: A Systematic Review, 2021, doi:10.1007/978-3-030-58669-0_62.

[6] Alshurideh et al., "Understanding the Quality Determinants that Influence the Intention to Use the Mobile Learning Platforms: A Practical Study," International Journal of Interactive Mobile Technologies (IJIM), **13**(11), 157–183, 2019.

[7] S.A. Salloum, K. Shaalan, Adoption of E-Book for University Students, 2019, doi:10.1007/978-3-319-99010-1_44.

[8] S.A. Salloum, M. Alshurideh, A. Elnagar, K. Shaalan, "Mining in Educational Data: Review and Future Directions," in Joint European-US Workshop on Applications of Invariance in Computer Vision, Springer: 92–102, 2020.

[9] S.A. Salloum, M. Alshurideh, A. Elnagar, K. Shaalan, "Machine Learning and Deep Learning Techniques for Cybersecurity: A Review," in Joint European-US Workshop on Applications of Invariance in Computer Vision, Springer: 50–57, 2020.

[10] S.A. Salloum, R. Khan, K. Shaalan, "A Survey of Semantic Analysis Approaches," in Joint European-US Workshop on Applications of Invariance in Computer Vision, Springer: 61–70, 2020.

[11] K.M. Alomari, A.Q. AlHamad, S. Salloum, "Prediction of the Digital Game Rating Systems based on the ESRB."

[12] S.A. Salloum, A.Q. AlHamad, M. Al-Emran, K. Shaalan, A survey of Arabic text mining, Springer, Cham: 417–431, 2018, doi:10.1007/978-3-319-67056-0_20.

[13] C. Mhamdi, M. Al-Emran, S.A. Salloum, Text mining and analytics: A case study from news channels posts on Facebook, 2018, doi:10.1007/978-3-319-67056-0_19.

[14] A.S. Alnaser, M. Habes, M. Alghizzawi, S. Ali, "The Relation among Marketing ads, via Digital Media and mitigate (COVID-19) pandemic in Jordan The Relationship between Social Media and Academic Performance: Facebook Perspective View project Healthcare challenges during COVID-19 pandemic View project," Dspace.Urbe.University, (July), 2020.

[15] M. Alshurideh, B. Al Kurdi, S. Salloum, "Examining the Main Mobile Learning System Drivers' Effects: A Mix Empirical Examination of Both the Expectation-Confirmation Model (ECM) and the Technology Acceptance Model (TAM)," in International Conference on Advanced Intelligent Systems and Informatics, Springer: 406–417, 2019.

[16] M. Alghizzawi, M. Habes, S.A. Salloum, M.A. Ghani, C. Mhamdi, K. Shaalan, "The effect of social media usage on students'e-learning acceptance in higher education: A case study from the United Arab Emirates," International Journal of Information Technology and Language Studies, **3**(3), 2019.

[17] M. Habes, S.A. Salloum, M. Alghizzawi, C. Mhamdi, "The Relation Between Social Media and Students' Academic Performance in Jordan: YouTube Perspective," in International Conference on Advanced Intelligent Systems and Informatics, Springer: 382–392, 2019.

[18] M. Habes, S.A. Salloum, M. Alghizzawi, M.S. Alshibly, "The role of modern media technology in improving collaborative learning of students in Jordanian universities," International Journal of Information Technology and Language Studies, **2**(3), 2018.

[19] B.A. Kurdi, M. Alshurideh, S.A. Salloum, Z.M. Obeidat, R.M. Al-dweeri, "An empirical investigation into examination of factors influencing university students' behavior towards elearning acceptance using SEM approach," International Journal of Interactive Mobile Technologies, **14**(2), 2020, doi:10.3991/ijim.v14i02.11115.

[20] S.A. Salloum, M. Al-Emran, M. Habes, M. Alghizzawi, M.A. Ghani, K. Shaalan, "Understanding the Impact of Social Media Practices on E-Learning Systems Acceptance," in International Conference on Advanced Intelligent Systems and Informatics, Springer: 360–369, 2019.

[21] M. Alghizzawi, M.A. Ghani, A.P.M. Som, M.F. Ahmad, A. Amin, N.A. Bakar, S.A. Salloum, M. Habes, "The Impact of Smartphone Adoption on Marketing Therapeutic Tourist Sites in Jordan," International Journal of Engineering & Technology, **7**(4.34), 91–96, 2018.

[22] S.F.S. Alhashmi, S.A. Salloum, S. Abdallah, "Critical Success Factors for Implementing Artificial Intelligence (AI) Projects in Dubai Government United Arab Emirates (UAE) Health Sector: Applying the Extended Technology Acceptance Model (TAM)," in International Conference on Advanced Intelligent Systems and Informatics, Springer: 393–405, 2019.

[23] M. Alghizzawi, M. Habes, S.A. Salloum, The Relationship Between Digital Media and Marketing Medical Tourism Destinations in Jordan: Facebook Perspective, 2020, doi:10.1007/978-3-030-31129-2_40.

[24] R.S. Al-Maroof, S.A. Salloum, A.Q.M. AlHamadand, K. Shaalan, A Unified Model for the Use and Acceptance of Stickers in Social Media Messaging, 2020, doi:10.1007/978-3-030-31129-2_34.

[25] K.S.A. Wahdan, S. Hantoobi, S.A. Salloum, K. Shaalan, "A systematic review of text classification research based ondeep learning models in Arabic language," Int. J. Electr. Comput. Eng, **10**(6), 6629–6643, 2020.

[26] J.D. Burger, J. Henderson, G. Kim, G. Zarrella, "Discriminating gender on Twitter," in Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 1301–1309, 2011.

[27] D. Gamal, M. Alfonse, E.-S.M. El-Horbaty, A.-B.M. Salem, "Twitter Benchmark Dataset for Arabic Sentiment Analysis," International Journal of Modern Education and Computer Science, **11**(1), 33, 2019.

[28] M. Nabil, M. Aly, A. Atiya, "Astd: Arabic sentiment tweets dataset," in Proceedings of the 2015 conference on empirical methods in natural language processing, 2515–2519, 2015.

[29] Z.J. and G. Xiaolini, "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis," Digital Object Identifier, 2017, doi:10.1109/ACCESS. 2017. 2672677.

[30] Q.C. Nguyen, D. Li, H.-W. Meng, S. Kath, E. Nsoesie, F. Li, M. Wen, "Building a national neighborhood dataset from geotagged Twitter data for indicators of happiness, diet, and physical activity," JMIR Public Health and Surveillance, **2**(2), e158, 2016.

[31] A. Amolik, N. Jivane, M. Bhandari, M. Venkatesan, "Twitter sentiment analysis of movie reviews using machine learning techniques," International Journal of Engineering and Technology, **7**(6), 1–7, 2016.

[32] G. Gautam, D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," in 2014 Seventh International Conference on Contemporary Computing (IC3), IEEE: 437–442, 2014.

[33] M. Furini, M. Montangero, "TSentiment: On gamifying Twitter sentiment analysis," in 2016 IEEE Symposium on Computers and Communication (ISCC), IEEE: 91–96, 2016.

[34] A. Porshnev, I. Redkin, A. Shevchenko, "Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis .," 2013 IEEE 13th International Conference on Data Mining Workshops, 440–444, 2013, doi:10.1109/ICDMW.2013.111.

[35] V. Ramanathan, "Twitter Text Mining for Sentiment Analysis on People ' s Feedback about Oman Tourism," 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), 1–5, 2019.

[36] V. Martin, "Predicting the french stock market using social media analysis," Proceedings - 8th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2013, 3–7, 2013, doi:10.1109/SMAP.2013.22.

[37] Q. Li, B. Zhou, Q. Liu, "Can twitter posts predict stock behavior?: A study of stock market with twitter social emotion," Proceedings of 2016 IEEE International Conference on Cloud Computing and Big Data Analysis, ICCCBDA 2016, 359–364, 2016, doi:10.1109/ICCCBDA.2016.7529584.

[38] Indrawati, A. Alamsyah, "Social network data analytics for market segmentation in Indonesian telecommunications industry," 2017 5th International Conference on Information and Communication Technology, ICoIC7 2017, **0**(c), 2017, doi:10.1109/ICoICT.2017.8074677.

[39] H. Dai, J. Hao, "Mining social media data for opinion polarities about electronic cigarettes," Tobacco Control, **26**(2), 175–180, 2017, doi:10.1136/tobaccocontrol-2015-052818.

[40] A. Husnain, S.M.U. Din, G. Hussain, Y. Ghayor, "Estimating market trends by clustering social media reviews," Proceedings - 2017 13th International Conference on Emerging Technologies, ICET2017, **2018**–**Janua**, 1–6, 2018, doi:10.1109/ICET.2017.8281716.

[41] K. Kandasamy, P. Koroth, "An integrated approach to spam classification on Twitter using URL analysis, natural language processing and machine learning techniques," 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2014, 1–5, 2014, doi:10.1109/SCEECS.2014.6804508.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1690-1697 (2020)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Analysis and Evaluation of Competitiveness in Medical Tourism Industry in Taiwan

Yen-Hung Chen*, Tin-Chang Chang

Department of Business Administration, Asia University, 500, Lioufeng Rd., Wufeng, Taichung, 41354, Taiwan

ARTICLE INFO

Article history:
Received: 09 November, 2020
Accepted: 20 December, 2020
Online: 28 December, 2020

Keywords:
Medical Tourism
Hospital Management
VIKOR
Entropy
SWOT
COVID-19

ABSTRACT

Tourism is relatively high profit, low-cost and high work opportunity industry. Medical tourism is special kind of business type in tourism industry. The characteristic of medical tourism is high threshold and high profit for hospital and government. But, fewer literatures research about competitiveness of medical tourism industry until now. Especially, medical tourism industry faces big challenge when COVID-19 occurs which stop patient to go abroad for treating in the world. This research arranges relative literatures and invite some expert's opinion to design criteria for evaluating competitiveness of medical tourism industry in each area. After that, linguistic VIKOR and Entropy will be integrated to analyze and evaluate performance of medical tourism industry among China, Taiwan, Japan and South Korea. Finally, some reasonable suggestion will be provided for Taiwan government and hospital.

1. Introduction

Compare with other area in Asia, medical service in Taiwan is relatively good whose price–performance ratio is high in the world. Based on health care index from CEOWORLD magazine, the rank of Taiwan is in top levels among eighty regions in the world (Please refer to Figure 1). This magazine investigates and believe that Taiwan's health care standard ranks first in medical equipment, medical personnel technology and medical expenses fields. Based on this information, Taiwan government can consider to apply this strength of Taiwan medical service to develop economics.

Business volume of tourism relative industry such as accommodation industry, transportation industry, catering industry and amusement industry will increase when government develop tourism industry [2]. Besides, a huge amount of work opportunity can be generated by tourism industry. Based on above reason, medical tourism become the main development business activity in Taiwan. However, medical tourism industry faces big challenge when COVID-19 occurs which stop tourist to travel in the world and foreign patient can't go to Taiwan for treatment [3]. It needs a novel criteria and framework to evaluate competitiveness of medical tourism industry for satisfying the sustainable development goal. The goal (research objectives) of this research is to develop a framework to evaluate and understand competitiveness of medical tourism industry under the sustainability development condition in Taiwan.

The content of this research is organized as follows. Some literatures about medical tourism evaluation will be arranged and discussed in chapter 2. After that, some operation content about linguistic variable and proposed method will be explained in chapter 3. The competitiveness of medical tourism industry among China, Taiwan, Japan and South Korea will be analyzed and evaluated based on proposed method and SWOT analysis about Taiwan medical tourism industry will be executed in chapter 4. Some management implementation will be token over in chapter 5. Finally, conclusion and future research will be discussed as ending.

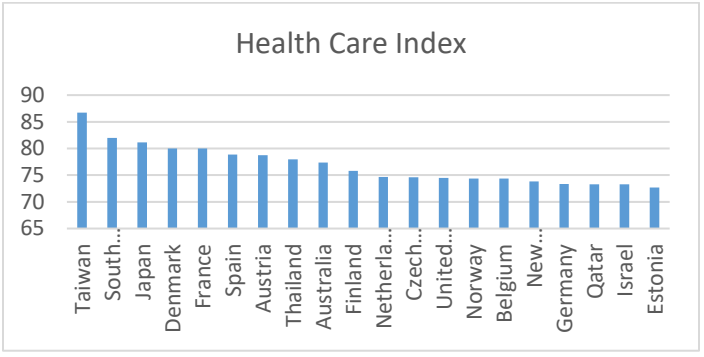


Figure 1: Rank of medical service in the world [1]

*Corresponding Author: Yen-Hung Chen, chenyenhung1234@gmail.com

2. Literature Review

In 2005, the business volume of global medical market and healthcare market reached US$200-300 billion. In the same time point, the tourist volume of the global medical tourism industry market exceeds 19 million person per year and business volume of medical tourism industry approaches US$20 billion [4]. Enterprise in medical tourism industry should manage and evaluate its performance. So, there are various literatures research about medical tourism evaluation has been executed under this development trend. Below literatures are relative research about medical tourism evaluation.

In [5], the author collected relative literatures and surveyed relative experts' opinions to build 33 evaluation criteria. And then, fuzzy VIKOR were applied to evaluate service quality of public medical tourism hospital. In [6], the author developed a comprehensive framework to assess the factors affecting health tourism development in Yazd province in China. Their method integrated interpretive structural modeling (ISM) with fuzzy TOPSIS to evaluate medical tourism competitive ability. Research result showed that patient satisfaction is the most important factors for developing medical tourism. In [7], the author tried to validate the measurement model for medical tourism research. In their research, the research dimension about human capital, physical infrastructure and patient satisfaction had been validated by confirmatory factor analysis (CFA). Future scholar can use their questionnaire to test performance of medical tourism industry. In [8], the author used strengths, weaknesses, opportunities and threats (SWOT) analysis to evaluate business status of medical tourism industry in Turkey. After that, analytical hierarchy process (AHP) were applied to analyze and rank the importance of each factor which will influence the development of tourism industry in Turkey.

In [9], the author took Indian medical tourism hospital as research target. This scholar applied SWOT analysis to evaluate the development status of each organization. And then, technique for order preference by similarity to ideal solution (TOPSIS) were combined with AHP to rank the performance of each hospital. In [10], author considered that medical tourism industry is normal business phenomenon because some country face long waiting queues, high treatment costs or lack of insurance in some country. They applied SWOT model and best-worst techniques to develop and analyze medical tourism industry development strategy in Yazd province of Iran. In [11], the author collected relative information of different city in India and used rough number to integrate experts' opinions. Analytic hierarchy process was applied to evaluate weigh of each medical tourism criteria and multi‐attributive border approximation area comparison methods were used to rank performance of medical tourism site. According to analysis result, we can know that "quality of infrastructure of healthcare institutions", "supply of skilled human resources" and "new job creations" were three most important factor to select medical tourism destination. In [12], the author collected relative information to build medical tourism evaluation criteria and investigated fuzzy TOPSIS with decision making trial and evaluation laboratory (DEMATEL) to analyze the key factor for developing medical tourism industry in Malaysia. According to their research result, human and technological factors are most important key factors for medical tourism industry. In [13], author used SERVQUAL methodology to evaluate service quality of hospital and designed twenty-eight criteria to build hospital service quality. Fuzzy AHP and Elimination and Choice Translating Reality (ELECTRE-I) were employed to evaluate and rank performance of each hospitals. According to experiment result, trustworthiness and empathy were the most important dimension for patient to select hospital.

According to above information, we can know that a lot of literatures discuss about medical tourism in each dimension. They used each kind of multi criteria decision making technology to handle each kind medical tourism evaluation problem (such as importance factor analysis, medical service evaluation, hospital performance evaluation and hospital destination decision etc). However, past literatures investigative about medical tourism evaluation problem who usually lack to consider the dimension about sustainable development of medical tourism industry when some special event will stop development of medical tourism such as COVID-19. This study will fill this research gap.

3. Preliminary and Entropy-VIKOR Method

In this study, we will introduce the research tool – SWOT analysis, linguistic variable, entropy method and Entropy-VIKOR Method.

3.1. SWOT analysis

Strengths, Weaknesses, Opportunities and Threats Analysis (SWOT Analysis) are the most useful analysis tool. This tool has been applied in each kind of fields such as hospitality industry evaluation [14], tourism development strategy [15], medical tourism assessment [8] etc. SWOT is the comprehensive analysis tool which think over relative conditions of relevant competitors and external environmental conditions of research target. This method allows research target to analyze, examine and investigate external opportunities and potential threats. After that, research target will understand and recognize their own advantages and disadvantages for analyzing its competitiveness (Refer to Table 1).

Table 1: SWOT analysis concept

		Impact on Research Target	
		Good for Research Target	Harmful to Research Target
Sources of Influencing Factors	Within the Research Target	Strengths	Weaknesses
	External Environment	Opportunities	Threats

3.2. Linguistic Variable

Linguistic variable is useful tool. This tool can let expert to express their opinion under the same evaluation scale and integrate experts' opinion easily [16]. Below content are the definition, notation and operation about linguistic variable.

Definition 1. The η scale linguistic term set S is the set of linguistic variables. S can be described by following equation [17] and [18].

$$S=\{s_0^{\eta}, s_1^{\eta}, \dots, s_{\eta-1}^{\eta}\} \quad (1)$$

where $s_{\text{ɿ}}^{\eta}$ ($\text{ɿ} = 0,1,\dots,\eta-1$) is η scale linguistic variable and η is the total volume of linguistic variable in S.

Definition 2. △ is the linguistic transfer function. This function can transfer linguistic variable into crisp value n (n is between 0 and 1). This function can be described by following equation [17] and [18].

$$\Delta(s_z^{\eta}) = \frac{z}{\eta-1}=n \quad (2)$$

Definition 3. $\triangle^{-1}$ is the linguistic transfer inverse function. This function can transfer crisp value n (n is between 0 and 1) into linguistic variable. This function can be described by following equation [17] and [18].

$$\triangle^{-1}(n)=s_{n*(\eta-1)}^{\eta} \quad (3)$$

3.3. Entropy Method

Entropy Method is one kind of weight evaluation technology. Under this method, importance of each criterion will be decided according to information volume in each criterion. So, multi criteria decision making (MCDM) technology integrated with entropy method can distinguish the performance of each alternative easily [19]. Below content are notation and operation about entropy method.

Definition 4. Suppose that φ_q is entropy value for criterion q. If information in criterion q is low, then φ_q is low. Entropy function should be described by following equation [20] and [21].

$$\varphi_q=-1*\frac{1}{lnM}\sum_{i=1}^{M} f_{iq} ln(f_{iq}) \quad (4)$$

where $f_{gq}=\frac{x_{gq}}{\sum_{i=1}^{M} x_{iq}}$.

Definition 5. Weight (importance) of each criterion should be decided by entropy value. The weight of criterion q will be acquired by following equation [20] and [21].

$$w_q=\frac{1-\varphi_q}{\sum_{z=1}^{N}(1-\varphi_z)} \quad (5)$$

where N is the total volume of criteria.

3.3. Entropy-VIKOR Method

In this research, Entropy-VIKOR Method are selected to cope with medical tourism competitiveness evaluation problem. In Entropy-VIKOR Method, some notation will be applied and can be described as Table 2.

Table 2: Notation in Entropy-VIKOR Method

Set	Notation	Detail Description
Research target	$T=\{T_1, T_2, \dots, T_m\}$	m represents volume of research target
Criteria	$C=\{C_1, C_2, \dots, C_n\}$	n represents volume of criteria
Weight criteria	$W=\{w_1, w_2, \dots, w_n\}$	n represents volume of criteria
Decision matrix	$X = \begin{array}{c} \\ T_1 \\ T_2 \\ \dots \\ T_m \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \dots & C_n \end{array} \\ \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \dots & \tilde{x}_{1n} \\ \tilde{x}_{21} & \tilde{x}_{22} & \dots & \tilde{x}_{2n} \\ \dots & \dots & \dots & \dots \\ \tilde{x}_{m1} & \tilde{x}_{m2} & \dots & \tilde{x}_{mn} \end{bmatrix} \end{array}$	$\tilde{x}_{ij}$ represents performance of research target T_i with respect to C_j.
Expert	$E=\{E_1, E_2, \dots, E_K\}$	K represents volume of experts

At first, research targets should be decided by organization. Organization should organize the evaluation project team. After that, some experts will be invited into this team. After experts discuss with each other, criteria for evaluating competitiveness of medical tourism industry can be arranged and selected. And then, experts' opinion will be gathered for evaluating performance of each research target respect to each criterion. At last, Entropy-VIKOR method will be used to evaluate competitiveness of medical tourism industry among each research target in Asia. SWOT analysis will be executed for detailly evaluating competitiveness of medical tourism industry among each research target. Research result will be provided to organization for reference. Relative execution process can refer to Figure 2.

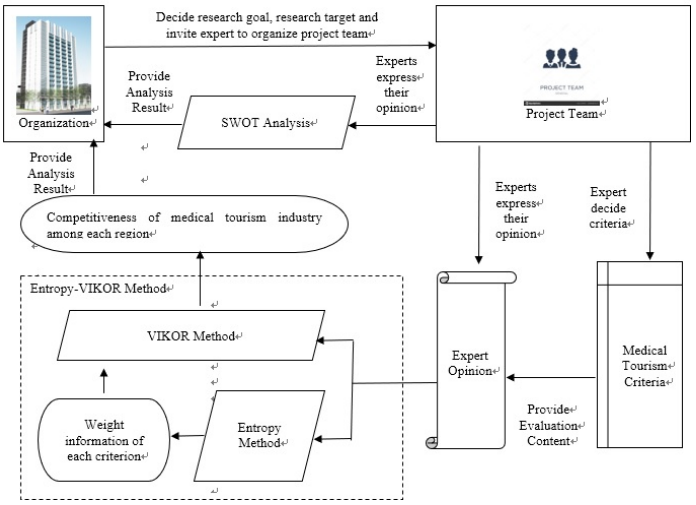


Figure 2: Execution process of proposed method

Let $\tilde{x}_{ijk}$ be expert k's opinion for performance of region i with respect to criterion j. Experts' opinion will be integrated by following equation.

$$\tilde{x}_{ij}=\triangle^{-1}\left(\sum_{k=1}^{K}\frac{\Delta(\tilde{x}_{ijk})}{K}\right) \quad (6)$$

In this research, weight of criteria will be calculated and acquired by entropy method. Weight evaluation function can be describing according to following equation.

$$w_q=\frac{1-\varphi_q}{\sum_{z=1}^{n}(1-\varphi_z)} \quad (7)$$

where $L_{iq}=\frac{\Delta(\tilde{x}_{iq})}{\sum_{i=1}^{m}\Delta(\tilde{x}_{iq})}$, $\varphi_q=-1*\frac{1}{\ln(m)}\sum_{i=1}^{m} L_{iq} ln(L_{iq})$.

The advantage of using entropy method to acquire the weight of each criteria is that entropy method can decide the weight of each criteria automatically for reducing the evaluation burden of experts.

Below content is relative execution step in VIKOR method.

Positive ideal solution (PIS) is the importance index in VIKOR method. PIS represents the ideal performance of virtual best research target (alternative). PIS can be calculated according to following equation.

$$f_j^*=max_i \triangle^{-1}(\tilde{x}_{ij}) \quad (8)$$

Negative ideal solution (NIS) is another importance index in VIKOR. NIS represents the poorest performance of virtual worst research target (alternative). NIS can be calculated according to following equation.

$$f_j^-=min_i \triangle^{-1}(\tilde{x}_{ij}) \quad (9)$$

Total performance of research target (the largest group utility) S_i represents the overall performance of research target T_i. S_i can be calculated according to following equation.

$$S_i=\sum_{j=1}^{n} w_j * \frac{f_j^*-\triangle^{-1}(\tilde{x}_{ij})}{f_j^*-f_j^-} \quad (10)$$

The individual regret value of research target (the smallest individual regret) R_i represents the maximum regret degree of decision maker if he/she select research target T_i. R_i can be calculated according to following equation.

$$R_i=max_j\left[w_j * \frac{f_j^*-\triangle^{-1}(\tilde{x}_{ij})}{f_j^*-f_j^-}\right] \quad (11)$$

Total performance of research target S_i and individual regret value of research target R_i respectively means different consideration for making decision. So, expert can rank the performance of research target according to largest group utility S_i and smallest individual regret R_i simultaneously. Based above view point, it generates the compromise result Q_i. Q_i should be calculated for ranking performance of each research target based on following equation.

$$Q_i=v*\frac{S_i-S^*}{S^--S^*}+(1-v)*\frac{R_i-R^*}{R^--R^*} \quad (12)$$

where v represents decision parameter. This parameter is used to make decision for adjusting the importance between S_i and R_i. $S^*=min_i S_i$, $S^-=max_i S_i$, $R^*=min_i R_i$, $R^-=max_i R_i$.
Research target T_i will be better if Q_i is fewer.

4. Competitiveness Analysis of Medical Tourism Industry Among Asia Region

Competitiveness of medical tourism industry in Taiwan will be evaluated by an organization. After discussing and considering, four regions have been chosen as research target to evaluate their competitiveness. Those research targets are South Korea (T_1), Japan (T_2), Taiwan (T_3), China (T_4). Above research target are regions in Asia which are main medical tourism competitor for Taiwan.

The execution process of proposed method is as follows:
Step 1. Organize evaluation project team

In the beginning, Taiwan government organize evaluation project team and invite nine experts to execute this evaluation project. Experts in this team include government officer, employees in traditional hotel, employees in travel agency, nurse leader, doctor (hospital dean) and university professors. The investigation target covered already include tourism industry, medical industry, government and university. So, respondents in this survey are representative (Refer to Table 3).

Table 3: Interviewee related information

no	attribute	Content
1	Industry	Hospital dean
2	Industry	Nurse leader
3	Industry	High level employees in travel agency
4	Industry	High level employees in hotel
5	Government	Staff in the Ministry of Economy
6	Government	Staff in Tourism Bureau
7	Government	Staff in the Ministry of Health and Welfare
8	Academia	Professor in department of business management
9	Academia	Professor in department of nursing management

Step 2. Execute evaluation criteria decision

After collecting literatures and experts discuss with each other, eight criteria are employed to evaluate performance of research target. The criteria include medical industry image (C_1), medical ability (C_2), telemedicine ability (C_3), medical cost (C_4), financial status in hospital and travel agency (C_5), volume of medical tourism traveler per year (C_6), medical service speed in hospital (C_7) and maximum service volume in region (C_8). Above criteria are suitable to evaluate competitiveness of medical tourism industry because of below reason.

(a) medical industry image

Patient's medical industry image in this region will influence his/her intention to go to the region [22] and [23], foreign hospital does not have opportunity to service this patient if he/her do not go to this region.

(b) medical ability

Some patients want to receive high quality medical service, so medical ability of medical industry in this region is the main consideration when patients select medical tourism destination region [24] and [25].

(c) telemedicine ability

Hospitals need to treat their foreign patients and maintain relationship with those patients. However, COVID-19 causes that medical tourism patient cannot go abroad to receive face-to -face medical service. So, telemedicine ability is more and more important for region to develop medical tourism industry in order to maintain relationship with foreign patients and continuously treat those patients.

(d) medical cost

In some region, medical cost is very expensive for citizen [26] and [27]. So, patient has the intention to cure his/her disease in foreign region. Medical cost is the main consideration for choosing medical tourism destination region.

(e) financial status in hospital and travel agency

Some hospital's business volume decreases because COVID-19 occurs and foreign patient cannot go abroad to receive medical service. So, robust financial status in hospital and travel agency is key factor if medical tourism industry wants to sustainable development.

(f) volume of medical tourism traveler per year

Volume of medical tourism traveler in this region is the important index to evaluate medical tourism development in this region. High volume of medical tourism traveler not only can reduce medical cost but also can increase potential foreign patient because medical tourism traveler will execute "buzz marketing" if this traveler satisfies medical service in this region.

(g) medical service speed in hospital

Foreign patients go abroad for receiving medical service will request service quality. In some literatures, it justified that medical service time will negative influence patient safety and service quality perception [28]. So, controlling medical service speed in the reasonable range is the important competitiveness of hospital and it will influence the development of medical tourism industry in this region.

(h) maximum service volume in the region

The maximum medical service volume in the region is important index because this index will decide maximum business volume of medical tourism industry in this region. According to scale economics view point, high business volume of medical tourism industry can reduce medical cost in this region because hospital can purchase a huge amount of medicine and medicine price will reduce by bulk purchase [29].

Step 3. Collect information

In this study, nine scale linguistic variables are employed for experts to express their opinions [30] and [31]. Expert's opinion about competitiveness of each research target (region) in medical tourism industry respect to each criterion can refer to Table 4.

Table 4: Performance of each research target

	Expert 1's opinion				Expert 2's opinion			
	South Korea	Japan	Taiwan	China	South Korea	Japan	Taiwan	China
C_1	EG	EG	P	MG	MG	F	EG	VG
C_2	MP	MG	MP	EG	MG	EG	EG	EG
C_3	MP	MG	MG	EG	MG	F	EP	EG
C_4	G	EG	EG	MG	MP	MG	VG	MG
C_5	MG	F	MP	P	F	MG	EG	EP
C_6	EG	F	MP	EP	EG	F	EP	EP
C_7	EG	F	VG	F	G	MP	MP	EG
C_8	MG	G	MG	MP	MG	EG	EG	EP
	Expert 3's opinion				Expert 4's opinion			
	South Korea	Japan	Taiwan	China	South Korea	Japan	Taiwan	China
C_1	VG	MP	F	EP	MP	F	VG	F
C_2	EG	MG	MG	EP	F	G	VG	EP
C_3	MG	MG	G	G	MP	G	VG	MG
C_4	EP	MG	G	MG	F	G	G	F
C_5	MG	MP	EP	F	MG	MP	EG	VG
C_6	VG	F	VP	MP	G	MG	F	MP
C_7	MP	EG	MG	VG	EP	VG	F	P
C_8	EG	MG	F	MG	G	G	MG	G
	Expert 5's opinion				Expert 6's opinion			
	South Korea	Japan	Taiwan	China	South Korea	Japan	Taiwan	China
C_1	F	MG	F	MP	P	VG	VG	F
C_2	EG	VG	MP	F	EG	EG	MP	EG
C_3	MG	F	F	MG	MP	MG	EP	MG
C_4	EP	P	EG	VG	VG	VG	MP	MG
C_5	EG	VG	EP	EP	MP	P	MP	EG
C_6	MG	MG	P	EP	EP	MG	P	EP
C_7	F	EG	MG	F	F	MP	F	P
C_8	MP	F	MG	VG	VG	F	P	MG
	Expert 7's opinion				Expert 8's opinion			
	South Korea	Japan	Taiwan	China	South Korea	Japan	Taiwan	China
C_1	MP	F	VG	P	VG	G	P	MP
C_2	MG	VG	MP	MP	F	EG	MG	MP
C_3	EG	F	MG	EG	MG	VG	MG	VG
C_4	EP	P	F	F	F	F	EG	VG
C_5	EG	MG	MP	EP	MG	EG	F	VG
C_6	EP	F	EP	P	VG	MG	EP	MP
C_7	EP	MP	EG	P	MP	EG	VG	MG
C_8	EP	VG	P	VG	P	VG	F	P
	Expert 9's opinion							
	South Korea	Japan	Taiwan	China				
C_1	EG	G	MG	MG				
C_2	F	EG	MP	MP				
C_3	F	MG	F	VG				
C_4	MP	MP	VP	MG				
C_5	G	MP	MP	P				
C_6	VG	EG	EP	EP				
C_7	P	P	F	F				
C_8	MG	VG	VG	EG				

Step 4. Calculate weight (importance) of each criterion

In this work, we use entropy method (Refer to equation 7) to calculated weight of each criteria. Analysis result can refer to Table 5.

Table 5: Weight of each criterion

Criteria	C_1	C_2	C_3	C_4
Weight	0.1246	0.1902	0.1600	0.1190
Criteria	C_5	C_6	C_7	C_8
Weight	0.0656	0.1043	0.0906	0.1457

In this study, the rank of each research target (region) in medical tourism industry is decided by VIKOR method. The advantage of VIKOR is that VIKOR can solve decision problems with conflicting criteria. In this research, medical ability (C_2) and medical service speed in hospital (C_7) are conflicting criteria. Industry image (C_1) and medical cost (C_4) are conflicting criteria.

Step 5. Calculate PIS and NIS

Positive ideal solution (PIS) is the ideal status for medical tourism industry in Asia. On the opposite direction, negative ideal solution the worst status for medical tourism industry in Asia. PIS and NIS are reference points used to evaluate performance of each medical tourism industry (research target). Equation 8 and equation 9 are employed to calculate PIS and NIS. Analysis result can refer to Table 6.

Table 6: PIS and NIS

PIS	C_1	C_2	C_3	C_4
	0.7344	0.9688	0.9219	0.7969
	C_5	C_6	C_7	C_8
	0.7656	0.7500	0.7344	0.8438
NIS	C_1	C_2	C_3	C_4
	0.5156	0.5781	0.5625	0.4219
	C_5	C_6	C_7	C_8
	0.4688	0.1719	0.4688	0.6406

Step 6. Calculate total performance of research target S_i and the individual regret value of research target R_i .

Total performance of research target S_i means the entire performance of research target T_i. Individual regret value of research target R_i means the maximum regret degree of experts if he/she select research target T_i. S_i and R_i separately represents performance analysis result of research target based on different analysis direction. Equation 10 and equation 11 are employed to calculate S_i and R_i. Analysis result can refer to Table 7.

Step 7. Calculate Q value

In this research, decision parameter v is set up as 0.5. We consider that the evaluation index S_i and R_i is important equally. Equation 12 are employed to calculated Q value. Analysis result can refer to Table 7. According to analysis result, the rank of each region is T_2>T_1>T_3>T_4. The competitiveness rank sequence of medical tourism industry in Asia is Japan, South Korea, Taiwan and China.

Table 7: Total performance, individual regret value, Q value and rank among each medical tourism industry in Asia

	South Korea (T_1)	Japan (T_2)	Taiwan (T_3)	China (T_4)
Total performance of research target S_i	0.5795	0.1897	0.6310	0.6758
Individual regret value of research target R_i	0.1457	0.0974	0.1674	0.1902
Q value	0.6612	0.0000	0.8310	1.0000
Rank	2	1	3	4

Step 8. SWOT Analysis

According to experts' opinion (Refer to Figure 3), Taiwan medical industry possesses some strength is the dimension of medical cost (C_4) and medical service speed in hospital (C_7). However, Taiwan medical industry should take effort in the dimension of medical industry image (C_1), medical ability (C_2) and telemedicine ability (C_3). Medical ability and telemedicine ability of Taiwan medical industry will influence foreigners' medical industry image for Taiwan. Because volume of medical tourism traveler per year (C_6) in Taiwan is relatively small, Taiwan possesses high growth space to develop in the medical tourism market. Compare with south Korea, Japan and China, Taiwan lack scale economics in medical industry because financial status in hospital and travel agency (C_5), maximum service volume in region (C_8) is relatively small. The disadvantage of scale economics in Taiwan will cause that it is hard to reduce Taiwan's overall medical costs and is not easy to enhance Taiwan's overall medical industry image (Refer to Table 8).

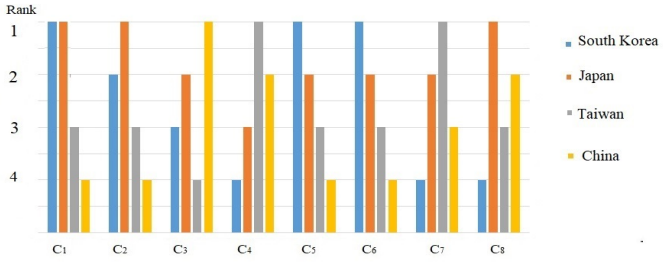


Figure 3: Rank of each region among different dimension

Table 8: SWOT analysis for Taiwan medical tourism industry

Strengths (1) Medical cost is relatively cheap. (2) Medical service speed is relatively fast.	Weaknesses (1) Taiwan medical industry image from tourism traveler is relatively less well-known. (2) Medical ability is weak. (3) Telemedicine ability is weak.
Opportunities (1) Volume of medical tourism traveler is relatively small.	Threats (1) Financial status in hospital and travel agency is relatively small (2) Maximum service volume in region is relatively small

5. Management Implementation

According to experts' opinion, Taiwan possesses high growth space to develop in the medical tourism market because medical cost in Taiwan is relatively cheap which can attract some foreigner whose economics ability is not good enough to receive high price medical service in Japan or South Korea. There are some policies which can promote the competitiveness for Taiwan tourism medical industry.

5.1. Increase volume of medical students each year

There are only 1,300 new students who can enter to study department of medicine in Taiwan each year. Under this total control policy, although Taiwan can maintain high medical standards, it cannot improve Taiwan's medical energy to meet the needs of additional medical services for foreign patients.

5.2. Build overseas medical zone

Taiwan has implemented universal health insurance. Therefore, Taiwan citizen can use low-cost medical services. However, the purpose of developing medical tourism industry is to increase Taiwan's overall GDP. It is unfair to Taiwan citizens if charging standards for foreign patients and Taiwan citizens are the same. Doctor and nurse should use English to serve foreign patients.

This is also a language threshold to develop medical tourism industry.

In order to compete with foreign medical institutions, this study suggests that Taiwan should set up an overseas medical zone. The hospitals in the zone are dedicated to treating foreign patients. It is beneficial to the development for Taiwan's medical tourism industry.

5.3. Development of telemedicine technology

The occurrence of COVID-19 in 2019 has caused the stagnation of the tourism industry in various countries around the world. This event has also affected the development of medical tourism. The medical tourism industry needs to maintain relationship with foreign patients. Telemedicine technology is an important means for hospitals to maintain relationship with foreign patients when international tourism activity cannot be implemented during 2019-2020. Taiwan's medical tourism industry should strengthen and improve telemedicine technology.

5.4. Develop overseas free clinic activities

Taiwan's medical industry is less well-known than south Korea and Japan. Taiwan should go to various countries in southeast Asia and central Asia for free consultations. This will not only enhance Taiwan's medical industry image but also increase the potential patience source for Taiwan's medical tourism industry. The world's low-end industrial manufacturing order may move from China to various countries in southeast Asia and central Asia. Therefore, people from countries in southeast asia and central asia will gradually have the ability to travel to Taiwan for receiving medical service.

6. Conclusion

Medical tourism is very important industry in many countries. A novel framework for evaluating competitiveness of medical tourism industry has been designed in this research. South Korea, Japan, Taiwan and China are selected as research target to evaluate their competitive ability. According to analysis result, we find out that Taiwan's medical tourism competitiveness falls in backward group. So, Taiwan government should provide more resource to improve its medical tourism competitive ability for developing medical tourism industry especially some citizen in other country such as America has high motivation to go abroad for medical treatment because medical expenses is too expensive in America. So, this research provides four suggestion to Taiwan government.

In the future, relative scholar can use this model to evaluate competitiveness of medical tourism industry among Asian countries and European countries. Scholar can analyze how the cultural difference can influence the rank of medical tourism competitiveness.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

I state that the data in this research collect from experts in each field and experts' occupation includes government officer, employees in traditional hotel, employees in travel agency, nurse leader, doctor (hospital dean) and university professors. Other scholars can use the data to execute extended research.

References

[1] I. Sophie, "Revealed: Countries With The Best Health Care Systems. CEOWORLD Manazine," https://ceoworld.biz/2019/08/05/revealed-countries-with-the-best-health-care-systems-2019/, 2019.
[2] H. M. Chen, C. H. Tseng, "The performance of marketing alliances between the tourism industry and credit card issuing banks in Taiwan," Tourism Management, **26**, 15-24, 2005.
[3] S. Gössling, D., Scott, C. M. Hall, "Pandemics, tourism and global change: a rapid assessment of COVID-19," Journal of Sustainable Tourism, 1-20, 2020. https://doi.org/10.1080/09669582.2020.1758708
[4] L. Qi, Y. X. Zhuo, S. Cai, "Taiwan's international medical tourism development strategy," Taiwan Medical Circle, **55**(8), 48-51, 2012. https://doi.org/10.11634/216837861403126
[5] T. H. Chang, "Fuzzy VIKOR method: a case study of the hospital service evaluation in Taiwan," Information Sciences, **271**, 196-212, 2014. https://doi.org/10.1016/j.ins.2014.02.118
[6] A. M. Sharifabadi, F. A. Ardakani, "A model for health tourism development using fuzzy TOPSIS and interpretive structural modeling in Yazd province," Journal of Health Administration (JHA), **17**(55), 73-88, 2014.
[7] Z. Awang, A. Afthanorhan, M. Mohamad, M. A. M. Asri, "An evaluation of measurement model for medical tourism research: the confirmatory factor analysis approach," International Journal of Tourism Policy, **6**, 29-45, 2015.
[8] A. Görener, "A. swot-ahp approach for assessment of medical tourism sector in Turkey," Alphanumeric Journal, **4**, 159-170, 2016.
[9] P. Ajmera, "Ranking the strategies for Indian medical tourism sector through the integration of SWOT analysis and TOPSIS method," International Journal of Health Care Quality Assurance, **30**, 668-679, 2017.
[10] F. Abadi, I. Sahebi, A. Arab, A. Alavi, H. Karachi, "Application of best-worst method in evaluation of medical tourism development strategy," Decision Science Letters, **7**(1), 77-86, 2018.
[11] J. Roy, K. Chatterjee, A. Bandyopadhyay, S. Kar, "Evaluation and selection of medical tourism sites: A rough analytic hierarchy process based multi-attributive border approximation area comparison approach," Expert Systems, **35**, 12232, 2018.
[12] M. Nilashi, S. Samad, A. A. Manaf, H. Ahmadi, T. A. Rashid, A. Munshi, W. Almukadi, O. Othman Ibrahim, O. H. Ahmed, "Factors influencing medical tourism adoption in Malaysia: A DEMATEL-Fuzzy TOPSIS approach," Computers & Industrial Engineering, **137**, 106005, 2019.
[13] A. Singh, "An integrated approach towards ranking hospitals using Fuzzy AHP and ELECTRE-I technique," International Journal of Healthcare Management, 1-10, 2019. https://doi.org/10.1016/j.ins.2014.02.118
[14] J. Takacs, S. A. Vaduva, "A swot analysis of the global hospitality industry," Revista Economica, **69**(6), 105-119, 2017.
[15] A. Reihanian, N. Z. B. Mahmood, E. Kahrom, T. W. Hin, "Sustainable tourism development strategy by SWOT analysis: Boujagh National Park," Iran. tourism management Perspectives, **4**, 223-228, 2012.
[16] P. Liu, F. Teng, "An extended TODIM method for multiple attribute group decision-making based on 2-dimension uncertain linguistic Variable," Complexity, **21**, 20-30, 2016.
[17] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning," Learning systems and intelligent robots. Springer, Boston, MA, 1-10, 1974.
[18] F. Herrera, L. Martinez, "The 2-tuple linguistic computational model: Advantages of its linguistic description, accuracy and consistency," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, **9**, 33-48, 2001.
[19] A. I. A1 Maghsoodi, G. Abouhamzeh, M. Khalilzadeh, E. K. Zavadskas, "Ranking and selecting the best performance appraisal method using the MULTIMOORA approach integrated Shannon's entropy," Frontiers of Business Research in China, **12**(1), 2018.
[20] R. Levy, J. P. F. LeBlanc, E. Gull, "Implementation of the maximum entropy method for analytic continuation," Computer Physics Communications, **215**, 149-155, 2017. https://doi.org/10.1016/j.cpc.2017.01.018
[21] W. Huang, B. Shuai, Y. Sun, Y. Wang, E. Antwi, "Using entropy-TOPSIS method to evaluate urban rail transit system operation performance: The China case," Transportation Research Part A: Policy and Practice, **111**, 292-303, 2018.
[22] A. De Nisco, G. Mainolfi, V. Marino, M. R. Napolitano, "Tourism

satisfaction effect on general country image, destination image, and post-visit intentions," Journal of Vacation Marketing, **21**(4), 305-317, 2015.
[23] S. Chaulagain, J. Wiitala, X. Fu, X. "The impact of country image and destination image on US tourists' travel intention," Journal of Destination Marketing & Management, **12**, 1-11, 2019.
[24] C. R. Peters, K. M. Sauer, "A survey of medical tourism service providers," Journal of Marketing Development and Competitiveness, **5**(3), 117-126, 2011.
[25] J. Connell, "Contemporary medical tourism: Conceptualisation, culture and commodification," Tourism management, **34**, 1-13, 2013.
[26] W. Bies, L., Zacharia, "Medical tourism: outsourcing surgery," Mathematical and Computer Modelling, **46**(7-8), 1144-1159, 2007.
[27] D. York, "Medical tourism: the trend toward outsourcing medical procedures to foreign countries," Journal of Continuing Education in the Health Professions, **28**(2), 99-102, 2008.
[28] D. S. Kc, C. Terwiesch, "Impact of workload on service time and patient safety: An econometric analysis of hospital operations," Management Science, **55**(9), 1486-1498, 2009.
[29] H. E. Frech III, L. R. Mobley, "Resolving the impasse on hospital scale economics: a new approach," Applied Economics, **27**(3), 286-296, 1995.
[30] L. Martı, F. Herrera, "An overview on the 2-tuple linguistic model for computing with words in decision making: Extensions, applications and challenges," Information Sciences, **207**, 1-18, 2012.
[31] J. H. Wang, J. Hao, "A new version of 2-tuple fuzzy linguistic representation model for computing with words," IEEE transactions on Fuzzy Systems, **14**, 435-445, 2006.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1698-1703 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Education Value Chain Model for Examination, Grading, and Evaluation Process in Higher Education based on Blockchain Technology

Meyliana[1,*], Yakob Utama Chandra[1], Cadelina Cassandra[1], Surjandy[1], Erick Fernando[1], Henry Antonius Eka Widjaja[1], Harjanto Prabowo[2]

[1]*Information Systems Department, School of Information Systems, Bina Nusantara University, Jakarta, 11480, Indonesia*

[2]*Computer Science Department, Bina Nusantara University, Jakarta, 11480, Indonesia*

ARTICLE INFO

Article history:
Received: 20 November, 2020
Accepted: 08 December, 2020
Online: 28 December, 2020

Keywords:
Blockchain
Examination
Evaluation
Grading
Higher education
Student

ABSTRACT

Blockchain is a solution to improve data integrity and minimize data manipulation. Blockchain technology can be used in student examination management in universities. The process of managing student examination produces some information that can be stored in a blockchain database such as information about the lecturer as the exam creator, exam proctor, student attendance, student exam scores, and the student study continuity status. Blockchain technology ensures student data is valid and reliable, therefore, the company's or organizational trust the teaching and learning process in higher education. This research was conducted using a qualitative approach, which is user center design (UCD) the goal is building a blockchain technology model that can be used by universities in the process of managing student examination, grading, and evaluation. This process will be repeated every semester until students graduate from the university. The propose of this research is to create a business process model based on blockchain technology that can improve data integrity in universities.

1. Introduction

In teaching and learning process in higher education, students will pass through midterm exams, final exams, lab exams, and certification exams [1]. This examination is carried out to measure the student comprehension during the teaching and learning process. Lecturer can also measure how far learning objectives achieved during the process [2].

After examination, students will get the score according to the criteria determined by the department or university. All grade will be recorded in Student Academic Transcript. This Academic Transcript is a reference for company to hired new graduate working in their company [3]. The company will filter the grade and choose the relevant position according to the student ability [4].

Valid and reliable grade will increase the company or organizational confidence and trust in the teaching and learning process that occurs in higher education [5]. However, in 2019 until now, there have been several cases of diploma certificate forgery. As the impact, the company or organization doubts about the data originality stated on the academic transcript and they need to crosscheck back the data to the university [6].

This is challenge by the university to keep the data integrity and security from any manipulation. According to the case happen in 2019, employee working in the company can also be a threat who manipulate the data [7].

This problem requires universities to updates the technology used, especially those concerning the data integrity and security. Blockchain technology is immutable, unchangeable. It also peer to peer, distributed (transparent), validated, and secured, which are very suitable for solving data integrity and security issues [7], [8].

One of the main processes in producing student grades is student examinations management process. This process has several sub processes, they are examination creation process, exam proctor attendance recapitulation, Student attendance recapitulation, the grading process (collecting exam answer and grading from lecture, and the evaluation process (to determine the student status for next semester).

All of the processes will produce data and information in academic transcript which are very important for students, the

*Corresponding Author: Meyliana, Meyliana@binus.edu

validity of data and information is a top priority for universities to For this reason, blockchain solutions are needed in managing student value data [9].

Based on the phenomena, the research question for this research is how to build an education value chain model in student exam management process based on blockchain technology in universities? The goal is to build an education value chain model for exam management process to ensure the data integrity.

2. Literature Review

2.1. Education Value Chain

This section discusses the Education value chain in general, starting from prospective students applying to the university, pass the test, and accepted as students. Students follow two types of activities during their study at higher education, they are intra-curricular and extra-curricular activities.

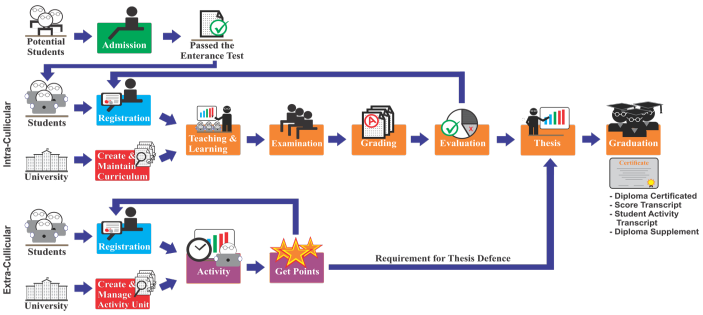


Figure 1: Education Value Chain

The intra-curricular activity starts from course registration for each semester. Then students will follow the teaching and learning process, students will take mid and final exams in each semester. For each exam, lecturer will give the score to students. Based on the results, students will be evaluated by the department and academic advisory unit so student status can be known every semester. In the final semester, students will take a final project or thesis, and thesis defense will be carried out at the end. Apart from following intra-curricular activities, students also take part in extra-curricular activities. Students will register to student organizations and participate in the organization activity. As a result of participating in the activities, students will get points. The points are collected into Student Activity Transcript (SAT). This SAT points is a pre-requisite for thesis defense [7].

2.2. Examination

In general, in one semester there are two exams, mid exam and final exam. With examination, the department or university can find out how far students can understand the overall material delivered and how far the learning objectives achieved. In addition, exam answers can provide feedback for lecturers, departments, and universities [10].

2.3. Grading

After the student takes the exam, the lecturer will give a score in accordance with the assessment measurement set by the university. Grading is the starting point for students to get a cumulative achievement index measured from the cognitive side of students during the teaching and learning process. Assessment may vary according to the type of exam being performed. The theory exam can be assessed based on the answers written by students. Oral exams or presentations are assessed directly based on the results of student presentations. Meanwhile, practical lab exams can be assessed based on the results of the practicum work made [11].

2.4. Evaluation of Study Progress

After students finish their semester, the department and academic supervisor unit will evaluate the learning progress. If students get the satisfactory results, the student may be got appreciation. However, if the evaluation process result is not good, Student will follow various coaching and guidance determined based on student cases and discussions between departments and student advisory units. On the other hand, if students get poor grades and less motivation, then the department and student advisory unit will try to provide solutions according to student interests [12].

2.5. Blockchain in Education

Blockchain research in education can provide new insights to education practitioners. Blockchain technology can be used to improve data integrity to minimize crimes such as diploma certificate forgery. In addition, previous research related to blockchain implementation is the application of smart contracts to business processes in universities [9][13].

3. Research Methodology

Research methodhology used in this research is user-centered design (UCD) [14]. In UCD methodology, research begins with an understanding of the context of use, at this stage a literature study is carried out related to the implementation of blockchain technology in universities. The next stage is to specify user requirements by conducting observations, interviews, and focus group discussions (FGD) to determine the teaching and learning process, especially in examination, grading, and evaluation.

Then a design solution was carried out by creating an education value chain model for the examination, grading and evaluation processes that apply to universities based on blockchain technology. At this stage, it is also described in detail the business processes and identify which activities need to store data and information to the blockchain database. The next stage is evaluate the requirements, FGD is conducted to see whether the model can be accepted and implemented in each university. This evaluation becomes input for validation on the design solution, specify user requirements and understand context of use. If there are things that are still not suitable, then it becomes feedback for the improvement of each stage in this research methodology.

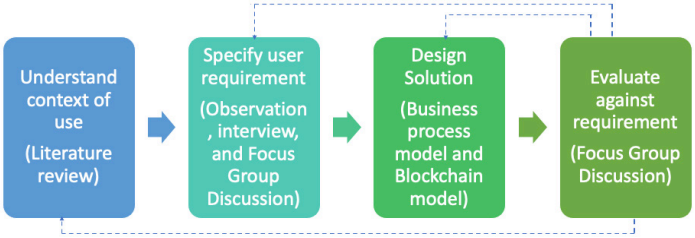


Figure 2: Research Methodology

4. Result and Discussion

The teaching and learning process in higher education which is discussed in this study is the examination, grading and evaluation process as shown in Figure 3.

These three processes are focused on the process of student examination management and evaluations starting from the preparation, execution, and evaluation of student study. All data generated in this process will be stored in a blockchain database so that universities can guarantee the data they have cannot be manipulated and have a good integrity.

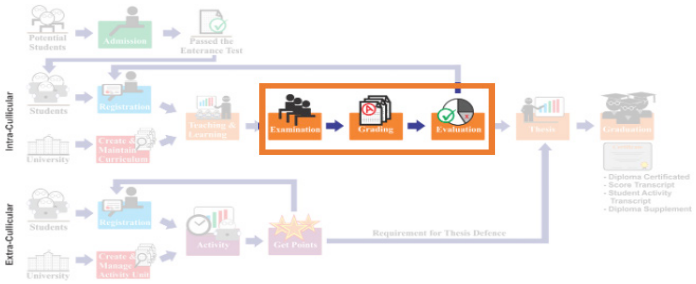


Figure 3: Focus of Research

The examination process begins with exam schedule by the student scheduling department. From the exam schedule, the department will determine the lecturer as the exam question creator. The lecturer will fill out an Exam Agreement Form that contains the topic of the questions and the learning outcomes that will be achieved by students after completing the exam. This form is filled to get the standardized exam questions and approval from the department. The data of the lecturer will be recorded on the blockchain database for honorarium.

The questions that have been completed will be examined by the department and given to the academic operations section to be distributed according to the exam schedule to students.

The exam scheduling section will also schedule the lecturers as exam proctor and the attendance realization data will be stored in a blockchain database as proof of payment of exam proctor fees.

Students take the exam twice per semester (mid exam and final exam). For a practicum course, students take a practicum exam. In each exam, the attendance will be recorded, and this data is stored in a blockchain database as proof of student attendance in the exam.

If the student is absent or cannot attend the exam, student can apply for a make-up exam in accordance with the applicable terms and conditions.

The exam answers are collected by the exam proctor to academic operations section. The academic operational section will validate the answer and distribute it to the lecturer. The lecturer will check the answers and gives score to each student. The score is entered by the lecturer through the system and will be validated by the department in the score judiciary. Student scores consist of several components, such as assignment scores, mid-test scores and final exam scores. This score is stored in the blockchain database. Lecturers will receive the examiner's honorarium according to the number of exam files corrected and proof of payment of this honorarium is stored in the blockchain database.

All grades entered in one semester will be evaluated by the academic supervisor and department. The results of the evaluation will determine the status of students in the continuity of the studies. The status will be stored in the blockchain database.

Figure 4 shows the transactions illustration from the education value chain model for the examination, grading and evaluation processes based on blockchain technology. Users who play a role in the examination, grading, and evaluation processes are students, lecturers, academic operations, scheduling units, student advisory units, and faculty / departments.

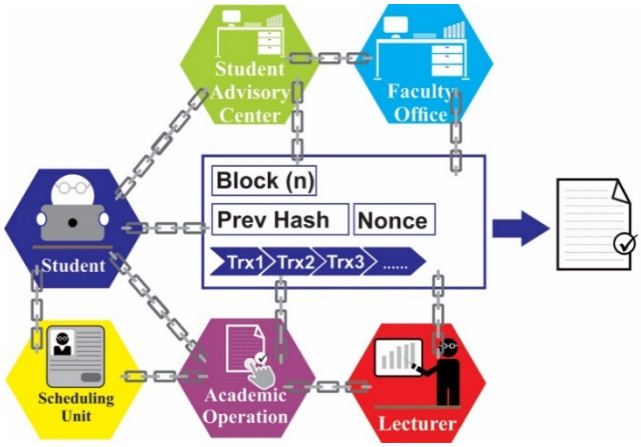
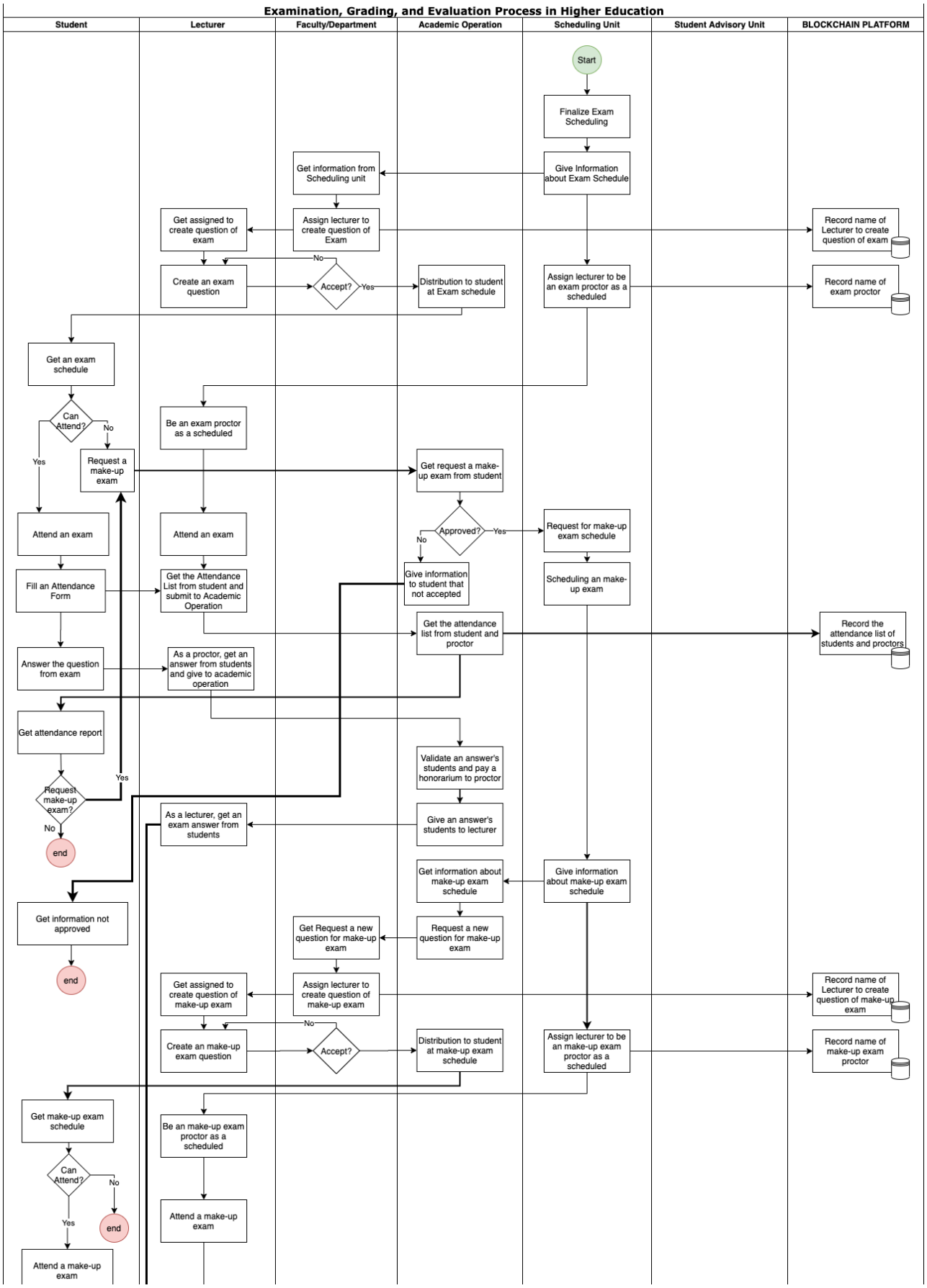


Figure 4: Transaction Illustration of Education Value Chain Model for Examination, Grading, and Evaluation for Higher Education based on Blockchain Technology

Figure 5 (a and b) shows the college's business process in teaching and learning activities focusing on the examination, grading, and evaluation processes. From the business process described in Figure 4 above, the data storage in the blockchain database is important in connection to the financial audit and payment process. The data are:

- Record name of lecturer as the exam creator
  - This data contains the semester period, lecturer code, course code, class code, exam date, and honorarium amount.
- Record name of Exam Proctor
  - This data contains the semester period, lecturer code, course code, class code, exam date, exam room.
- Record the attendance list of students and exam proctor
  - This data contains the semester period, lecturer code, course code, class code, exam date, exam room, student code, and the amount of the exam proctor fee.
- Record a score of students and Record the new score.
  - This data contains the semester period, course code, class code, exam date, exam room, student code, exam component code, and student scores.
- Record a payment of scoring honorarium to lecturer
  - This data contains the semester period, lecturer code, course code, class code, exam date, exam room, number of students and the amount of honorarium
- Record the student's evaluation for each semester
  - This data contains the semester period, course code, student code, component code grades, grades, social studies, and GPA


Examination, Grading, and Evaluation Process in Higher Education
Student
Lecturer
Faculty/Department
Academic Operation
Scheduling Unit
Student Advisory Unit
BLOCKCHAIN PLATFORM
Start
Finalize Exam Scheduling
Give Information about Exam Schedule
Get information from Scheduling unit
Assign lecturer to create question of Exam
Get assigned to create question of exam
Record name of Lecturer to create question of exam
Create an exam question
Accept?
No
Yes
Distribution to student at Exam schedule
Assign lecturer to be an exam proctor as a scheduled
Record name of exam proctor
Get an exam schedule
Can Attend?
No
Yes
Be an exam proctor as a scheduled
Request a make-up exam
Get request a make-up exam from student
Approved?
Request for make-up exam schedule
Attend an exam
Fill an Attendance Form
Get the Attendance List from student and submit to Academic Operation
Give information to student that not accepted
Scheduling an make-up exam
Get the attendance list from student and proctor
Record the attendance list of students and proctors
Answer the question from exam
As a proctor, get an answer from students and give to academic operation
Get attendance report
Validate an answer's students and pay a honorarium to proctor
Request make-up exam?
Yes
No
end
As a lecturer, get an exam answer from students
Give an answer's students to lecturer
Get information about make-up exam schedule
Give information about make-up exam schedule
Get information not approved
Request a new question for make-up exam
Get Request a new question for make-up exam
Get assigned to create question of make-up exam
Assign lecturer to create question of make-up exam
Record name of Lecturer to create question of make-up exam
Create an make-up exam question
Distribution to student at make-up exam schedule
Assign lecturer to be an make-up exam proctor as a scheduled
Record name of make-up exam proctor
Get make-up exam schedule
Be an make-up exam proctor as a scheduled
Attend a make-up exam

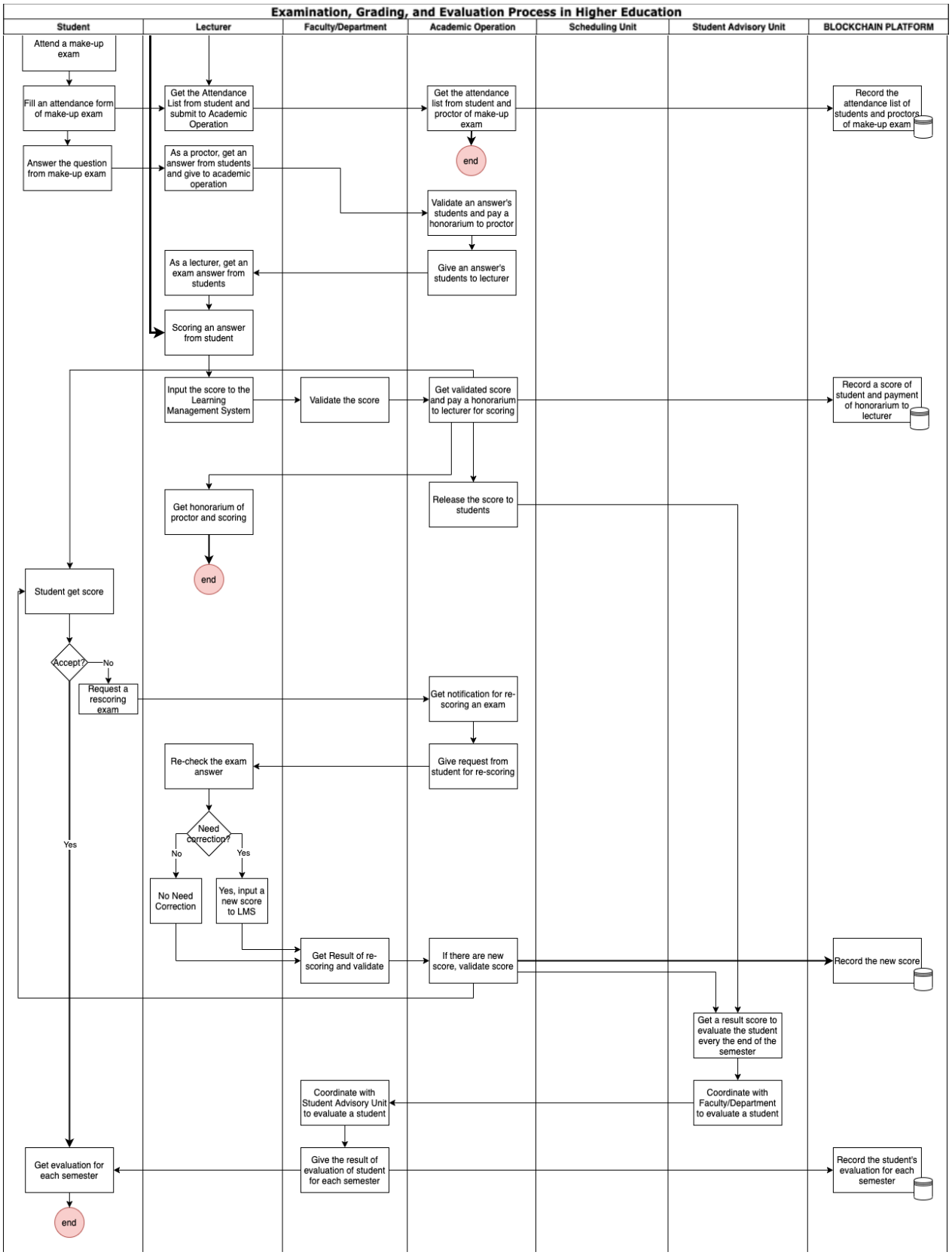


Figure 5: (a) Business Process of of Education Value Chain Model for Examination, Grading, and Evaluation for Higher Education based on Blockchain Technology (b) Business Process of of Education Value Chain Model for Examination, Grading, and Evaluation for Higher Education based on Blockchain Technology

5. Conclusions

By using blockchain technology, it can be concluded that data integrity will be guaranteed, and data storage is more secure because the data has been validated and cannot be changed. Data changes can still occur by making new transactions without eliminating previous transactions, for example when students protest the score which causes the previous score to change. The Both scores are stored in the blockchain database.

Following the research methodology, we have conducted a focus group discussion to validate the model that has been made and the results are that all participants agree with this model and they think this study can be easily applied in universities in recording the academic transcript.

The process of examination, grading, and evaluation is very appropriate using blockchain technology, because this process is very crucial and requires data validity to produce the academic transcript for students.

In the current condition of the Covid-19 pandemic, using blockchain technology is more needed because the process input and evaluation are done digitally. With blockchain technology, universities make a digital transformation in the education sector and make education technology more effective, especially during the Covid-19 pandemic.

The limitations of this study are the process of interviewing and focus group discussion using virtual conferencing to get answers from the speakers and there is limited time in digging deeper into the information needed. So it is necessary to do the validity of the resulting model several times. Future research will build an integrated overall model to produce a value chain with data integrity starting from students entering higher education until students graduate and get diplomas, transcripts, diploma companion documents, and student activity transcripts.

Acknowledge

This study is supported by the Directorate General of Strengthening for Research and Development, Ministry of Research, Technology, and Higher Education, Republic of Indonesia as a part of Penelitian Terapan Unggulan Perguruan Tinggi Research Grant to Bina Nusantara University entitled "Implementasi Blockchain Platform untuk Menciptakan "Good Governance" pada Perguruan Tinggi" or "The Implementation of Blockchain Platform to Create "Good Governance" in Higher Education" with contract number: 12/AKM/PNT/2019 and contract date: 27 March 2019.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M.M.A. Getso, Z. Johari, "the Blockchain Revolution and Higher Eucation," International Journal of Information Systems and Engineering, **5**(1), 57–65, 2017, doi:10.24924/ijise/2017.04/v5.iss1/57.65.

[2] R. Wuthisatian, "Student exam performance in different proctored environments: Evidence from an online economics course," International Review of Economics Education, **35**(August), 100196, 2020, doi:10.1016/j.iree.2020.100196.

[3] Z. Wu, T. He, C. Mao, C. Huang, "Exam paper generation based on performance prediction of student group," Information Sciences, **532**, 72–90, 2020, doi:10.1016/j.ins.2020.04.043.

[4] M. Apte, A. Bhave-gudipudi, "ScienceDirect Cooperative Learning techniques to bridge gaps in academia and corporate," Procedia Computer Science, **172**(2019), 289–295, 2020, doi:10.1016/j.procs.2020.05.046.

[5] A. Herrarte, R. Llorente-heras, A. Herrarte, R. Llorente-heras, "Competencies, occupational status, and earnings among European university graduates," 2017, doi:10.1016/j.econedurev.2017.10.006.

[6] Meyliana, Y.U. Chandra, C. Cassandra, Surjandy, E. Fernando, A.E. Widjaja, H. Prabowo, "A Proposed Model of Secure Academic Transcript Records with Blockchain Technology in Higher Education," (Conrist 2019), 172–177, 2020, doi:10.5220/0009907401720177.

[7] M. Meyliana, Y.U. Chandra, C. Cassandra, S. Surjandy, H.A. Eka Widjaja, E. Fernando, H. Prabowo, C. Joseph, "DEFYING THE CERTIFICATION DIPLOMA FORGERY WITH BLOCKCHAIN PLATFORM: A PROPOSED MODEL," in Proceedings of the International Conferences ICT, Society, and Human Beings 2019; Connected Smart Cities 2019; and Web Based Communities and Social Media 2019, IADIS Press: 63–71, 2019, doi:10.33965/ict2019_201908L008.

[8] S. Singh, N. Singh, "Blockchain: Future of financial and cyber security," Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics, IC3I 2016, 463–467, 2016, doi:10.1109/IC3I.2016.7918009.

[9] M. Turkanovic, M. Holbl, K. Kosic, M. Hericko, A. Kamisalic, "EduCTX: A blockchain-based higher education credit platform," IEEE Access, **6**, 2018, doi:10.1109/ACCESS.2018.2789929.

[10] L.M. Daniels, M.J. Gierl, "The impact of immediate test score reporting on university students ' achievement emotions in the context of computer-based multiple- choice exams," Learning and Instruction, 2017, doi:10.1016/j.learninstruc.2017.04.001.

[11] I. Cornelisz, M. Meeter, C. van Klaveren, "Educational equity and teacher discretion effects in high stake exams," Economics of Education Review, **73**(November 2018), 101908, 2019, doi:10.1016/j.econedurev.2019.07.002.

[12] H. Liao, J. Hitchcock, Reported Credibility Techniques in Higher Education Evaluation Studies that use Qualitative Methods: A Research Synthesis, Elsevier Ltd, 2018, doi:10.1016/j.evalprogplan.2018.03.005.

[13] R. Bdiwi, C. de Runz, S. Faiz, A.A. Cherif, "Towards a New Ubiquitous Learning Environment Based on Blockchain Technology," 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), 101–102, 2017, doi:10.1109/ICALT.2017.37.

[14] What is User Centered Design? | Interaction Design Foundation (IxDF), Nov. 2020.

ASTES Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1704-1709 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Analysis of Learning Difficulties in Object Oriented Programming in Systems Engineering Students at UNTELS

Teodoro Diaz-Leyva*, Omar Chamorro-Atalaya

Faculty of Engineering and Management, Universidad Nacional Tecnológica de Lima Sur Lima, Lima, 15047, Perú

ARTICLE INFO

Article history:
Received: 01 September, 2020
Accepted: 01 December, 2020
Online: 28 December, 2020

Keywords:
Object Oriented Programming
Learning difficulty
Programming Language I
Systems Engineering Professional Career

ABSTRACT

The learning difficulties of Object-Oriented Programming can be one of the causes of failure in college students, which can even lead to student dropout. This article has the purpose of analyzing the main learning difficulties of object-oriented programming using the Java programming language in Systems Engineering students of the Universidad Nacional Tecnológica de Lima Sur (UNTELS), research that arises because the students reflect low academic performance in the subject of Programming Language I (object-oriented programming) that is developed in the fourth academic cycle with six hours of practice and zero hours of theory, because this is determined by the study curriculum. To collect data, first, the possible causes associated with learning difficulties have been identified through direct dialogue with the students of the second academic semester of 2019, then with these causes a questionnaire has been prepared to apply the survey technique. establishing indicators classified in the dimensions Teaching methodology, Previous knowledge of programming courses, reading materials, learning concepts and Process of learning to program. This research yielded as a result that students perceive with greater difficulty the indicators corresponding to the dimension "Learning concepts", in second place, the dimension Process of learning to program is presented, in third place the dimension Teaching methodology, in fourth place the Reading materials dimension and fifthly the dimension Previous knowledge of programming courses, the results of which will serve as a basis for the teachers of the subject to improve the strategies in the teaching / learning process in the students of the professional career of Engineering Systems.

1. Introduction

In [1], the author expresses due to the generalized use of the object-oriented programming paradigm, teaching is not without difficulties, since it is based on a solid base of elementary concepts on which to add experiences of design of applications oriented to objects. Currently, one of the most widely used paradigms in software development is Object Oriented Programming (OOP), it implements programs in terms of objects, includes concepts of inheritance, cohesion, abstraction, polymorphism, coupling and encapsulation.

In [2], the author indicates, the academic programs of Engineering in Systems, the learning of the programming oriented to objects is seen in relation to the scene of the life of their professional tasks. Students are very concerned that their learning is an element that allows them to act effectively and efficiently outside the university environment. Likewise, in [3] the author points out that it is necessary for the student to learn OOP and not the peculiarities of the chosen language, to present the theoretical contents in a way that facilitates their understanding and progressive learning.

In [4], the author points out, many times, even using languages that support OOP, the mistake is made of not programming in this methodology, because there are no clear concepts in this regard. Also in [5], the author indicates that, for many years, the emphasis in teaching was to transmit concepts and procedures whose meaning was often not understood. Also in [6], the author expresses the object-oriented paradigm includes a large number of concepts that allow the development of robust applications.

In [7], the author points out that currently there is a variety of programming languages, learning generally requires knowledge of flow diagrams, pseudo-code and the syntax of the programming language itself to build programs under the structured approach,

*Corresponding Author: Teodoro Neri Diaz Leyva, Sector 3 Grupo 1A 03, Av Central, Villa El Salvador-UNTELS, 982716223 & is_diazl@hotmail.com

that is, , using sequential, conditional and repetitive structures. Thus, learning to program at a professional level, regardless of the language used, is not a one-day task. There are students who fail to acquire the necessary programming skills, even after completing a Computer Science Programming Fundamentals course.

For many students, learning difficulty occurs as they move from structured programming to object-oriented programming and integrate various software development technologies. Thus, in [8], the author points out that the problem of learning object-oriented programming is manifested since it is a complex issue that involves the integration of many elements such as the object-oriented paradigm, the programming language, the environment development, development methodology, modeling language, development patterns. and programming logic. Therefore, students are faced with an overwhelming number of concepts in a short period of time, which makes it difficult to assimilate and develop skills to generate lines of code.

When studying complex courses such as computer programming, one should begin by understanding the fundamental concepts of programming, in this sense, in [9], the author indicates the search for strategies that allow improving the teaching and learning of programming. an indispensable task. Among the causes of failure, it has been concluded that they do not reside in the student's difficulty in translating the solution of a problem into the phrases of a programming language, but rather have to do with a lack of methodology, habit and skill. to solve problems.

In [10], the author explains that the learning of the disciplines of Algorithmic and Algorithmic Programming presents, perhaps, one of the highest levels of difficulty in the careers of Computer Science and Computer Science. Historically, students have faced problems assimilating abstract mathematical notions, particularly when these include the dynamics of how algorithms manipulate data (Acm and Ieee-Cs, 2009). Object-oriented programming courses require prior knowledge of programming logic that is generally learned in algorithm courses, however, this is where students already have learning difficulties.

In [11], the author points out that the problem of university dropouts is complex and includes various causes; Institutional characteristics would be the first dropout factor, followed by vocational, academic and, finally, economic interests. Likewise in [12], the authors point out that dropout is associated with academic capital, the perception of performance, skills gaps to face studies, time management and the incorporation of learning techniques and habits that affect the decision to drop out of students.

At the Universidad Nacional Tecnológica de Lima Sur (UNTELS) the Programming Language I subject refers to Object-Oriented Programming with Java programming language, it is dictated in the fourth cycle of the Professional Career of Systems Engineering, which is develops in six (6) hours of practice and zero (0) hours of theory; After having taken two structured programming subjects, the academic performance of the students is not optimal, in the first academic semester of 2019 of 28 students enrolled 64% failed the course and in the second academic semester of 2019 of 25 students enrolled 40%.

In this context, the present research aims to describe the factors that hinder the learning of Object-Oriented Programming with Java Language, since the student goes from studying programming with a structured approach to object-oriented and with little theory about the different topics addressed during its development. For this, a direct dialogue was established on the possible problems associated with learning difficulties, which have been ordered and reflected in a physical questionnaire regarding the teaching methodology, previous knowledge of programming courses, reading materials, learning concepts. and the process of learning to program with the purpose of improving the teaching and learning strategies in the students of the Programming I language course in the professional career of Systems Engineering at the National Technological University of Lima Sur.

2. Methodology

2.1. Research level

The research level of this article is descriptive / correlational. Descriptive because it specifies the characteristics of a group of students with respect to the learning difficulties of Object-Oriented Programming and is correlational since it seeks to know the degree of association between the dimensions Teaching methodology, Reading materials, Previous knowledge of courses Programming, Learning Concepts and Learning Process for Programming. In [13], the author describes the purpose of this type of study is to know the relationship or degree of association that exists between two or more concepts, categories or variables in a particular context.

To achieve these results, the following objectives have been set: to describe the main difficulties in learning object-oriented programming. Determine the dimension with the greatest learning difficulty of Object Oriented Programming.

2.2. Population and sample

The population is made up of 25 students of the Programming Language I subject of the second academic semester of 2019 of the Systems Engineering professional career of the National Technological University of Lima Sur.

The sample is equal to the population, because it was only considered to apply the questionnaire to students in the second academic semester of 2019, which corresponds to a number of 25 and who take the subject of Programming Language I (Object-Oriented Programming).

In [14], the authors in their article "The Proportionality of Women Graduated from the Professional Career of Mechanical and Electrical Engineering at UNTELS: Analysis of their Academic Performance and Field of Labor Action", point out that the sample will be equal to the population; Because the sample is less than 50, that is, the weighted average and the weighted average by specialty (Mechanical Subjects and Electrical Subjects) of the 38 women graduated from the Faculty of Mechanics and Electrical Engineering, until 2018 II. In this sense, in this research, the sample is equal to the population, since there is no other group of students studying the same Programming Language I subject in the second academic semester of 2019 that constitute a greater number in the population and therefore the sample is larger. Likewise, in [15], the authors highlight that the UNTELS University is an institution with a few years of operation, specifically it was founded in 2007, and its first promotion of graduates is registered

in 2011, to date the University does not have any report or information that allows relevant actions to be taken to improve the teaching and learning process of the subject Programming Language I.

2.3. Instrument used in data collection

The instrument used in data collection is the "survey", with questions about difficulties associated with learning object-oriented programming at UNTELS, considering the dimensions Teaching methodology, Reading materials, Previous knowledge of programming courses, Learning of concepts and Process of learning to program.

3. Result

Next, we proceed to describe the results obtained from the processing of the data collected in the survey "Learning difficulties of the Programming Language I course", of the second semester of 2019.

In Figure 1, the graphical representation of the results obtained with respect to the indicators that hinder the learning of object-oriented programming of the dimension "Teaching methodology" is shown.

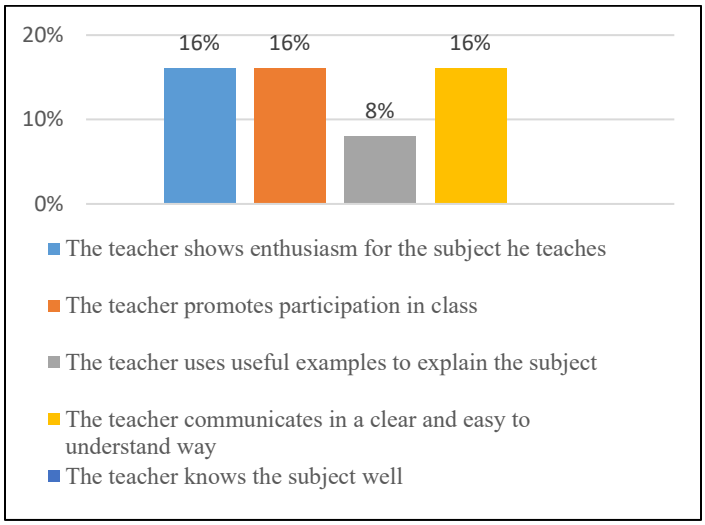


Figure 1: Higher indicators of the Teaching Methodology dimension that hinder the learning of OOP

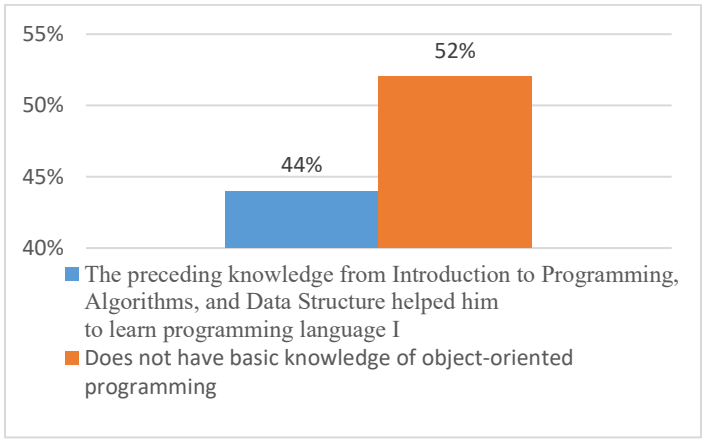


Figure 2: Indicator of the dimension Previous knowledge of programming courses that hinder the learning of OOP

As can be seen in the previous figure, the indicators that show the highest percentage are "The teacher shows enthusiasm for the subject he teaches", "The teacher promotes participation in class", "The teacher communicates in a clear and easy way understand "equivalent to 16% of the dimension" Teaching methodology "where students consider that it hinders learning OOP.

In Figure 2, the graphical representation of the result obtained with respect to the indicator that hinders the learning of object-oriented programming of the dimension "Previous knowledge of programming courses" is shown.

As can be seen in the previous figure, the indicator that shows the highest percentage is "Does not have basic knowledge of object-oriented programming", equivalent to 52% of the dimension "Previous knowledge of programming courses" where students consider it difficult to learn of the OOP.

Figure 3 shows the graphical representation of the results obtained with respect to the indicators that hinder the learning of object-oriented programming of the dimension "Reading materials".

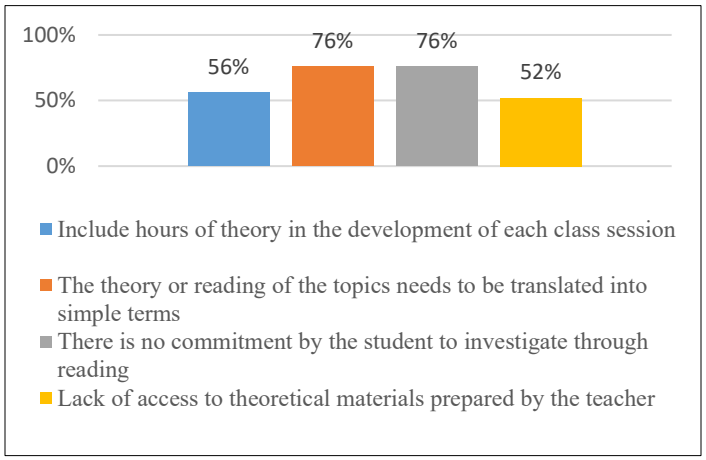


Figure 3: Higher indicators of the Reading Materials dimension that hinder the learning of OOP

As can be seen in the previous figure, the indicators that show the highest percentage are "The theory or reading of the topics needs to be translated in simple terms", "There is no commitment by the student to investigate through reading" equivalent to 76% of the dimension "Reading materials" where students consider that it hinders the learning of OOP.

Figure 4 shows the graphic representation of the results obtained with respect to the indicators that hinder the learning of object-oriented programming of the "Concept learning" dimension.

As can be seen in the previous figure, the indicator that shows the highest percentage is "Collections: Java Collection Framework, List and Array List, Map and Hash Map, Set and Hash Set, Iterator" equivalent to 60% of the dimension "Learning of concepts "where students consider it difficult to learn OOP.

Figure 5 shows the graphic representation of the results obtained with respect to the indicators that hinder the learning of object-oriented programming of the dimension "Process of learning to program".

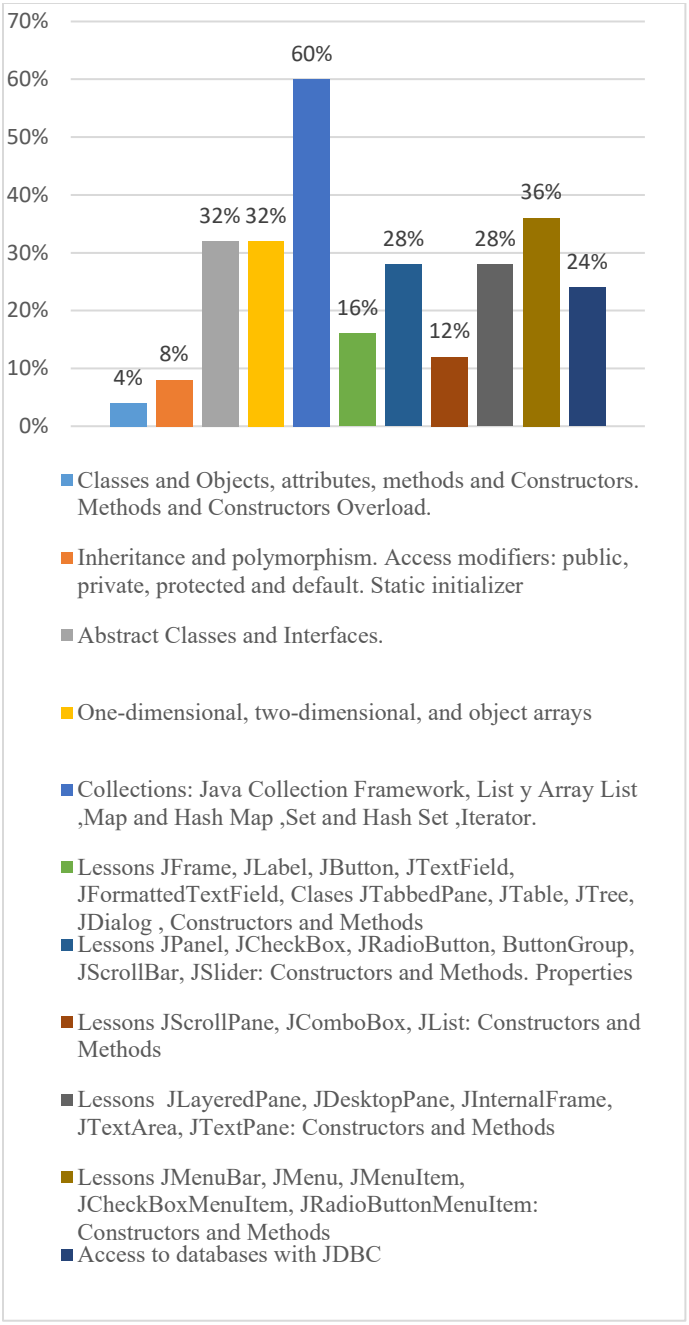


Figure 4: Highest indicator of the dimension "Learning concepts" that hinder the learning of OOP

As can be seen in the previous figure, the indicator that shows the highest percentage is “Develop programs using class libraries to create graphical user interfaces: AWT and Swing” equivalent to 48% of the dimension “Process of learning to program” where students consider it difficult to learn OOP.

To determine the dimension with the highest level of learning difficulty in OOP, we obtain the average for each dimension.

From the previous figure, we can determine that the highest percentage of difficulty associated with learning OOP is in the dimension "Learning concepts" equivalent to 8.20.

For the correlational analysis, we will classify the dimensions into two groups: Teaching Process Dimension and Program Development Dimension. The following table shows this classification.

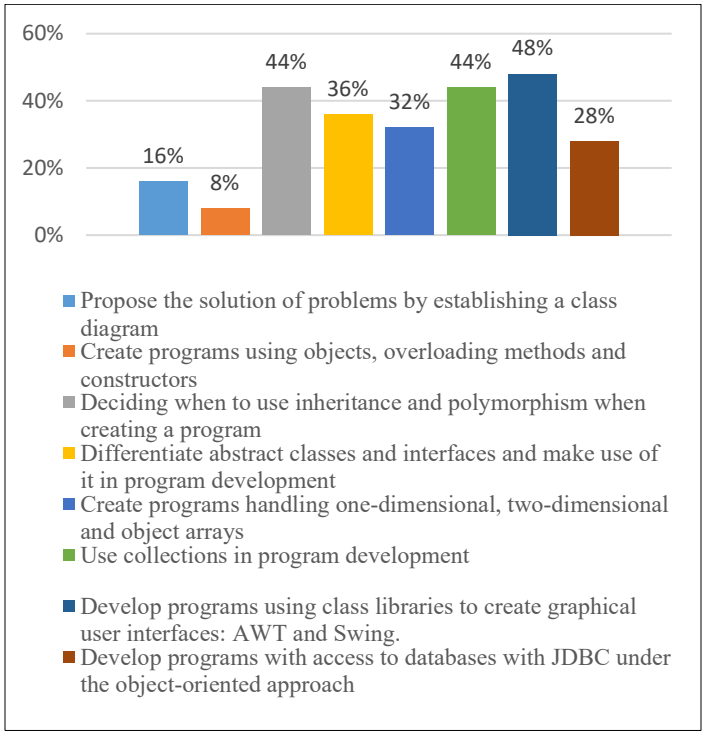


Figure 5: Highest indicator of the dimension “Process of learning to program” that hinder the learning of OOP

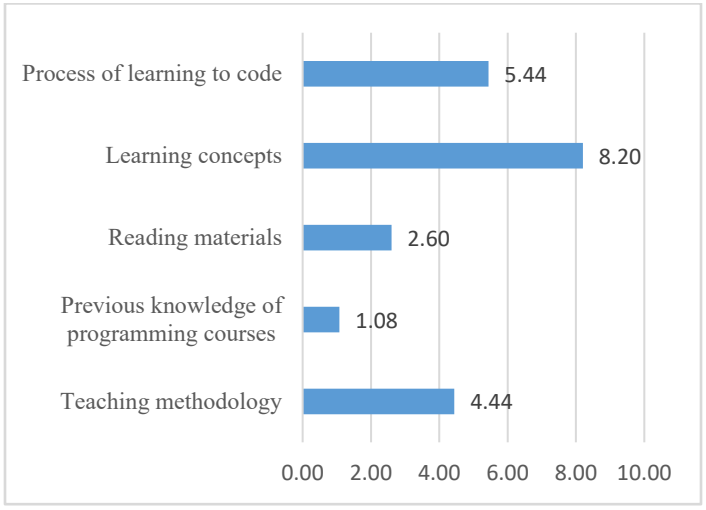


Figure 6: Average degree of difficulty for each dimension

Table 1: Classification of dimensions.

Teaching process	Program development
Teaching methodology	Process of learning to code
Reading materials	
Previous knowledge of programming courses	
Learning concepts	

In [16], the author points out to analyze the correlations, we have the interpretation of values expressed by various authors in scales, one of the most used being the Spearman rank correlation coefficient, which can score from -1.0 to +1.0 as detailed in table two

Table 2: Degree of relationship according to correlation coefficient.

Rank	Relationship
-0.91 a -1.00	Perfect negative correlation
-0.76 a -0.90	Very strong negative correlation
-0.51 a -0.75	Considerable negative correlation
-0.11 a -0.50	Mean negative correlation
-0.01 a -0.10	Weak negative correlation
0.00	There is no correlation
+0.01 a +0.10	Weak positive correlation
+0.11 a +0.50	Mean positive correlation
+0.51 a +0.75	Considerable positive correlation
+0.76 a +0.90	Very strong positive correlation
+0.91 a +1.00	Perfect positive correlation

In order to identify if there is a relationship or degree of association between the dimensions specified in Table 1, the Pearson Correlation coefficient is used, using the SPSS V25 statistical software.

Table 3: Pearson's correlation between the dimensions Learning concepts and Process of learning to program.

		Learning concepts	Process of learning to code
Learning concepts	Pearson correlation	1	0,659**
	Sig. (bilateral)		0.000
	N	25	25
Process of learning to code	Pearson correlation	0,659**	1
	Sig. (bilateral)	0.000	
	N	25	25
**. The correlation is significant at the 0.01 level (bilateral).			

In Table 3, we observe that the dimensions Learning concepts and Process of learning to program are directly related, that is, as there is more learning of the concepts of object-oriented programming with Java, there will be greater capacity in the process of learning to programming in Systems Engineering students, since a considerable positive correlation of 0.659 was obtained as a result.

Table 4: Pearson correlation between the dimensions Teaching methodology and Process of learning to program

		Process of learning to code	Teaching methodology
Process of learning to code	Pearson correlation	1	,235
	Sig. (bilateral)		,258
	N	25	25
Teaching methodology	Pearson correlation	,235	1
	Sig. (bilateral)	,258	
	N	25	25

In Table 4, we observe that the dimensions Teaching methodology and Process of learning to program present a mean positive correlation of 0.235.

Table 5: Pearson's correlation between the dimensions Reading materials and Process of learning to program.

		Process of learning to code	Reading Materials
Process of learning to code	Pearson correlation	1	-,014
	Sig. (bilateral)		,948
	N	25	25
Reading Materials	Pearson correlation	-,014	1
	Sig. (bilateral)	,948	
	N	25	25

In Table 5, we observe that the dimensions Reading materials and Process of learning to program are inversely related by presenting a mean negative correlation of -0.014, this implies that the number of reading materials on object-oriented programming can be increased, however , does not necessarily lead to increase their capacity of the process of learning to program in the students of Systems Engineering.

Table 6: Pearson correlation between the dimensions Previous knowledge of programming courses and Process of learning to program.

		Previous knowledge of programming courses	Process of learning to code
Previous knowledge of programming courses	Pearson correlation	1	0,036
	Sig. (bilateral)		0,863
	N	25	25
Process of learning to code	Pearson correlation	,036	1
	Sig. (bilateral)	0,863	
	N	25	25

In Table 6, we observe that the dimensions Previous knowledge of programming courses and Process of learning to program present a mean positive correlation of 0.036.

4. Discussion

In relation to the results of the present investigation, the following discussions are held below:

In [17], the author points out that starting in the world of object-oriented programming is a complex task for many students, the teacher plays an important role in student motivation, as evidenced by this research in its indicator "The teacher promotes participation in class ", where 16% of the students consider that it hinders their learning of OOP, it can be said that this result is consistent with the research entitled" Initiatives to motivate Programming students "in which it is concluded that motivation is low due to the novelty of the subject, the difficulty of students in learning abstract concepts and the lack of capacity for continuous work.

In [1], the author describes regarding the indicator “The theory or reading of the topics needs to be translated into simple terms”,

76% of the students consider that it hinders the learning of Object-Oriented Programming, coincides with the thesis doctoral degree entitled "Virtual environments based on active learning for the teaching of object orientation" in which he concludes that it is necessary for those who are starting in this paradigm to reach a clear understanding of its concepts, abstract in most cases, which represent in software development.

Regarding the indicator that students "do not have basic knowledge of object-oriented programming" before taking the Programming Language I course, 52% indicate that it makes learning difficult for them, this evidence agrees with the study entitled "Methodology of teaching by learning object-oriented programming logic "pointed out in [18] by the author, in which he concludes that there is a tendency to use Java as the first language directly with the object-oriented concept, without deepening the development of the bases logic of programming, this can be very harmful, because we are going to generate good programmers to code using languages, but without logical bases, that is, programmers who do not know how to program.

Regarding the indicator "Collections: Java Collection Framework, List and Array List, Map and Hash Map, Set and Hash Set, Iterator", 60% of the students consider a difficult topic to understand since it is related to classes, interfaces, inheritance, polymorphism for its understanding can be considered consistent since it coincides with the study entitled "Learning object-oriented programming: Experiences in technical and upper secondary education" indicated in [19] by the author, in which it is concluded that the subject of greater difficulty of understanding for the control groups is inheritance, while for experimental groups it is polymorphism.

5. Conclusions

It has been possible to determine the main difficulties associated with the learning of object-oriented programming (Programming Language I) in the students of the School of Systems Engineering of the National Technological University of Lima Sur, which are "The teacher shows enthusiasm for the subject he teaches", "The teacher promotes participation in class", "The teacher communicates in a clear and easy to understand way" with 16% indicators related to the teaching methodology". The indicators "Lack of translating the theory or reading of the topics into simple terms", "There is no commitment by the student to investigate through reading" with 76% related to the reading materials. The indicators "Does not have basic knowledge of object-oriented programming" with 52% related to Have prior knowledge. The indicator "Collections: Java Collection Framework, List and Array List, Map and Hash Map, Set and Hash Set, Iterator" with 60% in relation to Learning concepts. The indicator "Develop programs using class libraries to create graphical user interfaces: AWT and Swing" with 48% related to the Learn to program process.

It has been possible to determine the dimension that has the greatest learning difficulty of Object-Oriented Programming, which is "Learning concepts" with an average of 8.20, we can mention that this expresses that currently the curricular plan related to programming courses at school Systems Engineering do not have hours of theory to help manage fundamental concepts of Object Oriented Programming.

6. Recommendations

While developing the Programming Language I (OOP) course in its entirety in a practical way using the Java language, it is recommended to develop manuals or self-instructional videos that help the student understand concepts of object-oriented programming.

Since the curricular plans at the Universidad Nacional Tecnologicas de Lima Sur are updated every three years, it is recommended to include hours of theory in programming courses or, failing that, develop intelligent applications that help understand object-oriented programming concepts.

References

[1] G. Jiménez, Virtual environments based on active learning techniques for teaching object-orientation, doctoral thesis, Complutense University of Madrid, Spain, 2018.

[2] I. Friss de Kereki, What is object-oriented programming? A phenomenographic approach. Educational Research Notebooks, 1998.

[3] D. Gayo, A. Cernuda del Rio, J. Cueva, M. Diaz, M. Garcia, J. Redondo, Reflections and experiences on the teaching of OOP as the only paradigm, University of Oviedo, Spain, 2003.

[4] J. Rios, Object-oriented programming in non-object-oriented languages: C, An Experience. Scientia Et Technica, **21**(29), 107-111. Technological University of Pereira, 2005.

[5] S. Rueda, I. Señas, A problem-solving approach to teaching OOP. National University of the South - Bahía Blanca, 2008.

[6] H. Flórez, Object-oriented programming using Java. Bogotá: Ecoe Ediciones, 2012.

[7] J. Insuasti, Teaching and learning problems of programming fundamentals. Education and Social Development Magazine, **10**(2), 234-246, 2016. DOI: org / 10/18359 / reds.1701.

[8] J. Sánchez-García, M. Urías-Ruiz, B. Gutiérrez-Herrera, Analysis of the Learning Problems of Object-Oriented Programming, Scientific article in Ra Ximhai, **11**(4): 289-304, 2015.

[9] J. Angulo & O. Palomar, LM1: a study methodology for the subject "Programming 1". Educere: Venezuelan Journal of Education, **73**, 635-648, 2018.

[10] J. Fuentes-Rosado & M. Moo-Medina, Difficulties of learning to program. Engineering education magazine, **12**(24), 76-82, 2017. https://doi.org/10.1093/ptj/79.4.371

[11] M. Sánchez, University dropout in students from a private university in Iquitos. Digital Journal of Research in University Teaching, 60-83, 2012.

[12] R. Carvajal & C. Cervantes. Approaches to university dropout in Chile. Educação e Pesquisa, 44, 2018.

[13] R. Hernández, C. Fernández, M. Baptista, Research Methodology, Mexico: Mcgraw-Hill / Interamericana editores, 2010.

[14] D. Arce, O. Chamorro, Y. Leandro, J. Castro, E. Martínez, The Proportionality of Women Graduated from the Professional Career of Mechanical and Electrical Engineering at UNTELS: Analysis of their Academic Performance and Labor Field of Action . Advances in Science, Technology and Engineering Systems Journal, **5**(1), 368-372, 2020. https://dx.doi.org/10.25046/aj0501273

[15] O. Chamorro, T. Diaz, D. Arce, J. Castro, D. Barrientos, A. Roman-Gonzalez, Satisfaction of the graduate for the continuous improvement of educational quality in untels. Advances in Science, Technology and Engineering Systems Journal, **4**(5), 151-157, 2019. https://dx.doi.org/10.25046/aj040503

[16] M. Mondragón, Use of Spearman's correlation in an intervention study in physiotherapy. Mov.cient.**8**(1): 98-104, 2014. https://doi.org/10.1093/ptj/79.4.371

[17] M. Salamó, J. Camps, C. Vallespí, D. Vernet, X. Llorà, E. Bernadó & X. González, Initiatives to motivate Programming students, Proceedings of the VII Conference on University Teaching of Informatics (JENUI 2001), 329, 2001.

[18] L. Román, Methodology for teaching and learning the logic of object-oriented programming.

[19] R. Tabares & H. Gómez, Learning object-oriented programming: Experiences in technical and higher secondary education.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1710-1718 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Efficient and Scalable Ant Colony Optimization based WSN Routing Protocol for IoT

Afsah Sharmin*, Farhat Anwar, S M A Motakabber

Faculty of Engineering, Department of Electrical and Computer Engineering, International Islamic University Malaysia, Kuala Lumpur, 53100, Malaysia

ARTICLE INFO

Article history:
Received: 31 August, 2020
Accepted: 21 December, 2020
Online: 28 December, 2020

Keywords:
Wireless Sensor Networks
IoT
Routing algorithms
ACO
Energy consumption

ABSTRACT

IoT integrates and connects intelligent devices or objects with varied architectures and resources. The number of IoT devices is growing exponentially. Due to the massive wave of IoT objects, their diversity and heterogeneity among their architectures, the existing communication protocols for wireless networks become ineffective in the context of IoT. Wireless Sensor Network (WSN) has the potential to be integrated to the internet of things (IoT). The issues of the routing of WSNs impose nearly similar prerequisites for IoT routing technique. Most of the traditional routing protocols are not appropriate for WSNs and IoT because of resource constraints, computational overhead and environmental interference and do not take into account the different factors affecting energy parameter and do not accommodate node mobility. Routing algorithms must ensure the data transmission in an efficient way, having proper knowledge of the IoT system. For this reason, many intelligent systems have been utilized to design routing algorithms to handle the network's dynamic state. In this paper, an ant colony optimization (ACO) based WSN routing algorithm for IoT has been proposed and analyzed to enhance scalability, to accommodate node mobility and to minimize initialization delay for time critical applications in the context of IoT to find the optimal path of data transmission, improvising efficient IoT communications. The proposed routing algorithm is simulated using MATLAB for performance evaluations. The evaluation results have recorded an improvement in conservation of energy, of almost 50% less consumed energy even with an increase in the number of nodes, by comparing with an existing routing technique based on ant system, a current routing protocol for IoT and the conventional ACO algorithm.

1. Introduction

With the widespread use of IoT devices, the issue of designing efficient routing protocol has attracted more attention in networking research. This paper is an extension of work originally presented in ICOM'19 [1].

According to the prediction of Cisco™ IBSG [2], by 2020, 50 billion devices will be connected to the Internet. IoT provides network connectivity between these smart devices everywhere and all times. The emergence of a new ubiquitous computing era has been created due to the evolution of wireless networks and sensor technologies, allied to the increasing demand for new IoT applications for the provision of smart services [3]. In this context, WSNs play an essential role to the expansion of IoT while providing ubiquity of networks with smart and low-cost devices that are easy to deploy. In an IoT system, a large collection of autonomous and dynamic sensor nodes are used to gather information by detecting physical parameters, communicate and cooperate with their environment and send their data to the internet. Sensor nodes have the ability of self-organization and sensor networks function in a distributed way [4]. IoT with the integration of WSNs, have a wide range of applications spaces that shape human life and also have impact on economic benefits. The physical domain of IoT is presented through the connected networks of objects and nodes utilizing wireless sensors. A large number of small devices surround the environment help to manage the physical world by sensing, processing, communicating and analyzing the data in the IoT network system [5]. The development of new applications and technologies are drawing attention from the research perspective, such as smart home, environmental monitoring, healthcare, transport, agriculture, offices, buildings and smart cities etc.

*Corresponding Author: Afsah Sharmin, Faculty of Engineering, Department of Electrical and Computer Engineering, International Islamic University Malaysia, 53100 Kuala Lumpur, Malaysia, Tel: +60167101538, afsahsharmin@gmail.com

Designing efficient routing protocol is the key factor to improve the energy efficiency, data transmissions, scalability and prolong network lifetime in WSNs and IoT. However, several considerations are required for resource-constrained network system, such as energy efficiency, scalability, security, autonomy, computational complexity, environmental constraints for wireless link, node mobility, the QoS (quality of service) requirement for a particular application, during IoT routing. The sensed information ought to be sent to the base station for further operations in various IoT applications through the competent forwarding mechanism while keeping in consideration different functionalities of IoT objects states.

Until now, there is a large number of routing algorithms have been proposed by the research community, considering the energy parameter, yet its exploitation is not well thought-out [6]. The existing node energy and the distinctive components affecting this energy parameter need to be considered to spare node energy and enhance the communication quality of the network. The node's lifetime relies upon the battery-life to a huge range, and the irrational energy utilization will effectuate the system to expire early and decrease the lifetime of the network [7]. Hence the key research issue is the designing of efficient route calculation algorithm that can ensure efficient information communication within IoT while maintaining scalability. Most of the traditional routing protocols are not adaptable other than energy efficient, if the varied difficulties in various applications are considered or due to the dense and complex conditions and a wide range of radio obstruction. Many intelligent systems including the working mechanism of the biological systems have been employed for designing routing algorithms to handle the network's dynamic state while keeping pace with energy efficiency during information communication [8]. Ant colony optimization (ACO) based algorithms emphasize the design of routing protocols that are robust, adaptable and scalable [9]. The coordination of ants depends on the ability of self-organization that ant colony optimization based swarm intelligence techniques possess [10]. The probabilistic approach of ACO is used to determine the routes, and the pheromone update formulation is used for further updating of the pheromone trail [11].

An ant colony based routing technique, named EICAntS, for effective communications within IoT is presented in [12]. This algorithm considers the energetic parameters. The pheromone estimations in the ant system is related here by the calculated global efficacy factor. An improvement with regard to network lifetime and conservation of energy is shown from the evaluation results. The algorithm does not include the heuristic information and pheromone update strategy consisting of pheromone evaporation rate and/or the amount of pheromone deposited. The energy effect precisely addresses the data class handled by the node and does not specify the different factors affecting energy parameter such as the energy consumption in free-space and multi-path fading standard of wireless communication as well as no particulars are provided on how to compute the nodes' energy level. The REL routing technique for IoT applications, which is focused on energy and link quality information, is proposed in [13]. Testbed experiments are used besides simulation for the evaluation where REL increases the system and network lifetime, reliability, energy-efficiency, QoS of IoT applications and reduces packet loss rate. In addition, it offers a path determination system which is basically an end to end route. For this purpose, it depends on cross-layer data with insignificant overhead. Furthermore, it permits information transmission with a reasonable appropriation of wireless and remote assets. A piggyback and on-demand system guides nodes to become energy proficient where the residual energy is sent to their neighboring nodes. Nonetheless, no enhancement system is utilized in contradiction of ant based protocol, which utilizes the ant colony system.

In [14], LEACH-MA protocol is proposed which is based on modified ACO. The residual energy parameter is used along with LEACH (Low-energy adaptive cluster hierarchy routing) protocol for the selection of current cluster head. The measure of energy consumption is minimized by this algorithm. The energy and distances are combined in this approach for choosing the cluster head. But no facts have been provided on the threshold value for selecting the cluster head. In [15], a hybrid tree-based search approach, called ANT-BFS, is proposed to discover the optimum route for information communication. This scheme integrates breadth first search with ACO to minimize the amount of energy consumption. However, the memory and computational time requirements may arise some issues for this technique in case the cluster head and the sink are located far from each other, so it might not be suitable for large scale. Content based routing (CCR) protocol is presented in [16] that ensures reliability of information transmission for IoT applications. It utilizes the process of data aggregation and ensures good load balancing. This method transmits data based on the message content and incorporates link quality information. The traffic reduction gain is achieved by this technique, where an objective function is used by the nodes for routing of the heterogeneous sorts of content. Each node builds a distinct routing entry by using the content to select the next node for transferring the data via the certain reliable communication link. The energy consumption has been conserved by this method by forwarding aggregated data to selective nodes. However, no details were provided on the estimation of the reliability while taking into consideration of the parameters that are also unspecified for this reliability.

In [17], an ant-based routing algorithm considering the energy supervision is presented. Reward and punishment technique are adopted with the pheromone update rules. The energetic parameter is considered and this technique increases network lifetime and energy-efficiency. However, various factors of energy utilization are required to be addressed. The node delay parameter is used here, but no details were found on the delay factor. An improvised energy saving ant colony based routing protocol for sensor networks is proposed in [18]. Three phases are introduced in this multipath protocol, such as discovery of neighbor through link information, transmission of packets through EWMA (exponentially weighted moving average) technique and efficient & reliable end to end delivery. The simulation results indicate that the proposed routing protocol for dynamic networks stands efficient in general performance, particularly regarding energy efficiency and throughput when contrasted with standard and other novel proposed routing algorithms. A multi-constrained technique that ensures QoS, known as IAMQER focused on ant colony, is proposed in [19]. The simulation results show that average energy consumption is minimized and packet delivery ratio is increased by utilizing this technique. Also, a route evaluating function is presented here. But IAMQER strategy might suffer longer processing delay as it is based on traditional ACO method.

This paper is an extension of work initially reported in [20], where the necessary parameters for communication process have been taken into account by the proposed method, such as mobility and energy parameters. Most of the proposed routing algorithms have addressed merely some communication parameters that are not adequate enough to enhance the communication quality needed by the energy constrained IoT applications. The important communication parameters must be considered by the mechanisms that mediate in the correspondence procedure, for instance routing, for resolving the issues of routing and for efficient communications in the IoT network system. The proposed protocol has utilized the advantages of the ant system to discover the optimum path for information communication and to improve communication quality within an IoT system.

2. System Model

2.1. The Network Model

The considered network model consists of *M* sensor nodes with random distribution in an $L \times L$ rectangular area. The proposed algorithm comprises the system model for IoT communications which is designed with several nodes or sensors M_i. The adopted graph G with the nodes and connecting links has constructed the network. It is assumed that the sensor nodes have the same computation and preliminary energy. The nodes can update the information about its neighbors. The received signal strength indication (RSSI) is used to compute the approximate distance of the senders by the nodes. For that calculation, the transmitted power of the objective is acknowledged.

2.2. Proposed ACO Algorithm

The most pheromone path is chosen as the shortest and optimal path in the traditional ACO algorithm [11] and the energy parameter is not considered. Hence the network's node energy on that path reduces abruptly and lessens the entire network's lifespan. Furthermore, novel routing protocols are required to manage the overhead of mobility. As the topological changes for the mobility of the sensor nodes and the sink nodes generate frequent updates in the network, which may drain the node energy extremely in an energy-constrained system. This paper expands our research works presented in [1] in terms of network performance to improve communications within IoT. The proposed system of ours [21] has been investigated more here to enhance scalability, to accommodate node mobility, and to minimize initialization and processing delay for time critical applications in the context of IoT. An improved network routing algorithm based on ACO is proposed here by analyzing the nodes' balanced consumption of energy.

At the place of node *i* and time *t*, each ant *m* will adhere to the following probabilistic formula to select the next node *j* being the forwarding node of the subsequent route for the enhancement that is proposed of the ACO algorithm of ours [20] for the next hop routing:

$$P_{ij}^{m} = \begin{cases} \dfrac{[\tau_{ij}(t)]^{\alpha}[\eta_{ij}(t)]^{\beta}[\vartheta_{ij}(t)]^{\gamma}E_j}{\sum_{s\subset allowed_m}[\tau_{is}(t)]^{\alpha}[\eta_{is}(t)]^{\beta}[\vartheta_{is}(t)]^{\gamma}E_s}, & j \subset allowed_m \\ 0, \quad others \end{cases} \quad (1)$$

here,
$$\eta_{ij}(t) = \frac{1}{d_{ij}} \quad (2)$$

where $\tau_{ij}(t)$ and $\eta_{ij}(t)$ are the amount of pheromone and heuristic information on edge (*i, j*) respectively and $\eta_{ij}(t)$ is typically $1/d_{ij}$. α and β are two parameters that control the influence of the pheromone intensity and heuristic information respectively. d_{ij} is the distance between *i* and *j*. The average mobility parameter is used to calculate the stability factor, $\vartheta_{ij}(t)$, where γ is the mobility constant. E_j is the node residual energy that ant *m* will visit.

To avoid faster local convergence in case large amount of pheromone deposition happens on the routes, the pheromone update is required to improve as well. So, the amount of pheromone is updated and limited by incorporating a threshold value and can be obtained as follows:

$$\tau_{ij}(t+1) = \begin{cases} T, & \tau_{ij}(t+1) > T \\ (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij}(t) & else \end{cases} \quad (3)$$

$$\Delta\tau_{ij}(t) = \sum_{k=1}^{m} \Delta\tau_{ij}^{k} \quad (4)$$

where ρ is the pheromone coefficient for evaporation, $\rho \in (0,1)$, T represents the threshold value used to limit excessive pheromone deposition, $\Delta\tau_{ij}(t)$ is the increased pheromone concentration of edge (*i, j*), usually given by,

$$\Delta\tau_{ij}^{k} = \begin{cases} \dfrac{R}{L_k} & if\ kth\ ant\ uses\ the\ edge\ (i,j) \\ 0 & otherwise \end{cases} \quad (5)$$

where *R* represents strength of pheromone, L_k denotes the path length of the kth ant.

2.3. Energy Model

The energy model of wireless communications that is presented in [22] is incorporated by the proposed system. The free-space and multi-path fading model are utilized here depending upon the distance *d* in between the sending and receiving nodes and a threshold value, d_0. The following equations for energy consumption for the transmission (E_{tran}) of an *S*-bit data by the sensors are used:

$$E_{tran}(S,d) = \begin{cases} SE_{elec} + S\epsilon_{fs}d^2 & if\ d < d_0 \\ SE_{elec} + S\epsilon_{mp}d^4 & if\ d \geq d_0 \end{cases} \quad (6)$$

where E_{elec} indicates the energy dissipated to run electronic devise circuitry. The energy consumption in free-space and multi-path fading model are given by ϵ_{fs} and ϵ_{mp} respectively. *d* denotes distance and d_0 is the threshold value. The receiving energy, $E_{Rx}(S)$, for an *S*-bit data for a node is provided as follows:

$$E_{Rx}(S) = SE_{elec} \quad (7)$$

To calculate the residual energy of a node n_i, the following equation is used:

$$E_{res_i} = E_{tot_i} - E_{tran_i} \quad (8)$$

where E_{res_i} is the residual energy, E_{tot_i} is the total initial energy and E_{tran_i} is the transmission energy.

2.4. Fitness Function

A function for the evaluation of the routes, path assessing index, is provided here considering the current energy of the nodes, and the path of routing. If remaining energy is not assessed as in the conventional ACO algorithm, which causes early demise of some nodes and ultimately affects the whole network lifetime. The routing path is related to each particular ant after all the ants get to the destination node. The fitness value for the path can be computed as follows:

$$f_{(fitness)}{}_m^k = \frac{E_{res_i}}{L_m^k} \quad (9)$$

where the residual energy level of a sensor node n_i is E_{res_i}. L_m^k is the length of the route for m^{th} ant and k^{th} iteration. The pheromone is updated on the optimum path, having the highest fitness value.

2.5. Planning of Route Phases

Step 1. The route's arrangement phase. At first an initialization signaling is broadcasted by the sink node. Each node acquires its own neighboring node, then updates and adds it to its own routing table. The adjacent node link pheromone value is set to 1. The current node's ant packet is generated by each node that contains the number of nodes and the routing table. N_{max} is set to the maximum number of iterations, and the initial number of iterations is set to 1.

Step 2. The route's organization phase. The next node will be selected by the ant that locates in node *i* according to (1)-(2). Upon receiving the ant package, a node forwards ant package in accordance with the probability P_{ij}^m .

Step 3. The route's optimization phase. Once all ants get to the destination node, i.e. the sink node, the fitness function is used for optimal route selection where the fitness value is computed for the route according to (9) and the route with highest fitness value is chosen for optimal route for data transmission. The concentration of pheromones on this path is then updated according to (3)-(5).

In this route's optimization phase, once the ant package is received by the sink node, it will count how many ants packets each node sends. Let's assume that the total node number is *n* and the number of ants package sent by, for instance, node *i* is X_i (i =1,……., n).

Then the total sum of the network ant package can be expressed by means of:

$$T = \sum_{i=1}^{n} X_i \quad (10)$$

Back ant package is generated by each node. When back ant package is received by the nodes, the back ant package's adjacency linkage information is updated by (3)-(5). The node is selected as the next hop which resides in the back ant package. Following the information in the back ant packets, it is determined whether to send a new packet of ants at that same instance. When new ant package need to be sent, the route is established.

2.6. Proposed Improvement

Forming clusters associated with the group of nodes is deliberated on achieving scalability and robustness. For this reason, the routing protocol presented in [20] can be integrated with a clustering routing technique, such as LEACH (Low-energy adaptive clustering hierarchy) protocol [22]. It improves the selection strategy of optimal cluster head (CH), which is based on probability, node residual energy, and the distance of a node from the base station (BS). For selecting the cluster head node *j*, if it is assumed that node *i* is the current cluster head and next node is *j*, an ant *m* will use the probability calculation given by:

$$Prob_i(\mathrm{t}) = \frac{dist_i{}^{*\alpha} + [P_{ij}^m(t)]^{*\beta}}{\sum_1^{N_i} dist_i{}^{*\alpha} + [P_{ij}^m(t)]^{*\beta}} \quad (11)$$

Where $Prob_i$ provides the probability of each node to be selected as a CH (cluster head), $dist_i$ is the distance of node, α and β are two control parameters and N_i denotes the set of cluster nodes. P_{ij}^m is found from (1). Figure 1 shows the flow diagram of this suggested enhancement.

Using the proposed ACO based routing algorithm and the proposed improvement given above the initial optimal CH (cluster head) is selected. Then the final optimal CH is elected using the maximum fitness function value provided in (9). The optimal initial cluster head sends data to the optimal final cluster head. The optimal final cluster head transmitted data to the sink node.

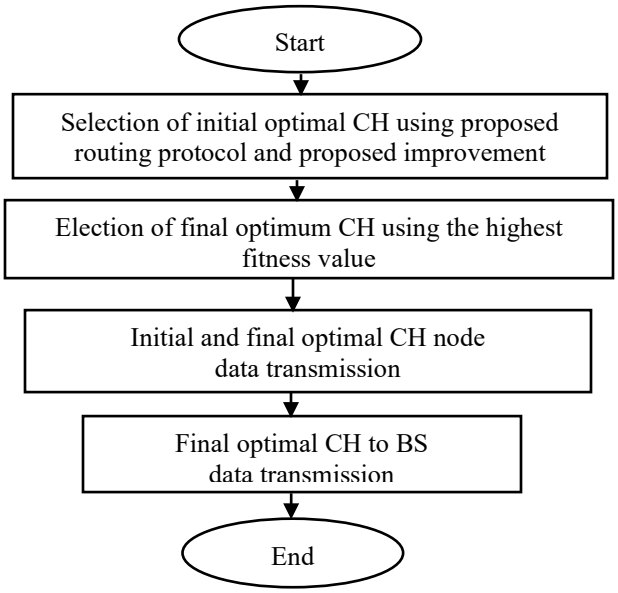


Figure 1: The flow diagram of the proposed improvement

Figure 2 shows the flow diagram of the proposed routing protocol operation.

3. Result and Discussion

The simulation is performed utilizing MATLAB where the proposed routing protocol is evaluated according to some important parameters, i.e., the consumed energy, the lifetimes of nodes, best cost, end-to-end delay, throughput and packet delivery

rate. The method represented here is evaluated against the conventional ACO, a current routing protocol for IoT, e.g., RPL (routing protocol for low power lossy network) [23] and an existing ant colony based algorithm, EICAntS algorithm [12] as benchmark protocols.

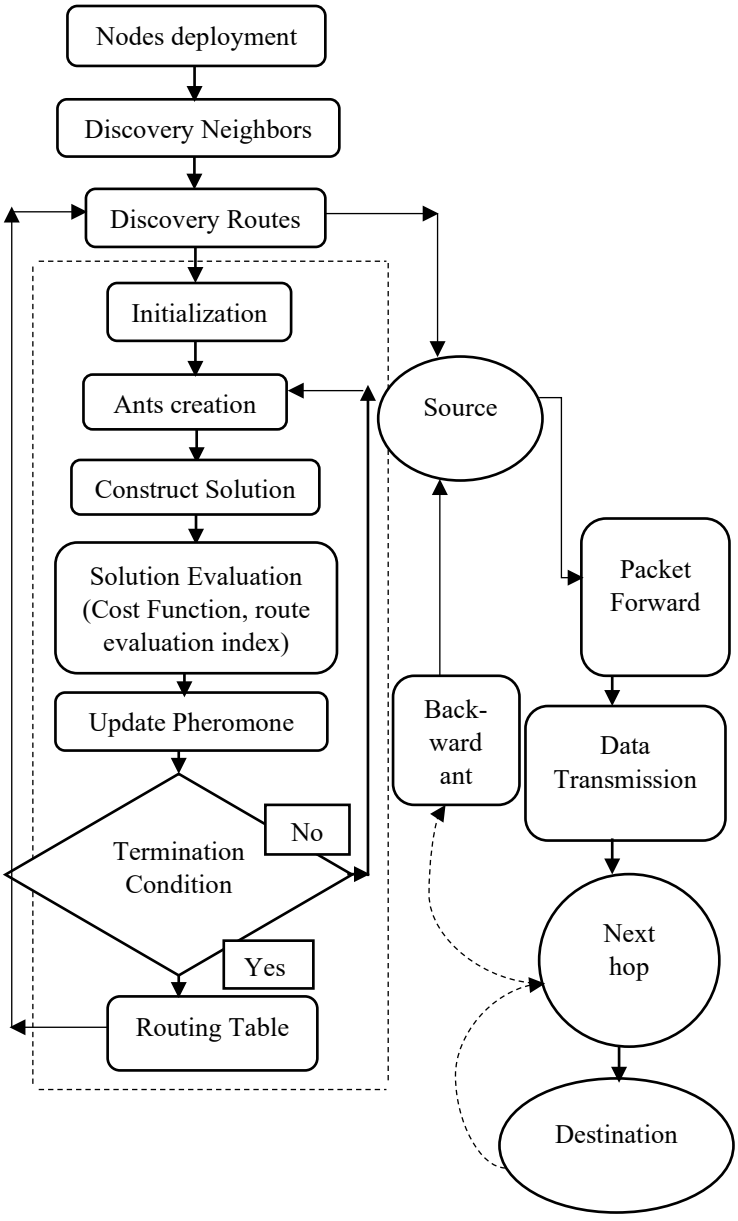


Figure 2: Overview Block Diagram demonstrating the Proposed ACO based Routing Algorithm Operation

The simulation parameters are based on the repeated tests and on the basis of simulation analysis of various parameters and set as: $\alpha = 1$, $\beta = 1$, $\gamma = 1$, $\rho = 0.05$. More parameters are provided in the Table 1.

The proposed system is supporting a mechanism permitting the selection of the best cost route for information communication from the source node to the destination, the base station. The network's parameters are dynamically adjusted to the network evolution, the growing number of nodes, the changed environment issues due to mobility, and outstanding node energy.

Table 1: Simulation Parameters

Parameters	Values
Simulation Region	100m×100m
Number of Nodes	40,50,60,70,80,90,100
T	100
No. of ants	40
Initial energy per node	0.5 joule
Node speed	2 m/s ~ 5 m/s
Message bits transmitted	4000 bits
Transmission distance	50 m

In Figure 3, for the proposed ACO based routing protocol and for the traditional ACO based protocol, the comparison of the node energy consumption per transmission is shown. The considerable improvement is shown for the proposed method. For majority of the nodes, contrasted with the conventional protocol, the proposed method has smaller energy consumption, almost 50% less. The proposed calculation has presented noticeable refinement for the total energy consumed of each node for individual discovery for the routes, contrasted with the benchmark protocol.

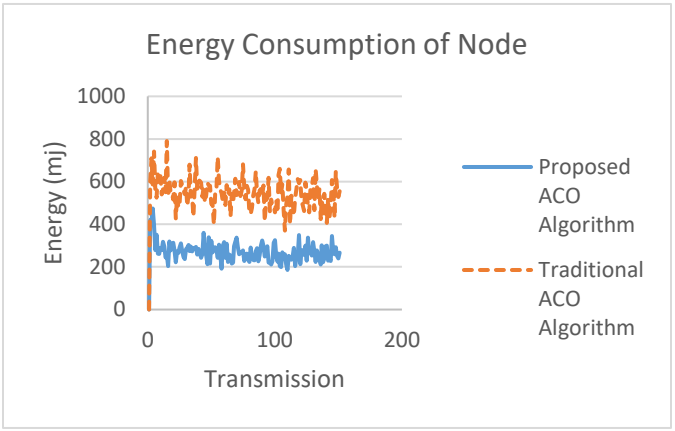


Figure 3: Comparison of node energy consumption per transmission

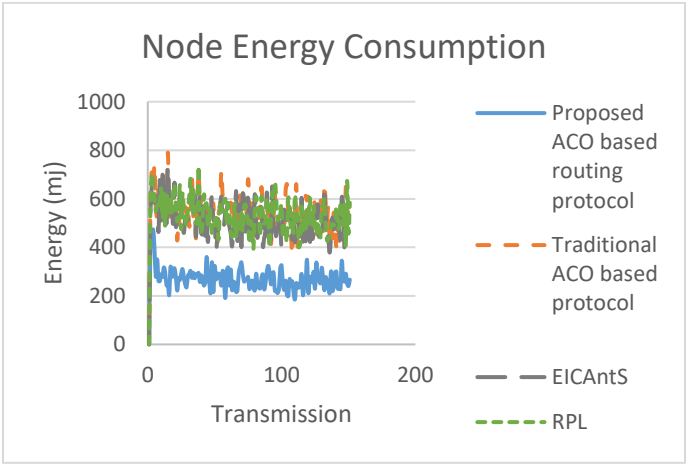


Figure 4: Comparison of node energy consumption per transmission

From Figure 4, we can see the node energy consumption per transmission based on improved method proposed here and other benchmark protocols, based on original ACO, RPL and EICAntS protocol. It is demonstrated from the evaluation results that

compared with the other two algorithms, the proposed ACO algorithm has attained a refinement, a much smaller energy consumption, almost 50% less.

The average energy consumption for the increasing quantity of nodes is represented in Figure 5 and Table 2. Figure 5 shows that the system using the proposed ACO based routing protocol as a resolution has minimized the average energy consumption, almost 50% less, by comparing with the original ant colony-based system, RPL and EICAntS technique. Since the proposed solution allows the determination of the optimal path for packet transmissions, and the retransmissions are avoided as well. Also, this presented method optimizes path selection strategy and minimizes updating phases. The scalability is maintained by the proposed system that has attained lower energy utilization contrasted with the benchmark protocols even when the number of nodes increases in the network.

Table 2: Average Energy Consumption (mj) with the Number of Nodes for 150 transmissions

No. of Nodes	Proposed ACO based routing protocol	Traditional ACO based protocol	EICAntS	RPL
40	270.49	545.89	516.65	532.65
50	281.67	570.07	566.42	574.78
60	303.76	620.61	606.97	632.29
70	330.97	650.51	660.42	679.46
80	350.65	686.47	695.87	692.01
90	372.42	739.92	744.55	735.55
100	401.21	791.33	799.01	768.49

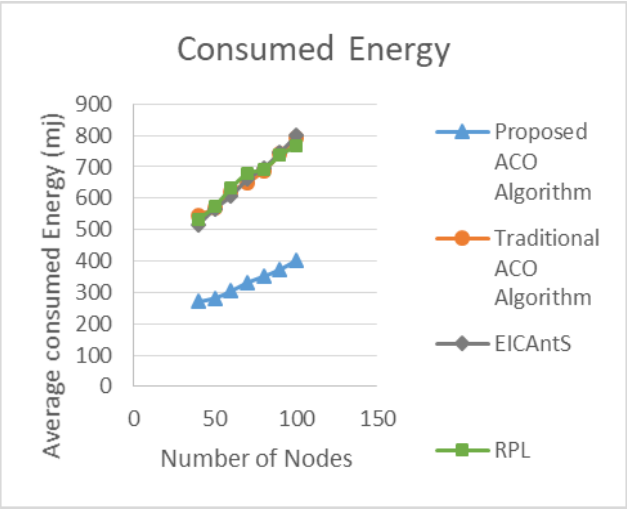


Figure 5: Comparison of the average energy consumption

The effective use of energy lead to increase of network lifetime. The percentage of survival nodes based on improved protocol proposed here and the traditional ACO based protocol are shown in Figure 6 and 7. The issues of ineffectiveness and uncertainty of the routes are alleviated as the protocol proposed here takes into account the energetic parameters. The energy utilization needs to be balanced and the present node energy needs to be deliberated; otherwise, the node energy is depleted and some nodes expire. The network lifetime is affected by this. The proposed approach deliberates the existing energy of the nodes besides the route path. Employing the suggested enhancement according to (11), the longevity, i.e. the nodes' lifetime and lifetime of the network, is increased by the offered ACO based technique.

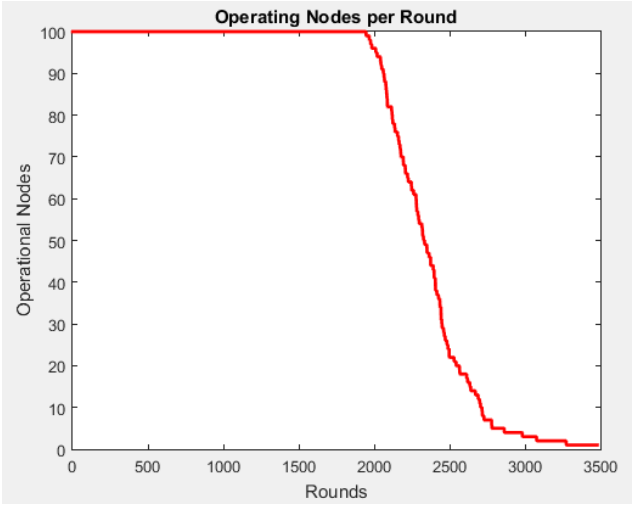


Figure 6: The percentage of survival nodes for the proposed method

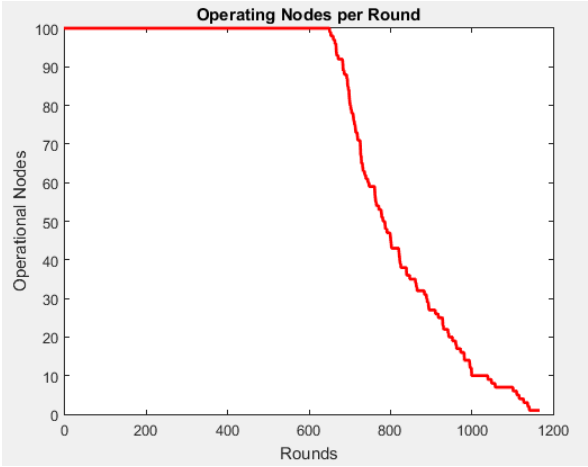


Figure 7: The percentage of survival nodes for the traditional ACO based protocol

We can see from the Table 3 and in the Figure 8 that the average end-to-end delay increases with an increase in the number of nodes. Route discovery and path selection stages affect this parameter, which are instated all the more regularly with an imperative number of nodes and with the topology evolution because of mobility of the nodes. Using the proposed algorithm the route discovery and selection stages are improved and the repetition problem is minimized when many packets should be resent to the destination in case the route might not be the best

always. The proposed system achieved better results with regard to average end to end delay, almost 40% reduced end to end delay, compared to the other benchmark techniques.

Table 3: Average End-to-End Delay (ms) with the Number of Nodes using Fitness Function

No. of Nodes	Proposed ACO based routing protocol	Traditional ACO based protocol	EICAntS	RPL
40	159.78	367.41	364.42	370.67
50	203.59	397.57	394.64	382.81
60	250.81	413.22	412.96	410.75
70	263.37	427.28	414.09	425.13
80	299.25	464.68	460.82	455.42
90	312.43	478.52	465.4	466.35
100	327.46	485.86	479.72	488.96

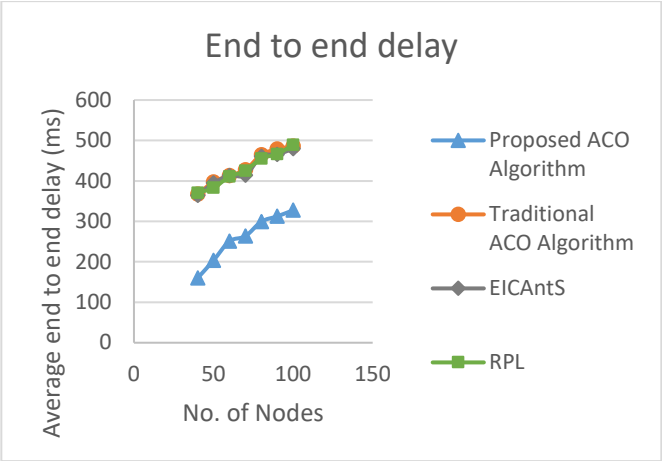


Figure 8: Evolution of average end to end delay with the number of nodes using fitness function

Throughput results are presented in Table 4 and depicted in Figure 9. The quantity of digital data, over a physical or logical connection per time unit, is the measure of the throughput in sensor networks. It is measured in bits/s or bps (bits per second), occasionally in per-second data packets or in per time-slot data packets. It can be defined mathematically, as the total number of packets delivered over the total simulation time:

$$\text{Throughput} = N / \text{Total simulation time} \quad (12)$$

where N is the number of bits received successfully by all destinations.

We can clearly see from the contrast that the results obtained are better than the results recorded by the other network.

Table 4: Throughput with the Number of Nodes over the Total Simulation Time

No. of Nodes	Simulation Time (Proposed Algorithm)	Simulation Time (EICAntS)	Proposed Algorithm	EICAntS
40	42 sec.	42 sec.	3809.52 (bps)	975.24 (bps)
50	45 sec.	45 sec.	4445.45 (bps)	1137.78 (bps)
60	48 sec.	49 sec.	5000 (bps)	1253.88 (bps)
70	51 sec.	51 sec.	5490.196 (bps)	1405.49 (bps)
80	54 sec.	53 sec.	5925.93 (bps)	1545.66 (bps)
90	55 sec.	55 sec.	6545.45 (bps)	1675.63 (bps)
100	57 sec.	57 sec.	7017.54 (bps)	1796.49 (bps)

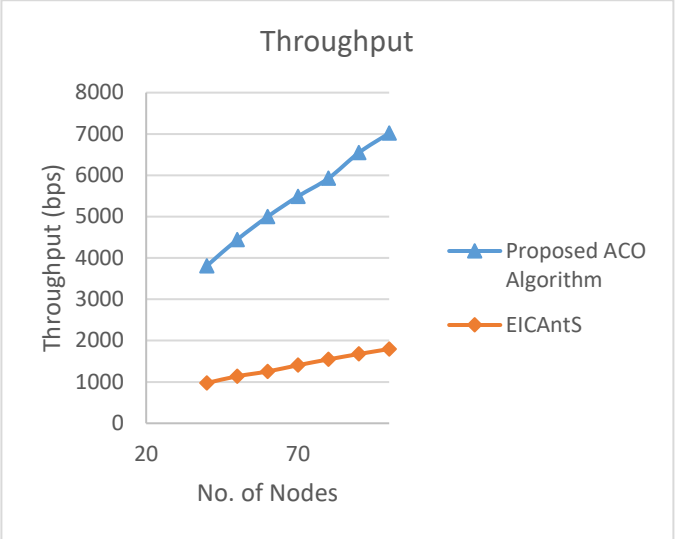


Figure 9: Throughput for the Proposed ACO and EICAntS algorithm

The calculation of the proposed algorithm in accordance with the delivery rates of the packets is provided in the Figure 10. The good results found with the proposed approach, even with an imperative number of nodes because the system has the capability that let many packets to send to get to the destination node in short time.

We can see from the comparison of the best cost as presented in Figure 11 that the network using proposed protocol has minimized initialization delay compared to the benchmark protocols. The network system achieved faster convergence for the initialization delay to discover the best cost route, which is about 30% improvement in convergence. Current node energy, mobility, route path, and cost function including route assessment index are

considered by the proposed method, making it preferable as a solution for the network system to improve communication quality. Figure 12 shows that utilizing the route evaluation calculation, about 60% enhancement can be recorded to find the optimum path with faster convergence in initialization delay.

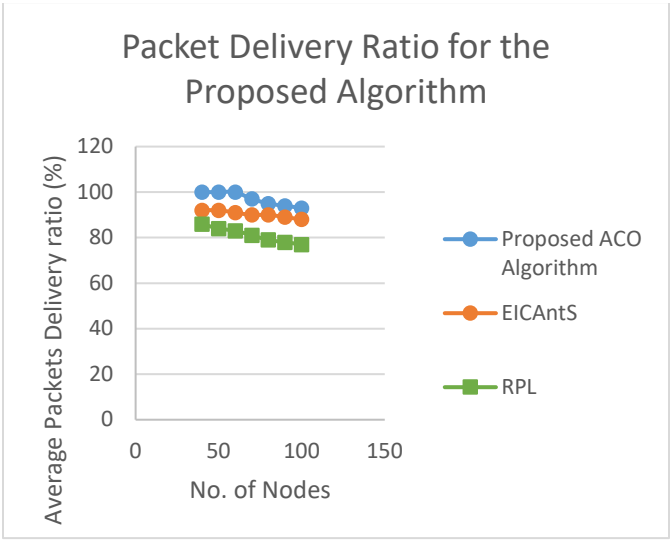


Figure 10: Packet Delivery Rates with the Number of Nodes

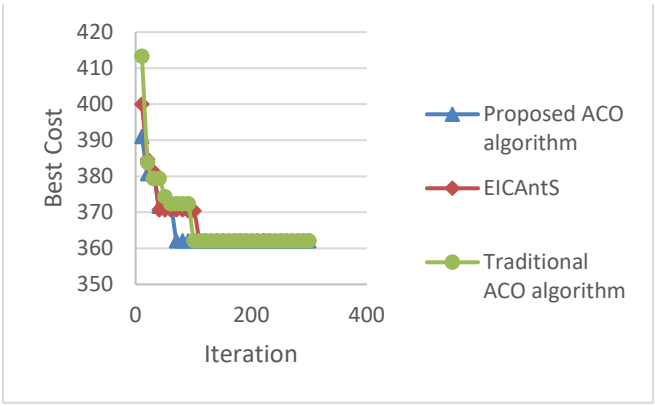


Figure 11: Comparison of the best cost with respect to the number of iterations (not utilizing fitness function)

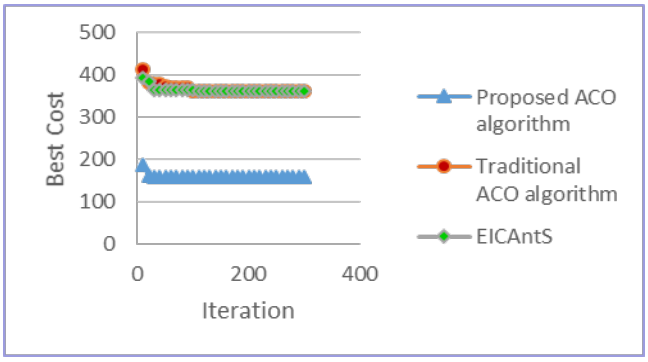


Figure 12: Comparison of the best cost with respect to the number of iterations using fitness function

4. Conclusions

The popularization of IoT-connected devices enabling development of IoT applications possess multiple and complex aspects of IoT while designing efficient communication protocols. The proposed system has utilized the beneficial features of an ant colony system to improve the route determination process and communication quality and mechanism in the context of IoT and to make IoT network system more scalable under varying load conditions. The parameters used in the selection of the routes are dynamically attuned to manage the network's dynamic condition. Current node energy, mobility, route path, and cost function including route assessment index are considered by the proposed method to enable efficient information communications within IoT. In the probability formula of ACO algorithm, the energy factor and the average mobility of the nodes are incorporated as well. It is demonstrated from the evaluation results that the proposed ACO based routing algorithm attained faster convergence for the initialization delay, about 30% improvement to discover the optimum route, and reduced energy consumption, of almost 50% less consumed energy even with the increasing number of nodes, compared with the benchmark algorithms. The proposed system achieved better results in terms of average end to end delay, almost 40% reduced end to end delay, compared to the other standard protocols. Also, it is confirmed that the proposed protocol as energy efficient method even for scalable networks, where the performance does not deteriorate abruptly as the number of nodes increases. Rather, it maintained improved network performance comparing with the other algorithms from the benchmark papers.

References

[1] A. Sharmin, F. Anwar, S.M.A. Motakabber, "Energy-Efficient Scalable Routing Protocol Based on ACO for WSNs," 2019 7th International Conference on Mechatronics Engineering, ICOM 2019, 1–6, 2019, doi:10.1109/ICOM47790.2019.8952053.

[2] Cisco, How the Next Evolution of the Internet Is Changing Everything, White paper, 2011, CISCO White Paper, 2011.

[3] Y.A. Malkani, L. Das Dhomeja, "Secure device association for ad hoc and ubiquitous computing environments," in 2009 International Conference on Emerging Technologies, ICET 2009, 2009, doi:10.1109/ICET.2009.5353132.

[4] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "Wireless sensor networks: A survey," Computer Networks, 2002, doi:10.1016/S1389-1286(01)00302-4.

[5] K. Sohraby, D. Minoli, T. Znati, Wireless Sensor Networks: Technology, Protocols, and Applications, 2006, doi:10.1002/047011276X.

[6] G. Anastasi, M. Conti, M. Di Francesco, A. Passarella, "Energy conservation in wireless sensor networks: A survey," Ad Hoc Networks, 2009, doi:10.1016/j.adhoc.2008.06.003.

[7] W. Dargie, C. Poellabauer, Fundamentals of Wireless Sensor Networks: Theory and Practice, 2011, doi:10.1002/9780470666388.

[8] M. Saleem, G.A. Di Caro, M. Farooq, "Swarm intelligence based routing protocol for wireless sensor networks: Survey and future directions," Information Sciences, 2011, doi:10.1016/j.ins.2010.07.005.

[9] T. Gui, C. Ma, F. Wang, D.E. Wilkins, "Survey on swarm intelligence based routing protocols for wireless sensor networks: An extensive study," in Proceedings of the IEEE International Conference on Industrial Technology, 2016, doi:10.1109/ICIT.2016.7475064.

[10] R.A.L. Rabelo, J.V.V. Sobral, H.S. Araujo, R.A.R.S. Baluz, R.H. Filho, "An approach based on fuzzy inference system and ant colony optimization for improving the performance of routing protocols in Wireless Sensor Networks," in 2013 IEEE Congress on Evolutionary Computation, CEC 2013, 2013, doi:10.1109/CEC.2013.6557967.

[11] M. Dorigo, V. Maniezzo, A. Colorni, "Ant system: Optimization by a colony of cooperating agents," IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 1996, doi:10.1109/3477.484436.

[12] S. Hamrioui, P. Lorenz, "Bio inspired routing algorithm and efficient communications within IoT," IEEE Network, 2017, doi:10.1109/MNET.2017.1600282.

[13] K. Machado, D. Rosário, E. Cerqueira, A.A.F. Loureiro, A. Neto, J.N. de

Souza, "A routing protocol based on energy and link quality for internet of things applications," Sensors (Switzerland), 2013, doi:10.3390/s130201942.

[14] V. Gupta, S.K. Sharma, "Cluster head selection using modified ACO," Advances in Intelligent Systems and Computing, 2015, doi:10.1007/978-81-322-2217-0_2.

[15] R. Khoshkangini, S. Zaboli, "Efficient Routing Protocol via Ant Colony Optimization (ACO) and Breadth First Search (BFS)," International Conference on Internet of Things (IThings 2014), (March), 375–381, 2014, doi:10.1109/iThings.2014.69.

[16] M.D. Devi, K. Geetha, K. Saranyadevi, "Content Based Routing Using Information Centric Network for IoT," in Procedia Computer Science, 2017, doi:10.1016/j.procs.2017.09.145.

[17] R. Bao, H. Pan, Q. Dong, L. Yu, L. Shao, "Ant colony-based routing algorithm for wireless sensor networks," Chinese Journal of Sensors and Actuators, **24**(11), 1644–1648, 2011, doi:10.3969/j.issn.1004-1699.2011.11.025.

[18] A. Nayyar, R. Singh, "IEEMARP- a novel energy efficient multipath routing protocol based on ant Colony optimization (ACO) for dynamic sensor networks," Multimedia Tools and Applications, 2019, doi:10.1007/s11042-019-7627-z.

[19] Y.L. Wang, M. Song, Y.F. Wei, Y.H. Wang, X.J. Wang, "Improved ant colony-based multi-constrained QoS energy-saving routing and throughput optimization in wireless Ad-hoc networks," Journal of China Universities of Posts and Telecommunications, 2014, doi:10.1016/S1005-8885(14)60267-3.

[20] A. Sharmin, F. Anwar, S.M.A. Motakabber, "A Noble Approach of ACO Algorithm for WSN," Proceedings of the 2018 7th International Conference on Computer and Communication Engineering, ICCCE 2018, 152–156, 2018, doi:10.1109/ICCCE.2018.8539295.

[21] A. Sharmin, F. Anwar, S.M.A. Motakabber, "A novel bio-inspired routing algorithm based on ACO for WSNs," Bulletin of Electrical Engineering and Informatics, 2019, doi:10.11591/eei.v8i2.1492.

[22] W.B. Heinzelman, A.P. Chandrakasan, H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," IEEE Transactions on Wireless Communications, 2002, doi:10.1109/TWC.2002.804190.

[23] S.S. Solapure, H.H. Kenchannavar, "Design and analysis of RPL objective functions using variant routing metrics for IoT applications," Wireless Networks, **26**(6), 4637–4656, 2020, doi:10.1007/s11276-020-02348-6.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1719-1726 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Design of Power Efficient Routing Protocol for Smart Livestock Farm Applications

Shahenda S. Abou Emira[1,2], Khaled Y. Youssef [3,*], Mohamed Abouelatta[1]

[1]*Faculty of Engineering, Electronics and Electrical Communication, Ain Shams University (ASU), Cairo, 39827, Egypt*

[2]*Faculty of Engineering, Electronics and Electrical Communications, October University Modern Sciences and Arts (MSA), Cairo, 39827, Egypt*

[3]*Faculty of Navigation Science and Space Technology, Beni-Suef University (BSU), Beni-suef, 62521, Egypt*

ARTICLE INFO

Article history:
Received: 16 September, 2020
Accepted: 27 December, 2020
Online: 28 December, 2020

Keywords:
AODV
IoT
Livestock
MANET
Routing
Sustainability

ABSTRACT

The demand on livestock as a source of protein is increasing with the dramatic increase of world population which approaches 8 billion. As a result, the monitoring of the performance of livestock breeding is becoming essential. The technique which could increase livestock production and improve the efficiency of operation is digitalization of livestock management .This could be achieved by models' realization that relates the rates of productivity and the parameters of operation which count on collecting big volume of data that results from digitalization process. In fact, the adoption of IoT technology in livestock environments is usually challenged by energy problems and power efficient communication technologies. In this paper a modified AODV protocol is proposed to enhance the traditional AODV protocol used in MANET networks that was mainly concerned with optimal route selection based on shortest path to send the data packets through it. The proposed technique models and simulates operational parameters of intermediate nodes as a factor to decide on the optimal path selection in IoT based MANET networks in addition to the shortest path that leads to transmission power gain of the source node. The results show a significant power performance enhancement that reaches average power saving of 33.3% compared to the traditional AODV protocol.

1. Introduction

This paper is an extension of work originally presented in 14th International Conference on Computer Engineering and Systems (ICCES) [1]. The increase in world population that reaches 8 billion increases the demand for livestock production. As a result, it is necessary to avoid livestock losses and protect it from diseases as this could threaten the countries' economy and food security. The process of observing livestock was difficult as it took a lot of time to detect if they had medical issues or not. Accordingly, researchers illustrated nowadays different advanced monitoring systems for livestock, these systems depend on remote sensors. Different types of sensors were used as: accelerometer, rumination, body temperature and humidity, and sound sensors. They were placed on the cattle to collect data and send it to the processing unit to take the appropriate action [2].

Today, Different technologies play an important role to increase livestock production and to save their losses as well as improving the operation efficiency in large farms. One of the recent technologies that offers the measure, actuate, sense and infer of physical and physiological parameters is the Internet of Things (IoT) technology. Livestock performance monitoring systems using IoT technology require special type of networks. Adhoc network is one of these special networks as it is energy efficient network and elongates the devices' lifetime. In the Adhoc network, different sensors are used to monitor behavior, stress, diseases and temperature. These sensors provide the IoT network with the collected data which allow the early detection of diseases. As a result, early treatment is offered which saves money and time.

The main problem in IoT technology is the energy source as it needs to work for a long time to reduce the need for replacement of sources or human collaboration. This paper concentrates on the study of suitable network topology that provides sensor network operation using energy efficiently. Accordingly, different

*Corresponding Author: Khaled Y. Youssef, Khaled.youssef@nsst.bsu.edu.eg

techniques against farms' environment are illustrated and simulated. Additionally, a Mobile Adhoc network (MANET) algorithm is introduced to decrease the source energy drainage and assure energy sustainability.

2. Literature Review

Several literatures are presented in previous work that proved the significance of IoT and sensor technologies to improve the livestock farm performance especially the large animal farming. In [3], the author presented a group of micro-sensors which were non-invasive sensors. They were mounted on the livestock using a collar that sent data via the internet wirelessly. As a result, a continuous connection between animals and veterinarian was established.

In [2], the author proposed a system using various sensors which were mounted on the cattle body. These sensors were used to monitor different parameters which allowed to detect the cattle body issues constantly. In [4], the author developed a system using misting systems for heat stress reduction. It was also developed to monitor humidity and temperature ranges. The system assured the avoidance of cattle death and the increase of producing high quality milk by preserving the ranges of the monitored parameters constant.

Although the literature above highlighted the use of rechargeable batteries as a main power source of the sensor nodes used in their studies, the researchers did not consider the battery recharging methodology especially for Adhoc wireless sensor networks. On the other side, the following literature developed techniques that enhance wireless sensor networks' performance.

In [5], the author focused on wireless sensor networks and proposed a gateway based Geographical Energy Aware Routing (M-Gear) protocol. In the system, the sensor nodes were distributed according to their location into four regions. Outside the sensing area, the base station was placed and at the center of the sensing area, the gateway node was located. Their developed protocol performance was compared to the performance of Low Energy Adaptive Clustering Hierarchy (LEACH) protocol. The compared performance was in terms of energy consumption and network lifetime. The comparison proved that M-Gear protocol was better than LEACH protocol.

In [6], the author developed a Modified Sensor Protocol for Information via Negotiation which is called M-SPIN. They made a comparison between the performance of M-SPIN and the traditional SPIN protocol using the broadcast communication. In the developed protocol, the hop distance between nodes and sink node was measured using a distance discovery phase. Additionally, the total number of transmitted packets was reduced as a result of transmitting data only to the sink node. As a result the consumption of total energy was minimized.

In [7], the author proposed an algorithm to modify the Dynamic Source Routing (DSR) protocol. The number of Route Reply (RREP) packets was decreased using the proposed algorithm and it also decreased the size of the header for the data packets which resulted in decreasing the overhead. They also developed an algorithm to reduce the energy consumed in data transmission.

In [8], the author presented a Dynamic MANET On-demand (DYMO) protocol and ant colony optimization algorithm which was capable of managing environmental changes. Graphs were used in ant colony optimization algorithm which was based on probabilities' method to find routes. Two factors were measured to evaluate the discovered routes. The measured factors were the route reliability and delay time. The software used to implement the algorithm was NS-2.

Accordingly, the proposed modified AODV technique capitalize on extending the enhancements in the literature to develop IoT operation aware Adhoc wireless sensor network that is suitable for use in livestock farm applications.

3. Overview on Energy-Efficient Routing Protocols

Mobile Adhoc network (MANET) is a group of self - configuring mobile nodes. These nodes can be used at the same time as router and end node, therefore a fail in the operation of some nodes affects the performance of the network. Energy deficiency of nodes is one of the main problems which affects the performance of network in MANET. As a result, it is important to use energy efficiently in MANET in order to increase the lifetime of node's battery [9]. There are several routing protocols in MANET such as Destination Sequenced Vector (DSDV), Dynamic Source Routing (DSR) and Adhoc On-demand Distance Vector (AODV). These protocols are used to select the optimal route between source and destination in order to send data through it.

3.1. Destination Sequenced Distance Vector

DSDV is one of the proactive routing protocols in MANET. It is a table driven routing protocol and it is based on Bellman-Ford algorithm. In this protocol, a routing table is used at each node to store all available destinations. Additionally, advertisements are made as a severe change may exist in the table's data. In this protocol, a sequence number is also sent to the destination in order to discriminate the old routes from the new ones which avoids the formation of loops. It isn't recommended for highly dynamic networks because it needs regular updates in its routing tables which lead to battery drainage [10].

3.2. Dynamic Source Routing

DSR protocol creates routes only when requested, therefore it is one of the reactive routing protocols. It doesn't use routing tables at the intermediate nodes as it uses source routing. In this protocol, the intermediate nodes send a Route Request (RREQ) packet with its own address identifier. The available routes in the DSR protocol are being tracked by a route cache. Two main mechanisms are used in DSR: route discovery and route maintenance. They are used to discover routes and to notice if any change occurs in the network topology [11].

3.3. Adhoc On-demand Distance Vector

AODV is one of the most known reactive routing protocols in MANET. It creates only a route when a source node needs to send data to a destination node therefore it is better than DSDV as it minimizes the number of broadcasts. It has a good performance in dealing with large number of mobile nodes as it creates dynamically at each intermediate node a routing table. It also has

a sequence number counter which prevents the formation of loops in its routing tables. This sequence number also keeps the nodes updated by comparing their routing tables to check how updated information they have about other nodes [12].

4. Performance Evaluation of Adhoc networks

There are different parameters measured to assess the performance of MANET protocols as summarized in Table 1.

4.1. Packet Delivery Fraction

The ratio between the number of broadcasted packets by the source node and the number of received packets at the destination node is known as PDF. In case of low number of nodes, DSR performs better in this metric but the performance declines with the increase of the number of nodes. On the contrary, DSDV performs better in case of large number of nodes but AODV showed that its performance is uniform in both cases [13].

4.2. Average end to end delay

It is the time it takes from the packet being created at the source node to the packet being delivered at the destination node. In this metric, DSDV performance declines with the increase of the number of nodes. However, the DSR and AODV performance is uniform [13].

4.3. Number of Packets Dropped

It is the failed data packets that couldn't reach the destination node. In case of large number of nodes, DSR and AODV have a good performance. On the other hand, DSDV has the worst performance as it declines with the increase of the number of nodes. The performance of AODV in this metric is the best as it exchanges information periodically between nodes which ensures connection [13].

Table 1: Performance of MANET Protocols

Metrics	AODV	DSDV	DSR
Average end to end delay	Uniform performance	Performance declines when number of nodes increase	Uniform Performance
PDF	Uniform Performance	Good performance when number of nodes increase	Performance declines when number of nodes increase
Number of packets dropped	Good performance when number of nodes increase	Performance declines when number of nodes increase	Good performance when number of nodes increase

5. Proposed Architecture for Livestock monitoring IoT networks

Figure 1 shows the architecture of a group of mobile nodes connected in MANET network. Each node in this network has different sensors such as accelerometer, skin temperature, humidity and ambient temperature, and respiratory rate. The data sensed from these sensors will be transmitted to a central processing server (CPS) to perform the proper action. In such a network, Power should be used efficiently to elongate the battery lifetime of each node. Accordingly, AODV protocol which is an energy efficient routing protocol is used to select the optimal route to send the collected data to the destination node. As a result, the total transmitted power of source node is reduced.

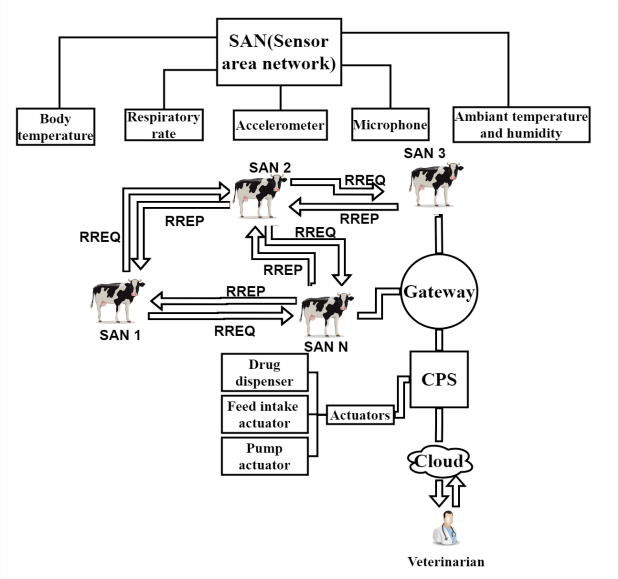


Figure 1: System Architecture for Monitoring livestock Performance

Figure 2 shows the flowchart of how AODV protocol is implemented. When a source node starts to search for a route to send its data through it to the destination node, it searches its routing table. If there is a route in its routing table to destination, it sends its data packets through this route. If there is no route to the destination node, it broadcasts a Route Request (RREQ) packet to all the neighbor nodes. If the node that received the RREQ packet is the destination node or it has in its routing table a route to the destination node, it sends a Route Reply (RREP) packet to the source node. If the receiving node isn't the destination node or there is no route for the destination in its routing table, it broadcasts a RREQ packet to all its neighbors. When the source node receives a RREP packet, it transmits its data packets through the same route where it received the RREP packet. If several RREP packets are received by the source node, it selects the RREP packet received from the shortest route and discard other packets. As a result, the power for transmitting the data packets is reduced which increases the source node battery lifetime.

A modified AODV protocol is proposed to enhance the overall network power performance. The optimal route in the modified protocol is selected depending on the index weight of the path rather than depending only on the shortest path. The index weight is a new factor measured to indicate the operation time of each node in a certain path. Each node is given a certain index

which represents its operation time. The index weight of the path is measured as the total index of nodes in a certain path divided by the number of nodes in this path. The path with the highest index weight has the lowest priority to be selected while the path with the lowest index weight has the highest priority to be selected.

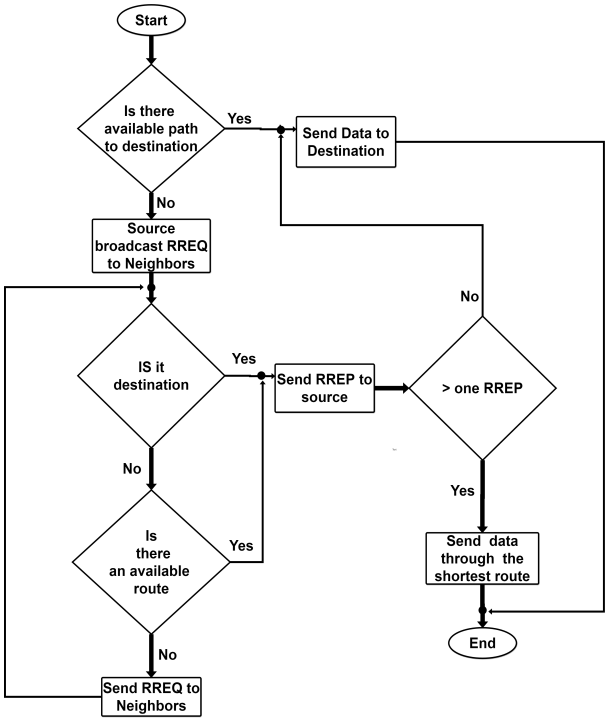


Figure 2: AODV Flowchart

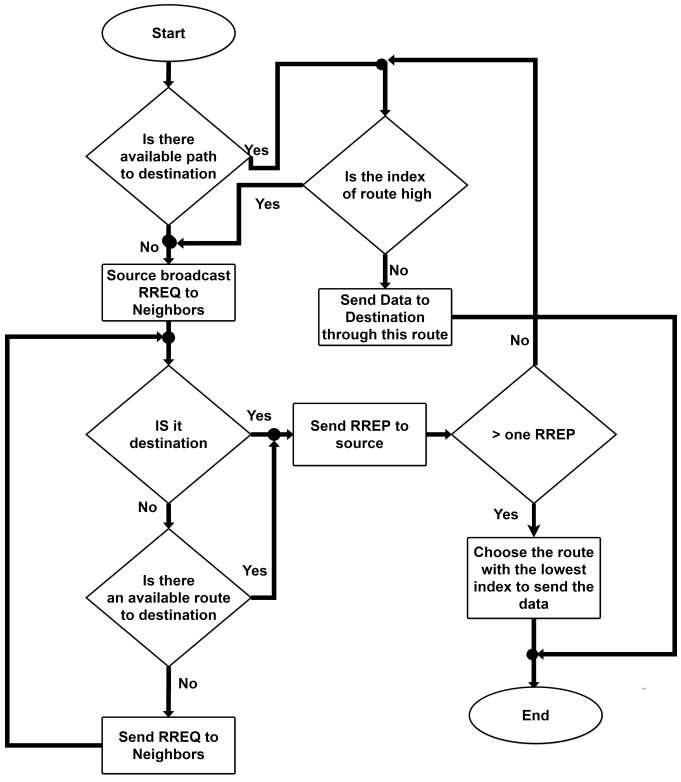


Figure 3: Modified AODV Flowchart

The flowchart in Figure 3 shows how the modified AODV protocol is executed. When a source node starts to search for a route to send its data through it to the destination node, it searches its routing table. If there is a route in its routing table to destination, it checks its index weight. If it is low, it sends its data packets through this route. If there is no route in its routing table or the index weight of the available route is high, it broadcasts a Route Request (RREQ) packet to all the neighbor nodes.

If the node that received the RREQ packet is the destination node or it has in its routing table a route to the destination node, it sends a Route Reply (RREP) packet to the source node. If the receiving node isn't the destination node or there is no route for the destination in its routing table, it broadcasts a RREQ packet to all its neighbors. When a RREP packet reaches the source node, it checks the index weight of the route. If it is high, it sends RREQ packet to all its neighbors searching for other routes with lower index weight. If it is low, it starts to send data packets to the destination through the same route where it received the RREP packet. If the source node receives more than one RREP packet. It selects the route with the lowest index weight to send data packets. As a result, it increases the lifetime of the serving nodes which enhances the overall network power performance.

6. Simulation Model Results

In this simulation, MATLAB software is used to simulate AODV protocol. In Figure 4, it is shown how the source node, destination node and the number of nodes are entered in this model. It is assumed in this simulation that the network has 20 mobile nodes randomly distributed in a farm size of 100 m^2 and a 20 m transmission range between nodes. The source node and the destination node are assumed to be node 2 and node 10 respectively. AODV protocol is implemented in this model to enable the source node to request different available paths to the destination node. The source node will receive RREP packet from the intermediate nodes if they have an available route to the destination node. In Figure 4, the available routes are shown.

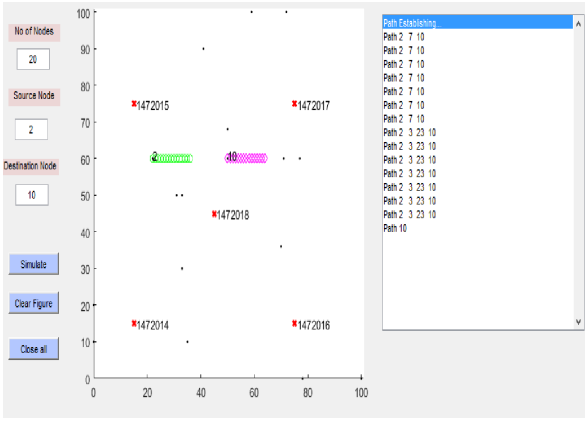


Figure 4: Simulation Model

In our work, two main performance indicators are used to measure the impact of the proposed advanced AODV protocol on the IoT network performance mainly the energy consumption and network lifetime.

i) Energy consumption is based on MATLAB model that is governed by the following equation

Energy Consumption (Joule) = Energy Consumption Rate (Joule/bit) * Transmission rate (packets per sec) * Packet size (bits) (1)

The multipath fading is also accounted in the simulator model given that the amount of Energy consumption rate considered in the model for the transmitter or receiver circuitry is 50 x 10^{-9} and amount of energy consumption for multipath fading is 0.0015 x 10^{-12}.

ii) Network lifetime that is defined as the ratio between its residual energy and the energy consumed to transmit data packets (i.e. the time spent until the node run out of energy). It is the third parameter calculated at each route as shown in Figure 7. In this case, the route with the longest network lifetime is the optimal route as it has the lowest energy consumption and the shortest distance. As a result, route 12 is the optimal route in terms of distance, energy consumption, and network lifetime.

The simulation results for applying AODV protocol in cows' performance monitoring system are shown in this section. The first parameter measured as shown in Figure 5 is the total distance between the source node and the destination node at each route. The optimal route is found based on this parameter as the optimal route is the shortest route which is selected to transmit data through it. In this case, the route with the shortest distance is route 12 so it is the optimal route.

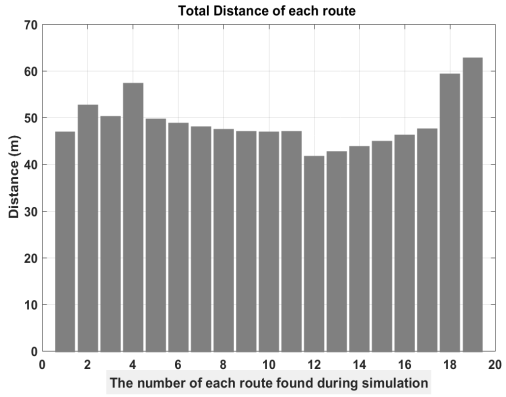


Figure 5: Total Distance versus The Number of Each Route Found During Simulation

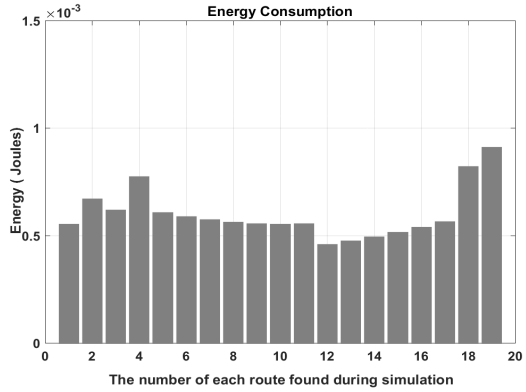


Figure 6: Energy Consumption versus The Number of Each Route Found During Simulation

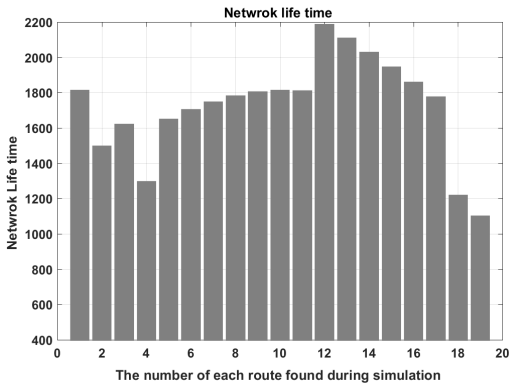


Figure 7: Network Life Time versus The Number of Each Route Found During Simulation

In this simulation, the energy consumption is the second parameter calculated at each route as shown in Figure 6. Equation 1 represents the energy consumption model. The route with the shortest distance has the lowest energy consumption. In this case, route 12 is the optimal route in terms of energy consumption and distance.

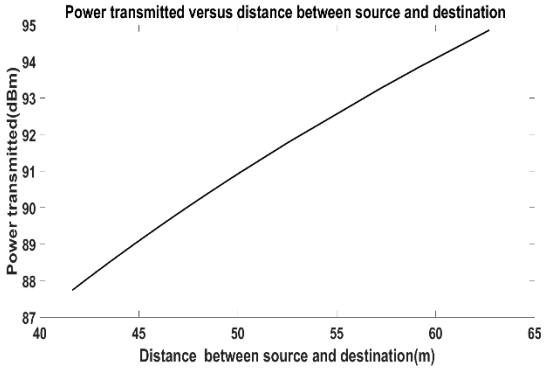


Figure 8: Transmitted Power versus Distance between Source and Destination During Simulation

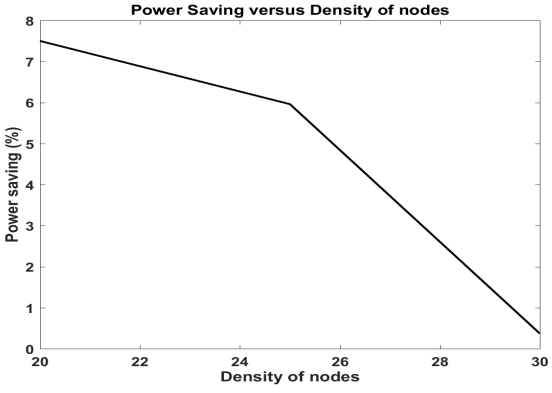


Figure 9: Power Saving versus Node Density in Mobile Adhoc Network

The fourth parameter calculated in this simulation as shown in Figure 8 is the transmitted power at each route. The data are transmitted with minimum power when transmitted through the shortest optimal route to the destination node. In this case, route 12 is the optimal route in terms of distance, energy consumption, network lifetime, and transmitted power.

The fifth parameter measured is the power saving as shown in Figure 9. It indicates when selecting the shortest optimal route in MANET how much transmitted power is saved. It is measured at different number of nodes as shown in Figure 9 to indicate that when the node density increases the power saving decreases. Figure 10 shows the power saving which indicates how much transmitted power is saved when selecting the shortest route in MANET compared to the transmitted power in fixed network. It is measured at different number of nodes to indicate that when the node density increases the power saving decreases as shown in Figure 10.

Table 2: Simulation Results

Route Number	Distance (meters)	Transmitted Power (dBm)	Power Gain
1	46.84294241	89.78211633	
2	52.58886799	91.79210254	
3	50.14213562	90.96446273	
4	57.192105	93.24979289	
5	49.60330529	90.7767743	
6	48.66124402	90.44367794	
7	47.90305917	90.1708796	
8	47.33955199	89.96531538	
9	46.97970233	89.83276007	
10	46.83013443	89.77736582	
11	46.894632	89.80127492	0.924
12 Optimal Route	41.64331698	87.73816223	
13	42.60107524	88.13317207	
14	43.66889973	88.56323963	
15	44.83914468	89.02264243	
16	46.10374679	89.50579849	
17	47.45489265	90.00758945	
18	59.20525888	93.85076102	
19 Worst Route	62.72062331	94.85276426	

Figure 11 shows different paths between source and destination which are path1 (P1), path2 (P2), path3 (P3), path4 (P4) and path5 (P5). It is assumed for each path that the distance between nodes is equal and the total transmitted power of the source node is 30 dBm. It also shows the transmitted power of each intermediate node in each path. There are some nodes which are common in more than one path such as nodes: 3, 4, 7 and 9. These nodes are called busy intermediate nodes as they could be used several times in the network which will lead to the drainage of their batteries. As a result, as mentioned earlier in section 5 a modified AODV protocol is proposed to save the battery of these busy nodes. There are different scenarios for node process shown in Table 3 as each intermediate node will have an index to indicate how busy this node in order to select the optimal path in terms of its index. In Table 4, it shows the transmitted power of each node and its index of processing. It also shows their index weight which is the total index of nodes divided by the number of nodes in each path. The selection of the optimal route will be based on its index weight, the route with lowest index weight is the optimal route to send data through it.

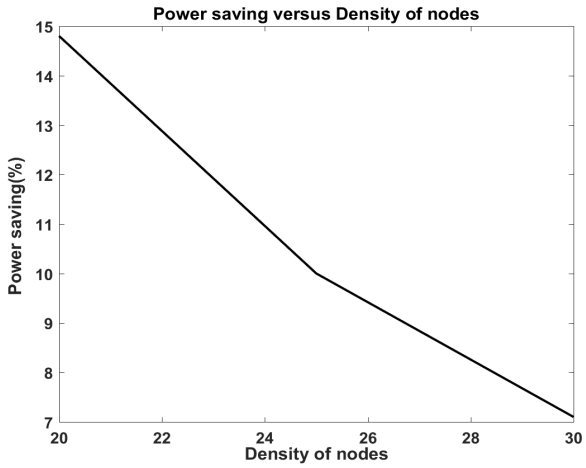


Figure 10: Power Saving versus Node Density in Fixed Network

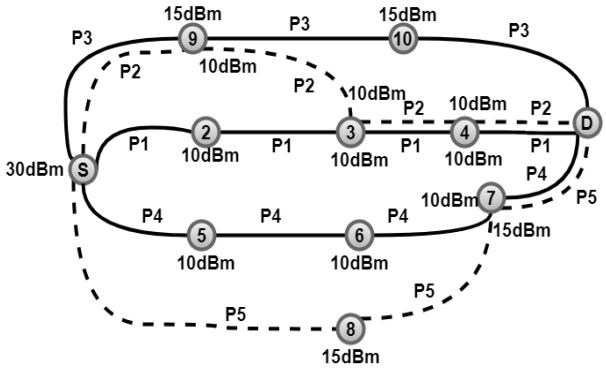


Figure 11: Different Paths Available to Send Data Packets from Source Node (S) to Destination Node (D)

Table 3: Different Scenarios for Node Process and Its Index

Node process	Duration of not processing node (T in hours)	Index(i)
Event Based	T >= 12	0
	6 =< T <12	1
	3 =< T < 6	2
	1 =< T < 3	3
	T < 1 hour	4
Periodic	Periodic Time (Tp in hours)	Index(i)
	Tp >= 12	5
	6 =< Tp <12	4
	3 =< Tp< 6	3
	1 =< Tp < 3	2
	Tp< 1 hour	1
	Data size (D in kilobyte(KB))	Index(i)
	D>=200	5
	100=<D<200	4
	50=<D<100	3
	10=<D<50	2
	D<10	1

Table 4: Parameters of Intermediate Nodes

Path	Intermediate Nodes	Power(dBm)	Index	Index Weight
P1	2	10	1	2.33
	3	10	3	
	4	10	3	
P2	9	10	5	3.67
	3	10	3	
	4	10	3	
P3	9	15	5	4
	10	15	3	
P4	5	10	1	2
	6	10	0	
	7	10	5	
P5	8	15	2	3.5
	7	15	5	

Figure 12 shows the index weight of each path which indicates the status of the intermediate nodes. If the index weight of a path is high, it indicates that its intermediate nodes were processing for long time therefore this path can't be selected to transmit data through it. As a result, this will save the lifetime of their batteries.

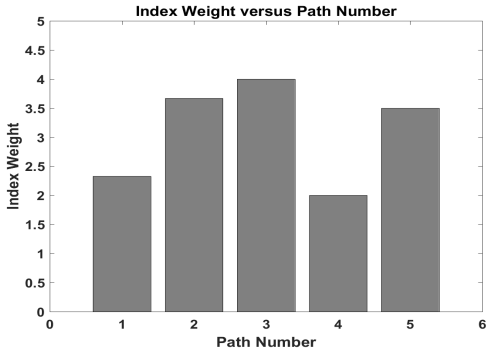


Figure 12: Index Weight versus Path Number

Figure 13 shows the power gain of each path which indicates how much power is saved by selecting a certain path. The optimal path shown in Figure 13 is path 4 as it has the highest power gain.

Figure 14 shows the power gain versus the index weight to indicate how much power is saved when selecting the path with the lowest index weight. As shown in Figure 14, the path with the lowest index weight has the highest power gain.

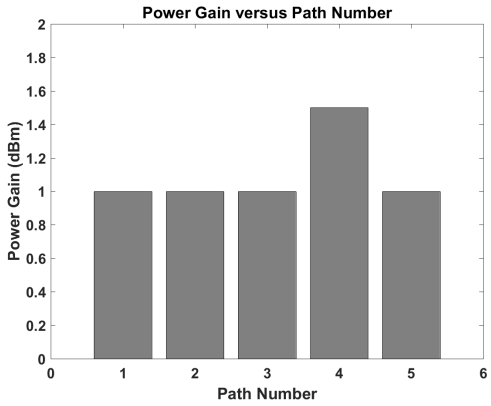


Figure 13: Power Gain versus Path Number

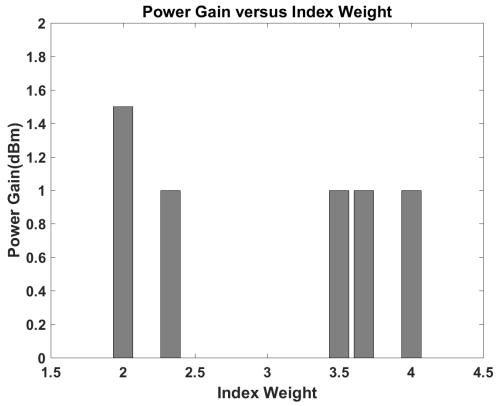


Figure 14: Power Gain versus Index Weight

7. Discussion and Conclusion

The results from applying the AODV routing protocol to select the shortest route showed that when the total distance between source and destination decreases, the energy consumption decreases which increases the network lifetime. It also showed that a decrease in the transmitted power occurs when the total distance decreases. For MANET and classical fixed networks, the power saving is calculated for each route found at different node densities. It showed that when the node density increases, the power saving decreases.

Table 5: Power Gain for Different Path Selection Scenarios

Busy intermediate nodes	Common in Path	Total Power(dBm) on busy node	Scenario	Power gain (dBm)
3	P1	20	1 (P1 selected)	10
	P2		2 (P2 selected)	10
4	P1	20	3 (P1 selected)	10
	P2		4 (P2 selected)	10
7	P4	25	5 (P4 selected)	15
	P5		6 (P5 selected)	10
9	P2	25	7(P2 selected)	15
	P3		8 (P3 selected)	10

Table 6: Traditional AODV versus Proposed Modified AODV Protocol

Routing Protocol	Optimal Path	Hop Count	Optimal Path Selection	Power Gain (dBm)	
Traditional AODV	3	3	Shortest Path	10	By comparing the traditional AODV and the proposed modified AODV, the results show a significant power performance enhancement that reaches average power saving of 33.3%
Proposed Modified AODV	4	4	Lowest Index Weight	15	

The traditional AODV protocol focused only on how to enhance the battery lifetime of the source node depending on selecting shortest optimal route. It didn't highlight the importance of the intermediate nodes which are serving nodes and their batteries could be drained quickly due to their operation time in the network. As a result a modified AODV protocol is proposed, this protocol added a new feature to select the optimal route. The selection of the optimal route in the modified AODV protocol is depending on the operation time of each node in the path. Each node is given an index which specifies its operation time as shown in Table 3. An index weight which is the total index of the nodes in a certain path divided by the number of nodes in this path is measured for each path. This factor indicates if the available path could be used or not. The path with a high index weight has the lowest priority to be selected because it indicates that the operation time of the intermediate nodes in this path is very high and thus the probability to have lower battery state of charge is higher.

Accordingly, the optimal path is the path that has the lowest index weight.

By applying the modified AODV protocol to select the optimal route depending on the index weight of the route as a priority to selection of the shortest route (i.e. if two routes have the same index, then the shortest one is selected). The protocol supports to give the intermediate node with longer operation time the opportunity to focus on sensed data processing. The networking processing function is then off loaded from this intermediate node to other nodes with lower operational time and accordingly higher probability of battery charge lifetime.

As observed in Table 5 the optimal path in terms of power gain is path2 and path4. It is also observed in Table 4 that the best path in terms of index weight is path 4. By comparing the power gain when selecting the shortest route (measured in hop counts) which is path 3 in Figure 11 to the power gain when selecting path4 which has the lowest index weight, it proved that the power is enhanced by 33.3% as shown in Table 6.

As a summary, a convenient network topology is studied in this paper to provide a network of sensors which has an energy efficient operation. Accordingly, in order to decrease the source energy drainage which assures energy sustainability, MANET algorithm is proposed. In MANET, AODV is the most common known routing protocol as its performance is the best compared to other MANET protocols. In this simulation, AODV was applied to allow the selection of the optimal route between source node and destination node. The optimal route is selected as it has the shortest distance, the lowest energy consumption and the longest lifetime.

A modified AODV algorithm is proposed in order to enhance the overall network power performance. The protocol depends on giving intermediate nodes with lower operation time a higher priority to share in the network routing even if the path could be longer which could lead to a power saving that reaches 33.3%. As a result, the paths in the modified AODV are selected based on the index weight of the path. The selected path is the path with the lowest index weight.

References

[1] S.S. Abou Emira, K.Y. Youssef, M. Abouelatta, "Power Efficient Routing Protocol (AODV) for IoT Based Livestock Applications," in 2019 14th International Conference on Computer Engineering and Systems (ICCES), 407–412, 2019, doi:10.1109/ICCES48960.2019.9068149.

[2] M. Meenakshi, S.S. Kharde, "Advance Cattle Health Monitoring System Using Arduino and IOT," **5**(4), 3365–3370, 2016, doi:10.15662/IJAREEIE.2016.0504162.

[3] J. Edwards, D. Gibson, "Novel Technology for the Remote Monitoring of Animals," Companion Animal Society Newsletter, **23**(2), 56–59, 2012.

[4] M. M, S. A, "IOT-BASED MEASURING TEMPERATURE AND HUMIDITY IN THE CATTLE DIARY," International Journal of Sustainable Development, **7**(11), 7358–7360, 2015, doi:10.24327/IJRSR.

[5] Q. Nadeem, M.B. Rasheed, N. Javaid, Z.A. Khan, Y. Maqsood, A. Din, "M-GEAR: Gateway-based energy-aware multi-hop routing protocol for WSNs," in Proceedings - 2013 8th International Conference on Broadband, Wireless Computing, Communication and Applications, BWCCA 2013, 164–169, 2013, doi:10.1109/BWCCA.2013.35.

[6] Z. Rehena, S. Roy, N. Mukherjee, "A modified SPIN for wireless sensor networks," in 2011 3rd International Conference on Communication Systems and Networks, COMSNETS 2011, 2011, doi:10.1109/COMSNETS.2011.5716469.

[7] M. Tamilarasi, S. Chandramathi, T. Palanivelu, "Efficient energy management for mobile ad hoc networks," Ubiquitous Computing and Communication Journal, **3**(5), 12–19, 2001.

[8] A.M. Oranj, R.M. Alguliev, F. Yusifov, S. Jamali, "Routing Algorithm for Vehicular Ad Hoc Network Based on Dynamic Ant Colony Optimization," International Journal of Electronics and Electrical Engineering, **4**(1), 79–83, 2016, doi:10.18178/ijeee.4.1.79-83.

[9] M. Tarique, K.E. Tepe, M. Naserian, "Energy saving dynamic source routing for ad hoc wireless networks," in Proceedings - WiOpt 2005: Third International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 305–310, 2005, doi:10.1109/WIOPT.2005.21.

[10] P. Manickam, T. Guru Baskar, M. Girija, D. Manimegalai, "Performance Comparisons of Routing Protocols in Mobile Ad Hoc Networks," International Journal of Wireless & Mobile Networks, **3**(1), 98–106, 2011, doi:10.5121/ijwmn.2011.3109.

[11] V. Denkovski, B. Stojcevska, T. Dovenski, V. Pachovski, A. Bozinovski, "Performance Evaluation of Routing Protocols in a Wireless Sensor Network for Targeted Environment," JITA - Journal of Information Technology and Applications (Banja Luka) - APEIRON, **10**(2), 2016, doi:10.7251/jit1502110d.

[12] I.D. Chakeres, E.M. Belding-Royer, "AODV routing protocol implementation design," in Proceedings - International Conference on Distributed Computing Systems, 698–703, 2004, doi:10.1109/icdcsw.2004.1284108.

[13] S. Ade, P.. Tijare, "Performance Comparison of Routing Protocols in Mobile Ad Hoc Networks," International Journal of Information Technology and Knowledge Management, **2**(2), 545–548, 2010, doi:10.1007/978-981-10-6890-4_4.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1727-1733 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

New Properties of Crimes in Virtual Environments

Roman Dremliuga[1,*], Natalia Prisekina[1], Andrei Yakovenko[2]

[1]*School of Law, Far Eastern Federal University, Vladivostok, 690922, Russia*

[2]*National Technology Initiative Center for Neurotechnology, Virtual and Augmented Reality Technologies, Far Eastern Federal University, Vladivostok, 690922, Russia*

ARTICLE INFO

Article history:
Received: 27 August, 2020
Accepted: 04 December, 2020
Online: 28 December, 2020

Keywords:
Virtual reality
Criminal law
Virtual crime

ABSTRACT

Virtual reality is a technology that literally allows constructing of a new reality for its users. VR has huge potential, it is able to change social life, communications, and the academic sphere, but VR may also be applied by criminals. In the present paper properties of potentially committed crimes in VR are analyzed. It is concluded that VR provides criminals with some advantages. The first advantage is the multijurisdictional problem in prosecuting the offender that is usual for most of IT. The second advantage is that a virtual environment gives a realistic experience to the user and literally replaces the real world. The third advantage is an integration of virtual-world environments with haptic devices that may factually affect the user in the real world. All of these allow the commission of new kinds of crimes where properties of conventional and cybercrimes are combined. Among such kinds of crimes analyzed in the paper are cybercrimes that affect the body and mental state of victims. VR is researched as a first tool that in combination with haptic devices gives criminals the opportunity to commit crimes against sexual freedom and to cause physical harm. Thus, VR crimes need a special legal framework that will take into consideration this kind of crime.

1. Introduction

This paper is an extension of work originally presented in 2019 International Conference on Cybersecurity (ICoCSec) [1]. Virtual reality (VR) is one of the most promising breakthroughs. It literally replaces the real world. VR dips humans into an alternative reality where different aspects of human life are imitated. The tasks of introduction of any information technology may vary, for instance, technology could be applied to make the world around us better. At the time of the pandemic, it is the only way to visit new spaces that are so similar to the real world [2]. Virtual reality is not a single technology but a set of technologies that permit a developer to construct an environment that would be recognized by our body and mind as real. VR is a tool to create new realistic worlds. Sometimes this environment is used for work, very often for gaming and social interaction. There are some forecasts that soon people will prefer virtual worlds over reality.

Virtual reality is a widely used technology. The market of VR is anticipated to grow from US$27 billion in 2018 to US$209.2 billion in 2022 [3]. Some researchers noted that VR attracts a lot of investors' attention and large companies invest in the technology [4].

Virtual reality is technology that literally allows constructing a new reality for its users. VR has great potential, it is able to change our everyday life, communications, and academic sphere. Nowadays virtual reality is actively applied in different spheres. The biggest driver is the entrainment and computer games. Also, a VR application is noticeably beneficial in teaching [5,6]. Sometimes virtual reality helps to simulate real situations [7]. VR is widely applied in medicine [8]. Some cases of rehabilitation of felons in VR exist [9,10]. In addition, a virtual world can be tuned in accordance with the needs of the user that makes it more attractive than real.

Virtual reality technology is a variety of hardware and software tools, which help to imitate a real or fictional world for a user. It interacts with our senses (vision, hearing, touch, smell, sense of balance, etc.) and misleads the body and mind that they are in the real world [11]. Virtual reality is an environment that is imitated through technology. The main virtual reality device is a special

*Corresponding Author: Roman Dremliuga, Email: dremliuga.ri@dvfu.ru

helmet, that may mimic the outside world through rendering video and audio. There are also numerous additional devices such as controllers, joysticks, steering wheels, haptic interfaces which increase a realism of VR illusion. All devices are designed to make an imitated world as real as possible.

"Virtual reality" has a common-used definition of the term. It is considered that VR is an information technology with special characteristics. For example, the following definition is suitable for the goals of the research. Virtual reality term is interpreted as "a technically constructed interactive environment with the help of computer facilities; as the generation and operation of objects, real or imaginary, on the basis of their graphic representation, simulation of their physical properties and ability to influence and independent presence in space, as well as the creation of such objects by means of special computer equipment" [12].

Development and introduction of any technology depend on a legal framework that is applied. Investors, developers, and users have to understand current and future legal rules that concern application of technology. In terms of business, the legal expertise is necessary on the step of making the initial decision. Wrong assessment of legal risks could lead to a total misapplication of funds, time, and efforts. Thus, to understand development of VR it is necessary to understand the current and future legal frameworks in this sphere.

The question is whether VR is another technology to which existing cyberspace laws apply or whether entirely new rules need to be developed. In order to form the correct legal framework, it is necessary to understand the conceptual differences between this technology and other technologies.

The sufficient legal risk is to face criminal cases with new properties. Some technologies provide tremendous opportunities for commitment of crimes. If a developer understands which challenges would exist in terms of criminal application, he will design his products by taking into consideration potential risks. The key is understanding how VR could be used by criminals and how it could be treated from a legal point of view.

The idea of this paper is not directly to enumerate possibilities of illegal use of VR but assess future prevalence and social consequences. To achieve this goal, it is necessary to utilize some assumptions that research will rely on.

- If VR maintains to achieve the same illegal result as some other technology, this automation will be replaced by VR in case of higher effectiveness of VR.
- If some property bolsters to spread VR use in society, this property will be the key characteristic that makes VR useful for criminals.
- Enforcement is always behind leaders of the crime 'industry' since it is bound by strict regulation and bureaucracy.

Thus, three aspects were carried out in the present paper. Properties of VR as a tool of crime were analyzed; challenges for the development of regulation that is caused by the characteristics of VR were reviewed. Some recommendations for the development of a legal framework were provided.

2. Research methodology

Research methodology for studying legal challenges of VR technology that goes beyond both dogmatic analysis has to be applied in the research. Analysis in the research follows positions developed within the philosophy of law and legal theory. Among relevant methodologies are the hermeneutic and argumentative approaches.

As the research is interdisciplinary, there is a synergy between methodlogies of disciplines. Interdisciplinary research must extend beyond mere multiplication of disciplinary methods and perspectives, to incorporate some element of integration or, mutual dialogue between disciplines. Interdisciplinarity is absolutely essential for the research as its object lies on the intersection of disciplines. As fairly commented some authors "it is important to have a multidisciplinary approach inconstructing such a methodology as the tools analysed stem from different disciplines and domain of expertise" [13].

Factually the research is based on a comparison of current and future properties of VR technology with current and potential legal framework. The focus on the legal and societal impacts of VR use requires us to observe the results of researches in the sphere of influence of virtual environment on humans. It is clear that such an important question cannot be accepted only on the basis of one or several studies, therefore, it is necessary to use the results of a whole set of studies to draw conclusions about the characteristics of VR, important for the formation of legal norms for its regulation.

The expected results of the study are directions where the legal framework applied to VR has to be upgraded. Recommendations with a high level of abstract should be developed. In future, de lege ferenda proposals could be based on our results.

Despite analysis and conclusions of the paper could be applied in different legal systems, the study is mostly focused on legal practice and conceptions of the United States and EU states as leaders in introduction of VR. The reseacrh is less applicable to asian technological leaders due to sufficicent differences in legal frameworks. Some ways of VR applications that are usual in the United States and EU are prohibited in other countries. For instance, virtual sex or VR pornograhpy are illegal in China, Indonesia and Malaysia.

Thus, the methodology can be shortly described as the application of a legal case study to the research of technology with an enormous social impact. The paper as some others [14] utilized a case study methodology. It is manifested in gathering and analyzing proven effects of VR on a user in terms of social, ethical, and legal risks. As the area of risk is quite wide, the paper focuses on risks related to the commission of criminal offences in VR.

3. Legal challenges

3.1. Increasing significance of VR

The main purpose of VR is to provide an environment for spending time. The application of VR for gaming dominates nowadays but VR is also used for workshops and business meetings. It is actively applied for educational purposes and simulation of difficult tasks. The advantage of VR is clear: a

customized reality that provides a high level of similarity of human reactions with the real nature. It is effective for games where people pass through all spectrum of emotions as well as for training where people become ready for the stress of a real situation. Most VR applications in business communications or games imply audio and video information exchange in some virtual space that increases the level of participation.

The level of realism could be even improved when real appearance is applied. In virtual reality, a user is frequently seen by others as an anthropomorphic avatar whose appearance may be similar to the user's appearance in the real world or may be completely different. Despite the voice and personal appearance of the virtual world can be altered, this does not exclude the possibility that voice and visual appearance of the user would be the same as in the real world.

The big advantage of VR is that people are able to communicate in a realistic way even if they are separated in the real space. Thus, people, in the form of relatively realistic avatars, communicate as they would be face-to-face. VR provides the possibility to imitate eye contact and to manipulate virtual objects that participants see together. It is similar to a video call, but participants of VR meetings have a higher level of involvement.

VR technology is far from perfection. Headsets cannot track eye movement with high accuracy. VR Graphics is far from a full imitation of the representation of the real world and is rather applied to construct some simple spaces with a low quantity of details. Facial expressions are still rendered in virtual reality quite poorly.

Despite the above, virtual reality is a very popular environment for spending time and human interaction. Under different studies, the amount of VR's application for spending time and social interaction is fastly growing [15]. There are approximately 170 million users of virtual reality worldwide. A double increase in the number of VR consumers is evidenced in recent years.

The growth of legal applications of VR implies an increase in criminal use. It is anticipated that the number of virtual reality users should correlate with the number of crimes committed in VR, as this was observed with the growth of the number of Internet users. As more people use VR worldwide, there are more potential victims and criminals in VR. Some researchers found the proliferation of the use of the global network (the Internet) as a major factor in the growth of cybercrime. Thus, the number of criminal acts in VR is expected to increase due to the growth of the number of its users [16].

Wide use of VR leads to a classic dilemma. A legislator has to choose between the need for a free exchange of computer information, as a factor of digital economy growth and the need to defend state and social security by restrictions on IT use. To find the balance between public security and the necessary freedom of technology application for its development, a lot of different factors have to be taken into consideration.

As with any technology, VR may facilitate criminal activity. Virtual environment may be a tool or instrument to commit a crime as well as an additional factor that facilitates it. Recently, the use of high technology to commit crimes increases, and it concerns virtual reality. An opportunity how criminal law is applicable in case of crime in virtual environment should be studied.

3.2. VR as a potential tool to commit a crime

VR is in some sense a unique technology that makes crimes with the application of that technology dissimilar with others. It may constitute difficulties for the investigations of such offences. VR as a tool confers crimes by new properties. Virtual reality is not simply another new technology, which requires to design technical standards. Rather VR application requires to introduce a new legal framework. There are several reasons why a traditional legal framework does not qualify. Basically, due to its special characteristics, VR requires fundamentally new legal frameworks and solutions.

The first, commonly known problem is cross border jurisdiction. It is difficult to identify a person with whom the user interacts in VR. Virtual environment and its users are often in different jurisdictions. A world that is undivided in VR may exist regardless of borders of the concrete state in terms of its users. Millions VR users have no idea where the virtual reality platform is physically located, where their data are processed and stored. Thus, virtual space may be outside the jurisdiction of the state where the victim applies to law enforcement.

Some authors commented that: “It will take a lot to get domestic police interested in investigating a crime where the criminal is in a foreign country. … Still, the greater difficulties of extradition are likely to exceed the greater ease of proof. And many VR street crimes might thus be practically ignored by traditional police departments” [17]. Thus, the multijurisdictional challenge is the multidimensional problem. There are a lot of questions as from legal assessment of committed offence as well from criminal procedure rules.

Any digital cross border space poses a challenge to law enforcement and the legislator. People from states with different legal systems meet in a virtual environment. Cross border character of VR spaces creates a situation where the crime has anonymous, remote in space, and global character. When virtual reality is applied to commit a crime, deanonymization and pursuing criminals become very difficult and even impossible.

The second feature of virtual reality as a place, tool, or instrument of crime is the capability of a virtual environment to create an illusion of reality. Sometimes this property forces real consequences for the potential victims of crime in VR.

Studies demonstrate that people uploaded at VR may be afraid of walking across a wooden plank that is on the ground in the real world but placed in virtual reality between roofs of skyscrapers. The experimental environment forces similar psycho-emotional reactions of a probationer as if it happens in the real world [18]. Dizziness, nausea, stomachache, and shaking body are just a few of the reactions to virtual heights [17]. Thus, experience in a virtual environment, in turn, may bring real consequences for body and mind. The VR customer may be literally frightened to a state of heart attack by a game in which he (she) feels quite real [19]. In case of the absence of consequences for the body, it may be a difficult psychologically traumatic experience. Thus, a virtual experience could cause permanent physical or psychological harm.

The realism of virtual environments has also positive effects, for example, VR is successfully applied for autism control [20] or in treatment of phobias [21]. But the side effect of VR realism is

that it is a perfect tool for hackers or other cybercriminals to "hack" into the victim's heads. Earlier, cybercriminals only had distant access to property, business reputation, and other values, but now they can literally get inside the brain. Harm also could be caused by the application of VR in medical treatment without direct intention [22].

In comparison with the real world, VR provides certain advantages for criminals in terms of the range of psycho-emotional reactions that could be provoked. In the physical world, it is very difficult to place a victim into the "Grand Canyon" and is even impossible to organize a meeting with a monster from science fiction. Hence, it is possible to distantly manipulate consciousness and emotions through VR by imitating any real or fictional environment. [23]. The character of such intervention is limited only by the imagination of the criminals. Current studies demonstrate that with the help of VR any emotion for up to 8 minutes could be forced [24].

Virtual worlds become more realistic due to the development of VR software and hardware. Objects are rendered with a higher degree of detail. More sophisticated models are applied to imitate interaction between objects in virtual reality. The link between events in VR and the body of the user gets more channels. Some researchers underline that higher realism implies a more significant psycho-emotional response of the customer on events in the computer environment [25].

As some researchers claim “Multimodal inputs will be combined in unprecedented patterns of stimulation that will ultimately make a VR experience more real than reality. Strong social emotions, in turn, will likely reinforce the sense of presence and make it easier to ascribe mental states to other virtual characters” [26]. Thus, intervention into human mental state is conceivably even more aggressive in future because new developments in sphere of VR technology.

The next challenge is development of haptic devices or other devices that are capable to affect the customer in the reality. Haptic technology enhances, new ones appear each year [27]. For example, there are enough investments in technologies for remote sexual intercourse. In the case of virtual environment application, skin sensations are networked between the users. The widespread use of such technologies implies that a criminal is potentially able to commit illegal intimate contact.

In some studies, conducted in the United States, it is underlined that hacker’s replacement of another user in a sexual act or unauthorized access to the user’s body in VR with the use of a haptic interface has similar characteristics with crimes against sexual freedom. In the legal research papers, such act committed with intention is considered to be either sexual harassment or rape [28]. Before haptic technology, it would be impossible to commit a rape distantly. Thus, haptic devices and similar equipment provide the capability to intervene in body inviolability.

The opinion exists that the above challenge could be overcome by a technical solution. Some authors state that “the ability to define default consent in software can make VR safer than the real world—for instance, well-designed software may let me consent in advance to certain types of touching but not others, or touching by some people but not others, and touching that isn’t consented to won’t even be felt. The question that remains is to what extent, if at all, it should be viewed as the victim’s job to set software consent boundaries; more on that below” [17]. Thus, the questions here: who would be responsible for the introduction of defense from body encroachments, what would be the liability of the developer or the seller for negligent informing about such kind of risk? And the main question is whether it is still crime or not?

All discussed above presuppose that VR allows to committing crimes which are impossible to commit with traditional IT. For instance, homicide could be committed on the Internet without VR, but not rape. A victim who has been raped in VR may not immediately realize what happens. Thus, a victim may be in a helpless state. As the victim is unable to understand that some illegal actions happen, he (she) is unable to take measures against the criminal. In some sense, it implies that rape or harassment in VR is probably even more dangerous because it is latent even for the victim.

Moreover, software and hardware for distant sex have low security against cyberattacks [29]. Often developers in this sphere are focused on new user engagement than on fixing vulnerabilities. It indicates that such kind of products could be easily hacked. Potentially criminals even could provoke VR users to move outside their secure boundaries in the reality through application of malware [30].

Due to the innovations in the VR sphere criminals are able to encroach not only on the mind but also on the body of victims. It is a complicated question whether the sexual assault in VR is comparable with real-world crime in terms of consequences for a victim. Of note, in the case of VR, there is a great number of difficulties in the prosecution of the offender. In addition, criminal significantly reduces the sum of risks related to the act (risk to be caught at the crime scene, risk of suffering if the victim resists, risk of public censure even if the guilt is not proven, and some others).

Development of haptic suits with feedback poses a risk of increasing harm-related crimes. The suit by electrical impulses informs the body of what happens with it in VR. The haptic device stimulates skin when some object touches the user in virtual world. Currently, there are no regulations on the permissible power of impact on the body. It is not legally defined what the degree of pain and harm is possible in VR. Of note, such a rule has to depend on the individual characteristics and other parameters. Due to the absence of regulation a risk of inflicting real suffering from haptic suits in virtual reality exists.

For instance, «new peripherals and technologies enter the market supporting a variety of interaction fidelities, such as data glove controllers, skeletal motion capture suits, tracking systems, and haptic devices» [31].

3.3. Immoral acts

As described above some manifestations of VR are not covered by modern legal regulations but there are some difficulties in addressing the mentioned challenges even in terms of moral rules. As law and morality are both means of conducting control, they are strongly linked. It does not imply that both social regulators have to correlate each other, but at least legislator has to take into consideration the norms of molality. And if a gap in the legal framework exists, VR has to be assessed from a moral point of view. A moral system of norms is clearly often very dynamic and opaque, but it is often helpful to know what is good or what is bad from a moral point of view. Moreover, the last social initiative in countries of Western Europe and Northern America as MeToo and

BLM demonstrate that moral norms could have even bigger impact on society than legal rules.

The first question is arising from identity flexibility in a virtual environment. In VR any user can choose any appearance that is provided by a software, risk exists that a person will apply the wrong appearance. At the same time, avatars and names influence on the perception of the user by other persons. These characteristics that are very significant for the attitude of other persons could be easily changed. It raises some ethical questions that are significant from ethical and sometimes legal point of view.

Imagine a teenager who uses VR poses as an adult and engages in an intimate conversation or even virtual sexual contact with an adult person. Who would be liable for such actions? Who has to take precautionary measures? Who must be in charge of the customer's appearance? What would the legal consequences of similar action be? Probably the legislator has to oblige VR developers to incorporate technical solutions for this kind of case. If one person misleads another person by his/her appearance in VR, how does it correspond with moral or even legal rules?

Even if users represent themselves in VR without misleading other persons, appearance also could be an issue. A demonstrative example is an avatar that looks like a naked body. Public nudity is banned by ethical and legal norms in most of the states, but a video demonstration of a naked body is rather not. Whether the naked body in VR is regarded as pubic nudity or is it a demonstration of video content from an ethical and legal perspective? Does the legislator have to oblige the developer to add measures against the use of such avatars?

A prosecution of public nudity in VR is harder from a legal and ethical point of view. What is a nature of this act? Factually it is a rendering of video content. However, a demonstration is accompanied by emulating personal presence. Thus, nudity in VR is not the same as nudity in games or movies. At the same time in terms of the intent of a person, it could be considered as public nudity. Virtual avatar is able to be real human size, with real human movements, seemingly standing close to other users. Mentioned avatars are perceived much more like a real person than a picture on a screen. Such actions may offend someone's feelings or cause psychological trauma in a demonstration for minors.

The issue of a proper avatar in VR is not limited to the nudity problem. For instance, an avatar may offend someone's feelings in other ways. It may be intolerant of gender, race, religion, political beliefs. More ways to offend or to cause suffering exist in VR. Imagine that someone's deceased relative's face is used in VR. Some users are able to use Hitler or famous serial killer's face for their avatars.

It is able to harm, disgust, or provoke users which see the avatar in VR. Appearance could be chosen as intentionally as well as recklessly. Developers can clearly provide technical measures to filter or block graphical representation of a user in virtual reality. But as it is linked with additional costs of software development, special regulation obliging VR companies has to be enacted.

Another problem is compliance with ethical norms applied in reality. As users all come to VR from the real world, some expectation exists on the conduct of the user in a virtual environment. There is an enormous temptation to transpose the moral norms of the ordinary world into the virtual world. Is it appropriate from a moral point of view to scream, create a mess, etc. in the VR art gallery?

There is a suggestion that "VR environment operators can easily implement code that can deal with the screamers. The operator could, for instance, allow each user to control the perceived volume, for that user, of any other user. That is good not just to silence the screamers, but also to quiet down acquaintances who are a bit too loud, or to amplify acquaintances who mutter. And this should be technically trivial to code" [17]. But it provokes the same question who will oblige the operator or the developer of the application to bear additional expenses on the implementation of these rules. Any state that implements such rules will automatically decrease chances in the race for technological dominance in the sphere of VR.

Of note, the issue of content filtering in VR is within the scope of the free speech agenda. When the legislator imposes additional legal requirements on VR developers, it implies reducing–or elimination–of the ability of individuals to speak freely in VR [32]. It is difficult to differentiate harmful behavior from an expression of someone's freedom of speech right on the level of development, as any case depends on context, viewer, or environment. Thus, a restriction for the developer or the operator is a threat to the freedom of speech in a virtual environment.

Another issue related to legal restrictions in VR is enforcement. The developer that designs or supports VR software or hardware is able to incorporate restrictions in its products; however, it is inevitably linked with real-time tracking of the user's conduct. It is presumably a high price for secure VR.

Development of restrictive measures will face also technical issues. Experience of companies which control and process of big data shows that "training moderators to overcome cultural biases or emotional reactions in the application of rules to facts can be analogized to training lawyers or judges. Regardless of the label, training content moderators involves a repetitive process to override cultural or emotional reactions and replace them with rational valid resolutions" [33]. So, an automatic realization of rules could produce unjust decisions, discriminative ruling, and other negative effects. It is probably more difficult to automatically make just decisions about the character of VR scenes than to create a transparent judicial system. This part of companies' work is not limited by the assertion that technical decisions have to be developed, the company has to create a division to manage disputes with customers which are not enjoyed by rule application.

Big IT companies are often criticized for a lot of power, but the capabilities of a virtual environment will allow private companies to have almost unlimited power over customers and their data. Factually, companies get control over all channels of human perception in different spheres from gaming to sexual relations. "VR settings mimicking as close as possible the multisensory nature of real-world human interactions would be a welcome and powerful addition to the field. Experiments aligning auditory, vestibular, and tactile signals with the visual domain are already improving VR immersion and realism" [26]. Thus, some private companies will control our reality, minds, and bodies.

4. Discussion

The study is just the beginning. This paper describes the key challenges related to managing risks of virtual environment use. VR is in the spotlight of social attention and attracts more attention from researchers. The study could be developed in different directions. There are several topics for further research. First of all,

it is necessary to systematize theoretical and practical legal knowledge related to the regulation of VR. It could include but not be limited to the research literature, legal practice, and international public agreements, as well as case-study.

Also, it is necessary to define the term "VR" and differentiate its categories and types, depending on the depth of immersion. It is helpful in terms of systematization and unification of results of current and future studies. It is necessary to categorize types of crimes that could be committed in VR. As some authors fairly comment on significance of legal definition and categorization of VR "it will serve as a basis for choosing the optimum regime of their legal protection" [34].

In terms of multidisciplinary dialogue, it is necessary to find out categories of the permissible technical parameters of VR functioning. It is very supportive in terms of preventing the destructive impact of technology on the psychophysiological state of the subject. For instance, there is a need to regulate the frame rate to avoid: headache, nausea, dizziness, epilepsy, and other disorders. It is especially critical for the application of VR by minors and juveniles.

Some authors claim that it is necessary "to add additional investigation and analysis testing stages to the development of virtual reality technologies in efforts to protect the public. These tests might not focus on physical health and safety concerns, but rather on physiological and social influences" [35]. However, in accordance with the analysis, the main concerns are mental and physical health. Moreover, any measures that take into consideration social influences would be very local and depend on the volatile social environment.

There is a need to define and regulate the allowable limits of the subject's functional permissibility in virtual reality (virtual ethical standard). Moreover, it is necessary to study the potential consequences of non-compliance with the prescribed rules of conduct, in order to prevent the direct projection of actions in VR in real life, such as the loss of boundaries of objective reality and the unconscious killing of a person in the real world.

In some papers, authors conclude that current law is very flexible and could be also applied to VR [22]. These studies do not take into consideration the dual nature of VR, i.e. this cybertechnology is rapidly merged with real world. The question arises in this study that current law is not applicable in cases where VR manifests itself in cyberspace and space of the real world. For instance, illegal access to haptic devices is not the same as usual "hacking".

There is research on restrictions of speech in VR in terms of US legislation. However, there is no conclusion how it corresponds to the First amendment and leaves a lot of space for further discussion [36]. In accordance with the analysis of current research in the sphere of medicine and psychology, it is difficult to conclude whether it is appropriate to apply the law of real world or virtual worlds for VR. The main reason is that virtual space and behavior of users there are often perceived as reality by other users.

In addition, there are several problems related to VR regulation that have to fall within the scope of criminal law. For instance, rules for the regulation of rendering in VR, application of real-world standards in VR, total control of users' conduct by IT giants, and appearance that misleads other users.

5. Conclusion

In summary, our study demonstrates that the offender gets new opportunities in VR. It literally helps criminals to hack the head or body of the victim. VR allows to provoke the emotions and to manipulate the consciousness of the victim at a completely new level. It is proved in the research that VR use could not be regulated as any other IT technology as the effect of perception by the victim is comparable in strength with the effect of events in the real world. In the paper characteristics of the virtual environment that have to be taken into consideration during a process of VR, regulation development is underlined.

Moreover, VR combined with haptic devices requires even more sophisticated regulation as such environment merges with the real world. Due to the integration of skin-feedback devices into the virtual environment, the consequences of VR's actions also have an influence on the user in the real world that could not be considered as just illegal manipulation with computer information. As many types of criminal acts with harm to the victim's body can now be committed remotely, a new kind of crime with dual nature is originated.

The article shows that there are not enough criminal law measures to combat crimes in VR. First of all, the legislation should provide for the obligations of the developer of VR technologies in order to minimize the possibility of criminals. Development and adoption of such measures is not an easy task, it is necessary to minimize economic costs in the implementation of these measures, and to prevent serious restriction of freedom of speech.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

The study was carried out with the financial support of the Russian Foundation for Basic Research in the framework of research project No. 18-29-16129

References

[1] R. Dremliuga, R. Iakovenko, N. Prisekina "Crime in virtual reality: Discussion", 2019 International Conference on Cybersecurity, ICoCSec 2019, 8970947, 81-85, 2019.

[2] A.O.J. Kwok, S.G.M. Koh "COVID-19 and Extended Reality (XR)", Current Issues in Tourism, 2020. https://www.tandfonline.com/doi/full/10.1080/13683500.2020.1798896

[3] Forecast augmented (AR) and virtual reality (VR) market size worldwide from 2016 to 2022 (in billion U.S. dollars). Statista, 2018. Retrieved October 26, 2020 from https://www.statista.com/statistics/591181/global-augmented-virtual-reality-market-size/.

[4] M. Meißner, J. Pfeiffer, C. Peukert, H. Dietrich, T. Pfeiffer "How virtual reality affects consumer choice", Journal of Business Research, 117, 219-231, 2020.

[5] M. Figueiredo, R. Mafalda, A. Kamensky "Virtual reality as an educational tool for elementary school", Smart Innovation, Systems and Technologies, 198 SIST, 261-267, 2021.

[6] Y.P. Zinchenko, P.P. Khoroshikh, A.A. Sergievich, S.A. Gutnikov, K.S. Golokhvast "Virtual reality is more efficient in learning human heart anatomy especially for subjects with low baseline knowledge", New Ideas in

Psychology, 59,100786, 2020. DOI: 10.1016/j.newideapsych.2020.100786
[7] S. Ahn, J. Bostick, E. Ogle, K.L. Nowak, K. McGillicuddy, J. Bailenson "Experiencing Nature: Embodying Animals in Immersive Virtual Environments Increases Inclusion of Nature in Self and Involvement with Nature", Journal of Computer-Mediated Communication, **21**(6), 399-419, 2016. https://doi.org/10.1111/jcc4.12173
[8] M.C. Howard "A meta-analysis and systematic literature review of virtual reality rehabilitation programs", Computers in Human Behavior, **70(**1), 317-327, 2017.
[9] B. Ticknor Virtual Reality and Correctional Rehabilitation: A Game Changer", Criminal Justice and Behavior, **46**(9), 1319-1336, 2019.
[10] B. Ticknor "Using virtual reality to treat offenders: An examination", International Journal of Criminal Justice Sciences, **13**(2), 316-325, 2018.
[11] G.P. Otjutsky "Understanding the virtual reality: methodological difficulties", Historical, philosophical, political and legal sciences, cultural studies and art criticism. Issues of theory and practice, **2-2** (52), 161-170, 2015.
[12] E.A. Solodkina "To the definition of the concept of virtual reality", Bulletin of the Peoples' Friendship University of Russia. Series: Philosophy, (1), 190-194, 2004.
[13] V. Jaillot, M. Istasse, S. Servigne, M. Rautenberg, I. Lefort, "Describing, comparing and analysing digital urban heritage tools: A methodology designed with a multidisciplinary approach", Digital Applications in Archaeology and Cultural Heritage, 17,e00135, 2020.
[14] P.P. Shrestha, B. Davis, G.M. Gad, "Investigation of Legal Issues in Construction-Manager-at-Risk Projects: Case Study of Airport Projects", Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, **2**(3),04520022, 2020.
[15] E. Fritsch "Growth in wearable technologies", Screen Printing, 110(5), 22-26, 2019.
[16] R. Dremliuga "Subculture of Hackers in Russia", Asian Social Science, **10**(18), 158-162, 2014. DOI: 10.5539/ASS.V10N18P158
[17] Mark A .Lemley, E. Volokh "Law, Virtual Reality, and Augmented Reality", University of Pennsylvania Law Review, **166**, 1051-1138, 2018. Available at SSRN: https://ssrn.com/abstract=2933867 or http://dx.doi.org/10.2139/ssrn.2933867
[18] Q . Wang, F. Hu, H.Wang "Are you afraid of heights and suitable for working at height?", Biomedical Signal Processing and Control, 52, 23-31, 2019.
[19] T.W. Cleworth, R. Chua, J.T. Inglis, M.G. Carpenter "Influence of virtual height exposure on postural reactions to support surface translations", Gait and Posture, 47, 96-102, 2016.
[20] S.N.V. Yuan, H.H.S. Ip "Using virtual reality to train emotional and social skills in children with autism spectrum disorder", London Journal of Primary Care, **10**(4), 110-112, 2018.
[21] J.-M. Park, G.-Y. Noh "The effects of vr simulation game on driving fear and efficacy", JP Journal of Heat and Mass Transfer, **15**(Special Issue 2), 161-169, 2018.
[22] Y.Y. Lim, P. Poronnik, T. Usherwood, B. Reeve "Medical negligence laws and virtual reality in healthcare", Australian journal of general practice, **49**(8), 525-529, 2020.
[23] R. Ueoka, A. AlMutawa, "Emotion hacking VR: Amplifying scary VR experience by accelerating actual heart rate", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10904 LNCS, 436-445, 2018.
[24] B. Meuleman, D. Rudrauf "Induction and profiling of strong multi-componential emotions in virtual reality", IEEE Transactions on Affective Computing, 2018, Available at: https://ieeexplore.ieee.org/document/8432120/
[25] J. Hvass, O. Larsen, K. Vendelbo "Visual realism and presence in a virtual reality game", 3DTV-Conference 2017-June, 1-4, 2018.
[26] A. Monti, S.M. Aglioti "Flesh and bone digital sociality: On how humans may go virtual", British Journal of Psychology, **109**(3), 418-420, 2018.
[27] A. Leleve, T. McDaniel, C. Rossa "Haptic Training Simulation", Frontiers in virtual reality, **1**(3), July 2020. https://doi.org/10.3389/frvir.2020.00003
[28] L. Strikwerda "Present and Future Instances of Virtual Rape in Light of Three Categories of Legal Philosophical Theories on Rape", Philosophy and Technology, **28**(4), 491-510, 2015.
[29] M.J. Faustino "Rebooting an Old Script by New Means: Teledildonics—The Technological Return to the 'Coital Imperative'", Sexuality and Culture, **22**(1), 243-257, 2018.
[30] M.U. Rafique, S.-C.S. Cheung "Tracking Attacks on Virtual Reality Systems", IEEE Consumer Electronics Magazine, **9**(2), 8977814, 41-46, 2020
[31] D. Duggan, C. Kingsley, M. Mazzeo, M. Jenkins "Exploring Extended Reality as a Simulation Training Tool Through Naturalistic Interactions and Enhanced Immersion", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11575 LNCS, 272-282, 2019.
[32] Jack M. Balkin, "Information Fiduciaries and the First Amendment", UC Davis Law Review, **49**(4), 183-1234, 2016. Available at SSRN: https://ssrn.com/abstract=2675270.
[33] K. Klonick. "The new governors: The people, rules, and processes governing online speech", Harvard Law Review, **131**(6), 1599-1670, 2018.
[34] T.A. Kochubey, "Legal protection of virtual and augmented reality technologies", Advances in Intelligent Systems and Computing, 1100 AISC, 184-191, 2020.
[35] B. Kenwright "Virtual Reality: Ethical Challenges and Dangers [Opinion]" IEEE Technology and Society Magazine, **37**(4),8558774, 20-25, 2018.
[36] M.J. Blitz "The First Amendment, video games, and virtual reality training" in "Research Handbook on the Law of Virtual and Augmented Reality", Edited by W. Barfield and M.J. Blitz, Edward Elgar Publishing, 712 p.2018.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1734-1740 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

A Machine Vision Approach for Underwater Remote Operated Vehicle to Detect Drowning Humans

Yaswanthkumar S K*, Keerthana M, Vishnu Prasath M S

Department of Electrical & Electronics Engineering, Kumaraguru College of Technology, Coimbatore, 641049, India

ARTICLE INFO

Article history:
Received: 15 September, 2020
Accepted: 29 December, 2020
Online: 30 December, 2020

Keywords:
Remote Operated Vehicle
Image Processing
Drowning People
Semi-Autonomous

ABSTRACT

Today, Drowning is the 3rd major cause for unintentional injury death accounting for 7% of deaths of all injury deaths. Drowning is a state of suffocation when water or other fluids accumulate the lungs, resulting in respiratory impairment, ultimately leading to death. The predominant problem during rescue operations of such accidental drowning is to locate or track the person underwater, facilitated with the help of our bare eye sight, which makes the process too cumbersome. Also, in case of moving water bodies such as rivers and sea, the process of tracking the person becomes too difficult. Thus, there is pressing need for development of an engineered solution to solve this problem. This research involves development of such a robotic technology which will facilitate the work of locating the position of the drowning person underwater. This method comprises of Image processing along with GPS tagging embedded on a Remote Operated Vehicle maneuvering underwater at the site of accident to detect the exact location of the drowning person along with the GPS coordinates in order to ease the process of rescue. This robotic system was tested on a real time environment in order to demonstrate the efficacy of the system under practical circumstances. Thus, this system can be used at places where emergency rescue operations are needed.

1. Introduction

Drowning accounts for 7% of all deaths that occur due to injury. It is also the 3rd major cause for unintentional injury death. As drowning is an unexpected event, there are neither any precautionary steps to prevent it, nor any existing technology to initiate rescue operations rapidly. According to stats from 2016 nearly 350000 people are reported dead due to drowning. In India, Drowning is the second largest cause of death for children below age of 15. The situation is even worse at Russia. The drowning mortality rate is of more than 3.9 per 10000 individuals.

Nowadays, few assistive technologies are available in the market to provide safety for humans in or under water. But these solutions are based on scenarios where people must use this technology before nearing any water bodies. But since Accidental Drowning are unpredictable and the underwater environment varies from place to place a solution adaptable to all water bodies is the main focus of this work. Development of this technology can be used not only on rescue operations of accidental drowning, but also to mitigate the risk of loosing track of people during natural calamities like heavy floods or Tsunamis. Thus, there is a urgent need to focus on development of a solution for the above mentioned scenario.

Our research work is to embedded Image Processing methodologies to detect drowning people underwater and to use GPS coordinates of their exact location with the use of a Robotic Remote Operated Vehicle which can by controlled autonomously or semi autonomously to navigate through water bodies during the rescue process. Also, this system uses a on board camera, live camera feed can also be surfaced over radio communications to the rescue team, which would prove to be great advantage to the people on the field for the rescue operations. All the existing methods are either not cost-effective or have installation issues at underwater environments.

2. Survey on Previous Researches

This paper is an extension of work [1] originally presented in 2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), where a different method for the detection of humans drowning underwater was illustrated.

In [2], is a research work that involves development of an automated system that detects and tracks objects that are of potential interest for human video annotators. By pre-selecting salient targets for track initiation using a selective attention algorithm, we reduce the complexity of multi-target tracking, in

*Corresponding Author: Yaswanthkumar S K, Email: yaswanthsk4@gmail.com

particular of the assignment problem. Detection of low-contrast translucent targets is difficult due to variable lighting conditions and the presence of ubiquitous noise from high-contrast organic debris ("marine snow") particles.

In [3], the paper presents an attentional selection system for processing video streams from remotely operated underwater vehicles (ROVs). The system identifies potentially interesting visual events spanning multiple frames based on low-level spatial properties of salient tokens, which are associated with those events and tracked over time. Using video cameras, it is possible to make quantitative video transects (QVTs) through the water, providing high-resolution data at the scale of the individual animals and their natural aggregation patterns.

In [4], this paper explains an agent system is strategy to enhance the underwater manipulation. If the location of the agent can be measured, the end effector is able to be place to any position. To implement this system, the method of an agent vehicle localization is proposed. The method uses the sonar images of moving agent obtained by forward-looking sonar. To detect the location of the agent in the sonar images, the convolutional neural network is applied. Through field experiment, we confirm the proposed method can detect and track the agent in the successive sonar images.

In [5], This paper presents methods applied for automated detection of fish based on cascade classifiers of Haar-like features created using underwater images from a remotely operated vehicle under ocean survey conditions. The images are unconstrained, and the imaging environment is highly variable due to the moving imaging platform, a complex rocky seabed background, and still and moving cryptic fish targets. Several Haar cascades are developed from the training set and applied to the validation and test video images for evaluation.

In [6], This paper presents a robotics vision-based system for an underwater pipeline and cable tracker. This system is introduced to improve the level of automation of underwater remote operated vehicles (ROVs) operations which combines computer vision with an underwater robotics system to perform target tracking and intelligent navigation. This study focuses on developing image processing algorithms and nonlinear control method for tracking the pipeline or cable. By using the adaptive sliding mode controller, the ROV can easily track the cable or pipeline.

In [7], is a proposed system of a multi-step and all-round underwater image processing specially for the underwater images taken in succession to improve the image quality, remove the dynamic interference and reconstruct the image. For images with complex texture, the inpainting result brings out much smoothness, at the cost of massive details. For images with low resolution, the transition of color near the edge of the vacancy is abrupt

In [8], The paper presents a wavelet-based fusion method to enhance the hazy underwater images by addressing the low contrast and color alteration issues. The publicly available hazy underwater images are enhanced and analyzed qualitatively with some state-of-the-art methods. Green color dominates the Image in the evaluated set of images. In the results, it can be observed that the natural color profiles of Image are preserved by correcting the dominated color and it was hard before to distinguish the color of the water and objects

In [9], it shows that man-made objects in the image can be distinguished from natural objects. This paper proposes that a control strategy to adjust image processing parameters, and sequence multiple applications of tools, is especially important for digital processing of underwater images. This paper highlights the key difficulties in processing underwater images such as obscuration of objects of interest by fish / sand / coral / algae in the water and change of illumination, color balance and scale of fish images due to refraction, absorption in water and large range of distances of observation and different sizes of fish

3. Schematic & Working of ROV

As shown in Figure 1. The devised system consists of a on board camera embedded to the ROV to communicate the live video feed to the operator and also for the video to be processed by the on-board Processor to detect for humans underwater. The main system consists of sub units such as:

- Transmission & Receiving Unit
- Processor Unit
- Sensing Unit
- Maneuvering Unit
- Powering Unit

Each of which will be discussed on the following sections.

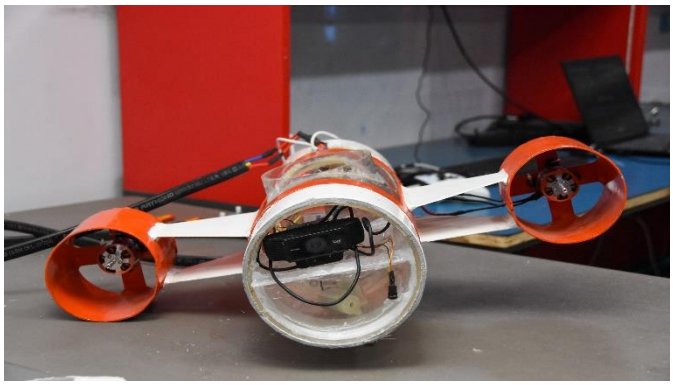


Figure 1: Hardware assembly of ROV

3.1. Transmission & Receiving Unit

A). Introduction

The transmission Unit consists of a Two axis controller to control the movement of the ROV along with a Screen or a monitor to view the live feed from the camera on-board which is interfaced to a Raspberry Pi through USB communication. A PC Joystick controller with TWO axis was used to provide the control signals as a input ls to the motor to control the movement of the ROV. Common PC monitor was used along with the raspberry PI to view the live feed from the ROV's on-board camera. The Receiver Unit Consists of a Raspberry PI which is on-board in the ROV. The Video data from the camera is given as an input to this raspberry pi from which is then transferred to the other raspberry pi through wireless communication.

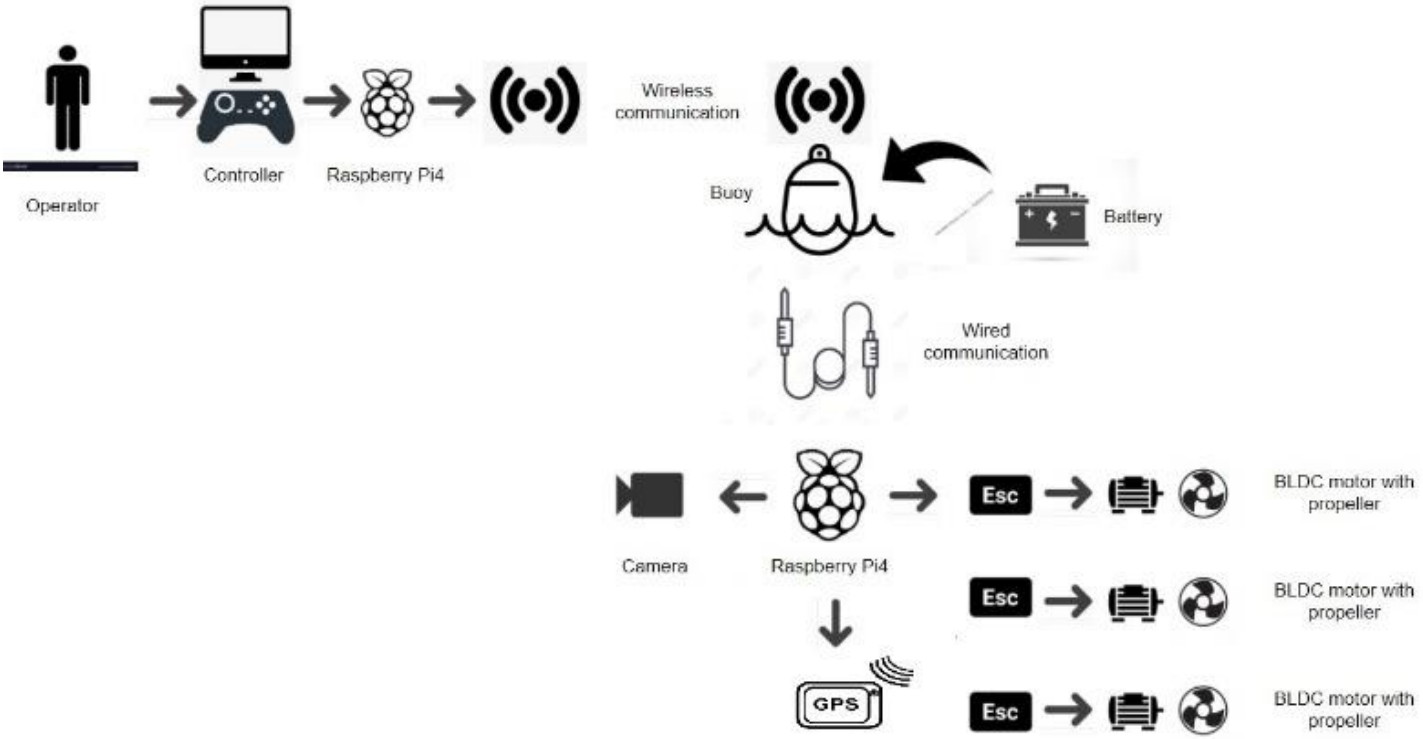


Figure 2: Schematic Diagram

B). Working

The control signals from the operator are provided as joystick input through USB communication to Raspberry PI on the Transmitter side. These signals are then transmitted to the other Raspberry PI on the Receiver side through WIFI Communication. This transfer of data can also be performed using RF transmission or MQTT via internet which will provide further distance range between the operator and the ROV. The operator can view the live feed in the monitor, to control the ROV effectively. The video feed from the USB Camera is also transmitted wirelessly through WIFI Communication from the Receiver Raspberry Pi to Transmitter Raspberry Pi.

3.2. Processor Unit

A). Introduction

The Processor unit consists of a Raspberry Pi (receiving unit) on-board in the ROV. This Receives the control signals from the Operator which is then used to control the movement of the ROV. This also receives data from the on-board camera and on-board GPS for further transmission.

B). Working

As the PWM from the Transmitter unit are received by the Processor Unit (receiving unit), these signals are the compared with the on-board accelerometer to measure the difference in the values. These values are again corresponded to respective PWM signals and then is given as an input to the respective Electronic Speed Controllers (ESC's) which control the spin of the motors. The Processor also performs the function of receiving the input data from camera and GPS on the ROV and transfer them to the operator.

3.3. Sensing Unit

A). Introduction:

The sensing unit consists of a IP68 waterproof high megapixel camera to be interfaced with the Processor. This will provide the video data for Image Processing and as well as send the live data from underwater to the operator. It also consists of a waterproof GPS chip to be enclosed in an air tight chamber of ROV. This provides the GPS coordinates of the ROV underwater.

B). Working

The camera transmits the data to the processor through USB communication, which is then used as the input for Image Processing algorithm and is also transmitted to the operator. The GPS module sends the signal to raspberry pi through SPI protocol which is also sent to the operator through wireless Communication.

3.4. Maneuvering Unit

A). Introduction

Maneuvering unit consists of 60 Amps ESC's, each individual for respective 500 kV BLDC motors. As the PWM control signals are given as an input into ESC, the speed of the motor varies accordingly the position and moment of the ROV also varies.

B). Working

As the operator provides the input signals through the JOYSTICK controller, this input data from the Transmitter Raspberry pi, reaches the Processor Raspberry pi (receiver unit). Those input signals are then provided as an input to the respective ESC after being compared with the current accelerometer value &

then the required amount of power will be delivered to the respective motor which is now being controlled by ESC.

3.5. Powering Unit

ROV is completely powered through wires as it is being tested for its buoyancy when battery pack is being placed inside the air tight enclosure of the ROV.

4. Imaging Methodology to Detect Drowning Humans in a - underwater environment

Image Processing has overcome the limits of its own discipline. Today, it is possible to process a live video feed right after few trained models. Due to this rapid increase in its uses, this technology is pressured to produce strong methodologies which would offer a wide variety of solutions for real time problems. Our methodology involves usage of Faster Region Convoluted Neural Network (FRCNN) algorithm.

4.1. Hardware Specification

The hardware specifications are illustrated in table 1.

Table 1: Hardware Specification

System	Dell G3 3579
GPU	Nvidia GeForce GTX 1050Ti
CPU	Intel Core i7-87507H
Memory (RAM)	8192 MB
Display Memory (VRAM)	4018 MB
Operating System	Windows 10 64-bit

4.2. Dataset

As only humans are to be detected underwater a dataset consisting of 800 images is used as dataset. These 800 images comprise a variety of underwater images with varying light intensity, water color, and humans in it. This dataset also contains images of halfway human pictures along with overlapping images too.

A). Data Preprocessing:

Reading

Using the matplotlib library on a python script is used to reading all the images from the dataset. This helps us to know whether all the all the images from the dataset are loadable and could be read by the algorithm.

Re-sizing

As larger images would take longer time to process, all the images collected are res-sized to be less than 200KB in size.

B). Data Labelling

The datasets must be labelled, since labelled images must be provided as input to train the training algorithm. In order to label the datasets, the images are first loaded onto the LabelImg tool. Now, humans in each image are to be bounded by drawing bounding boxes around them as shown in Figure 3. Then, the labelled data of all these images are to be stored in an XML format. Each image file would create individual XML files containing the File name, path, the label and coordinates for the bounding boxes. These are also called annotation files which will be used to train the algorithm.

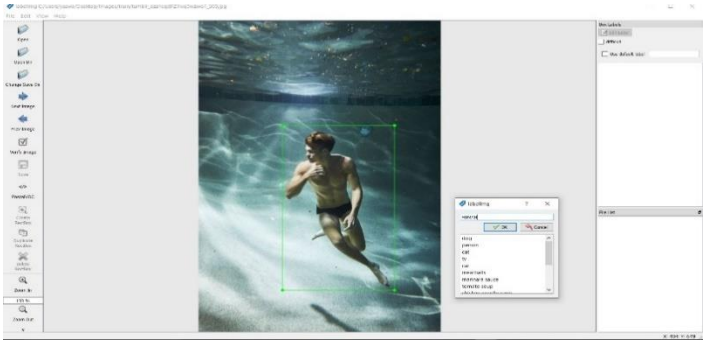


Figure 3: Labelling the datasets

4.3. Implementation

In order to bring out the best outcome, two algorithms were used 1). Faster RCNN and Yolo V3. Both these algorithms were implemented and trained using the TensorFlow Object Detection API. This makes the training process faster as TensorFlow object detection API comes with the model called ZOO which is a collection of models and algorithms.

A). Configuring the training algorithm

In order to detect the best algorithm which can be used, both the Faster RCNN and YOLO algorithms were configured to the same values.

Data Slicing

Both the algorithms were trained and tested using the same data sets. So, for both the algorithms 75% of the images and 25% of the images were used as training datasets and testing datasets respectively.

Classes

Since, humans were the only object to be detected by the algorithms, the class for these algorithms was set to 1.

Learning Rate

Learning rate is a parameter which allows us to control the model's response to estimated error each time the model weights are updated. The learning rate was set to a default value of 0.0002 while training both the algorithms.

Batch Size

The batch size corresponds to the consumption of VRAM, after looking into Table 1. The value set of batch size of the algorithms is set as 4.

4.4. Training Procedure

Figure 4 shows the training process of Faster RCNN algorithm along with the classification loss and the time taken after each iteration.

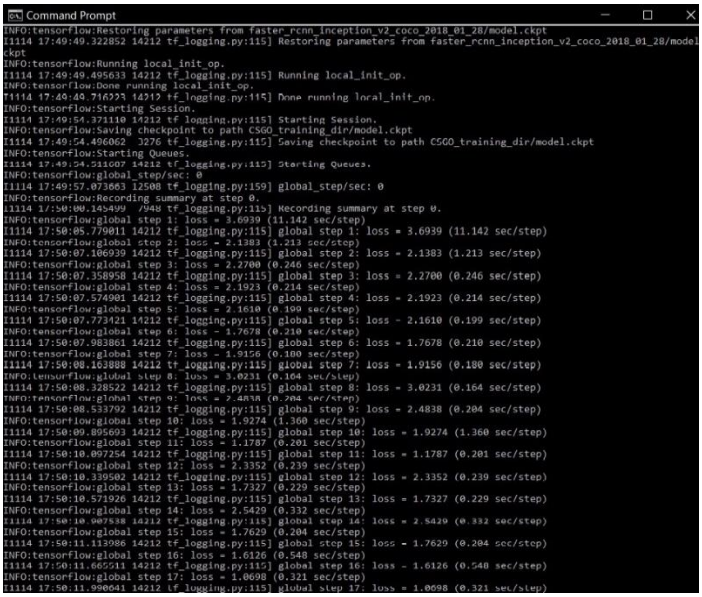


Figure 4: Training Process of Faster RCNN

As you can see, Classification loss is high when the training processes started. But after each step, there is a gradual decrease in the classification loss as the algorithm kept learning in each step. Figure 4. shows the exact decrease in classification loss during training process compared along with the testing process.

4.5. Performance Metrics

In order to compare the results of these two models, accuracy of these two models are to be contrasted. To calculate the accuracy few parameters are to be considered as metrics:

1. True Positive TP: A human is present in the Image and, the algorithm detects the human.
2. False Positive FP: No human is present in the Image, but the algorithm detects human.
3. False Negative FN: A human is present in the Image, but the algorithm does not detect the human.
4. True Negative TN: No human is present in the Image, and nothing is being detected.

A). Accuracy:

It is a measure to check whether the model or the algorithm is trained correctly to detect humans in an underwater environment or not. The formula to calculate the accuracy as follows:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

B). Precision

It is a ratio of positively predicted images that contain humans in it to that of which actually contains humans in it. The formulae to calculate precision as follows:

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

C). Recall

It is the ratio of images that actually contain human in it to that of which were predicted positively. The formula to calculate recall as follows:

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

D). F_1 Score

It is the measure of accuracy of the model combining both precision and recall. A model or algorithm is considered to be perfect if it's F_1 Score is 1. The formula to calculate F1 Score as follows:

$$F_1 \text{ Score} = 2 * \left(\frac{\text{Percision} * \text{Recall}}{\text{Persion} + \text{Recall}}\right)$$

5. Results

The results for both the algorithms are summarized in this section. The analysis of results as follows:

5.1. Faster RCNN Results

Figure 5, Figure 6 & Figure 7 are a collection of test data from 3 different samples under varying light conditions and other environmental factors for faster RCNN algorithm.

Figure 5: RCNN Output 1

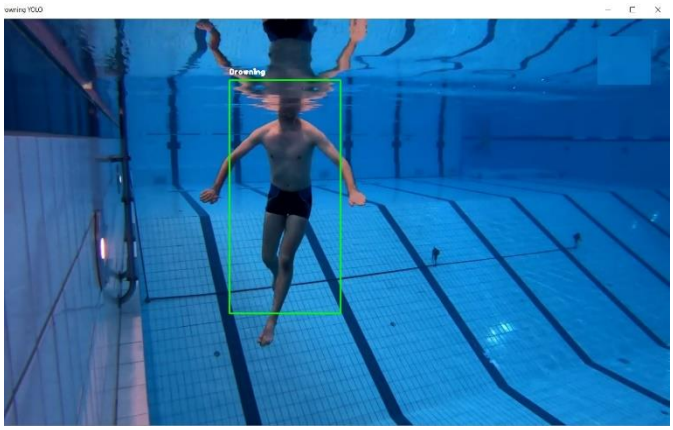


Figure 6: RCNN Output 2

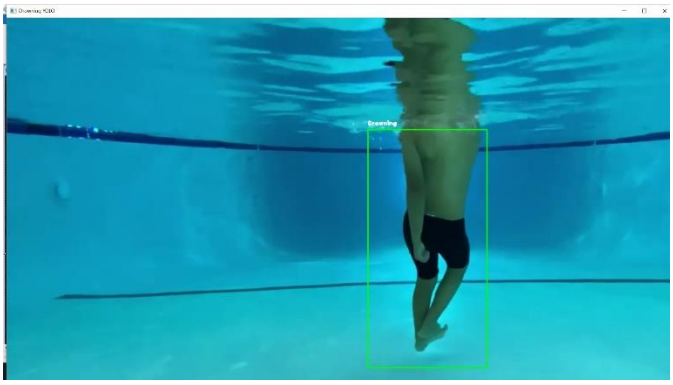


Figure 7: RCNN Output 3

The number of TP, TN, FP, FN from RCNN is tabulated in the table 2. These are further used to analysis the algorithms.

Table 2: RCNN Output Tabulation

#	True Positives	True Negative	False Positive	False Negative
Total	148	14	9	29

5.2. Yolo V3 Results

Figure 8, Figure 9 & Figure 10 are a collection of test data from 3 different samples under varying light conditions and other environmental factors for faster RCNN algorithm.

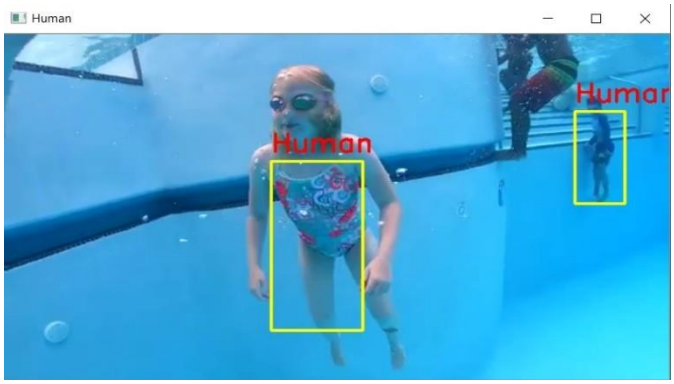


Figure 8: Yolo V3 Output 1

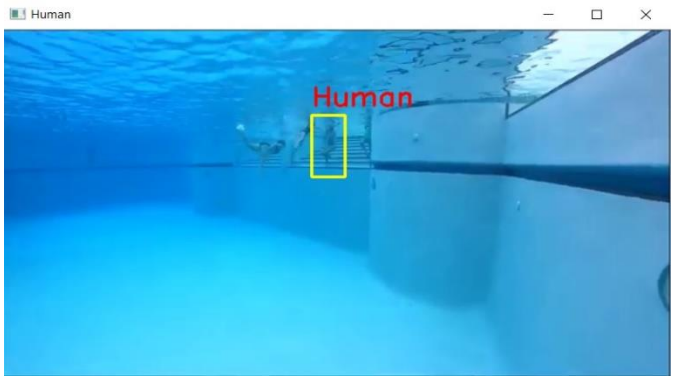


Figure 9: Yolo V3 Output 2

The number of TP, TN, FP, FN from YoloV3 is tabulated in the table 3. These are further used to analysis the algorithms.

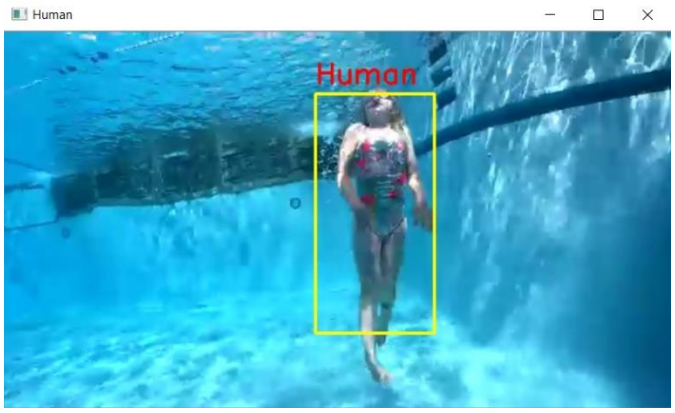


Figure 10: Yolo V3 Output 3

Table 3: YoloV3 Output Tabulation

#	True Positives	True Negative	False Positive	False Negative
Total	125	12	3	60

5.3. Analysis of Results

Table 4, provides us the collection of results containing the number of TP, TN, FP, FN parameters obtained by Faster RCNN and Yolo V3 algorithms.

Table 4

#	True Positives	True Negative	False Positive	False Negative
F - RCNN	148	14	9	29
YOLO V3	125	12	3	60

From this table, the accuracy of Faster RCNN and YOLO V3 algorithms are 81% and 68.5% respectively. Clearly, Faster RCNN is more Accurate than YOLO V3 in detecting humans in a underwater environment. It is also to note that Faster RCNN was able to consistently detect Humans if overlapping occurs. While, YOLO V3 failed to detect humans during overlapping frames.

Table 5

Algorithms	Precision	Recall	F_1 Score
Faster RCNN	0.94	0.83	0.881
YOLO V3	0.97	0.67	0.792

Table 5, Shows us that YOLO V3 is more precise compared to that of Faster RCNN, but then Faster RCNN has a high recall value than that of YOLO V3. To compare both the models, F_1 Score was calculated which also points out that Faster RCNN algorithm is the Best fitted algorithm to detect humans underwater. The results contribute to some preliminary findings to the growing field of underwater robotics in real time environments.

6. Future Work

The experimentation has been tested only on clear water bodies. The implementation of this ROV under real time circumstances such as lakes and river where the opacity will be translucent or opaque, it is better to use the thermal imaging as the

video input to this image processing models. As the body temperature of the aquatic species in that water body will be the similar, it will be easy to differentiate human body using thermal sensors and video feed of thermal imaging which will be performed in the near future.

7. Conclusion

This paper illustrates, an idea to develop a robotic system which will detect drowning people underwater at the times of an emergency. It also concludes that Faster RCNN is the best fit algorithm which can be used for this practical application. In our further research we plan to our work from a single ROV into a swarm of ROV's collectively working at a site of emergency to detect every drowning human underwater. We are also planning to enable full autonomous nature of the ROV in order to minimize the reaction time and communication time between the ROV and the Operator.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We would like to thank Forge Accelerator for providing the necessary support to complete this research work. We also thank Mr. Ravi Shankar, Mr. Koushal & Miss. Poovika for contributing to this research work by building the ROV from Scratch.

References

[1] S.K. Yaswanthkumar, O.K. Praveen, R. V. Rohit, "Autonomous Utility Vehicle (Auvs) Based Emergency Human Drowning Detection System Using Sonar and Thermal Detection Methods," in 2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics, DISCOVER 2019 - Proceedings, 2019, doi:10.1109/DISCOVER47552.2019.9007992.

[2] D. Walther, D.R. Edgington, C. Koch, "Detection and tracking of objects in underwater video," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, doi:10.1109/cvpr.2004.1315079.

[3] D.R. Edgington, K.A. Salamy, M. Risi, R.E. Sherlock, D. Walther, C. Koch, "Automated event detection in underwater video," in Oceans Conference Record (IEEE), 2003, doi:10.1109/oceans.2003.178344.

[4] J. Kim, S.C. Yu, "Convolutional neural network-based real- Time ROV detection using forward-looking sonar image," in Autonomous Underwater Vehicles 2016, AUV 2016, 2016, doi:10.1109/AUV.2016.7778702.

[5] G. Cutter, K. Stierhoff, J. Zeng, "Automated detection of rockfish in unconstrained underwater videos using haar cascades and a new image dataset: Labeled fishes in the wild," in Proceedings - 2015 IEEE Winter Conference on Applications of Computer Vision Workshops, WACVW 2015, 2015, doi:10.1109/WACVW.2015.11.

[6] R. Trahan, R. Kiang, "An analysis of the die testing process using taguchi techniques and circuit diagnostics," in Proceedings - International Test Conference, 1992, doi:10.1109/TEST.1992.527832.

[7] C. Cai, Y. Zhang, T. Liu, "Underwater image processing system for image enhancement and restoration," in 2019 IEEE 11th International Conference on Communication Software and Networks, ICCSN 2019, 2019, doi:10.1109/ICCSN.2019.8905310.

[8] A. Khan, S.S.A. Ali, A.S. Malik, A. Anwer, F. Meriaudeau, "Underwater image enhancement by wavelet based fusion," in USYS 2016 - 2016 IEEE 6th International Conference on Underwater System Technology: Theory and Applications, 2017, doi:10.1109/USYS.2016.7893927.

[9] R.E. Blake, "Digital image processing of underwater images," in 2006 IEEE US/EU Baltic International Symposium, BALTIC 2006, 2006, doi:10.1109/BALTIC.2006.7266195.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1741-1749 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

An Evaluation of some Machine Learning Algorithms for the detection of Android Applications Malware

Olorunshola Oluwaseyi Ezekiel[*,1], Oluyomi Ayanfeoluwa Oluwasola[2], Irhebhude Martins[3]

[1]*Cyber Security Department, Faculty of Computing, Air Force Institute of Technology, Kaduna, 800283, Nigeria*

[2]*Electrical Electronics Engineering Department, Air Force Institute of Technology, Kaduna, 800283, Nigeria*

[3]*Computer Science Department, Nigerian Defence Academy, Kaduna, 800283, Nigeria*

ARTICLE INFO

Article history:
Received: 08 October, 2020
Accepted: 25 December, 2020
Online: 30 December, 2020

Keywords:
Mobile phone
Malware
Algorithms
Operating System
Benign

ABSTRACT

Android Operating system (OS) has been used much more than all other mobile phone's OS turning android OS to a major point of attack. Android Application installation serves as a major avenue through which attacks can be perpetrated. Permissions must be first granted by the users seeking to install these third-party applications. Some permissions can be subtle escaping the attentions of the users. Some of these permissions can have adverse effects like spying on the users, unauthorized retrieval and transference of the data and so on. This calls for the need of a heuristic method for the identification and detection of malware. In this discourse, testing of classification algorithms including Random forest, Naïve Bayes, Random Tree, BayesNet, Decision Table, Multi-layer perceptron (MLP), Bagging, Sequential Minimal Optimization (SMO)/Support-Vector Machine (SVM), KStar and IBK (also known as K Nearest Neighbours classifier (KNN)) was carried out to decide which algorithm performs best in android malware detection. Two dataset was used in this study and were gotten from figshare. They were trained and tested in the Waikato Environment for Knowledge Analysis (WEKA). The performance metrics used are Root Mean Square Error (RMSE), Accuracy, Receiver Operating Curve (ROC), False positive rate, F-measure, Precision and recall. It was discovered that the best performance with an accuracy of 99.4% was the multi-layer perceptron on the first dataset. Random Forest has the best performance with accuracy, 98.9% on the second dataset. The implication of this is that MLP or random forest can be used to detect android application malwares.

1. Introduction

This work is an extension of [1] in which this work considers another dataset in carrying out the research. This research answers the questions from the comments received about the biasness of the algorithms used.

There has been a considerable increase in the usage of mobile phones year on year. Mobile phones have served as a medium of communication through calls and short mailing services (SMS) [2]. Technological advancement has increased the usage of mobile phones for various purposes. The operating system (OS) of mobile phones have evolved from java to Symbian with the current ones being android OS, windows OS and Apple OS – iOS [3]. The usage of mobile phones spans from individuals to organizations as well. This study focuses on Android OS. Mobile phones with Android OS comprises of some components which are; the

Android OS, the middleware and key applications [4]. The android OS has the biggest and fastest growing smartphone OS market, hence it has become a practical target for cyber criminals [5]. Large support to third party applications is also gotten from android. These third party applications can easily be downloaded on the popular Google Play store [6]. The developers of the Android OS do not develop third-party applications; it is done by software developers. Hence the name of the third party applications [6]. In [7], there are over 2.7 billion mobile users globally, and in 2022, 258.2 billion apps are projected to be downloaded. Android phones are useful for crucial functions ranging from financial transfers to e-commerce transactions

[*]Corresponding Author: Oluwaseyi Ezekiel Olorunshola, Electrical Electronics Engineering Department, Air Force Institute of Technology, Kaduna, 800283, Nigeria, +2348068341127 & seyisola25@yahoo.com

(goods payments) and so on [8]. It also allows other features, such as photo editing, game play, social media and messaging, internet searching, typing, listening, reading, and so on. Any of all these procedures are feasible due to the implementation of third-party apps.

The growth of android application malware has been made possible by the free application market for the Android OS [9]. The most severe and important threat to the Android OS arises as a result of malicious applications especially from unauthorized applications market that can easily be installed on any phone [9]. Android applications depend on users to grant permission to work optimally. Such permissions can forewarn users to the applications' secrecy and/or security [10]. Android malware can impersonate a legitimate program, but with intrusive permissions for which an agreement is required from the users before installation can be done [10]. The malware executes activity(s) that are malicious through these permissions. SMS spy pro is an example, masking itself as a tip calculator but forwarding all sent SMS and received SMS by the user to a third party [11]. Android malware can access and steal user data such as browsing history, SMS, location of GPS, phone records, e-mails, other application information (Like Facebook log-in details) and more [12]. Tasks such as snapping and sending photographs, sending unauthorized e-mails, conducting bank transfers without user awareness can also be executed [2]. The degree to which both malicious and innocuous applications can execute their functions is primarily dependent on the form of permissions that users grant them.

Considering the sensitivity of the data that can be colleted from users via the malwares combined with the kind of damage that can be done by these malwares (for example, spying, identity theft, extortion, logic bomb) [13], these malwares need to be identified effectively. Due to zero-day attacks, the conventional malware detection techniques that uses the malware signature to store them in a repository has proved to be ineffective [9]. Zero-day attacks are based on new flaws found by cyber attackers that the creator is unaware of and thus has not issued a patch [14]. Therefore, until the developers are aware, the attackers make the most of these vulnerabilities. Android malware with its corresponding permission requests must be found [9]. This study is focused on Olorunshola and Oluyomi's recommendations in [1]. An evaluation of the classification algorithms with regard to their performance in the identification of malicious software based on their manifested permission, command signature, API call signature and intents, is the aim of this study. This study aims to verify the conclusions drawn by Olorunshola and Oluyomi by training and testing the algorithms with additional datasets in this research. Two android datasets based on manifest permission, command signature, API call signature and intent are obtained from figshare created by [15] to achieve this aim. The algorithms are trained and tested on the datasets. The remaining portion of this work is structured as follows: review of literature, methodology, result and discussion, conclusion and future work.

2. Literature Review

2.1. History of Android

Android was created using the name of Android Inc company by Rich Miner, Nick Sears, Chris White, and Andy Rubin in 2003 [16]. Android began as a camera OS intended to enhance connectivity between personal computers (PCs) and cameras, allowing communication wireless between cameras and PCs. After a couple of months of the company's formation, the OS was adapted towards making mobile cell phones smarter [17]. Android was purchased by Google in August 2005. Android OS was created using the Linux kernel to make it free for mobile phone manufacturers [16,18]. The first version of Android OS (version 1.0) was released in November 2007, and the novel Android handset, T-Mobile, was deployed in September 2008. While it has a lot of feedback and appraisal, it ran with Android 1.0 OS, and it has other Google apps such as play store, charts, YouTube, etc. built on it. [16,18]. As the year passes by, android OS has evolved into improved models. Android 10 is the latest edition of Android, launched in September 2019 [19, 20].

The stack layers of Android apps contain applications, application framework, libraries and the Linux kernel, with each layer closely incorporated to offer stronger smartphone applications. Interoperability has also been provided for to make it easier for design developers to work and develop their applications at a cost-friendly level [4]. It made it possible for anyone to obtain due to the open source nature of the Android framework thus growing its users over the years. Android sales stood at 115 million units in 2011, and Android OS accounted for around 52.5 percent of the overall number of mobile phones [21]. The Apple App Store reportedly has 2.2 million applications [7]. Recently, about 68.71 percent of people use android OS while 29.60 percent use iOS, 0.30 percent use Windows devices, while 0.06 percent use Symbian [22]. The heavy use of android OS is attributed to several factors that include the reduced price relative to iOS, open source enabling third-party software installation unlike iOS. The sum total of applications available in the Android store (play store) as of 2017 was around 2,800,000 [22]. In March 2018, the applications in the play store crossed 3.600,000 [23].

2.2. Security in Android

IOS as a closed framework that offers more data protection and is more pristine than Android which is an open source. In the play store, developers can host the applications they have built giving rise to Android vulnerabilities [22]. The security enhancements in Android were clarified in [24]. They are public vs. private components, components open implicitly, permission for broadcast intent, permission for content provision, secure APIs, service hooks, pending intents, permission protection level, URIs permission. Security refinements such as these, aid the protection of the data of users from manipulation and abuse.

2.3. Android Malware

The skill to debone the code used in the development of Android applications and the open source nature of the applications also enabled the easy encoding of malicious codes and increased number of attacks [25]. In [26], Networks recorded that mobile malware rose to 155 percent in 2011, with a rise of 614 percent from March 2012 to March 2013, of which 92 percent of the malware is Android. When android devices are upgraded to the latest OS, about 77 percent of threats to android can be eliminated. Just one threat infected about 5,000 devices in the third quarter of 2018. This accelerated through phishing using download and

installation of a voice message application that is fake thereby sends all the user's data to a server or using text messages [27]. Report has it that 2,47 million of the latest mobile malware were discovered in 2013, indicating a 197 percent increase over 2012 [28].

From a simplistic SMS sending Trojan, Android malware has advanced to more futuristic codes that can corrupt other applications, encrypt user data, gain root privileges, download and install more applications that are malicious bypassing the knowledge of the users, and load a payload from a remote server to cellular phones [9,29]. In [29], the authors gave a complete malware for android analysis looking at the past, observing the present and predicting the potential android malware cases into the future. In [30], the author gave a thorough analysis of the models of mobile malware dissemination and explored potential future developments as well.

2.4. The dangers of android malware

Users are exposed to multiple attacks and may suffer from them if an android OS is compromised with malware. According to [25]. Many of the threats they face include but are not restricted to:

- Personal information theft which can lead to theft of identity
- Loss of privacy
- Monitoring of the users
- Financial loss through ransomware
- Remote operation of the phones

2.5. Android malware detection

Detection of Android malware can broadly be divided into;

- Signature-based: Detection using signature-based detection uses the malware's identity to determine the genuineness of the malware. The downside of this method is some form of attacks may be eluded. E.g., byte code level transformation [28].
- Machine-learning focused: The detection based on machine-learning uses an analytical method that includes extracting features from the application's behaviour [28]. These behaviours (permission request, API calls) are then developed into a dataset and then a machine learning algorithm is used to test and train. The measurement metrics used in the evaluation include: precision, accuracy, false positive, to mention a few.

Several approaches to android malware detection have been put forward; the use of applications' requested permissions, specific application programming interface for detection, the use of the applications' fundamental code (since most codes for malware are modifications of existent codes), sandboxing, components encapsulation, optional access control or signing of application [31,32].

This study is aimed at the usage of manifest permission, API call signature, intents and command signature for the android malware detection. The rights defined by the developer in the application to allow system interrelationship: to access system components (such as camera and GPS) or modules of other applications is permission [33]. There are four types of android permission with two classifications, according to [2,34]. These are:

- Signature permissions: These applications are authorized only if it is certified by the creator of the stipulated consents. It is used for component access restriction to a limited collection of application controlled and trusted only by the creator.
- System permissions: Once the application satisfies the signature criteria, these applications are authorized. Applications requesting these permissions are pre-installed by an advanced customer or system manufacturer.

The two permissions (Signature and System) are booked for applications that have signatures with keys accessible only to the creator of the firmware. This is not available for implementations from third party application.

- Normal permissions: These are software that do not need the involvement of the customer. They are immediately accepted.
- Dangerous permissions: These are software that prompts the user what system facility will be used before it can perform its function. Prior to installation, they must be issued by the consumers. [2, 34].

In [34] and [35], the studies titled "PScout: Analyzing the Android Permission Specification" and "Android Permissions Demystified" evaluated the different android OS permission specification. In [2], the author also examined certain motives for some evasive consents. Some are broadcast theft, malicious service launch, hijacking of activity, malicious activity launch, hijacking of service.

2.6. Review of Past Works

Some research has been conducted in the detection of malware, part of which includes:

A Naïve Bayes algorithm model for android malware detection system was proposed by Shang, Li, Deng, & He in [36]. By using the new permissions for malware and training permissions impact as the weight, the detection accuracy was enhanced. In order to strengthen the model, a detection model of information theory and permissions based on the enhanced Naive Bayes algorithm was used also. As a rate of detection for non-malicious applications, the proposed model earned 97.59 percent.

In [37], the authors proposed a hybrid intelligent model which evolves and uses support vector support (SVM). Genetic algorithm (GA) and particle swarm optimization (PSO) were used to resolve the optimization challenge in the SVM. This helped boost the classifier's accuracy. With the use of GA, 95.60 percent accuracy was obtained as the highest.

In [38], the authors performed a fixed evaluation of applications for android, tested the existence and occurrence of words that ae key in the manifest file of the applications, and established static feature sets from a dataset of 400 applications to yield superior results for malware detection. KNN and SVM classification algorithms were applied getting an average accuracy of 79.08 percent and 80.50 percent respectively.

In order to detect malware, Shohel Rana, Gudla, & Sung in [39] improved and analyzed certain algorithms through the implementation of a static analysis that is classifier based. When random forest was used on the dataset, 94.33 percent was obtained as the accuracy.

In [31], the authors proposed a technique of malware detection by first extracting the function call graphs from android applications and then embedding the function call graphs with a direct map feature. This was inspired by a linear-time graph kernel. They used the structural features of 12,158 samples of malware. SVM classifier was trained to differentiate between malicious and benign applications. 89% of the malware was detectable using this technique with minimal false alarms pin-pointing the code structures which are malicious within the android applications.

In [28], the authors extracted a weighted contextual API dependency graphs as program semantics which was used to create the feature sets. Authors also introduced graph similarity metrics to discover identical application behaviours for the purpose of discovering zero-day malware. In their work, a system called DroidSIFT was implemented to evaluate 2200 malwares and 13500 benign samples. It was able to correctly classify 93% malware instances while the detection of zero-day malware had a low false negative rate of 2% and false positive rate of 5.15%.

In [25], the authors presented a method of detecting malware in android by examining the manifest files of the android application since there are differences that are significant in the manifest files of benign and legitimate applications. Manifest files carry important information of the application like the name of the package, the version number of the application, the intent filter (action, category and priority), the API level, required permission, this approach was discovered to be effective by the use of real samples of android malwares. A total of 365 samples were used and the total correct detection was 90.0%.

In [5], the authors presented an automated detection system for the android platform (AMDA) where the behaviour analysis of the application was utilized. The features of both malicious and benign applications were first extracted to form the dataset used. Various algorithm in WEKA which include, Navies Bayes, decision tress and so on were used for the analysis.

In [40], the authors proposed a technique of applying the Bayesian classifier in a novel way. 1,000 samples each for both benign and malicious application was gotten. The accuracy obtained was 0.921 and true positive rate of 0.906.

In [41], the authors proposed a machine learning approach to the detection of malware in android using 200 features. These features were taken out of both the static and dynamic analysis of android applications. The deep learning technique used had an accuracy of 96.5%.

In [42], the authors used 17 supervised learning techniques to analyse adware in android devices and compared the result. It was discovered that Random Forest was the best classifier with an accuracy of 0.9838 and a false positive rate of 0.017.

This research was inspired by the recommendations in [1]. Their research study comprises of commonly and hardly used machine learning algorithms and it was found that multilayer perceptron (MLP) outperforms the other algorithm having an accuracy of 0.994. It was therefore recommended that other dataset should be put into consideration to confirm if MLP is dataset-specific.

3. Methodology

3.1. Environment Description

This research was conducted using the Waikato Environment for Information Analysis (WEKA) version 3.8.3 which was developed in [43]. Ten algorithms out of the available algorithms in WEKA were used in this research.

3.2. Dataset Description

From publicly available datasets, two datasets were used in the study. The first was gotten from figshare which was developed in [15]. The dataset contains a total of 215 characteristics which were extracted from 3,799 applications; 1,260, malicious and 2,539, benign. The second was also developed by Yerima in [44]. The dataset contains 215 attributes with 15,036 instances of which 5,560 was malware and 9,476 was benign.

Table 1: Description of the Android Malware Datasets

Description	Dataset	
	First	Second
Instances	3,799	15,036
Malware	1,260	5,560
Benign	2,539	9,476
Attributes	215	215

Attributes/characteristics are the applications' features that were extracted. The extracted features were gotten from the manifest permission, the API call signature, the commands signature and the Intents. Some features extracted under the manifest permission are SEND_SMS, RECORD_AUDIO, READ_PHONE STATE, WRITE_SMS, USE_CREDENTIALS, GET_ACCOUNTS, MANAGE_ACCOUNTS, RECEIVE_SMS. Some features under the API call signature are IBinder, Binder, createSubprocess android.os.IBinder, URLClassLoader. Some features under intents are android.intent.action.ACTION_POWER_DISCONNECTED, android.intent.action.NEW_OUTGOING_CALL, android.intent.action.CALL_BUTTON. The command signature has only the following features /system/app, mount, /system/bin, chmod, chown and remount.

3.3. Classification Algorithms application

The training and testing was carried out with the 10 algorithms chosen in the WEKA environment. The algorithms are Random forest, Naïve Bayes, Random Tree, BayesNet, Multi-layer perceptron (MLP), Decision Table, Sequential Minimal Optimization (SMO), Bagging, KStar and IBK (also known as K Nearest Neighbours classifier (KNN)). The 10-folds cross validation and 66% split were used to train and test in the condition. In the 10-folds cross-validation, the dataset is divided

into 10 folds or parts wherein each fold is divulged in almost equal sizes as the full dataset while the 9 remaining is used for training. This procedure is repeated for each fold, making it ten times. The mean for all the 10 folds is then calculated for each performance metric. The 66% split divides the dataset into 66% and 34%. It uses the 66% part of the data for training while using the 34% for testing. The 66% split was used to validate each algorithm.

3.4. Graphical Representation of the Methodology

Figure 1 shows the graphical representation of the how the research was carried out.

3.5. Performance Evaluation

Below are the performance metrics used in evaluating the algorithms.

- Accuracy: This calculates the rate of applications correctly classified, taking into account true positive, false negative, true negative and false positive [45].

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}$$

where TP = True positive
FN = False Negative
FP = False Positive
TN = True Negative

- False Positive Rate (FPR). This calculates the rate incorrectly classified instances as benign. A low FP-rate implies that the classifier is good. [46]

$$\text{FPR} = \frac{FP}{FP + TN}$$

- Precision, Recall and F-measure. Precision is the ratio of instances that are positively predicted among the retrieved instances. Recall is the ratio of instances which were positively predicted among all the instances and F-measure is the harmonic mean of recall and precision.

A high F-measure is essential since both precision and recall are desired to be at high levels [47,48]. Therefore

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FP}$$

$$\text{F-Measure} = 2\,x\,\frac{Precision}{Precision + Recall}$$

- Receiver Operating Characteristic (ROC) curve. A high ROC value means that the algorithms are good. The true positive rate is plotted against the false positive rate [46].

ROC Curve is Plot of FPR(x) vs TPR

where TPR is True Positive Rate

- Root mean square error (RMSE). This is the predicted error's standard deviation. The error between the training and the testing dataset is the predicted error. A low RMSE means that the classifier is good [49,50].

$$\text{RMSE} = \sqrt{1 - r^2}\; x\; SD$$

where SD = Standard Deviation

r = Predicted error

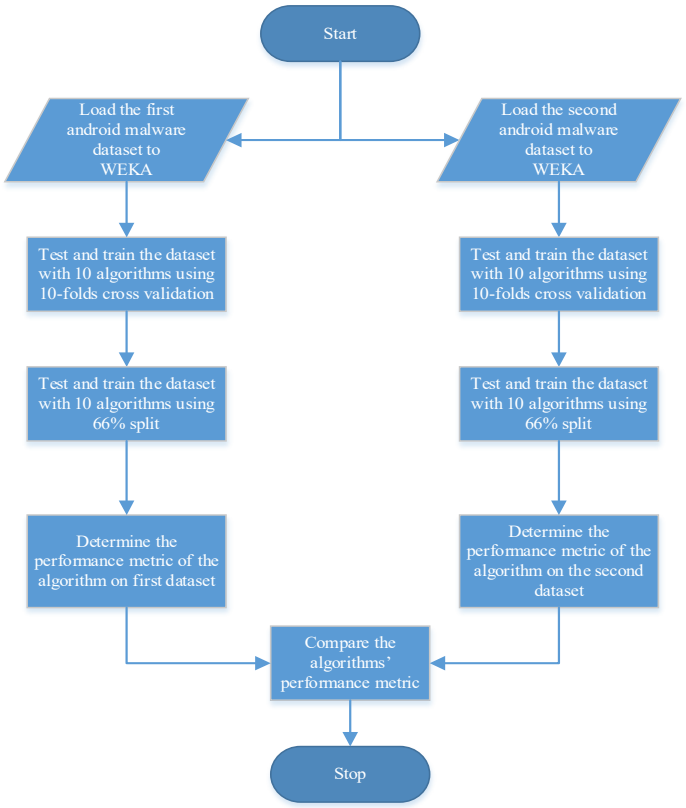


Figure 1: A graphical representation of the methodology

3.6. Algorithm Evaluated

Below are the description of two of the algorithms that stood during this research work

- Multilayer perceptron (MLP) is a perceptron that join up with extra perceptron, packed in many layers to figure out compound problems. Every perceptron in the input layer that is first layer to the left sends outputs to all the unseen layers that is the second layer perceptrons and all the second layer perceptrons send outputs to the output layer to the right that is the final layer [51].
- Random Forest: An additional dimension of randomness was introduced to bagging to create random forest in 2001. The algorithm creates tree predictors, and for all the trees in the forest, each tree depends on the random vector differently and autonomously sampled with the same dispersal. The best of these randomly chosen predictors is used in [52, 53].

4. Result and Discussion

This sections discusses the result gotten from the analysis done.

4.1. Analysis of the result

The dataset is made up of both malicious and innocuous applications. After the 10 algorithms have been trained and tested

with the 10 folds cross validation and 66 percent split in the WEKA environment, the obtained results is discussed in this section. Table 2 and table 3 show the performance metrics of dataset 1 and dataset 2 respectively. The performance metrics used are; accuracy, RMSE, false positive, ROC, precision, recall and f-measure. Figure 1 shows a display of the best performing algorithm in WEKA using the second dataset.

4.1.1. Discussion on the first dataset

While performance was optimal for all algorithms when trained and tested with the first dataset as seen in table 2, judging by accuracy, the top performing algorithm is the multilayer perceptron (MLP) having an accuracy of 99.4% when trained and tested under 66% split as previously stated in [1]. Under the 66% split, the MLP has the lowest false positive rate of 0.006 measured and also the highest recall (0.994), f-measure (0.994) and RMSE (00764).

Random Forest performs best measuring with respect to precision (0.993), accuracy (0.993), recall (0.993), f-measure (0.993) and ROC when trained and tested under the 10 folds cross validation [1]. MLP evaluated under the 66% split is, however, the best performing algorithm overall.

4.1.2. Discussion on the second dataset

In the second dataset, as seen in Table 3, while all algorithms perform well when evaluated with the dataset, the random forest had an accuracy of 98.9% when trained and tested under 10 folds cross validation making it the algorithm that performed best in terms of accuracy. The IBK has 0.013 as the lowest false positive rate. In comparison, the random forest has a reduced FPR of 0.016 and has the highest precision (0.989), recall (0.989), f-measure (0.989) and ROC (0.998). Kstar had the lowest RMSE of 0.0984.

Table 2: Showing the result of the performance metrics of the algorithms on the first datase

	Accuracy		**False Positive**		**Precision**		**Recall**		**F-Measure**		**ROC**		**RMSE**	
	10 Folds	66% Split	10 Folds	66% Split	10 Folds	66% Split	10 Folds	66% Split	10 Folds	66% Split	10 Folds	66% Split	10 Folds	66% Split
MLP	0.989	**0.994**	0.012	**0.006**	0.989	**0.994**	0.989	**0.994**	0.989	**0.994**	0.999	0.999	0.0938	**0.0764**
SMO	0.991	0.991	0.011	0.009	0.991	0.991	0.991	0.991	0.991	0.991	0.99	0.991	0.096	0.0923
IBK	0.992	0.991	**0.008**	0.01	0.992	0.991	0.992	0.991	0.992	0.991	0.995	0.996	0.0853	0.0851
KStar	0.992	0.991	0.01	0.01	0.992	0.991	0.992	0.991	0.992	0.991	0.999	**1**	**0.0811**	0.0823
Random forest	**0.993**	0.99	0.013	0.018	**0.993**	0.99	**0.993**	0.99	**0.993**	0.99	**1**	**1**	0.0913	0.0985
Bagging	0.986	0.983	0.02	0.023	0.986	0.983	0.986	0.983	0.986	0.983	0.999	0.998	0.1115	0.1186
Random Tree	0.972	0.972	0.033	0.029	0.972	0.972	0.972	0.972	0.972	0.972	0.97	0.972	0.1662	0.1669
NaiveBayes	0.958	0.96	0.045	0.039	0.959	0.961	0.958	0.96	0.959	0.96	0.994	0.995	0.1852	0.1821
Decision Table	0.938	0.947	0.066	0.059	0.939	0.947	0.938	0.947	0.938	0.947	0.977	0.979	0.2074	0.2032
BayesNet	0.927	0.93	0.046	0.046	0.937	0.938	0.927	0.93	0.929	0.931	0.992	0.993	0.2448	0.2482

Table 3: Showing the result of the performance metrics of the algorithms on the second dataset

	Accuracy		**False Positive**		**Precision**		**Recall**		**F-Measure**		**ROC**		**RMSE**	
	10 Folds	66% Split	10 Folds	66% Split	10 Folds	66% Split	10 Folds	66% Split	10 Folds	66% Split	10 Folds	66% Split	10 Folds	66% Split
MLP	0.972	0.97	0.037	0.034	0.972	0.97	0.972	0.97	0.972	0.97	0.992	0.992	0.1621	0.1676
IBK	0.988	**0.985**	**0.013**	**0.018**	0.988	**0.985**	0.988	**0.985**	0.988	**0.985**	0.994	0.99	0.1032	0.1167
KStar	0.988	**0.985**	0.014	**0.018**	0.988	**0.985**	0.988	**0.985**	0.988	**0.985**	**0.998**	**0.998**	**0.0984**	**0.1113**
SMO	0.978	0.972	0.027	0.035	0.978	0.972	0.978	0.972	0.978	0.972	0.976	0.969	0.147	0.1673

Random forest	**0.989**	0.984	0.016	0.024	**0.989**	0.984	**0.989**	0.984	**0.989**	0.984	**0.998**	0.997	0.1058	0.121
Bagging	0.978	0.973	0.029	0.034	0.978	0.973	0.978	0.973	0.978	0.973	0.996	0.995	0.1333	0.1465
Random Tree	0.973	0.968	0.03	0.037	0.973	0.968	0.973	0.968	0.973	0.968	0.972	0.965	0.164	0.1791
NaiveBayes	0.834	0.836	0.122	0.126	0.865	0.863	0.834	0.836	0.834	0.839	0.949	0.839	0.3972	0.3957
Decision Table	0.922	0.907	0.077	0.093	0.924	0.909	0.922	0.907	0.922	0.908	0.976	0.962	0.2322	0.2524
BayesNet	0.828	0.829	0.124	0.128	0.862	0.861	0.828	0.829	0.831	0.831	0.928	0.925	0.4082	0.4081

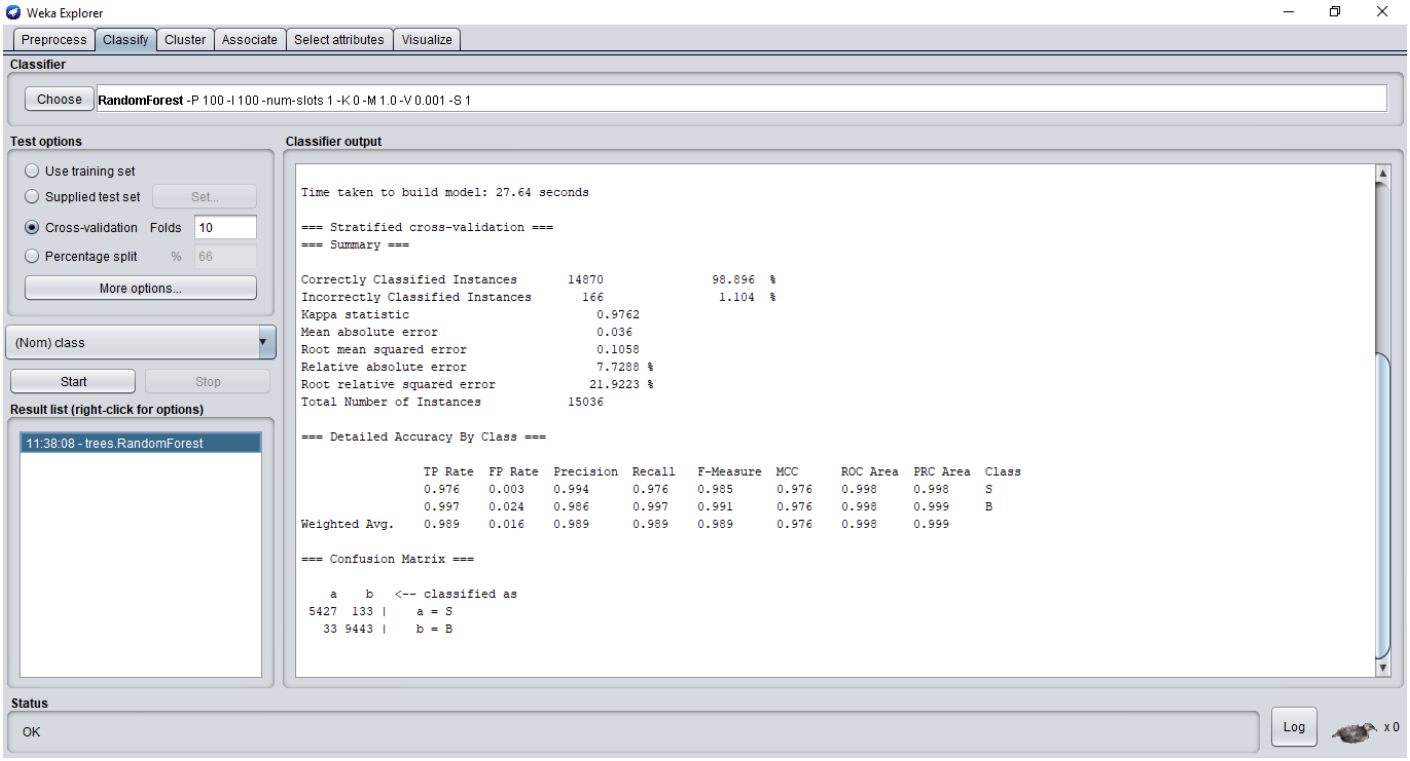


Figure 2: A display of Random forest in the environment using the second dataset.

Under the 66% split, KStar and IBK performs best with the same accuracy (98.5%), false negative (0.018), precision (0.985), recall (0.985) and f-measure (0.985). Kstar performs best in terms of ROC and RMSE with values of 0.998 and 0.1113 respectively. However, the overall best performing algorithm is Random forest tested under the 10 folds cross validation.

Although MLP was found to have the best performance under the 66% split when tested with the first dataset, it is seen that it was not able to maintain this performance when tested on the second dataset, this shows that MLP is data-set specific. This is as a result of the increased instances in the second dataset. The performance metrics of MLP was not maintained as a result of the increase in the instances. It is noteworthy in the classification of the first dataset that the analysis done under the 10 folds cross validation shows that random forest performs best with an accuracy of 99.3% having a difference of 0.001 when compared to MLP. The performance metric of random forest was further confirmed when analysis was done on the second dataset. It was observed that random forest performs best when trained and tested under the 10 folds cross validation. This therefore backs up the result of the first dataset analysis proving that random forest is not data specific.

5. Conclusion and Recommendations

Malicious applications has increasingly become sophisticated and complex with advancement in technological innovation. Models and methods for improved detection of such malware have been actively developed by various researchers in the field. The goal of this research study was to evaluate/appraise classification algorithms for malware in android detection. After the analysis in the WEKA domain, it was found that the performance metric of MLP was diminished with increased instances, making it dataset-specific. Random forest was found to perform better than the other algorithms with the first and second dataset when trained and tested under the 10 folds cross validation, having an accuracy of 99.3% and 98.9% respectively. Therefore, the contribution of this research is to determine that random forest

is the best performing algorithm in the detection of Android malware. Random forest is being proposed for use as the classification algorithm when building an android application for the purpose of android malware detection.

Further studies can consider selecting the best features of the dataset through feature selection or optimization for better performance. Other classification techniques under the reinforcement and unsupervised learning can be trained and tested. The studies should consider using Random forest as the classification and detection algorithm when developing anti-malwares, so as to provide a safe space for android users.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] O.E. Olorunshola, A.O. Oluyomi, "ANDROID APPLICATIONS MALWARE DETECTION: A Comparative Analysis of some Classification Algorithms," in 2019 15th International Conference on Electronics, Computer and Computation, ICECCO 2019, Institute of Electrical and Electronics Engineers Inc., 2019, doi:10.1109/ICECCO48375.2019.9043284.

[2] E. Chin, A.P. Felt, K. Greenwood, D. Wagner, "Analyzing Inter-Application Communication in Android," in Proceedings of the 9th international conference on Mobile systems, applications, and services, ACM: 239–252, 2011, https://doi.org/10.1145/1999995.2000018

[3] OS Timeline, Mobile operating system - Wikipedia, 2020.

[4] A. Developers, "What is android," 2011,

[5] K.J. Abela, D.K. Angeles, J.R.D. Alas, R.J. Tolentino, M.A. Gomez, "An Automated Malware Detection System for Android using Behavior-based Analysis AMDA," in International Journal of Cyber-Security and Digital Forensics (IJCSDF), Citeseer: 1–11, 2013.

[6] K. Nadia, The risks of third-party app stores, 2020.

[7] Ian Blair, Mobile App Download and Usage Statistics (2020) - BuildFire, 2020.

[8] P. Tracy, Smartphone Banking Definition & Example | InvestingAnswers, 2020.

[9] S. Verma, S.K. Muttoo, "An Android Malware Detection Framework-based on Permissions and Intents," Defence Science Journal, **66**(6), 618–623, 2016, DOI : 10.14429/dsj.66.10803.

[10] A.P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, D. Wagner, "Android Permissions : User Attention , Comprehension , and Behavior," in Proceedings of the eighth symposium on usable privacy and security, ACM: 3, 2012, https://doi.org/10.1145/2335356.2335360.

[11] M. Vennon, "Android malware: Spyware in the Android Market," Technical Report, SMobile Systems, 2010.

[12] G. Thompson, Android Malware Steals Data From Apps, Records Audio, Captures Screenshots - Binary Defense, 2020.

[13] Webroot, The Dangers of Hacking and What a Hacker Can Do to Your Computer | Webroot, 2020.

[14] Norton, Zero-day vulnerability: What it is, and how it works, 2020.

[15] S.Y. Yerima, Android malware dataset for machine learning 1 (Version 1), 2018, doi:https://doi.org/10.6084/m9.figshare.5854590.v1.

[16] J. Callaham, The history of Android OS: its name, origin and more, 2019.

[17] J. Alabaster, Android founder: We aimed to make a camera OS | PCWorld, IDG News Service, 2013.

[18] W. Bellevue, T-Mobile Unveils the T-Mobile G1 — the First Phone Powered by Android | T-Mobile Newsroom, 2008.

[19] Android version history, Android version history, Wikipedia, 2019.

[20] Android 10 | Android, 2019.

[21] M. Grace, Y. Zhou, Q. Zhang, S. Zou, X. Jiang, "RiskRanker : Scalable and Accurate Zero-day Android Malware Detection Categories and Subject Descriptors," in Proceedings of the 10th international conference on Mobile systems, applications, and services, ACM: 281–294, 2012, https://doi.org/10.1145/2307636.2307663

[22] D. Kavinya, R.S. Priya, R. Deebika, "Vulnerabilities attacks on mobile operating systems (Android versus iOS): A review," International Journal of Advance Research, Ideas and Innovations in Technology, **4**(5), 214–216, 2018.

[23] Statista, Number of available applications in the Google Play Store from December 2009 to June 2019, Google Play Store: Number of Available Apps 2009-2019, 2019.

[24] W. Enck, M. Ongtang, P. McDaniel, "Understanding Android Security," IEEE Security & Privacy, **7**(1), 50–57, 2009, https://doi.org/10.1109/MSP.2009.26

[25] R. Sato, D. Chiba, S. Goto, "Detecting Android Malware by Analyzing Manifest Files," Proceedings of the Asia-Pacific Advanced Network, **36**(17), 23–31, 2013, http://dx.doi.org/10.7125/APAN.36.4.

[26] J. Networks, Juniper networks third annual Mobile threats report, 2013.

[27] Alexandre Mundo Alguacil, C. Beek, C. Castillo, T. Dunton, J. Fokker, S. Grobman, T. Hux, N. Minihane, L. Munson, E. Peterson, M. Rivero, T. Roccia, C. Schmugar, R. Samani, R. Sims, D. Sommer, B. Sun, McAfee Labs Threats Report: December 2018, 2018, doi:10.1016/s1361-3723(19)30004-1.

[28] M. Zhang, Y. Duan, H. Yin, Z. Zhao, "Semantics-Aware Android Malware Classification Using Weighted Contextual API Dependency Graphs," in Proceedings of the 2014 ACM SIGSAC conference on computer and communications security, ACM: 1105–1116, 2014, https://doi.org/10.1145/2660267.2660359.

[29] C. Castillo, Android Malware Past , Present , and Future, 2011.

[30] S. Peng, S. Yu, A. Yang, "Smartphone malware and its propagation modeling: A survey," IEEE Communications Surveys and Tutorials, **16**(2), 925–941, 2014, doi:10.1109/SURV.2013.070813.00214.

[31] H. Gascon, F. Yamaguchi, D. Arp, K. Rieck, "Structural Detection of Android Malware using Embedded Call Graphs Categories and Subject Descriptors," in Proceedings of the 2013 ACM workshop on Artificial intelligence and security, 45–54, 2013, https://doi.org/10.1145/2517312.2517315.

[32] L. Davi, A. Dmitrienko, A. Sadeghi, M. Winandy, "Privilege Escalation Attacks on Android," in International conference on Information security, Springer, Berlin, Heidelberg: 346–360, 2010, https://doi.org/10.1007/978-3-642-18178-8_30

[33] F. Di Cerbo, A. Girardello, F. Michahelles, S. Voronkova, "Detection of malicious applications on android OS," Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **6540 LNCS**(January 2010), 138–149, 2011, doi:10.1007/978-3-642-19376-7_12.

[34] K.W.Y. Au, Y.F. Zhou, Z. Huang, D. Lie, "PScout: Analyzing the Android Permission Specification," in Proceedings of the 2012 ACM conference on Computer and communications security, ACM: 217–228, 2012, https://doi.org/10.1145/2382196.2382222.

[35] A.P. Felt, E. Chin, S. Hanna, D. Song, D. Wagner, "Android Permissions Demystified," in Proceedings of the 18th ACM conference on Computer and communications security, ACM: 627–637, 2011, https://doi.org/10.1145/2046707.2046779.

[36] F. Shang, Y. Li, X. Deng, D. He, "Android malware detection method based on naive bayes and permission correlation algorithm," Cluster Computing, **21**(1), 955–966, 2017, doi:10.1007/s10586-017-0981-6.

[37] W. Ali, "Hybrid Intelligent Android Malware Detection Using Evolving Support Vector Machine Based on Genetic Algorithm and Particle Swarm Optimization," IJCSNS International Journal of Computer Science and Network Security, 19(9), September 2019.

[38] M. Kakavand, M. Dabbagh, A. Dehghantanha, "Application of machine learning algorithms for android malware detection," ACM International Conference Proceeding Series, (February 2019), 2018, doi:10.1145/3293475.3293489.

[39] M. Shohel Rana, C. Gudla, A.H. Sung, "Evaluating machine learning models for android malware detection - A comparison study," ACM International Conference Proceeding Series, (March 2019), 17–21, 2018, doi:10.1145/3301326.3301390.

[40] S.Y. Yerima, S. Sezer, G. Mcwilliams, I. Muttik, "A New Android Malware Detection Approach Using Bayesian Classification," in 2013 IEEE 27th international conference on advanced information networking and applications (AINA), IEEE: 121–128, 2013.

[41] Z. Yuan, Y. Lu, Z. Wang, Y. Xue, "Droid-Sec : Deep Learning in Android Malware Detection," in ACM SIGCOMM Computer Communication Review, ACM: 371–372, 2014, https://doi.org/10.1145/2619239.2631434.

[42] J.Y. Ndagi, J.K. Alhassan, "Machine learning classification algorithms for adware in android devices: A comparative evaluation and analysis," in 2019 15th International Conference on Electronics, Computer and Computation, ICECCO 2019, Institute of Electrical and Electronics Engineers Inc., 2019,

doi:10.1109/ICECCO48375.2019.9043288.
[43] G. Holmes, A. Donkin, I.H. Witten, "Weka: A machine learning workbench," 357–361, 1994.
[44] S. Yerima, Android malware dataset for machine learning 2, 2018.
[45] J. Brownlee, Failure of Classification Accuracy for Imbalanced Class Distributions, 2020.
[46] J. Brownlee, ROC Curves and Precision-Recall Curves for Imbalanced Classification, 2020.
[47] Metacademy, F measure, 2020, Jun. 2020.
[48] Metacademy, Precision and recall, 2020.
[49] S. Glen, RMSE: Root Mean Square Error, StatisticsHowTo.com, 2020.
[50] S. Holmes, RMS Error, 2000.
[51] Y. Taguri, N. Gilboa, S. Gigi, I. Kober, A. Farhi, S. Awrabi, J. Freeman, Perceptrons & Multi-Layer Perceptrons: the Artificial Neuron - MissingLink, Dec. 2020.
[52] L. Breiman, "Random Forests," Machine Learning, 45(1), 5–32, 2001.
[53] A. Liaw, M. Wiener, "Classification and Regression by randomForest," R News, **2**(3), 18–22, 2002.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1750-1758 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Touristic's Destination Brand Image: Proposition of a Measurement Scale for Rabat City (Morocco)

Abdellatif Elouali[*,1], Smail Hafidi Alaoui[1], Noura Ettahir[2], Abderrazzak Khohmimidi[1], Nadia Motii[1], Keltoum Rahali[3], Mustapha Kouzer[3]

[1]*Management Sciences Research Team, Faculty of Law, Economics and Social Sciences (FSJES) Agdal, Mohammed V University in Rabat, 10090, Morocco*

[2]*Management Sciences Research Team, High School of Technology Salé, Mohammed V University in Rabat, 10090, Morocco*

[3]*Laboratory of Agrophysiologie, Biotechnology, Environment and Quality, Faculty of Science, Ibn Tofail University, Kénitra, 14000, Morocco*

ARTICLE INFO

Article history:
Received: 14 October, 2020
Accepted: 21 December, 2020
Online: 30 December, 2020

Keywords:
Brand Image
Touristic destination
Measurement scale
Morocco

ABSTRACT

The objective of this article is to highlight the existing relationship between the tourist's destination brand image of the city of Rabat (Morocco) and its choice by foreign tourists to be visited once again. We suggest a reliable measurement scale able to measure the dimensions and associations of the brand image most memorized by tourists, so as to further attract the researchers and stakeholders curiosity in the tourism sector which face, every day, new challenges of attractiveness and sustainability of tourist destinations in the new Moroccan regionalization context. The survey is carried out on a sample of 454 foreign tourists in the city of Rabat because of the importance that this city requires in terms of overnight stays, that is to say nearly one million in 2018 according to the statistics of the Moroccan Tourism Observatory. The results obtained from the survey show that the functional and abstract associations jointly constitute the brand image of the city of Rabat in the foreign tourist's memory and that the abstract associations are more significant in the destination's choice.

1. Introduction

Today, tourism in all its forms is marked by strong competition between tourist destinations at national and international levels. The emergence of a large number of tourist destinations has compelled tourism managers to adopt new managerial approaches [1]. Faced with this situation, it is no longer easy to attract so many tourists because these people are more demanding and seek quality or even total satisfaction regarding a precise destination. Henceforth, in order to market their tourist sites, tourism professionals will have an interest in mastering territorial marketing in order to position themselves clearly on the most buoyant markets by adopting strategies with a political brand trend [2].

Indeed, likewise several destinations, Morocco is positioned as a country with a purely tourist vocation with 13 million non-resident tourists in 2019, according to the Moroccan Tourism Observatory, this is because of the country's geographical position, its natural and human resources, and its civilizational and historical heritage. Nevertheless, Morocco has been facing a fierce competition over the last decade with the emergence of countries such as Turkey with 39 million foreign tourists in 2018, Egypt with more than 8 million foreign tourists in 2017 and Tunisia with more than 8 million foreign tourists in 2018 (the Moroccan Tourism Observatory).

This is why our study focused on the study of the brand image of the tourist destination with the aim to analyze the measurement scale of the variable of the brand image of Rabat city tourist destination in order to define the different dimensions linked to it regarding the functional and abstract brand image. Consequently, we will present the concept of the Brand Image of the Tourist Destination and the multiple measurement instruments related to this concept in order to answer the following central questions: By which forms of perceived images can the attractiveness of Rabat city tourist destination be improved? And what are the associations with the brand image that have a greater impact on the choice of destination by tourists?

2. Territorial marketing: theoretical foundations

Territorial marketing research has experienced unprecedented growth over the last ten years. What all this research has in common is the application of marketing

[*]Corresponding Author: Abdellatif ELOUALI, Avenue le Prince Héritier BP. 227-Salé, Maroc+212663173006 – abdellatif.elouali@gmail.com

principles to the sphere, the territory, the city, and the destination [2]. In fact, several designations exist for territorial marketing: city marketing or public service marketing, urban marketing.

Territorial marketing does not only include communication strategies to promote the territory, but all the strategies that allow the attraction of tourists in relation to other territories. Territorial marketing is therefore based on effective, coherent and well-thought-out communication. The territorial brand plays a crucial role in this communication approach. From this point of view, the territorial brand is a means to attract not only tourists, but also investors, businesses and residents, who are also involved in the process of tourism development of a territory [3].

Although it is different from the enterprise, the territory will gain in adopting the strategies of the enterprise, especially the marketing strategy to be able to attract tourists in a context of strong competition [4].

3. The brand image: a multidimensional concept

The brand represents an intangible asset in order to attract new customers and subsequently building customer loyalty [1]. The brand fills numerous functions designed to support consumption. In this way, the brand carries a value and brings certain notoriety [5]. The brand is a product communication tool. For the consumer, the brand confers a reference point with which he can identify [6]. It makes it possible to identify and make the city desirable in the eyes of visitors [7]. At the level of a tourist destination, the brand takes the form of a name and a logo. For the destination, the brand carries its identity and protects its reputation [1]. The territorial brand is used as a lever of attractiveness of a tourist territory, within the framework of territorial marketing [8].

In [9], the author explains that the mark is both a signifier and a signified. It is a signifier because it enables the goods to be distinguished. Distinction does not only derive from sight, but also from other senses: touch, hearing, taste, smell. But as a signifier, the trade mark evokes a sense in the consumer. In [10], the author reports the characteristics of a brand at the territorial level, arguing that the territorial brand refers to the perception of the reputation of the territory by internal and external actors. The territorial brand therefore aims to convey the meaning of the city object. Furthermore, the city itself has always been considered a mark by nature, inasmuch as its name enables it to distinguish itself from others. Moreover, the name of a city already evokes its reputation and image in the mind. The territorial brand therefore encompasses all perceptions relating to a place and its inhabitants [11].

3.1. The notion of brand versus the notion of brand image

A differentiation is to be made between the brand and the brand image. The latter corresponds to the mental image that emerges in the individual when he hears, sees or evokes the brand [12]. In [13], the author explains that brand image corresponds to "all the associations and impressions that a consumer has in his memory about a brand. In other words, the brand image designates everything that comes to mind when talking about a brand. It can be tangible elements such as the goods or services bearing the brand, but it is also possible to associate the brand image with abstract elements from the producer's communication or from other competing brands [14]. As a result, the brand image could be assimilated to consumer representations or perceptions of a brand. It has emotional, rational and memory dimensions. However, for the consumer to be able to remember an attribute of the trade mark, he must have already experienced it or at least have been exposed to it on many occasions. The brand image is therefore the result of experiences resulting from the use of the product or exposure to it [15].

Within an enterprise context, the brand image is the perception of the brand by all stakeholders and the public [16]. From the marketing point of view, the touristic territory is assimilated to a product for sale and is associated with the brand. The territory becomes a brand that has its own image to promote through branding, but the brand image could also be considered as the characteristics that allow the individual to evaluate the brand in relation to others. The characteristics taken into account are in most cases concrete attributes [17].

The brand image of a tourist destination corresponds to its mental representation by the consumer. From this point of view, symbolic meanings are associated with the characteristics of the destination in question. This representation varies according to the experience or communications made by marketers and/or acquaintances who have previously experienced the tourist destination [18].

3.2. Brand image psychological approach

The branding process formation is split into four phases [19]: The first phase acts to express the process of the image perception by the individual's senses. It is primarily the physical characteristics of the product and the consumers' expectations and interests that affect or not the consumers' attention. If the consumer doesn't grasp the stimulus of the product by the senses, then the perception is not possible and ultimately the purchase will not occur.

During the second phase the consumer decodes the stimulus and interprets it according to his learning. This phase is unique to each individual.

The third phase deals with the perception's mental representation. The perceived image embodies a meaning and is somehow translated.

It is at the fourth phase that the individual can, through his evocation capacity, make a judgment, express an opinion, feel about the perception or simply take to purchase decision. The image itself can be divided into a desired image, transmitted image and perceived image [20].

3.3. Related associations to the brand image

In order for a brand to trigger a consumer's positive attitude, the consumer must develop strong and positive associations with the brand. It is therefore essential to create a strong image [14]. The structure of an image depends on the key attributes of the tourism product as well as the associations with the brand. However, the attributes and associations to the brand are derived from the different information retained in the consumer's mind after a consumption experience. Interactions at tourist destinations structure a complex image of the destination. The associations derive from this fact, from the confrontation between the image of the destination as perceived by the visitor before the experience, and the one developed after the tourism experience. Moreover, this complex image corresponds to the brand image perceived by visitors [21].

Branding consists of associations linking the individual to a brand. These associations can be functional or abstract. Functional associations group together attributes situations of use, benefits of using the product. Abstract associations, on the other hand, involve psychological phenomena such as affects, the attribution of a symbol to the trade mark or the emotions that the individual experiences in relation to the trade mark [18].

The functional or symbolic association with the image depends on the intensity and frequency of the positive experiences that the consumer has with the product [6].

4. Perception of destination's brand image as a communication strategy

The strong growth of the tourism industry has prompted many destinations to position themselves on this market [22], not only capital cities (Paris, London, Vienna, ...) but also cities that have developed a scientific (Cambridge) or cultural (New York, Aix-en-Provence, Dijon, ...) image, art cities such as Venice or emerging cities that have developed their own identity (Prague, Istanbul, Seville) [23].

Indeed, the competition between territories and even between cities has triggered the rise to an action mode which assumes the rise of reflections on attractiveness at several levels. Territories and destinations are the first willing to produce discourses to attract but also to retain a certain category of investors, tourists or residents.

In present day, the challenge for tourist destinations consists in differentiating themselves by developing a specific identity. They will carry out communication campaigns which may be reinforced by heritage images (culture, local customs, festivals, architecture, nature, etc.) intended to stage and perpetuate local traditions and culture [24].

Tourism communication therefore plays a capital role in the construction of the touristic destination image. Nevertheless, the multiplication of communication campaigns by the different territorial strata and the media coverage of the various events can blur the image of destinations.

Besides, it is important to study the effectiveness of communication between various territorial entities towards an urban destination in the light of the theory of social representations. Indeed, the development of destinations brands becomes compulsory for the cities' survival in the face of international competition. In order to differentiate themselves, they must develop their own identity brand and consequently create a good image.

Let us point out that touristic destinations are no longer perceived as mere entry points, boarding locations or transit points during a trip, but also, they are seen as eye-catching sites (natural or cultural resources) and attractions (appropriate development for tourism of these resources) [25] of fully-comprehensive destinations.

Previously, destinations promoted themselves thinking that the creation of tourist attractions would suffice to attract tourists (European Travel Commission). Nevertheless, exposed to tourism development, to the multiplication of communication campaigns at the different territorial levels (national, regional and city), as well as to the media coverage of the various events (terrorist attacks, strikes; riots in some countries; Pandemic, for example (Covid-19), the images and representations of destinations may be blurred.

In other words, the image keeps a very important place in the choice of a destination, and the different information sources play a crucial role in the tourists' minds regarding this image formation [26]-[28]. Consequently, the communication issued by destinations and touristic organizations influences the determination of the traveler's attitudes and behavior as well as in the formation of the tourist's image of a destination [29].

Typically, researches related to tourist destinations deal with the destination's image, which is an essential element in the choice of a touristic destination [27], as it provides information on the specific and unique attributes related to the destination before the potential stay [28].

As a result, the destination's image can, through its communication with tourists, generate a strong and distinctive image compared to its competitors. Moreover, many studies distinguish two types of images resulting from this formation process: the induced image resulting from the information of tourism actors (advertising and promotion of the destination and tourism organizations) and the organic image resulting from the consumers exposure to non-tourist information sources, such as the media outside the tourism sector [27], [28].

The organic image is often rooted in the tourists' minds because it is made up of shared representations which convey stereotypes [30].

5. Measurement scales of the Brand Image of the tourist destination

The measurement of brand image is based on the associations that consumers attribute to the brand. The stronger are the associations, the stronger is the brand equity. The associations should be strong and varied [31]. The measurement of destination branding is complex due to the fact that it is relative and dynamic and, therefore, varies over time. Thus, the models developed so far to measure the brand image of a tourist territory are based on tangible and intangible elements. Tangible elements include price, accommodation, different tourist services. The intangible elements refer to the ideas and sensations that consumers experience when they see or hear about the destination [32].

The measurement of the image of the destination is based on tourist activities and their needs, including transport, various amenities, food, services and travel prices. The image can also be measured through the characteristics of the destination and its inhabitants, such as their hospitality, the political stability of the country, economic development and environmental management. The image can also be measured through the attributes of the destination and its inhabitants. These different elements make it possible to gauge the image of a destination not only in the eyes of tourists, but also in the eyes of the whole world [33].

5.1. Quantitative and qualitative approaches

The quantitative and qualitative approaches make it possible to measure the brand image of a destination. Four scales have been indicated by [16] to measure brand image: the attitude scale, the ranking scale, purchasing intentions and turnover evolution. The attitude scale refers to the attractiveness of the

brand. The rating scale considers brand preference, as well as the characteristics that distinguish the destination from its competitors. Purchase intentions, on the other hand, provide information on the consumer's conviction about the brand. The evolution of turnover gives indications concerning the volume of purchases of the brand and, hence, the realization of the purchase intention [16].

5.2. Direct and indirect approaches

The measurement of brand image can be done indirectly through the study of perceptions, or directly, through the analysis of preferences and direct measurements. For the indirect measurement approach, it is possible to base the measurement on mental representation and the individual's ability to remember the brand. The consumer recognition or recalling of the brand allows the assessment of the brand's strength. The stronger is the brand equity, the stronger and more positive are the associations. Moreover, brand image is considered strong in this indirect measurement approach when consumers associate many attributes with it. In other words, the indirect approach to brand measurement refers to the measurement of brand awareness. The latter has six dimensions: brand acceptance, brand memorization, first brands that come to the consumer's mind, dominant brand (only one brand memorized), brand awareness and brand opinion [16].

6. Psychometric approach

Branding image is a multidimensional concept, particularly since it is often measured indirectly through its sources and its consequences. Unlike the various techniques already mentioned, psychometric measurements display the advantage of being more practical, more operational and above all more direct. In fact, it is a question of requesting consumers' opinions, through a pre-established questionnaire. In other words, satisfaction's surveys should be carried out among tourist consumers. The main measurement scale developed in this respect is the initiative of [34]. Nevertheless, the representations' mental extraction from the consumer's mind requires qualitative methods that are more rigorous than metric scales.

7. Brand's image other measures

7.1. Explicit measures

When the consumer proceeds to a purchase decision, he tries to remember the advertising message and what the brand represents. It is above all the level of learning of the message that varies the result of this consumer research in classic or so-called Territorial Marketing. The effectiveness of an advertisement and the product or territory brand will be evaluated on the basis of the words memorized. This is an explicit measure, as it focuses on a single person and on what that person has retained from an explicit message.

Sometimes there may be a distortion between what the consumer has memorized, and the information given. In this case, the person does not correctly restitute what he or she has "learned" from the advertisement but adds other incorrect associations or fractions.

7.2. Implicit measures

Implicit measurement is useful to advertisers to determine what consumers know about their brand, regardless of the information source. Different types of measurement can be distinguished [18]:

- Memorization: measuring the effect of advertising exposure on brand memorization;
- Spontaneous notoriety: consists of measuring of advertising effectiveness by asking consumers to name brands for a product category.

8. Attitude study

Consists of determining the perception of the brand and not what consumers know about it. In this method, they make a judgment on the basis of the various criteria of the brand.

The aim of these pre-tests is to determine what makes consumers change their brand preferences. We strive to single out the elements that build consumer's loyalty. Furthermore, classical theories indicate that there are strong and positive associations in the consumer's mind that influence the general positive attitude towards the brand [35], [36]. The literature then suggests that the strength of the image, i.e. its ability to translate itself into utility and loyalty behavior, depends mainly on four attributes according to [37]:

- The strength of associations;
- The dominance of associations;
- The valence of associations;
- The cohesion of the associations.

Therefore, in [37] the author described that only associations considered by the consumer as positive, strong, unique, and coherent, build the strength of the image. For instance, they affect the overall preference for the brand or a re-purchase or recommendation behavior. By analogy, and mentioning in [38], we suggest that attitudinal fidelity to a destination depends on the strength and valence of its image. Ultimately, it is expected that a tourist with strong and positive associations with the destination brand would be more inclined to revisit it and recommend it to others.

After our review of the literature, it appears that there is no agreement on the dimensions that form the brand image of the tourist destination. In addition, many of the brand features of the touristic destination have remained vague and lacking of clarity.

This generalism causes a certain ambiguity between certain concepts close but different from the brand image of the tourist destination, which causes a problem in terms of its determinants and its scales of measurement. Therefore, the design of a tool (scale) to measure the image of the tourist destination presents a real challenge in order to guide future research and support the professionals in order to better manage their management system of the brand image of the touristic destination.

9. Research methodology

In order to develop a standardized measurement tool, we will adopt the advocated approach known as the Churchill's paradigm [39]. The objective of this methodological approach is incorporating the knowledge of measurement theory and the appropriate techniques for improving it into an automatic procedure.

Similarly, Churchill's paradigm offers the possibility of rigorously designing measurement instruments such as multi-scale questionnaires. To this end, we deemed it appropriate to use, first of all, a qualitative approach to explore the characteristics of the brand image of the touristic destination.

In our field investigation, we used the semi-directive interview for data collection. This method is favored given its great flexibility and the wealth of information it can produce [40].

According to the literature, we have elaborated an interview guide that includes about the respondent's identification, age, gender and the determinants of the brand image of the touristic destination; the reputation of the touristic destination and ultimately the expectations of tourists towards (Rabat city).

The purpose of all these questions is to understand how the touristic destination is to understand how the brand image of the touristic destination is perceived by foreign tourists, and consequently to identify the different characteristics of the brand image of the touristic destination. Ten interviews were conducted with random tourists in Rabat city with average time duration of twenty (20) minutes. The population sample was characterized by female predominance (06 women and 04 men). The average age span is 28 years.

Afterwards, we have used a quantitative method. Indeed, in order to proceed in a methodological logic, we have formulated items based on all the items used in the different studies examined in the literature and the attributes collected after processing the information from the interviews (the results of the qualitative analysis will be dealt with in another research paper). Hence, a first version of the questionnaire was drawn up, then it was examined by two teacher-researchers in management sciences, and finally it was tested with 16 tourists from the city of Rabat.

The comments collected made it possible, on one hand, to improve the questionnaire, and on the other hand, to underline the latest version of the questionnaire which consists of 26 items on a five-point Likert's scale. Also, we also used exclusively affirmative sentences (positive statements) following Devellis' recommendations. The data were collected in paper format. 454 usable questionnaires were collected. The representative population of the control sample is composed of 64% men and 36% women. The majority of the tourists are situated in the age group of 25 to 60 years old.

The following figures present the various information about our sample population (gender, age, country of origin):

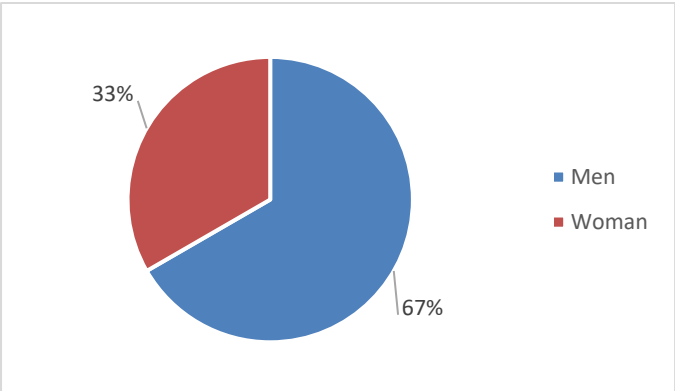


Figure 1: Tourists distribution according to the gender

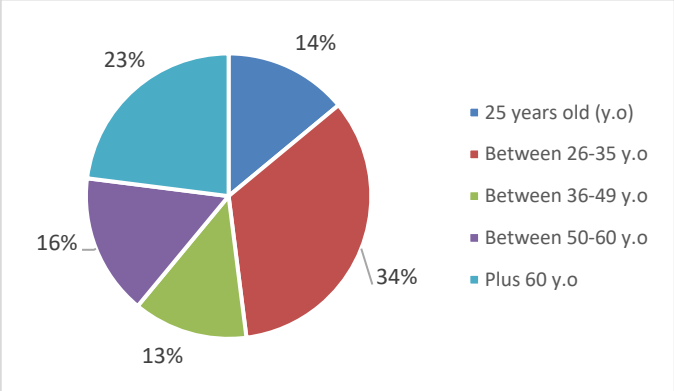


Figure 2: Tourists dispatching according to age

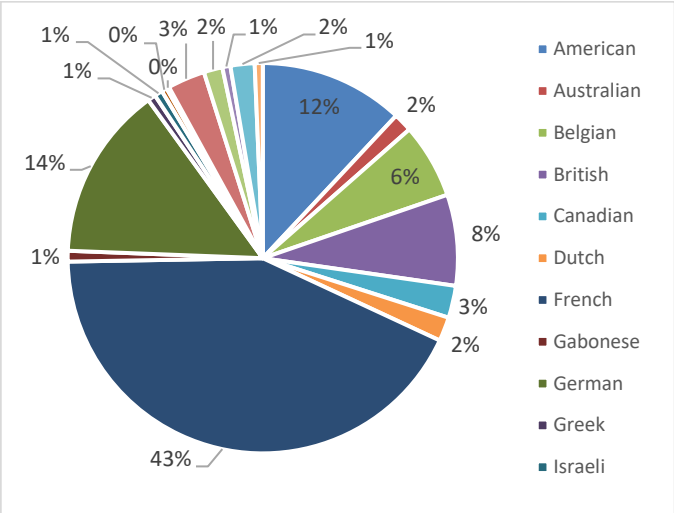


Figure 3: Tourists' dispatching according to their nationality

10. Exploratory Factor Analysis

To elaborate our measurement scale, we have gone through several steps. First, we have verified the content validity of the scale for measuring the brand image of the tourist destination. Content validity tests aims to check whether the questionnaire's different items are a good representation of the concept under consideration [41]. For this purpose, as explained before, we reviewed the academic literature to accompany and guide us in the questionnaire's elaboration and to reach a measuring scale of the brand image of the touristic destination beforehand. Indeed, the large number of items constituting the measurement scale (16 items), their heterogeneity and the way in which they can be taken out of the dimensions, offer the researcher a choice between three types of analysis' approaches:

- Independently, to examine each item by selecting only one item per theme;
- Aggregate all the items into a single scale;
- Undertake a factor analysis to identify the different underlying dimensions.

We have decided to introduce all items in a single scale without distinguishing between dimensions at the questionnaire level, in order to test the stability of the factor structure of the measurement scale we will use. The factor structure and the psychometric qualities of the scale of the tourist destination brand image were analyzed using SPSS (Statistical Social Sciences Package) 23 software, the data were subjected to

principal component factor analysis with Varimax rotation in order to test the dimensionality of the construct.

The Principal Component Analysis (PCA) is considered desirable in the development phase of a questionnaire in order to decrease the number of items and retain only keep those that allow the phenomenon to be characterized, particularly in order to identify the main factors [42].

The items purification deemed unsatisfactory was performed using the following elimination parameters: the rejection of items with a factor score of less than 0.3 and those with a high factor score on several factors [43]. The internal consistency reliability of the scale and its different dimensions was assessed by the Cronbach's alpha analysis as well as the correlation level for dimensions with at least two items.

11. Results and discussions

We undertook an exploratory approach to determine the dimensions of the tourism destination's brand. Firstly, we have prepared the number of variables, types of variables and finally the sample size.

Secondly, we looked at inter-item correlations. Then, we have measured the adequacy of sampling (KMO) and Bartlett's sphericity Test. The results showed that the KMO index is 0,758. It is qualified as good. This index shows that correlations between items are of good quality. After this, the result of the Bartlett's sphericity test is significant since $p < 0, 0005$ (Table 1).

Table 1: KMO Index and Bartlett's Test

Sampling precision measurement Kaiser-Meyer-Olkin	0,758
Approximate Khi-two	3664,384
Bartlett's sphericity test ddl	120
Bartlett's significance	0,000

In order to extract the appropriate of factors for our scale, we first have analyzed the table of total variance explained, and we have noticed that 4 components have an eigenvalue higher than 1. The 4 factors explain 66.38% of the variance (Table 2). These factors are considered significant.

Table 2: Initial eigenvalues

Component	Initial eigenvalues		
	Total	% of the variance	% cumulated
1	5,277	32,983	32,983
2	2,595	16,219	49,202
3	1,496	9,350	58,552
4	1,253	7,831	66,383
5	0,994	6,211	72,594
6	0,723	4,518	77,112
7	0,670	4,188	81,299
8	0,540	3,378	84,677
9	0,453	2,834	87,511
10	0,442	2,766	90,276
11	0,401	2,503	92,780
12	0,359	2,247	95,026
13	0,274	1,713	96,739
14	0,192	1,200	97,940
15	0,176	1,103	99,043
16	0,153	0,957	100,000

Secondly, and for more certainty, we have created a graphic from the eigenvalue using SPSS 23 (Figure 4) through the collapse layout and after examination of the Cattell's Elbow rupture. We have noticed that there is a change after the third factor. For this purpose, we retain three (03) factors that lie before the abrupt change in the slope of the Cattell's Elbow break for the analysis, since this criterion is precise than the eigenvalue criterion.

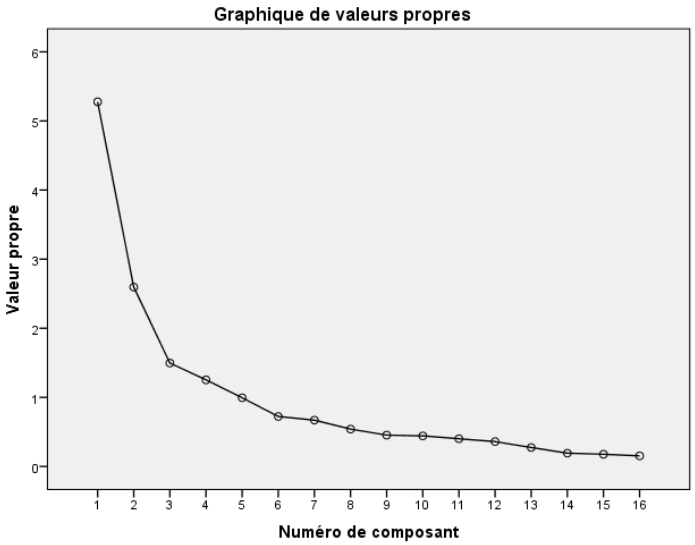


Figure 4: Initial eigenvalues

Table 3: Total variance explained through three dimensions (Extraction method: Principal Component Analysis)

Component	Initial eigenvalues		
	Total	% of the variance	% cumulated
1	3,537	32,153	32,153
2	2,125	19,320	51,473
3	1,327	12,064	63,537
4	0,901	8,193	71,731
5	0,808	7,345	79,076
6	0,630	5,731	84,807
7	0,463	4,210	89,017
8	0,382	3,470	92,488
9	0,340	3,087	95,575
10	0,303	2,754	98,328
11	0,184	1,672	100,000

The exploratory analysis in Table3, shows three (03) factors according to [44], which leads to the selection of the number of factors with a value of 1or greater. A clear factorial solution appears, without overlap. Factors are easily interpretable. In addition, the selected items appear to represent the brand image of the touristic destination in terms of content.

According to the component matrix (Table 4), the first factor consists of 5 items that represent the abstract brand image [45]. The latter characterizes the destination which is Rabat city as an adventures' destination judged by the foreign tourists interviewed, it is unique by its diverse touristic sites, A Sports destination because of the number of sporting events organized, namely the Rabat International Athletics Meeting, the Rabat International Marathon, the Hassan II Golf Trophy ,a city ranked second in the Maghreb region in terms of life quality after Tunis (capital of Tunisia), with a green belt of 1,063 hectares, Rabat

city managed to double the global area of green spaces for each individual given the standard of the World Health Organization set at 10 square meters per person (Tests garden), Nazhat Ibn Sina, Hilton forest, Maamora,... and ultimately, an events' destination (sporting, artistic and cultural).

In this regard, the author [46] argues that it is likely not all associations to the trademark have the same impact on behavior: associations involving symbolic benefits that are also abstract, would have a stronger influence than associations involving functional benefits.

The second factor represents the functional brand image, translated by the aspect of belonging to the territory and the local products, is composed of three items. These are related to the Peaks and Scents, Region City of the Rabat-Salé-Kénitra and the cultural activities namely Mawazine and the Rabat Jazz Festival for instance.

The third "Human and Intangible" factor is composed of three items. Rabat is a friendly city, a welcoming city, a wealthy city in terms of Historical Monuments (Mohammed-V Mausoleum, Oudayas Kasbah, Chellah, Rabat Saint-Pierre Cathedral, Hassan Tower, Mohammed VI Museum …)

Table 4: Components Matrix after rotation

	Component		
	1	2	3
IMA6	0,840		
IMA5	0,837		
IMF10	0,777		
IMA3	0,658		
IMF8	0,564		
IMF4		0,810	
IMA1		0,793	
IMF5		0,673	
IMA2			0,796
IMF9			0,769
IMF7			0,610

Table 5: Main Components' Analysis (Varimax Rotation) of the Rabat Destination Branding Scale (n=454)

Items	Components			Factors
	1	2	3	
IMA6. Adventures' destination	0,840			Brand image Marque Symbolic, Abstract
IMA5. Unique destination	0,837			
IMF10. Sports destination	0,777			
IMA3. Green city	0,658			
IMF8. Events (athletic, artistic, cultural).	0,564			
IMF4. Spices and scents		0,810		Belonging to the Territory and Local Products
IMA1. Region's city of Rabat Salé Kénitra		0,793		
IMF5. Cultural activities		0,673		
IMA2. Friendly city			0,796	Human capital
IMF9. Welcoming city			0,769	
IMF7. Historical monuments			0,610	Human, Capital Intangible

The three factors explain 63,537% (Table 3) of the total variance explained for a KMO of, 758 (Table 1). The internal consistency reliability for each component is mentioned in Table 5. After identifying the items that constitute our measurement scale (Table 5). We presently want to verify whether this scale is stable over time and whether it makes it possible to properly measure the brand image of the destination we have identified. So, we are going to perform an internal consistency analysis.

Table 6 displays the Cronbach's Alpha Index values for each component of the destination brand image. We observe that the value of the 1st component «Symbolic Brand Image» is 0,817, which is excellent, since it exceeds the minimum required threshold of 0,70 [47].

This beacon is arbitrary, but widely accepted by the scientific community. Consequently, we can claim that we obtain, for this scale composed of five elements, a satisfactory internal coherence. Similarly, the value of the 2nd component «Belonging to the Territory and Terroir Product» is 0,764, which is excellent. While the 3rd component's value «Human and intangible capital» is 0,673, which is acceptable.

Table 6: Cronbach's alpha coefficients

Components	Cronbach's alpha values
Symbolic brand image	0,817
Belonging to the territory and local product	0,764
Human and intangible capital	0,673

After examining the tables of total statistics of the elements corresponding to each dimension of the measurement scale of the Brand Image of Rabat Destination (Appendix 1), we have noticed that the Cronbach's alpha values of the first dimension improves (0,819 instead of 0,817) if we suppress item IMF8. Also, the Cronbach's alpha value of the third dimension improves (0,705 instead of 0,673) if we discard item IMF7. So our measurement scale of the Brand Image of the Final Destination of Rabat is presented as follows:

Table 7: City of Rabat Brand Image measurement scale

Factors	Items	Cronbach's Alpha
Symbolic brand image (abstract or functional)	IMA6. An adventures' destination	0,819
	IMA5.A unique destination	
	IMF10.Sports destination	
	IMA3. Green city	
Belonging to the territory and local products.	IMF4.Spices and scents	0,764
	IMA1.Region city of Rabat Salé Kénitra	
	IMF5. Cultural activities	
Human and intangible capital.	IMA2.Friendly city	0,705
	IMF9.Welcoming city	

11.1. Study results and implications

The results obtained show that functional and abstract associations jointly form the brand image of Rabat city in the foreign tourists' memory and that abstract associations are more significant in the choice of the destination. These results confirm that the brand image is constituted of several attributes linked to the associations, whether functional or abstract (the belonging to the territory and local product 0.764). The respondents consider that they are part of this destination and that local products remain a strong argument in the choice of destination. These data make it possible to mobilize a different approach in the management of the Rabat destination which prioritizes a diversified offer for foreign tourists capable of enriching this dimension of belonging and preference for local products and improving their quality.

The resources of the Rabat city destination mentioned in this study represent important distinctive and specific qualities in the overall satisfaction for the visit to a touristic location[48]. Moreover, the attributes linked to associations are a real means in order to develop and strengthen the brand image and, consequently, improving the attractiveness of the destination[49]. The results of this study represent a real reference for the managers of the Rabat destination and the Moroccan tourist sector for new prospects for the segmentation of the touristic offer and the opening up of new markets.

Conclusion

The aim of this research is suggesting a measurement scale of the touristic destination image in a Moroccan setting (Rabat city). To achieve this goal, we have conducted two empirical investigations. First, we have conducted a qualitative exploratory study in the form of ten semi-directive interviews, during July 2019, with a multi-nationality tourists visiting Rabat. Each interview lasted about twenty minutes. Then we have performed a quantitative study among 454 foreign tourists in Rabat during the time period from October to December 2019, and our results show:

- The existence of two different tourist's profiles based on the brand image they develop in their memory, a functional or abstract image and that the abstract brand image (Rabat unique city, city of adventure,0,819 of Cronbach's alpha) is more significant than the functional image (sports city...).These results are more compliant with those of [47] who suggested that not all associations with the brand have the same impact on tourist's behavior, with so-called symbolic or abstract associations having a stronger influence than those relating to functional benefits;
- The brand image, whether functional or abstract, constitutes an enormous potential for the Rabat destination (factor of belonging to the territory and terroir's product) of our model, 0,764);
- Our measurement scale suggests three (03) complementary factors, 9 items to measure the brand image of Rabat city. The three factors explain 63,537% (Table 3) of the total variance explained for a KMO of 0,758 (Table 1);
- At the managerial level, this study will enable the stakeholders of the national tourism sector to draw the necessary conclusions to segment the tourist's market at the level of Rabat city and at the national level according to the identified behaviors and preferences.

Likewise, all scientific research, our study displays some limitations which constitute as many threads of investigation. In fact, the scale has only been tested on one category of foreign tourists - those of Rabat city - and it is therefore appropriate to generalize the measure to several cities. Such an approach can improve the predictive validity of our scale at the national level.

Certain other research perspectives related to the results can be suggested. It would be interesting to study the structural relationships between the 3 factors of our model and to test their impact on other concepts such as attachment to the destination.

References

[1] M. Abakouy, M. Khatib, "Positionnement marketing et développement de la ville de Casablanca en tant que destination touristique," 13, 2016.

[2] B. Meyronin, J. Gayet, G. Collomb, Marketing territorial: enjeux et pratiques, Vuibert, Paris, 2015.

[3] A.E. Khazzar, H. Echattabi, "Les pratiques du marketing territorial dans le contexte marocain : Eléments de réflexion," **16**(1), 14, 2016. https://doi.org/10.48396/IMIST.PRSM/mjeim-v2i1.8989

[4] P. Marcotte, L. Bourdeau, E. Leroux, "Branding et labels en tourisme : réticences et défis," Management Avenir, **47**(7), 205–222, 2011.

[5] M.-U. Proulx, D. Tremblay, "Marketing territorial et positionnement mondial," Geographie, economie, societe, **8**(2), 239–256, 2006. https://www.cairn.info/journal-geographie-economie-societe-2006-2-page-239.htm

[6] C. Camelis, "L'influence de l'expérience sur l'image de la marque de service," Vie sciences de l'entreprise, **182**(2), 57–74, 2009. https://doi.org/10.48396/v2i1.8989

[7] F. Smaoui, "Les déterminants de l'attachement émotionnel à la marque : Effet des variables relationnelles et des variables relatives au produit," 27, 2008.

[8] F. Cusin, J. Damon, "Les villes face aux défis de l'attractivité. Classements, enjeux et stratégies urbaines," Futuribles, **367**, 25–46, 2010, doi:10.1051/futur/36725.

[9] T. Albertini, D. Bereni, G. Luisi, "Une approche comparative des pratiques managériales de la Marque Territoriale Régionale," Gestion et management public, **5**(2), 41–60, 2017. https://www.cairn.info/revue-gestion-et-management-public-2017-2-page-41.htm

[10] C. Lai, I. Aimé, La marque, 2016.

[11] H. Azouaoui, A. Ismaili, "L'identité de la marque de territoire et décision de localisation des entreprises : Approche par la littérature," 21, 2015.

[12] E. Belkadi, "Marketing Territorial de Casablanca : Etude de l'Image de Marque [Place Marketing: The Brand image of Casablanca]," **13**(3), 11, 2015.

[13] É. Tortochot, G. Alt, Design(s): de la conception à la diffusion, Bréal, Rosny-sous-Bois, 2004.

[14] A.-C. Marchat, C. Camelis, "L'image de marque de la destination et son impact sur les comportements post-visite des touristes," Gestion et management public, **5/3**(1), 43, 2017, doi:10.3917/gmp.053.0043.

[15] Malaval Philippe, Pentacom: communication corporate, interne, financière, marketing b-to-c et b-to-b / Philippe Malaval,... Jean-Marc Décaudin,... ; avec la collaboration de Christophe Bénaroya,... Jacques Digout,..., 3e édition, Pearson, Paris, 2012.

[16] A. ABYRE, Y. ALLAOUI, "Savoir mesurer l'image de marque pour pouvoir l'améliorer," Management Research, 7.

[17] J.-L. Leignel, E. Ménager, S. Tablonsky, Performance durable de l'entreprise. Piliers, évaluation, leviers - Jean-Louis Leignel,Emmanuel Ménager,Serge Tablonsky, Dec. 2020.

[18] H. Berrehail, E.H. Boukalkoul, "Concept de l'image de marque," . . May, **5**, 19, 2016.

[19] P.S. Manhas, L.A. Manrai, A.K. Manrai, "Role of tourist destination development in building its brand image: A conceptual model," Journal of Economics, Finance and Administrative Science, **21**(40), 25–29, 2016, doi:10.1016/j.jefas.2016.01.001.

[20] J. Lacœuilhe, "L'attachement à la marque : proposition d'une échelle de mesure," Recherche et Applications En Marketing, **15**(4), 61–77, 2000.

[21] G. Prayag, S. Hosany, K. Odeh, "The role of tourists' emotional experiences and satisfaction in understanding behavioral intentions," Journal of Destination Marketing & Management, **2**, 118–127, 2013, doi:10.1016/j.jdmm.2013.05.001.

[22] G. Dunne, S. Flanagan, J. Buckley, "Towards an Understanding of

International City Break Travel," International Journal of Tourism Research, **12**, 409–417, 2009, doi:10.1002/jtr.760.
[23] V. Minghetti, F. Montaguti, Cities to Play: Outlining Competitive Profiles for European Cities, Springer, Vienna: 171–190, 2010, doi:10.1007/978-3-211-09416-7_10.
[24] M. Gigot, "La patrimonialisation de l'urbain," 9, 2012.
[25] F.D. Grandpré, "Attraits, attractions et produits touristiques : trois concepts distincts dans le contexte d'un développement touristique régional," Téoros. Revue de recherche en tourisme, **26**(26–2), 12–18, 2007.
[26] T. Vo Thanh, V. Kirova, R. Daréous, "L'organisation d'un méga-événement sportif et l'image touristique de la ville hôte : une perspective par le concept de transfert d'image," Téoros : revue de recherche en tourisme, **33**(1), 87–98, 2014, doi:https://doi.org/10.7202/1036722ar.
[27] H. Kim, J.S. Chen, "Destination image formation process: A holistic model," Journal of Vacation Marketing, **22**(2), 154–166, 2016, doi:10.1177/1356766715591870.
[28] C. Dubois, M. Cawley, S. Schmitz, "The tourist on the farm: A 'muddled' image," Tourism Management, **59**(C), 298–311, 2017.
[29] P. Cottet, M.-C. Lichtlé, V. Plichon, J.-M. Ferrandi, "Image d'hospitalité des villes touristiques : le rôle de la communication," Recherches en Sciences de Gestion, **108**(3), 47, 2015, doi:10.3917/resg.108.0047.
[30] S. Zenker, E. Braun, S. Petersen, "Branding the destination versus the place: The effects of brand complexity and identification for residents and visitors," Tourism Management, **58**, 15–27, 2017, doi:10.1016/j.tourman.2016.10.008.
[31] H. Guizani, P. Valette-Florence, "Proposition d'une mesure psychométrique du capital client de la marque," Marche et organisations, **N° 12**(2), 11–41, 2010.
[32] P. Scorrano, M. Fait, L. Iaia, P. Rosato, "The image attributes of a destination: an analysis of the wine tourists' perception," EuroMed Journal of Business, **13**(3), 335–350, 2018, doi:10.1108/EMJB-11-2017-0045.
[33] H. Zhang, Y. Wu, D. Buhalis, "A model of perceived image, memorable tourism experiences and revisit intention," Journal of Destination Marketing & Management, **8**, 326–336, 2018, doi:10.1016/j.jdmm.2017.06.004.
[34] B. Yoo, N. Donthu, Developing a Scale to Measure the Perceived Quality of an Internet Shopping Site (PQISS), Springer International Publishing, Cham: 471–471, 2015, doi:10.1007/978-3-319-11885-7_129.
[35] C. Viot, "David Aaker: Efficacité publicitaire, capital marque, comportement du consommateur et lien marketing-finance," 21, 1994.
[36] K.L. Keller, "Conceptualizing, Measuring, and Managing Customer-Based Brand Equity," Journal of Marketing, **57**(1), 1–22, 1993, doi:10.1177/002224299305700101.
[37] S. Changeur, Le territoire de marque : proposition et test d'un modèle basé sur la mesure des associations des marques, These de doctorat, Aix-Marseille 3, 1999.
[38] K. ZENAIDI, J.-L. Chandon, "Une nouvelle mesure du capital marque par une approche formative," Revue Tunisienne Du Marketing, **1**(1), 8–21, 2009.
[39] G.A. Churchill, "A Paradigm for Developing Better Measures of Marketing Constructs," Journal of Marketing Research, **16**(1), 64–73, 1979, doi:10.2307/3150876.
[40] M.B. Miles, A.M. Huberman, Analyse des données qualitatives, De Boeck Supérieur, 2003.
[41] L. Dialogues, Market - 3ème édition - Études et recherches en... - Yves Evrard, Bernard Pras, Elyette Roux - Dunod, Dec. 2020.
[42] A.B. Costello, J. Osborne, "Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis," 7, **10**, 2005, doi:10.7275/JYJ1-4868.
[43] P. Roussel, F. Wacheux, Management des ressources humaines : méthodes de recherche en sciences humaines et sociales / sous la direction de Patrice Roussel et Frédéric Wacheux, de Boeck. Bruxelles, 2005.
[44] H.F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," Psychometrika, **23**(3), 187–200, 1958, doi:10.1007/BF02289233.
[45] J.L. Crompton, "Motivations for pleasure vacation," Annals of Tourism Research, **6**(4), 408–424, 1979, doi:10.1016/0160-7383(79)90004-5.
[46] M. Korchia, "Une nouvelle typologie de l'image de marque," 16ème congrès international de l'Association Française du Marketing, 18, 2001.
[47] J.C. Nunnally, Psychometric theory, 2nd Edition, McGraw-Hill, New York, 1978.
[48] A.F. Lecompte, M. Trelohan, M. Gentric, M. Aquilina, "Putting sense of place at the centre of place brand development," Journal of Marketing Management, 2017.
[49] R.H. Taplin, "The value of self-stated attribute importance to overall satisfaction," Tourism Management, **33**(2), 295–304, 2012, doi:10.1016/j.tourman.2011.03.008.

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1759-1777 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

The Designing of Institute's Educational Mascots for Brand Identity

Nop Kongdee*, Suparada Prapawong, Manissaward Jintapitak

Department of Animation and Visual Effects, College of Arts, Media and technology, Chiang Mai University, Chiang Mai, 50200, Thailand

ARTICLE INFO

Article history:
Received: 05 November, 2020
Accepted: 21 December, 2020
Online: 30 December, 2020

Keywords:
Character Design
Mascot Design
Toy and Souvenir
Representative
Graphic Identity
Advertising
3D Printing
Institute Marketing
Thai Cultural
Thai Character Design
University Advertising
Graphic Design
Symbol analysis
Concept Arts
Arts and Design

ABSTRACT

The purpose of this study is to show how to design the character used as the symbol or Mascot of the educational institution and how to apply the design results into advertising and Brand identity design such as printing art, poster, digital media, toy, and souvenir. In this study, the ideas of mascot design have been leaded to present a contemporary design under the conceptual framework from local culture, organizational culture, theories of society, arts and design through the analysis and synthesis of the organization, which is CAMT from Chiang Mai University, as the model of this study. CAMT has 7 bachelor's degree majors, namely Animation and Visual effects (ANI), Software Engineering (SE), Digital Film (DF). Modern Management and Information Technology (MMIT), Digital Game (DG), Digital Industry Integration (DII) and Knowledge Innovation Management (KIM). These Mascot design group were applied into Art product such as poster, Line sticker and end up with Toy product as a final stage, it was using digital 3D printing for action figure product in prototype level. This study including Toy creating concept designing and how the outcome product affected to Mascot identity, institution brand and the target groups of study which are widely from generation X to Z, mostly age range 18-23 years old. As a result, it founded that this Mascot package advertising becomes the distinctive character creating the new image of CAMT and the enhanced of Toy design product that raise CAMT remembered as a contemporary and digital organization in Chiang Mai University. Accordingly, it produced the impact results to receivers in many dimensions involving trend, creative behaviors such as Fan Art design, Doujinshi, or Caricature, Cosplay, Arts, and craft also Subculture from the public user.

1. Introduction

This paper is an extension of work originally presented in 2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering [1]. In the lights of arts and designing, a mascot[1] is an effective public relation, promotion and marketing means of building brand images. A mascot is an extension of a logo in order to create the deeper dimension of image by building graphic identities or brand identities, it bears the responsibility of lifting the spirits of the group as well [2] which is highly popular nowadays. The building of a brand can rely on different techniques. However, most of the time, it involves the design of products such as shirts, hats or even collectibles that are memorable and tangible.

As for this project, it concerns the creation of new mascots based on the rebranding of College of Arts, Media, and Technology (CAMT[2]) Organization, Chiang Mai University,

Thailand. This college offers courses that are relating to Digital Contents in fields of Sciences and Arts focusing on innovation and digital technology. This project involves the designing of the new mascots for CAMT that correspond to the vision of the organization that emphasizes on digital innovation and advancement. Thus, the direction for the designing of the mascots/characters focuses on modernity and concurrence with the vision of CAMT and images of Chiang Mai University while

[1] Mascot is any person, object, animal used to represent a group of common identity mostly use for professional sport team also involve with advertising and marketing of organization or team.

*Corresponding Author: Nop Kongdee, Email : nopmerzy@gmail.com

building new images that focus on the dimension if arts and designing of Chiang Mai province simultaneously. Thus, it is apparent that the scope for the design corresponds to culture, images and perspectives that are overlapping on one another in various dimensions. Nevertheless, a new and meaningful image, which is the core outcome of this project, depends on the designing method, analysis and definition.

Marketing and promotional extensions an important element that leads to the memorization of brand identities. To achieve this requires images to consist of multi-dimensional perspectives that do not go beyond the conceptual framework, such as the design of envelope, letterhead, graphics on shirts and even graphic posters in the forms of digital media and digital image communication (stickers of Line application), as well as the design of toys and souvenirs, which are included as parts of the creation in this project. It is expected that the designed mascots will lead to the added value to the brand and the wide publication of promotional media and images in unique manners.

2. Literature Review of Mascot and Brand

In [3], it is stated that brand identity is how a company is being identified. The consistency of this brand identity is formed by its features like culture, vision, personality, positioning, presentations, relationships and other meaning beliefs followed by the entity. In term of commerce, brand identity is required to create costumers' awareness and remembrance in order to get access and chosen from consumers. However, there are many methods of brand identity expression which mascot is one of the potential paths.

It is quoted that the origin of the word "Mascot" is derived from "Mascara" or mask in the etymological dictionary [4], Corominas (1980, 1985) collected the first identifiable written use of the word "Mascota" or mascot in the year 1233. The word "Mascot" comes from the provincial language "Mascoto" which means magical things or amulets. Afterwards, it has been abbreviated to be "Masco" which roughly means which that relates to supernatural [5]. This is the linguistic and meaning background of the word "Mascot" which leads to the perception that "Mascot" is the representative of mask and magic.

"Mascot" would be a real person or animal or a character (or a group) used continuously and in different poses and attitudes to represent, advertise or promote a product, a brand, a company or institution, an event or a specific cause [6]. Therefore, mascot could be made of and could be representation of things for gaining awareness and recall and also could make the product alive and cause positive attitude for consumers.

"Mascot is being heralded as the magic bullet in the contemporary marcomms armoury [7]. Visual representation of a brand by using brand mascots is a technique which has been applied by organizations for many years. As the visual 'ambassador' of a brand, the mascot's goal is to strengthen the brand identity. The visual properties which are translated by these mascots can cause the distinctiveness of a brand [8]. It could be concluded that mascot is one of the most powerful tools for communication of brand's identity as we could see many famous brands with their identities' representation by mascots for instance: Michelin Man by Michelin, Mr. Peanut by Planters, The Laughing Cow by The Laughing Cow, Mickey Mouse by The Walt Disney Company, Jolly Green Giant by B&G Foods, etc. Therefore, it has assured that mascot is appropriate with intimate image with consumers setting organizations [9]. Another reaffirmation is that Mascot usage in Thailand tends to be usage for public relations mainly in events [10]. It has been mentioned that 60% of tourists in Shiga Province, Japan has been increased by the Hikonyan mascot. All this information insist that mascot is a very important instrument for communication, especially for brand identity representation.

3. Analysis of the Symbol

"Meaning" is defined that while everything has its meaning, meanwhile, everything does not have a chance to form its meaning by itself. Human is the entire meanings former which is not under human nature in an individual context to create meaning for things. It takes at least two or more humans to perform to make it meaningful where perception of meaning of things is based on experiences from the five senses. Since the past, interpretation and meaning transmission have been hidden in many visual languages [11]. Symbol is one of the hidden visual languages which is significantly influence interpretation and perception of human for a long time.

In general, mascot is a mean to create brand identity, and has the objective to enable people to remember an organization through a unique character in addition to logo design because a mascot is created within the similar framework of art concepts, such as the use of colors, shapes and forms. A mascot is one of the channels representing semiotics which denote sign and symbol that has been produced and sent or transmitted to indicate how the meaning of representation was formed, including its understanding of a thing or a process of giving such meaning. Semiology was a tool that could make individuals understand the formation of meaning and its application [12]. Semiotic theory is an "Interpretive" theory that can be applied to most aspects of everyday life although most people would not realise it.

However, a character has the complicated story, as well as emotions, characteristics, identities, style and overall image through the characteristics of the character. Meanwhile, logo design has the limitation; but to create a mascot requires creator to consider the campaigns that the mascot is used for and how to make people remember the mascot. In addition, mascot power can build harmony and unity in an organization such as the use of a mascot of sport teams to promote the organization and tourism of the local area. For instance, the use of the raccoon mascot of American football team of Tennessee, USA. Raccoon is a local animal of Tennessee State; thus, it is used as the mascot to express

[2] CAMT or College of Arts, Media and Technology is a Digital School located at Chiang Mai University, Thailand.

cultural identity. Therefore, mascot creation is a mean to present organizational culture and images based on the concept of artistic marketing or 'Graphic Identity, which is a branch of marketing strategy. In the education and academic society, it is known that in Thailand, animals are used for representing, which is relating to the contemporary society, and has endless development in accordance with the ever-changing trend in the global society. As for Chiang Mai University, the representing symbol is an elephant. However, the University has no exact or clear mascot. However, based on the definition of the symbol, it can be said that the keyword that is significant for the creation of core identity of Chiang Mai University is a violet elephant, which has been used for a long period of time although there has never been any exact or official mascot. Therefore, it can be said that a symbol needs people to drive it and time for cultural absorption until it is memorized and defined eventually.

In this case study, College of Arts, Media and Technology (CAMT) is an organization of the university. Thus, the culture of the university dominates the identity of this organization. Consequently, this organization has 2 logos of 2 different types. The first logo is the core logo of the university that is jointly used by all faculties, with the name changed in accordance with each faculty, and the defined logo that is especially designed for the organization (as shown in Figure.1) in order to express the definition of the organization that is contemporary and specific through the use of art compositions and designing.

Figure 1: Logos of College of Arts, Media and Technology (CAMT): Core Logo and Defined Logo

The aforementioned logos reflect different definitions. The core logo of the university is composed of the symbolic animal and violet color as the focal points of the local, which create identities of the university. Meanwhile, the shapes and forms of components in the logo clearly reflect the Thai style. Meanwhile, the defined logo of CAMT is completely different from the former because it reflects the modernity of the organization. The definition of CAMT is the emphasis on inventions through digital contents. This definition concurs with the courses that this organization opens; hence, the logo is designed to be in a modern form that is minimized, without any implication or symbol of Thai culture used at all. However, the status as being Thai has been defined in the name that is another defined image. In other words, the status as being Thai or Lanna Thai has already been inserted in the contexts of the organization already. However, images or representing symbol that are shown have the aim to reflect the status as a contemporary organization. In addition, the use of color of the organization (faculty) reflects the status as a luxurious and international organization. This logo has been developed from the previous logo that was defined 10 years ago (as shown in Figure 2).

Figure 2: Previous logo of College of Arts, Media and Technology (CAMT)

It is apparent that the contexts of CAMT have changed and the logo has been minimized with the different aim and in a different manner. Originally, the concept of the logo was to make people memorize the alphabets, but the current logo emphasizes on making people memorize an image with contexts. The letter 'Can has the form like a beehive or geometric form, which concurs with the previous definition. Nevertheless, the contexts of CAMT are flowing and changing along with the changes in different eras and the expansion of scientific and art concepts that still create the complexity of the organization.

Since symbols are changed in accordance with societal ideas and trends that serve as driving forces, mascot designing has been changed accordingly. Originally, there had been 3 mascots to represent the Department of Animation and Games, the Department of Software Engineering and the Department of Modern Management (MMIT) because there were 3 departments back then, which were divided into 6 branches in the bachelor level, as mentioned before. Thus, an interesting question is whether the contemporary design of mascot to reinforce the images of an organization should be for the timeline of 10 years in advance, together with the analysis on the nature and ideas of university students nowadays and in the future (Generations Y-Z) in the scope of organizations of university students or not. However, the implication of CAMT from previous mascot still has core definitions that include freshness, curiosity, contemporaneity and ethicality of students still remain in the symbol. Therefore, it is necessary to use the previous mascot as the basis for the analysis and development for designing a new mascot.

3.1. Previous Mascot

Originally, the previous mascots of CAMT had been designed with the focus on the education and the teaching and learning that was clearly separated by departments. Thus, the designed characters were students in the uniform that precisely reflect the characteristics of the 3 departments, namely, Animation and Game, Modern Management and Software Engineering in the bachelor level (as shown in Figure 3).

Analyzed in the aspect of arts, the 3 previous characters were designed to have color theme which was like an old logo that reflects moods through facial expressions of a board character[3], the characteristics of which are reflected through facial expression and shape and form of the eyes. As for the proportion, each of the characters have the proportion of a kid, with a big head, big hands

and a small body. This style of design is called Super Deformation (SD)[4], which is the deformation of a human body into a downsized body like a deformed body of a small kid. The advantage of this style is that the designed characters are trendy in accordance with the drawing style of Japanese graphic novel that is widely popular. Consequently, the characters are eye-catching and easy to remember.

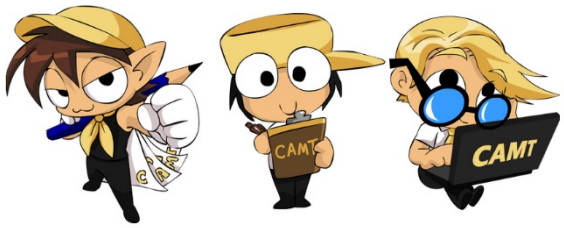


Figure 3: The previous Mascots of College of Arts, Media and Technology (CAMT)

Nevertheless, the mascots with the old definitions and policies have been used for more than 10 years since the establishment of the organization. With new policies and directions, there has been the need for extension and development of these characters so that they are contemporary while reflecting the identities of CAMT and CMU simultaneously.

3.2. Data Identification

College of Arts, Media and Technology has the educational approach that focuses on complexity which includes both sciences and arts, taught in a faculty that has learning method that is changed in accordance with eras, with the emphasis on innovation and digital technology as key principles. Therefore, the analysis and definition of the identities of the university are not fixed because the concepts and directions change annually in accordance with eras and trends. However, some keywords of the organization still remain. For instance, an organization that emphasizes on agriculture and technology may use lines and colors in organic art form in order to enable people to understand the definitions immediately. Alternatively, the Faculty of Political Science and Public Administration uses a scale and a lion as symbols that are related to Thai beliefs and symbolic definition which are combined together to form identities of the organization. As for CAMT, the symbol is designed to reflect the identities CAMT. However, such identities are under the domination of the identities of Chiang Mai University. Thus, there is the combination between identities of CAMT and identities of CMU, only for the identities from the social definition and graphic identities.

Chiang Mai University is an old complex university that provides courses of several subjects including medicine, sciences and arts in several branches, and extends and develops new curricula that fit modern era. From the view of the society, Chiang Mai University is a key University in the northern region of Thailand. At present, the university has the policy that focuses on digital technology and graduates of new generation with the competencies for diverse jobs, ethicality and promptness to the international level. However, such driving also leads to the change in the view of identities. In other words, this leads to the status as a contemporary university that accepts more foreign students, with the emphasis on linguistic aspect and tour system for foreigners in the university, or even the expansion of territory and branches in the northern region of Thailand, and the use of digital systems in the university, which turn the university Smart University, such as vehicle tracking system, game application for introducing places, class attendance (check-in) system with QR code and online method. Therefore, it is apparent that the current definition of the university is concurrent with the policy of CAMT because they are overlapping with each other and go in the same direction.

In addition, the definitions in the two aforementioned parts are also dominated by Lanna Thai status. It can be said that the Thai status and identities in terms of culture, thinking and traditions have control over the overall identity that cannot go out of this frame of social concept or definition. Consequently, the unique identity is created from the combination of several aspects such as media, people, policies, trends and politics, all of which are combined together. Therefore, the identity from the combination of the aforementioned aspects is a basis and component for designing a mascot with a clear scope that is different from a character that is a mascot of a private organization that has a broader scope. Nevertheless, this is no disadvantage. Rather, it enables the creation and designing of a character become clearer. However, there are several factors that are relevant, especially the concepts of the artist and the society which are key factors that lead to the building of the Character Style that are influenced by histories and the society as well.

3.3. Cultural Appropriation and Domination

Cultural Appropriation[5] is an approach concerning arts and society and is a key factor for the creation of art works. Therefore, it has both direct and indirect influences on the designing of mascots and characters based on experiences and societal perceptions of the artists or designers, as well as social trends, approaches referred to in a particular era and the implantation of ideas. For example, in this case, it is the creation of characters in the style of manga, which is a result from the great influences that Japanese manga works explored in Thailand to current. As for the designing, the thinking system of the artists focuses on applied art or extended arts, with the adherence to the origin. Therefore, the new identity reflects the combination among manga, original Thai arts and each artist's ideas. Not only does Thai status in arts and designing refer to lines and drawing but it also means identities, actions and stories that are added to a particular character. Thus, a character can be deemed to have identities that are based on Thai culture although it has culturally mutated or developed, which is called the 'mixed blood' or 'mongrel'. Nevertheless, based on the

[3] Board Character, according to the principle of character designing, is a character the characteristics and moods of which are mainly expressed with its look such as Scooby-Doo, 1969, Ruby J.

[4] Super Deformation Character (SD) is the style where by a character is deformed to look like a child, which originates from Japanese 'mangas'; and is sometimes called "Chibi" meaning a small thing, a petite person or a small kid.

approach of Culture Appropriation, the aforementioned phenomenon may be deemed a wrongdoing like plagiarism or copying of styles from any other culture in order to build and claim as our culture. Thus, the definitions of culture in various approaches are unclear. On the contrary, identity that is especially from the development of the aforementioned basic approach is accepted by people in the society and eventually becomes a symbol. Thus, the borrowing of culture that is caused from the domination over thinking occurs all the time, with the society acting like the judge and artists acting as presenters [13]. However, goals, objectives and approaches that newly occur have more significant implication. Hence, as for mascot designing at present, artists design their works to reflect the influences from mangas because of the aforementioned factors. It is not surprising that a newly designed character will look like a manga character. In addition, the key overlapping definitions, combined with styles, can create specific identity that serves as representation[6] of that society or organization as well. Therefore, cultural burrowing can occur all the time as a normal event in the society and in the field of designing.

3.4. Character Population and Demand

First of all, the target audiences for the mascot designing in this project are students in the age range of 15 – 25 years (Generation x) at present. From the survey, it is discovered that the characters that are popular are in the style of Manga[7] and have features of a kid (Manga Cute Characters). The advantage of this approach is the popularity of the style among the target audiences and audiences in other age ranges because cute characters do not mean manga style only; rather, it also means western characters such as Bug Bunny (W. Brothers, 1940). This character is used in western cartoon. It should be noted that this character is not originally created to serve as a symbol or logo but for producing animations for children audience. However, since the popularity of this character rises by dint of the expansion of media, Bug Bunny has become a symbol of western cartoon. Furthermore, Mickey Mouse (W. Disney, 1928) and Elza (C. Buck, 2013) have made Disney's cartoon characters memorable and great. In addition, may other characters from movies have also been popular and a societal factor. Examples of such characters are super hero group from Marvel Comics (1939), which have created trends that affect behaviors and thinking. Apart from memorable identities and looks, other attributes of these characters such as characteristics and thinking of the characters can generate new trends and thinking of people in the society. They can also trigger the building of social groups in accordance with personal preferences, which is concurrent with the concept of freedom of thinking nowadays. Looking back to Asian society that has manga characters as the foundations for thinking, one can say that manga characters were originally created to disseminate Japanese culture and thinking. However, with the popularity, manga characters have expanded to Thailand and become imprinted into many Thai artists during their childhood. Consequently, such artists design their works in the directions that they have been imprinted and in accordance with the popularity that evolves in each era.

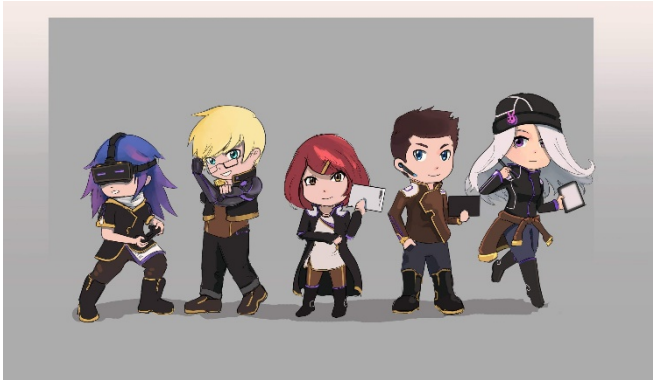


Figure 4: The Experimental Designs of Mascots:
DG – SE – MMIT - Dii – ANI draft 1

The mascots designed in this project belong to an educational organization. Thus, the outfits should be a uniform that reflects Thai culture in the improvised style and politeness in the society, mixed with the identities that the university wants to present, namely, digital technology, modernity and future world, with the basics of the identities from the identities of the university. In this case, violet, which is the color of the university, is used as a component of the design. Thus, in this designing experiment, culture domination and character population play prominent roles. As a result, there is a clear direction for the possible design.

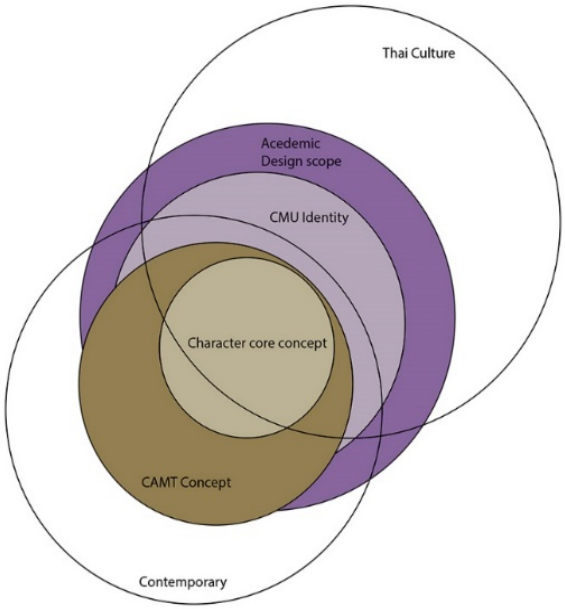


Figure 5: CAMT Mascot Design Scope, the overlapping area of contents analysis to finding of core concept design.

[5] Cultural Appropriation is the influenced of the artistìs background. In this case, it refers to Manga arts which applied to character design harmonious with Thai culture, CMU and CAMT.

[6] Representation in this case it intended merging of semiology and concept idea to presented something in specific purposed which approached to the meaning of Art work as a core concept.

[7] Manga are the Japanese comic or graphic novels style and become popular cultural in worldwide level since late 19th century.

Concerning the concept for designing the characters to represent the educational institute in this case, or CMU (as shown in Figure5), it is apparent that the scopes of the concepts are overlapping. Thai culture has partly taken part in the determination of the concepts of the characters because CAMT also has concepts in the international level. It cannot adhere to only Thai culture. Rather, foreign cultures have also partly played roles. At the same time, certain concepts that are the scopes of an educational organization, such as uniform, knowledge pursuit (curiosity), intelligence and friendliness. Such concepts belong to CMU and CAMT itself. The scopes of CMU have covered all the characters. At the same time, concepts of CAMT are slightly different from CMU because its identities are different from those of other faculties and are harmonious. In addition, such identities include contemporaneity that regards the directions of creation with all overlapping conceptual frameworks and scopes that cover the concepts of characters. Thus, it is apparent that the contemporary status and the Thai status are mixed together in the even ratio, with the concepts of characters as the cores for the designing and creation. In addition, feedback from audiences and users of the mascots is also taken into account. The most important aspect is the outcomes from the designing, which are regarded as the priority purpose.

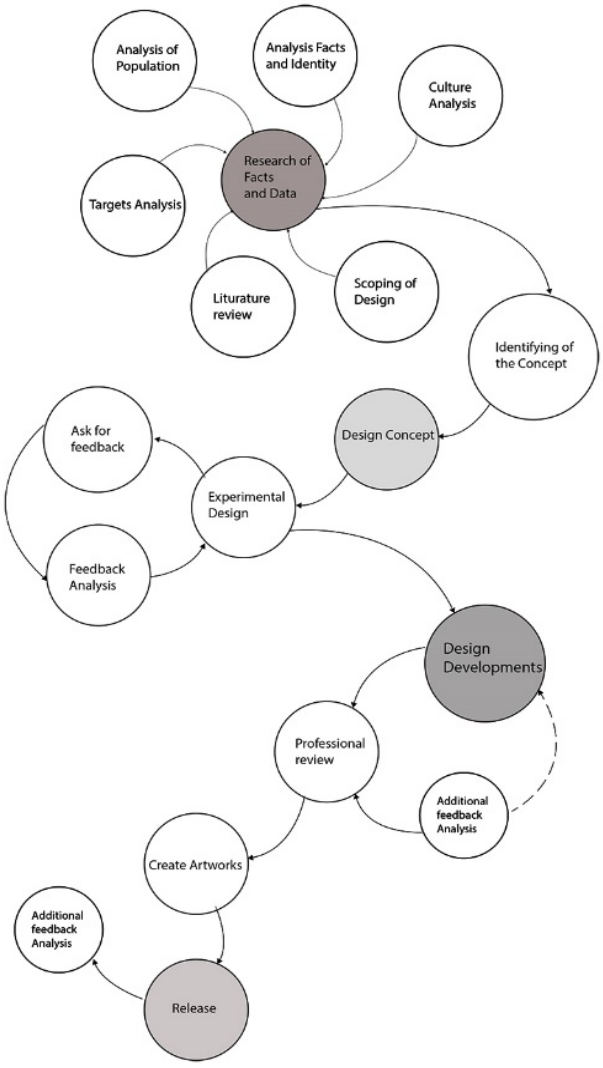


Figure 6: Design frameworks that specific for this project that require the audiences create an idea together with the artists.

4. Methodology

As for this study, the framework of the designing process consists of 4 phases that are 4 key steps, as follows. 1) Pre-production is the step that includes data research and analysis and determination for definitions and scope of the outputs and directions for designing, and study on the experiments by other artists. 2) Production is the step of the development of the designs that build images and patterns of the characters in ways that concur with principles of arts and designing and the concepts of contemporary design. Thus, this step incorporates experiments and creation of images in different dimensions. 3) Post-production is the step that includes the creation of media to present the outcomes from mascot designing, printed media to display in the building of CAMT, and promotional media including Line sticker and action figures. 4) Development and Conclusion is the step in which data are stored and products that have been released are refined in special occasions in order to add values to the brand in the consistent manner. In addition, new products are launched and new experiments in more diverse styles are conducted in accordance with social behaviors in each particular era.

In addition to product development, each of the aforementioned steps also includes the experiments that are based on the feedback from experts and some audience. The target audiences are divided into different age ranges, from people of generation Y or high school and university students to working adults. The designs of the mascots in this project must be neutral as possible whilst satisfaction and recognition should be supported by and concurrent with opinions and critics in social media as well.

5. Character Concept Identifying

The design definitions of each characters, Animation and Visual effects (ANI), Software Engineering (SE), Digital Film (DF). Modern Management and Information Technology (MMIT), Digital Game (DG), Industry Integration (DII) and Knowledge Innovation Management (KIM). The main concepts are contemporary and trendy including clothes and attitudes. The characters have their own distinctive characteristics, but work as a team. The same as CAMT concept that each major are related each other. The visions are combining which the words Honest, Discipline and Sharing (HOD)[8] which comes from the core values of CAMT (CMU CEO)[9] defining the current identity of CAMT. These definition and keywords are applied into Mascot mind and characteristic as a concept identity.

The definition of the Mascot of Animation and Visual Effect (ANI) is that this character has a high level of artistic features with individual thinking in digital technology-related arts. Therefore, the key instruments for creating the works in this project is the tablet pen painting plate, with the emphasis on creativity that allows the character to have high flexibility in terms of its appearances, which mean the character will not strictly adhere to any principle or rule.

As for the Mascot of Software Engineering (SE), this character is defined to be a scientific person who emphasizes on complex and systematic thinking with ability to create new software and

has literacy in special languages or digital ones (programming). Hence, this character relies mainly on left side brain (IQ) and truth.

As for the Mascot of Modern Management and Information Technology (MMIT), this character is defined as a person with leadership, innovativeness, capability of managing team members and planning, and ability to think and analyze, and intuition for forecasting things in the society through digital tools. This character is thus classified to have administrative skills and instant confidence for sound judgment.

In addition, the Mascot of Digital Game (DG) is defined separately from Software Engineering, but the emphasis is put on the creation of digital games. Thus, this character is both artistic and scientific and have more commitment than others. At present, there is a competition called E-sport. Hence, this character incorporates competitiveness, innovativeness and creativity.

Meanwhile, the Mascot of Digital Industry Integration (DII) is defined as a character who is based on facts and truth in the society, with the characteristics that are changing in accordance with the social trend. This character can learn from the issues in the digital industry; and thus, has diverse characteristic and ideas that are consistently changed. This character has diverse knowledge and ability to gather all knowledge in order to create new knowledge that is concise and easy to use.

The Mascot of Digital Film (DF) is defined as a character with creativity like the ANI Mascot, but this character focuses on the creation of cinematographic media, with combination between reality and imagination, and confidence. In addition, this character has capabilities and uniqueness in the field of entertainment arts because it has most prominent look of all members in the team. This character can be a director and an actor; thus, this character has as high confidence as MMIT Mascot, but in terms of entertainment only.

The definition of the Mascot Knowledge Innovation Management (KIM) is that it is the character that is regarded as the source of knowledge for the team because this subject is in Master Degree and Doctoral Degree levels only. The character is the eldest and most experienced of the team, with the characteristics to provide suggestions and advice to others in order to create new things from available information. Therefore, this character is prudent and does not show the emotion clearly. The special capability of this character is to analyze and file a great amount of information.

From the definitions of the characters, each character can be clearly and individually identified with values and policies of CAMT. Consequently, each character has specific definitions that are concurrent with the image and look of the character and different from another character depending on the subject. This is the concept for the designs of all the characters that has been combined with contemporaneity.

5.1. First Look of Design

After the data management is complete, in the Production Step, there will be the experiments of the art works by creating the mascots starting from forms, along with color designing, all of which are based on the definitions of the characters. In the image, it is apparent that gadgets that are related to each subject are added. For example, the ANI will hold a tablet and is floating all the time in order to show the implication of imagination and freedom in terms of arts. The experiments are still based in the feedback from the experts, which will be followed by the refinement of the designs because the designed characters are still not artfully perfect and are too complicated for the memorability.

Figure 7: CAMT early mascot part 2 design from artist A with digital gadget design e.g. tablet, gaming keyboard or smart phone.

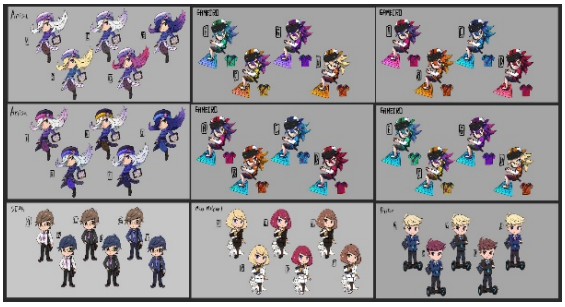


Figure 8: CAMT early Mascot design color alternative experiment

The using of color and shape, the trendy costume is also the key component by using the concept of Science fiction (Sci-Fi). The concept is that the costume design should indicate the futurism and the digital technology infused with CMU attributes and Lanna culture for representing the overall CAMT identity. In this process, there are experiments of color composition in various patterns (as shown in Figure7) as the attempts to clarify the identities of the mascot e.g. using red to portray danger and using blue to portray serenity.

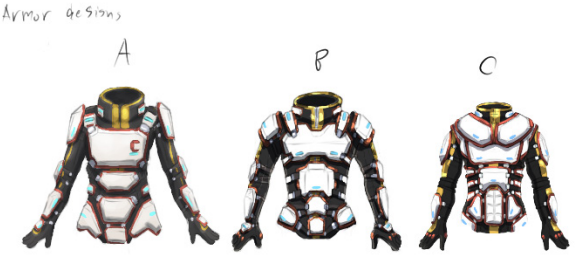


Figure 9: Early Digital CAMT Costume Design, Armor design

Concerning the designing of costumes of characters (as shown in Figure 9), initially, it is apparent that the look of the costume is too hard, which reduces the gentleness of the characters. In addition, the stories of characters are not agreeable. Thus, it is necessary that fabric materials are used as much as possible in order to create the costume that can be used in real life.

[8] HOD the definition of person who works at CAMT and part of CMU CEO representative words with honest, Discipline and Sharing used for personal mind at CAMT should have.

[9] CMU CEO the CAMT core values comprised with Community, Morality, Unity, Creativity, Excellence and Open for Challenge.

Furthermore, the proportion of the characters is changed from SD (Superhero Scale) in order to present charisma, confidence and implication of leadership, as well as the results from the analysis of the popular characters. As for this case, the characters from animation and game media are used. It is also apparent that characters from Ragnarok Online [10] Game have been highly popular for the past 20 years.

Figure 10: The example of ANI character transformation, *SD* re-scale design process.

Apart from rescaling, it is also vital to focus on the gestures of characters so that the identities and habits of each character will be presented more clearly. To do this, it is necessary to design the sketches in order to promote the mascots simultaneously by analyzing the characteristics of each character. For example, the Mascot of Digital Game (DG) (as shown in Figure 11) is reserved, loves playing games, and is not talkative. However, this character is keen on computer use and programming. This character is obsessed with some specific matters only. Such characteristics and habits are from the analysis of people of generation X who have complexity in thinking, individualism and introversion in high levels. However, with the light of the identities of CAMT people, this character also has optimism, which adds cheerfulness to this character until new identities are built. As a result, this character has gestures that show consistent movement and high responsiveness because most gamers have responsiveness rate that is faster than that of non-gamers. This is because of the main activities of the character.

Figure 11: The alternative of Character posing Design of CAMT DG Mascot

Unlike DG Mascot, Digital Industry Integration (DII) Mascot (as shown in Figure 12) is confident and capable of getting along with others easily. As a new character because the branch has been established this year, the character is designed to be still and prudent, with systematic thinking for management and confidence, but not in a great level as the confidence of MMIT and DF characters. In other words, this character can talk to, meet with and learn from other people all the time as reflected through the habit of the character whereby the character wears Bluetooth all the time with the promptness to discuss matters with people. At the same time, the character has still and stable gestures, with less movements than that of DG character.

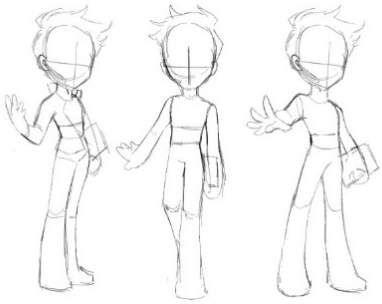


Figure 12: The alternative of Character posing design of CAMT DII Mascot: Confident Gestures.

Therefore, the creation of new postures that is carried out simultaneously with the rescaling is a significant step for adding personalities to each character. For instance, SE (Software Engineering) character (as shown in Figure 13) the uniqueness of which is the Smart Digital Arm. This character is the smartest one of the team, with expertise in language literacy and knowledge of programming. Hence, the gestures of this character represent confidence and uniqueness of the character. Initially, the postures emphasize on the Smart Digital Arm that is the unique feature of the character. Furthermore, the postures make the character look like a superhero but the child-like SD proportion make the character not much different from others in the team.

Figure 13: The alternative of Character posing design of CAMT SE Mascot

5.2. Detailing and Identifying

After designing the posing of characters, the next step is detailing and finishing the costumes by adding lines and details from the postures of the mascots. In this step, it is possible to create the graphics that can be actually used. In other words, there is the production of concept art which will be promoted in the step of publicity, which will shorten the period of time for media production. Nevertheless, there is a limitation that comes from a combination of organization synthesis. Besides, another limitation is that the costumes including hairs and faces have to be different and be infused with the gimmick[11] and the symbol. These are used to enhance the understanding of the receivers, e.g. Using of Digital suit look mixed Traditional Thai culture on fabric

[10] Ragnarok online is MMORPG game developed by Gravity interactive Company in Korea. The contents are based on manhwa Ragnarok by Lee Myung-jin. The game is the one of the most popular in the world and character design population.

form and texture with which can be seen in the details of the character, MMIT (as shown in Figure 14).

Figure 14: Detailing and Re-posing of design art concept, MMIT and ANI Characters

After the drawing process, more details are added to the characters. Each character is presented on different sides (in Character Sheet), together with expressions (in Expression Sheet) with details of gadget of each character in order to extend the story and details of the character in the future. In this step, the key action is to determine colors used for creating graphic identity without any deviation. The color system used is RGB that is concurrent with the color of the organization.

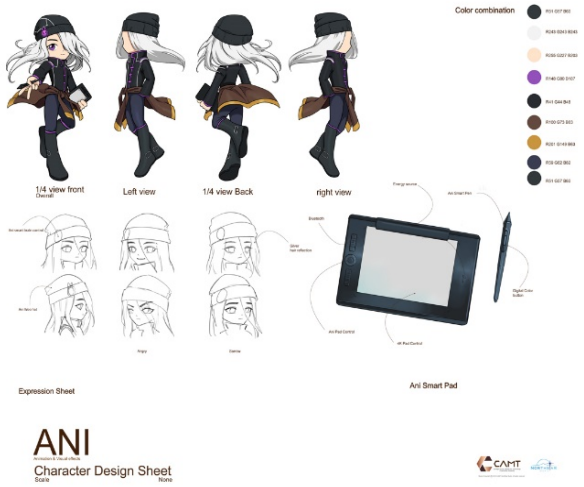


Figure 15: Character design Sheet of CAMT ANI (Animation and Visual Effects) Mascot

15, the ANI (Animation and visual effects) character is based on the personalities of the students of this branch and the body of knowledge from the analysis, the results from which reveal that the character is highly individual, especially in terms of art, or computer art, reserved, and outgoing, not for speaking but for presentation of arts and opinions. Therefore, the character is designed to have long silver hair, which reflects the obsession of independence. Thus, the character can fly, which is like the characteristics that is flexible, highly independent and sometimes introversive. The character carries a digital tablet as the important gadget that is used for drawing all the time. The eyes of the character are in violet color, which is a gimmick because violet is the color of CMU, which is inserted in the concept of the character. Meanwhile, the costume of the character is in the same direction as CAMT suit, with color tone limited to brown color and the character seems to be floating all the time.

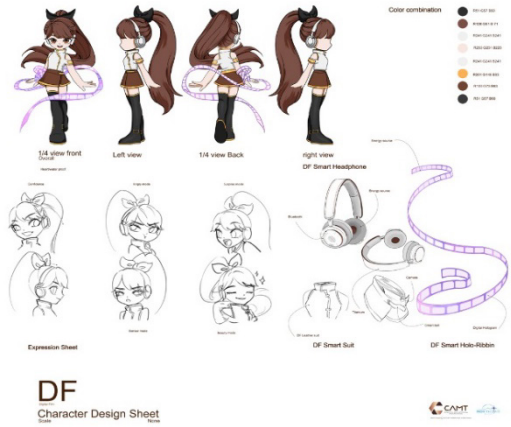


Figure 16: Character Sheet of CAMT DF (Digital Film) Mascot

Digital film) character is the combination of digital art and communication arts relating to entertainment art, especially cinematography. Therefore, the character has individualism as the ANI character and confidence to present self-identities because of capabilities and knowledge of acting and producing for films. Hence, this character is girly but confident, good looking and taking care of herself all the time, in terms of personalities and beauty. This is the only character of the team that is related to front stage. The accessory that represents this character is the Hologram Film Ribbon that is a traditional film roll. It symbolizes that although the character is in digital era, she still has knowledge of cinematography history. In addition, this symbol is easy to understand for propel in general, based on basic perception. In other words, when people see the symbol, they can immediately define the character. This character wears a miniskirt, but the miniskirt is also in the form of the contemporary uniform of the university, with the senses of fashion and sex appeal added. At the same time, the colors used are still related to the color of the organization, or brown in different shades.

Figure 17: Character Sheet of CAMT DG (Digital GAME) Mascot

[11] Gimmick, in this case is the way to input symbol identity into the design, including color, facial or accessories

DG (Digital Game) character is not much different from ANI one in terms of art statement with high levels of confidence and introversion. The differences are the love to competition and the capability of programming, which make this character inferior to ANI one in terms of arts. However, this character has complex thinking relating knowledge of digital interactive. This character is also inspired by e-sports that turn some game players to game producers. Hence, the character is highly obsessed and self-oriented. The key tool of this character is a traditional joystick that is linked to Oculus [2]. This tool is the key gimmick that presents a symbol that can hasten people's understanding on the concepts of the character. Meanwhile, the costume of the character has colors that are concurrent with the color of the organization. The costume is also designed to be agreeable to e-sport game team suit.

Figure 18: Character Sheet of CAMT DII (Digital Industry Integration) Mascot

Figure 19: Character Sheet of CAMT KIM (Knowledge Innovation Management) Mascot

DII (Digital Industry Integration) character is a male character and the only character of the team is has high confidence and skills of speaking and presentation, as well as concern with personality, which is reflected through the face and hairstyle that look polite and slick. The body of knowledge in this branch is related to the building of bodies of knowledge that can be implemented to contemporary industries in order to always catch up with current trends. Therefore, the character is confident and highly capable of communicating with others. The personal gadgets of this character are a smart Bluetooth and a notepad that the character carries all the time. This can manifest the implication of speaking and analytical skills, high are the key skills in this character. The costume of the character is adapted from suit jacket that makes the character looks more agile and active than other characters in the team.

Knowledge Innovation Management) character is the eldest character of the team. This character works at the back of the stager as the consultant of the team. This character collect and file all bodies of knowledge in every subject in the world. Thus, he is designed to have some gray hair as highlight, which is a tangible symbol to represent an aged person. Another unique point of this character is that he sits all the time, with the characteristics of a leadership who does not disclose himself or is the person secretly working at the back to ensure the success of the team. The significant skills of this character include the dissemination of knowledge, and the provision of advice in the fields of academy. The key gadget is the smart throne that is a tool to show and to store data like a hard disk. In addition, this gadget allows the character to communicate with others all the time. It is deemed as one of the most versatile digital gadgets because it can be used for creating, storing, and communicating. This throne is designed to have unique symbolic uniqueness of this character because it has violet color that is the thematic color of Chiang Mai University.

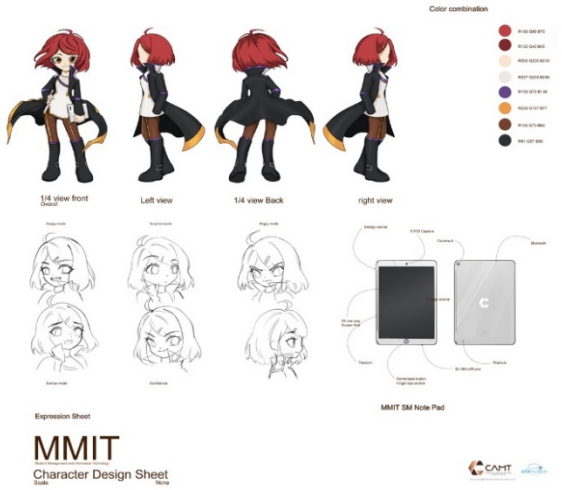


Figure 20: Character Sheet of CAMT MMIT (Modern Management and Information Technology) Mascot

As for MMIT (Modern Management and Information Technology) Mascot, this character is the leader of the team. She is capable of managing things quickly and efficiently. This character is also the key planner of the group and is also as capable of communication as DII character but has a greater level of leadership. Therefore, the character is designed to have the facial expression that reflects her confidence, with gestures that show high agility and delicacy, which are presented through the costume that consists of the cloak that waving all the time. The character has red hair which reflects boldness and confidence at the same

[12] Oculus is a VR game gadget from American technology company that used for visual reality in game specific support hardware and software in digital game.

time. Also, MMIT character has the key gadget of this character is the Smart Light Notebook that can store a lot of data but less than that of KIM character. However, the gadget of this character allows the character to have agility like a note pad does. Nevertheless, the concept of this character is inspired by the overall personality of the graduates of this branch, and the body of knowledge that has recently been developed.

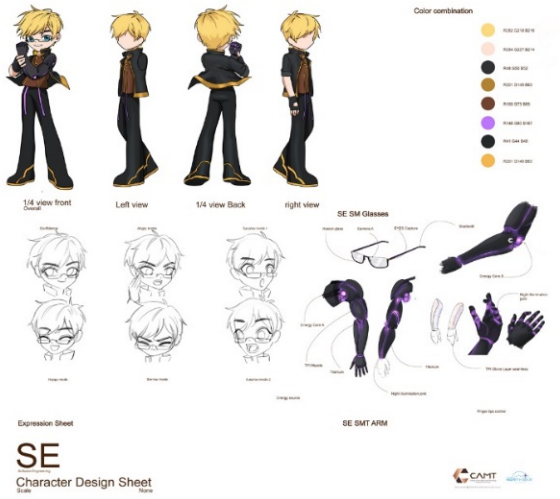


Figure 21: Character sheet of CAMT SE (Software Engineering) Mascot

Software Engineering) Mascot is deemed as the smartest character of the team, and has the knowledge of programming, production and language. This is the only character of the team that has mixed blood (a Thai – American person). The weakness of this character is the skills of arts. Even though being the smartest one, this character has to rely on ideas of ANI character as well. Therefore, SE character and ANI character are the closest friends of each other. The key tool of this is the Smart Digital Arms with AI at the finger tips, which allows the quick creation of programming works. This Smart Digital Arm was built by SE character himself, with color of CMU reflected through the violet light emitted from the tool. This gimmick has implication that can be widely understood with general perception in vision of color and arts.

Figure 22: Character Size Comparison when the character assemble as a team group.

6. Graphic for Advertising Product

The use of mascots for promoting an educational organization has more limitations than the use for a private organization in several aspects, including ethical images, definitions of education and organizational culture, all of which have to be taken into accounts and lead to the consideration of the types of media used. All media types, both physical and digital ones, are related to target groups and images of the organization. For instance, it may be deemed that the medium that is the symbol for memorization is a keychain souvenir because a general function of a keychain is to be a memorable symbol of an event or a place. However, at present, some functions may be added, such as to add a USB port flash drive to the keychain for multifunctional purpose. Likewise, designing the graphics that have mascots as the core element, it is necessary to present the images and tones that make the organization look like a lively and attractive place for every student like the students' second home that has modernity, and contemporaneity. In addition, the definitions of the term 'welcome' are reflected through the facial expressions of the characters. All the aforementioned concepts generate keywords that will be applied to all products under the principle of graphic identity in combination with the principle of graphic design in order to blend perception into arts perfectly without the need to sacrifice the identities of the organization. The building of such images needs the release of images in phases with the goals to allow audiences to send feedback and to enable all audience to perceive a little bit of new images of the organization at a time, starting from the display of the characters in the forms of posters in social media in special occasions such as New Year's Day, or in an event that has become a hot topic in the current society, such as the COVID-19 circumstance in 2020.

6.1. Poster Design

. rather, it includes the change of the head of the Facebook page of the organization, public announcements and question what the designed characters are representing. The chosen release day is the day for the celebration of 11th Anniversary of the University. This may have impacts on people in and outside of the University. In addition, this scheme al so makes students feel attached to the subjects they are majoring.

Figure 23: The Example of Character Poster Design

In 2020, there had been the outbreak of COVID-19 disease that caused Chiang Mai to be locked down for a certain period of time. At that moment, the professions that were worrisome were doctors and nurses who played key roles in taking care of patients, with the close encounter to the incurable disease. Therefore, the team designed a poster which was used in an online campaign to show the characters in the outfits of doctors and nurses in order to show the feelings of care and appreciation to doctors and nurses who worked hard. This design expressed the message that enabled people seeing the poster aware of the concern by the University with the situation, through the characters in the well-blended manner whilst expressing the perspectives that the organization had towards the country.

Figure 24: A Digital Poster using of Mascot on COVID-19 period in Thailand that also promoting a brand identity and policy at the same time.

Concerning the use of the poster design to deliver a message, at present the technique that has been found most effective currently is the creation of graphics that can be shown online. However, it is merely a one-way communication technique that can create awareness. Meanwhile, the use of mascots of the University to deliver messages to audience is a good way that reduces tension in certain situations along with presenting the images of the organization in new dimensions. Especially, in case of an educational organization, the delivered messages will be more reliable and accepted. All the aforementioned principles are related to social structure.

6.2. Line Sticker Design

To design characters to be used in social media communication is included in the marketing plan for building a brand in the modern era. This technique is to use characters that express emotions in the conversation to promote the key issue. In this case, the chosen media is Line[13] chat with is a popular mean of communication of people nowadays and is widely used in Thailand. Therefore, the team has created the mascots of CAMT to show emotions, appearances and symbols of human beings. This enables the users to choose the suitable characters that they like to complement their communications. Thus, designs of the gestures of the characters are related to currently popular messages such as 'congratulations', 'work hard' or 'thank you', all of which are still within the frameworks for designing the mascot of the organization without communicating messages that are too in-depth, allowing people of various age ranges, from high school students to middle-aged people, to use the designed characters and signs. In addition, the designed appearances and the used colors make users feel positive because the characters do not look too serious, which makes them fit the current trend.

The use of these character in digital communication (Line application in this case) has boosted the awareness of identity in deep and cognition because of each character has a differences personality for example, when they use MMIT emotion it can feel the leadership personality somehow from the face and body language, and that can show the identity of organization while using these symbols. The gender of the characters is also affected the user via using image or sticker chat, it makes the user more attractive on chat, same as using new patterns of language in

. Each version is designed to have a concept that is different from each other. The first version is for expressing common emotions such as 'thank you' and 'good' in 'good day'. At the same time, stickers in the other version are designed to be more specific with the works of the organization or professions, with new concepts and modern words, including slangs added, in order to make the designed characters become contemporary in the aspects of appearances, images and notions. In addition, facial expressions of the designed characters can look hilarious as well.

Figure 25: The example of Line Sticker CAMT mascot version 1-2 that already in the market.

-19 disease; hence, the research team has designed Line stickers in Thai language in order to use the designed mascots to promote and reinforce certain messages to the society. The mascots in this special version wear the outfits of doctors and nurses to show care and concerns with people in the society, which will spread positive vibe. Therefore, it is apparent that that the designed characters can be used for conveying diverse meanings and can be transformed into different styles whilst carrying the identities of the organization and their own identities at the same time, e.g., the KIM character, who is the source of knowledge or the team leader, wears a doctor uniform and has gray hair whilst other characters wears uniforms of PAs and nurses. At the same time, each character still shows his/her personal characteristics. For instance, the DF character, even though wearing a nurse uniform, shows cute and lively personalities as well. And that it makes people smile even in bad situation, this is a power of Mascot.

Figure 26: The examples of Line Sticker CAMT mascot COVID-19 version

[13] Line is a new communication application managed by a Japan company called Line Corporation. This application allows people to make free calls and free messaging. Invented by Naver Corporation Korea, it is the most popular

6.3. Architectural Decoration

The plan for the presentation of identities of the organisation is to increase the number of the woks consistently. One of the significant symbols is the L-shape building with glass wall at the main entrance. One thing that is indispensable is the decoration of the building, which will create the atmosphere and emphasize on the identities of the organization. The team designs big posters showing the designed characters to be displayed on the glass wall of the entrance in order to present new images of the organization and to renovate the building because the building is heated by the sunlight from the northern side. Pasted on the glass wall, the stickers will serve as blinds instead of curtains. The effects that happen are that the appearances of the building are completely changed because the colors and looks of the designed characters make the building of the organization look more exciting, active and delightful. The outcomes from this work lead to the discovery that to use bright and lively looks of the characters makes people visiting the building feel lively. These effects are like the ambience of Akihabara area in Japan, where some buildings are decorated with manga characters. Even though such manga characters are for advertising purposes, they make people in the space feel lively. Therefore, the decoration of the building, together with the design of landscape to have the building encompassed with a green garden, will also make students feel fresh and lively. At the same time, the decoration of the building builds unity in the organization. Also, the building of CAMT is the only building in Chiang Mai University, or even in Chiang Mai Province, that is decorated with images of contemporary characters in order to create the positive perception perfectly.

Figure 27 The use of Mascot for decoration, College of Arts Media and Technology, CMU.

7. Toy and Souvenir Product

As for the designing of characters, one of the indispensable elements is the creation of 3D models in order that the images and appearances of the designed characters will be more concrete. This concept of the creation of 3D models has led to the idea to create souvenirs to reinforce the identities of the organization through digital symbols and digital way of life. At present, many people, with artists and designers included, prefer to collect toys in the types of action figure, statue, Figma[4]. Therefore, the team has an idea to create souvenirs that are 3D models of the designed mascot characters with the framework to present the images of the organization along with the creation of aestheticism from small toys. The creation process starts with the creation of the characters in a computer before using a resin 3D printer which has the greater resolution than aplastic (PLA) 3D printer. The concept of this work is also based on some marketing principles concerning the presentation of images and symbols of the organization simultaneously.

7.1. Concept Design

This project is the creation of the set statue contained with 7 characters of CAMT in the sizes and scale were in accordance with the standard of Nendoroid[5], with the heights ranking between 5-15 cm, in the forms of SD characters [15]. Nendoroid is defined as a popular brand of small plastic figure replicas of manga and anime characters [16]. This is not too big statue but not too small based on population toys in the market. Mostly, for toys collector in worldwide so it become a standard size, scale and proportions. As a comparison of Nendoroid in the market place, there are the same scale and can replace the face and emotion but it lacks of group assembler product and light up system. This is an idea to create a difference type of Nendoroid figure. In this case, the works are created as statue figures, in order to set the gestures and looks to be presented that go in the same direction. The scopes of the design are as follows.

. Each collector must collect all the 7 models in order to get a complete logo.

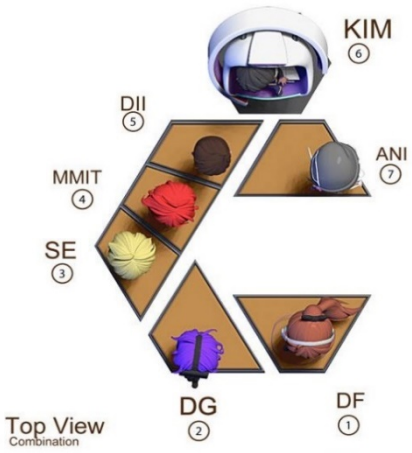


Figure 27: A Top view of 7 pieces of mascot combined into a CAMT logo

- The base is supported by a plate which is included in the model of KIM character, which is the team leader.
- The base of each model consists of the lighting system that related to the main plate, called LED light up base, which makes the display look more dimensional.
- The models are displayed in the stepped manner, with DF and DG characters sitting on the lowest steps whilst other characters are standing on the upper more steps. Thus, all characters will be seen from all angles without blocking one another.

Some characters, such as ANI, are floating all the time. Thus, the model of such a character consists of a metal pole that makes

[14] Figma Toy is a series of Japan toys invented by Max factory and distributed by Good Smile Company. The product focuses on highly articulated hand-sized figure of popular anime figure.

[15] Nendoroid is a series of plastic figures created by Japanese Good smile company in 2006. It becomes popular due to cute appearance. The figure has a function of articulation in appropriate scale for collectors.

the model look floating. In the future, the pole many be created with transparent material such as clear acrylic so that it will look more realistic. All the aforementioned ideas are complemented with the marketing design in order to make collectors desire to collect all the models in the set-in order to have a complete logo of the organization. In addition, the set of the models can serve as a lamp. As for toy designing, the most important elements are sizes and perspectives of the work in 3D view. Therefore, sizes are first set in the program of 3D image creation. The framework that is especially dedicated to this work is designed.

28: A Side view show the levels of the model steps

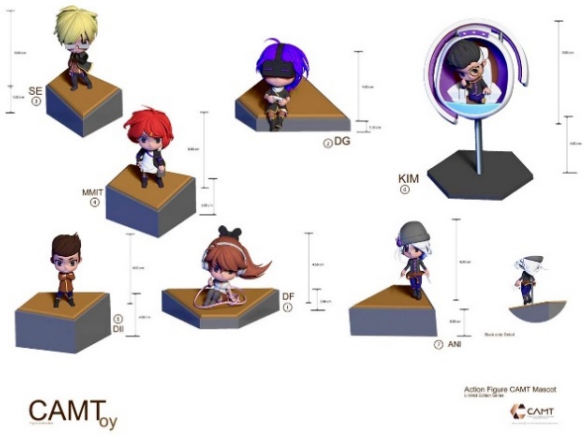


Figure 29: The separated details of each character

7.2. Methodology

The production of the toys can be divided into 3 main steps, namely, Pre-production, Production and Post-production, with the details of activities as follows.

- Pre-production is the step in which the initial design of each character and the concept for the assembly of the toys (Toy Mechanic Combination) are drafted. Also, details concerning used materials, postures, communication of the meanings and perspective are tested in the 3D space. In addition, the proportion and function for lighting systems can be adjusted in the step of the sketching in the 3D space. Furthermore, there is also the research for information of toys of the types of Statue and Nendoroid. If the size is bigger that 15 cm, the model set will need a great space, but a purpose of this model set is to be placed in any place such as at the space over the console of a car or a computer desk, which has no great space. Therefore, each model is designed to be 5-10 cm in accordance with the scale of each character.
- Production is the step that directly involves with the creation of the model set. This step starts with the creation of 3D model prototypes and the number of polygon is not taken into account. Thus, the design can have high quality subdivision, which will not affect the software for the printing. In this step, the focus is on the parts with joints that will be seriously taken into account because of the significant effects on the printing out. The consideration is also put on the designing of the colors for each character. Finally, the lighting system is added to the printed out model.
- Post-production is the step that still incorporates adjustment but the adjustment is more like the finishing of the details of each work in order to create the mold that will be used for mass production. This step also includes the taking of photographs to present the works and to get feedback which is used as statistical data of marketing demand. This step also includes activities in the production step if there is any suggestion from audience or expert. Most of the attained suggestions concern the use of mixed materials and the colors of the product.

From all the 3 steps as mentioned before, it is apparent that there are 2 keys activities in the production, which are the creation of products in 3D space and the printing out which require many experiments and adjustments.

This activity of the creation of the model is the experiment to determine the perspective and composition whereby all the models can be seen. From Figure 30, it is apparent that the SE character model is swapped with the DG character model in the final product. This is because SE model is in the standing posture whilst DG model is in the sitting posture. Hence, DG model is placed in the central area instead of SE model.

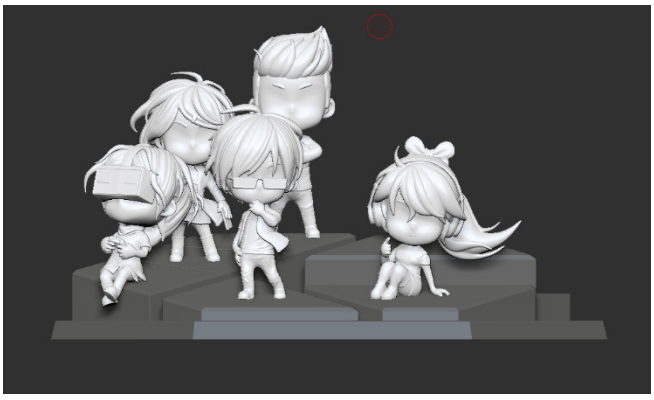


Figure 30: The First Draft 3D model in 3D Software.

This step also includes an important activity, which is the designing of separable parts such as faces, hair styles and bodies in order to let each model look seamless as possible. For instance, an SE model consists of 3 separable parts, namely, hair, head and body, which are easy for the assembly and the addition of details.

Figure 31: The test 3D model printing using resin separated 3 pieces.

. From the experiment, it is discovered that resin is very fragile especially in the parts of hair and small pieces in the model. Therefore, in this case, glasses are made of transparent plastic instead of resin.

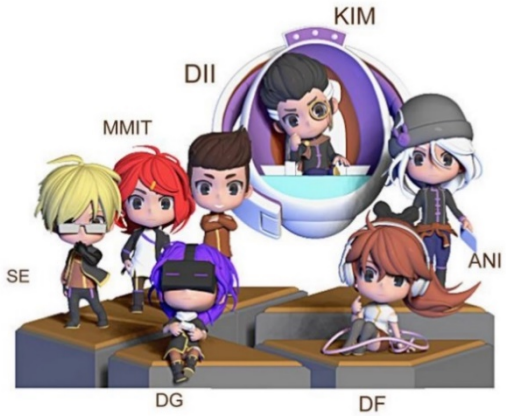


Figure 32: The Combination of the Mascot set in 3D vision.

After the test 3D printing of SE models, it has to make sure that when the character combine together, it in the right position and proportion. We use symmetrical in a feeling of mass and space, as you can see that KIM is blended to the right but on the left, there are 3 character standing on the step to make the composition look actually balance. The based system also designed in this step, in order to move on to printing a prototype step.

7.4. Printing a Product

This step involves the production of the models that have passed the test printing in 3D space already. All the pieces are put together, and the bases to support the works are also designed. The designing of the base to support each model is an important action in peeling the mold off the work. After the mold is peeled off, the work has to be scrubbed in order that the surface is smooth. In addition, a UV LED lamp is used for making the resin material hard and have smooth surface as expected. Instead of the lamp, sunlight can be used as well. Resin printing is different from PLA because it can show more details. In this case, a UV LED lamp is used for creating the prototypes, by using UV light to cause curing reaction on resin which makes resin hard.

. As shown in Figure 33, put together on the plate, the bases of the models do not fit well with one another. Therefore, the pieces of works have to be scrubbed with sand paper before the coloring process. Another problem is that the metal pole for holding the model is too high, disabling the model to stand by itself. In this step, there will be slight adjustment to the base. In the following step, the installation of lighting system will be focused on.

32: The Printed Resin Prototypes of Some Characters

Figure 33: The testing of LED light up system

As for the installation of lighting system, the switch is placed at the back of the plate to allow the connection with the adaptor for the light-up display. All the works in all the aforementioned steps have been described in posts on social media in order to attract feedback from message receivers. The attained feedback is satisfying to a certain degree but the light is too sharp and too bright. Therefore, the material of the base that cover the light bulb is changed to translucent one in white color in order to accumulate the light that hit the object. The opacity level of the material is 60 percent to deficit the light.

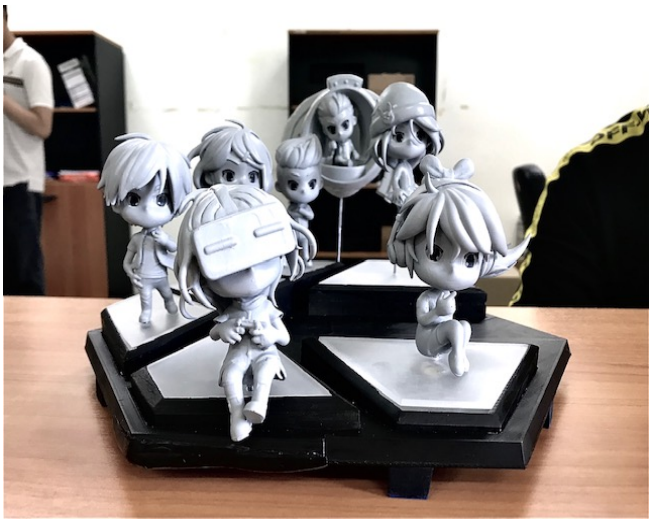


Figure 34: The Printed Prototypes Put Together on Plate in the Shape of the Logo of the Organization

7.5. Product Development

Based on suggestions and feedback from the experts and the audiences, the bases are modified to be in light gray color and the positions are changed, with KIM model pushed back of the plate. In case where a collector does not want to display all the models together, each model can be displayed separately. In addition, the KIM model is modified to show less seams because this model is the biggest and heaviest of all.

Figure 34: A prototype of the assembled product: picture for commercial released in October, 2020, on social media

After the product development leads to the satisfying outcomes, photographs of products are taken and shown on social media in order to get feedback in the Post-Production step. The results show that the satisfaction level is greater than 80 percent, but there are still requests for colored products, which are still in the course of experiment on different materials and colors.

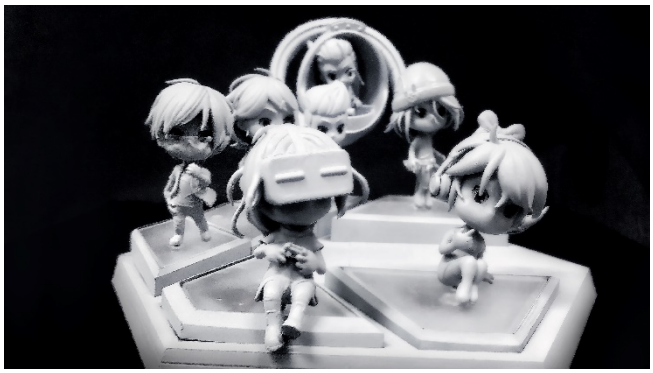


Figure 35: A prototype of the combination product and this picture is for commercial released in October, 2020 on social media.

7.6. Prototype

This final step involves the presentation of prototype products on social media. The presented prototype products are monotonously in the grayish white color. The products are presented in black and white color tone so that audiences can see the forms of all models, the lighting system of the base of each model and effects on the products. This style of presentation is popular among toy producers to present their works in a way to enable audiences to use additional imagination from experiences in primary perception and concept prior to the colored version.

Figure 36: A prototype of the combination product with LED light up test, released in October 2020 on social media.

In certain cases, consumers or audiences may want to see the prototypes in gray color, which can be optional. However, some audiences may p want to see just the colored version in order to consider the features of the products. Nevertheless, this style of prototype presentation will make audience consistently follow-up the updates of the products and give feedback. Afterwards, there is the primary virtual coloring so that audiences can imagine on the final products.

7.7. Product Publicity for Feedback

After the prototypes of the works are created, the next step is to create digital visualized images of the colored works to be presented to the audiences for their feedback. The colored works are presented with the concept of the overall product set in order to show clear identities of the designed mascots, which is the extension of the mascot designing.

Figure 37: A prototype in color version from digital visualization published for feedback

The issue relating to the works in this set is that they are just prototypes. This means the works can be modified in many other steps. There have been critics on material that the used material should be the mixed one that is composed of silicone because nowadays, silicone has been added as a component of materials to be used for creating works in the type of toy or statue in order to create the moods and dimensions of the surface that look more realistic. The extension of the product aims at the same goal, which is to extend the brand of the educational organization that is a choice for the promotion to make the brand sustainable.

Figure 38: A close-up shot of products of colored version

8. Results

As for the creation of elements of brand identity of the educational organization with various digital design-oriented media, namely, digital image, digital communication icon (sticker massage), poster, building decoration and toy (collectible) products, different feedback and opinions concerning different issues have been given by audiences and the experts.

In this project, many new characters (mascots) have been designed to have features different from original ones designed by artists/students. This means the characters in this project are re-versioned to have contemporaneity as other artists/students deem preferable. Theoretically, this style of drawing is called 'Dojinshi' [17] in Japanese language, which means the artists who are fond of certain characters to certain degrees re-create such characters in their own styles. Also, this style is called 'fan-art'.

The aforementioned fan-art phenomenon can be seen as the expression of feedback in the concrete and tangible manner.

Figure 39: Transformation of style of a character (ANI) into teen look can attract more favorable feedback.

Nevertheless, the designing of mascots for educational organizations still faces with certain limitations to designing, on the bases of target groups, organizational culture, spatial culture, and ethicality, which, all together, form the main scopes of the designing. Consequently, the designing cannot go beyond such scopes. This incident reflects the key difference between free comics and education comics because the designing has still had to embrace all the concept keywords. As a result, positive feedback and satisfaction are attained because more than 60% of the audiences (students and staff) deem that the designed mascots (characters) are related to themselves. Even though the uniform is designed to be more contemporary, the used symbolic colors of the organization, i.e., violet and brown, are implication that are hidden in the compositions of the designing and are comprehensible among target audiences.

The unique brand at CMU is a marketing concept for the creation in this project. The creation of the mascots that look contemporary and the extensive designing of other products in order to add values to the brand have led to the uniqueness of the University which results from this creation because the University is the only educational organization that applies platform and media usage in industrial level through the digital processing method in order to create versatile and tangible works. The ultimate result is qualitative and positive feedback that reflects a high level of satisfaction.

To create toy product is a key idea that enhances the strength of the brand of the organization because it reflects the versatile knowledge and skills acquire from learning. The concept of the University is to be a governmental university that is under regulations and conceptual frameworks but has been developed in the digital new normal era that follows the post-modern era. The designed products are tangible and create aestheticism while letting audiences absorb brand identities. This idea leads to a new direction of artwork creation that can be extended to industrial production. Furthermore, this experiment has received positive feedback from audiences in social media, which can be implied that it has market demand in a good level.

In techniques of creating these toys set, the using of digital methodology in ordinary which are follow step 1) 3D creating 2) Printing 3) Mass product creating is the same, but the idea of adding more option into the product such as light up figure or combination of set is the core of concept campaign if comparing with others Nendoroid in the marketplace and it keep develop and added more function into the product. However, the use of Toy

product in brand identity is a good idea of memorizing by touching, playing, and watching especially in institute's educational organization.

9. Conclusions and Discussion

From feedbacks from 3 groups of audiences, namely, CMU students, officer, staff and people in general, it is apparent that there are certain ideas that can be extensively applied to Institute's education brand designing and identity creation, with the main goal being to create tangible uniqueness for the organization through digital platform and digital media, in the form of products that people in the digital contemporary era can comprehend well an touched. Most of organization need a symbol present their identity. Logo is a good example, but it is lack of life and emotional in deep then that the idea of creating the Mascot was born. In this study, the scope idea does not include Digital Animation Film or game or any advance media, but it is the next step in advance once the brand is growing up and popular at certain level in social. Therefore, this study emphasize on to a new way of using Mascot design to enchanced and added value of the University brand in complicated cultural. It need many research of people mind, comments, feedback, population statices and demands, as well as analysis on behaviors of visitors who visited to the University and current students, which contributed to the creation of the new brand identity platform and the improvement of art identity in a future mindset. Key findings from the results analysis and presentation of the artwork product testing are as follows.

First, the designing of mascots for an education institute requires the connection among different dimensions of organizational culture, from the personal level that relies on the in-depth analysis on attitudes and specific personalities, the level of organizational image of the University, the level of provincial culture and the level of the country, under the conceptual framework that has limitations of dressing, ethicality, and purposes and policies of the organization. This is a contemporary study and creation project that creates new identities of the organization, which are reflected through characteristics of mascots (characters) that can be extensively designed as other products, based on the approaches of marketing strategies and advertising, all of which are systematically and sophisticatedly inter-related to one another. In addition, the trend of design is a factor for the creation. The results from the analysis point out that all the symbols designed in this project will be changed again in next ten years in order to sustain the values of the brand.

Secondly, the key factor for character designing is the target groups that are potential customers in term of marketing. Thus, it is significant to focus on students in order to study on trends and art population. From the study, it is discovered that 80 percent of the samples are interested in the styles of comic characters in animations and games rather than general photographs whilst 20 percent of the samples are interested in fashion in portrait style. It is also discovered that the cosplay and Korean pop (K-POP) fashion styles are most influential. Hence, the characters are designed in accordance with the aforementioned demands and trends, in combination with the spirits of digital era in order that they will have the ultimate looks that fit the trends while reflecting the organizational culture simultaneously.

Thirdly, the process of the creation of characters for an educational organization has conceptual framework that has more limitations than organizations of other types, including ethics, morals, and local culture. As for this case, it is Chiang Mai University which is a symbol of the city. Hence, it is necessary to create specific concept from analysis and synthesis. The aspect that must not be overlooked is the combination between organizational culture and local culture that is composed of ideas of people from local communities. Therefore, the outcomes are based on contemporary ideas that are more powerful than conservative perspective alone.

The fourth point of the findings is that toy product is a significant factor that affects imaginations for the audiences, people and provide beyond personal experience as we know that toys can make you imagine and more [18]. With this power, it is a powerful instrument for running promotional and marketing activities to promote brand identities perfectly. In general, people in this modern era, especially teenagers of generation Y people and younger people of following generations prefer to collect such products or any other products that features comic characters. From the analysis on the data of the market of Japan which known as the comic city as a culture, it is discovered that comic characters are used for presenting the tourism, and there have been new characters created until this style of art has become an element of culture since the post-modern era until the Bro-postmodern era when this art style is combined with the elements of the digital era. All the aforementioned cultural features have been absorbed by Thai youths. Consequently, modern artists can create manga-like art works, which can be regularly seen at present. From the analysis, it can be seen that this art style has also incorporated the features of traditional Chinese arts, especially the Chinese brush line drawing that has been developed to be digital line drawing that enables the creation of art works with new styles and contents. Eventually, the art works have been used as the bases for producing tangible mass products that, in turn, increase the popularity of the characters on which they are based. As for this project, the aforementioned product is the designed toy that also serves as a tool for measuring marketing outcomes, and effects in terms of art and culture simultaneously.

The final point from this study is that a perfect brand identity package should consist of tangible and intangible media that are contemporary, which will increase the values of brand and enable the brand to have versatility in terms of social statuses and positions so that the brand can effectively communicate with people in the current era. The findings from the analyses on consumer behaviors, products, trends, and positive ideas that have led to aesthetic perception which involve the awareness of artful beauty from sight and other senses. Hence, the design package for this study has incorporated digital art, and physical art and craft, which can reinforce the images of the organization in recognition and memories of audience in the easily comprehensible and quick manner, in addition to the use of ordinary graphic image or the logo of the organization.

These mascot design set and graphic brand identity project has been used in officially in Chiang Mai university, Thailand and its keep inspired to the people who visited CAMT as a school of new arts, media and technology at all the time.

References

[1] N. Kongdee, S. Prapawong and J. Manissaward, "The Designing of Institute's Educational Mascots," in 2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), 287-291, 2020, doi: 10.1109/ECTIDAMTNCON48261.2020.9090707.

[2] L.L. Adams , "The Mascot Researcher: Identity, Power, and Knowledge in Fieldwork," Journal of Contemporary Ethnography, **28**(4), 331-363, 1999. https://doi.org/10.1080/17430437.990688

[3] S. Mindray, A. Manolica, T. Roman, "Building Brands Identity," Procedia Economics and Finance, **20**, 393-403, 2015, doi: 10.1016/s2212-5671(15)00088-X.

[4] P. Medrano-Bigas, The Forgotten Years of Bibendum. Michelin's American Period in Milltown: Design, Illustration and Advertising by Pioneer Tire Companies (1900-1930), Ph.D Thesis, University of Barcelona, 2015: English translation, 2018.

[5] W. Nunthaporn, Graphic Design for Campaign of No Sweet for Thai Youths. M.A. Thesis, Chulalongkorn University, 2010.

[6] P. Medrano-Bigas, The Forgotten Years of Bibendum. Michelin's American Period in Milltown: Design, Illustration and Advertising by Pioneer Tire Companies (1900-1930), Ph.D Thesis, University of Barcelona, 2015: English translation, 2018.

[7] S. Brown, S.P. McCabe, Brand Mascots: And Other Marketing Animals, Routledge, 2014.

[8] V. Daniel, Does your mascot match your brand's personality?, M.S. Thesis, University of Twente, 2014.

[9] W. Nunthaporn, Graphic Design for Campaign of No Sweet for Thai Youths. M.A. Thesis, Chulalongkorn University, 2010.

[10] P. Prueangphong, Mascot Design for Cultural Events of Chiang Mai, M.F.A. Thesis, Chulalongkorn University, 2014

[11] N. Tongprasert, Meaning Design, Urgent Tag Co.,Ltd., 2016.

[12] W. Chiemvisudhi, "The sign and idenity signification through the officil Olympic game emblem," Veridian E-Journal, Silpakorn University, **10**(1), 2011-2092, 2017. https://doi.org/10.5749/12.2.0124

[13] C. Bacchilega, Postmodern Fairy Tales: Gender and Narrative Strategies, Philadelphia: University of Pennsylvania Press, 1997.

[14] S. Kapidzic, S.C. Herring, "Gender, Communication, and Self-Presentation in Teen Chatrooms Revisited: Have Patterns Changed?," Journal of Computer-Mediated Communication, **17**(1), 53-54, 2011. https://doi.org/10.1111/j.1083-6101.2011.01561.x

[15] S. Seibundo, My First doll: Clothing Patterns Creating in Nendoroid Doll Size, Seibundo Shinkosha, 2019.

[16] Y. Sone, Japanese Robot Culture: Performance, Imagination, and Modernity. Palgrave Macmillan, 2017.

[17] S. Seibundo, My First doll: Clothing Patterns Creating in Nendoroid Doll Size, Seibundo Shinkosha, 2019.

[18] S. Akiko, "Emerging 2.5-dimensional Culture: Character-oriented Cultural Practices and Community of Preferences as a New Fandom in Japan and Beyond," Journal Article Mechademia: Second Arc, **12**(2), 124-125, 2020. https://doi.org/10.5749/mech.12.2.0124